JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 11 (2023) ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoint

Reviews

Original Papers

Corrigenda and Addendas

Letters to the Editor

Research Letters

The Importance of Data Quality Control in Using Fitbit Device Data From the All of Us Research Program

Lauren Lederer[1], BS; Amanda Breton[2], MS; Hayoung Jeong[1], MS; Hiral Master[3], PhD; Ali R Roghanizad[1], PhD; Jessilyn Dunn[1], PhD

1

2

3

Corresponding Author:
Jessilyn Dunn, PhD

Abstract

Wearable digital health technologies (DHTs) have become increasingly popular in recent years, enabling more capabilities to assess behaviors and physiology in free-living conditions. The *All of Us* Research Program (AoURP), a National Institutes of Health initiative that collects health-related information from participants in the United States, has expanded its data collection to include DHT data from Fitbit devices. This offers researchers an unprecedented opportunity to examine a large cohort of DHT data alongside biospecimens and electronic health records. However, there are existing challenges and sources of error that need to be considered before using Fitbit device data from the AoURP. In this viewpoint, we examine the reliability of and potential error sources associated with the Fitbit device data available through the AoURP Researcher Workbench and outline actionable strategies to mitigate data missingness and noise. We begin by discussing sources of noise, including (1) inherent measurement inaccuracies, (2) skin tone–related challenges, and (3) movement and motion artifacts, and proceed to discuss potential sources of data missingness in Fitbit device data. We then outline methods to mitigate such missingness and noise in the data. We end by considering how future enhancements to the AoURP's Fitbit device data collection methods and the inclusion of new Fitbit data types would impact the usability of the data. Although the reliability considerations and suggested literature are tailored toward Fitbit device data in the AoURP, the considerations and recommendations are broadly applicable to data from wearable DHTs in free-living conditions.

(JMIR Mhealth Uhealth 2023;11:e45103) doi:10.2196/45103

KEYWORDS

wearable device; Fitbit; All of Us; data quality; noise; missingness; biometric monitoring

Introduction

Wearable digital health technologies (DHTs) have become increasingly popular in recent years, especially as DHTs offer better user experiences, more capabilities, and greater functionality to assess behaviors and physiology in free-living conditions. The *All of Us* Research Program (AoURP) is an initiative that is seeking to collect health-related information, including DHT data, from a diverse cohort of over 1 million participants in the United States. In the AoURP, DHT data are collected alongside electronic health records, biospecimens, surveys, and standardized physical measurements. The goal is to make these data accessible to both researchers and participants to advance precision diagnosis, prevention, and treatment [1].

In 2019, the AoURP expanded its data collection with the Fitbit Bring-Your-Own-Device (BYOD) project. This expansion has allowed participants to share their historical and ongoing Fitbit account data through the *All of Us* participant portal [2]. The AoURP's efforts to include Fitbit device data have continued to expand with the WEAR study, which gives eligible participants a no-cost Fitbit Charge 4 or Fitbit Versa 3 device [3]. Data from the BYOD project are available through the AoURP Researcher Workbench, which offers data access and analysis tools for DHT data, electronic health record data, biospecimens, surveys, and physical measurements [4].

As of 2023, over 15,000 *All of Us* participants have shared Fitbit device data [5]. The Researcher Workbench provides access to these data, including Fitbit-defined heart rate by zones that are based on percentages of estimated maximum heart rate, minute-level heart rate, daily activity summaries, minute-level intraday steps, daily sleep summaries, and sleep levels [6].

Digital biomarkers derived from DHTs can potentially be used to improve clinical diagnostics, predict disease status, and support personalized clinical decision-making [7]. With the increasing use of DHTs like Fitbit devices in the AoURP and other research and clinical settings, it is important that those working with these data consider the inherent limitations of Fitbit devices, given their underlying technology. This will enable improved data processing and fit-for-purpose implementations of Fitbit devices in research and clinical settings. Researchers might ask the following: "How reliable is the data from these devices? What are the sources of noise, error, and bias that should be accounted for when using this data? How can these be accounted for?"

In this viewpoint paper, we examine the reliability of Fitbit device data in the context of the AoURP's BYOD program. We focus on Fitbit devices, given their wide market share [8,9], the ongoing collection of data from Fitbit users in the AoURP (eg, BYOD program) [2,10], and these data's availability to registered Researcher Workbench users [5]. Data are currently not available from the AoURP regarding the Fitbit device models included in their data set. For this reason, the data cleaning strategies we present are device model agnostic. This paper focuses specifically on data considerations around physical activity (steps and movement intensities) and heart rate measurements generated by Fitbit devices on a daily and per-minute basis. Given that Fitbit device sleep data are derived from the same underlying sensors that determine heart rate and motion intensity metrics, the same fundamental considerations surrounding inherent measurement reliability should be considered when working with Fitbit device sleep metrics.

Sources of Noise

Measurement error can affect the reliability of Fitbit device data in both laboratory conditions and free-living conditions. In this section, we discuss the most commonly recognized sources of error that may be observed in Fitbit device data collected through the AoURP.

Inherent Measurement Inaccuracies

Fitbit devices include a 3-axis accelerometer and photoplethysmography (PPG) sensor, with more recent device models including additional sensors, such as an altimeter, a gyroscope, a skin temperature sensor, and multipurpose electrical sensors [11]. It may be helpful for researchers to consider the data supply chain as they work with DHT data [12]; the data that researchers generally have access to are processed, and the firmware that performs such data processing is regularly updated [13]. Aside from the data supply chain, there are also inherent limitations of the accelerometer and PPG sensors themselves that should be accounted for in data analysis.

All Fitbit models with Fitbit LLC's patented PurePulse technology (eg, Fitbit Charge, Fitbit Charge 2, Fitbit Charge 3, Fitbit Alta, Fitbit Versa, Fitbit Blaze, and Fitbit Ionic) use the same PPG hardware and software for heart rate estimation [14]. PPG sensors, which optically measure light absorption under the skin, may be affected by user motion and activity intensity, skin tone, and the wavelength of light used by the sensor [15,16]. When compared to gold-standard electrocardiography, Fitbit devices tend to underestimate heart rate [14,15,17]. Further, Fitbit device heart rate measurements have higher reliability under stationary conditions [14,18].

Fitbit devices use the 3-axis accelerometer to determine step count and categorize physical activity intensity (sedentary, light, moderate, vigorous, or moderate to vigorous) [19-21]. Comparisons between Fitbit device step counts using direct observation and gold-standard accelerometers, such as the ActiGraph GT3X+, demonstrate mixed reliability, depending on the type and speed of movement and the on-body placement of the Fitbit. During normal walking for example, the torso placement of the Fitbit device has resulted in the greatest accuracy, while ankle and wrist placement have been the most accurate in slow-walking and jogging, respectively [20]. Similar findings from other studies demonstrate that step count and physical activity intensity accuracy are affected by device placement and movement type [22-26].

Skin Tone

Skin tone may be another inherent source of error for DHTs that rely on optical measurements (eg, PPG or pulse oximetry) [27,28]. Both melanin and skin with tattoos absorb more green light, that is, wavelengths of around 530 nm, which are the LED wavelengths commonly used in PPG sensors [29]. There have been mixed findings in this area; a study by Shcherbina et al [30] on older generations (2014-2016) of consumer smartwatches found that darker skin tones positively correlated with increased heart rate error, whereas our study, in which we used more recent devices (2014-2018), did not find a relationship between heart rate measurement accuracy and skin tone across a subset of consumer smartwatches [18]. Clinical-grade pulse oximeters that rely on red and infrared optical measurement technology may also be affected by skin tone [31].

Fitbit devices may use a combination of both green and red wavelengths to estimate heart rate [15,32]. Although green wavelengths can enable more accurate heart rate measurements during movement when compared to red wavelengths, green wavelengths are more readily absorbed by melanin before reaching the photodetector [29]. Additional research is needed on whether and how the accuracy of optical-based DHT measurements, such as heart rate and saturation of peripheral oxygen (SpO_2), is affected by skin tone.

The data collected by the AoURP currently do not include data on skin color or the presence of tattoos under Fitbit devices; therefore, it is not possible to directly account for skin tone or wrist tattoos as a potential source of error in AoURP heart rate data. As a result, researchers working with *All of Us* data may need to take extra care when interpreting or translating results that may be influenced by skin tone or the presence of wrist tattoos, particularly when working with heart rate data.

Movement and Motion Artifacts

Motion artifacts can also be a source of error for heart rate, step count, and physical activity intensity data. Unexpected noise with random amplitudes and frequencies can be seen in raw sensor data and can cause the algorithms of Fitbit devices to falsely detect movement or a heart beat [33]. For example, the reliability of Fitbit Flex's step count and moderate to vigorous physical activity data was found to be dependent on the activity type (walking, stair stepping, jogging, and incline walking) [24], and step count error was shown to be higher during activity than during rest [20]. Therefore, it is likely that step count reliability varies, particularly during normal household activities, which may be logged by Fitbit devices as exercise movements.

Wearable device heart rate measurements are the most accurate under circumstances of rest, followed by physical activity and then rhythmic activity, such as walking or jogging. Our previous work demonstrated decreased reliability during rhythmic activities, such walking or typing. This was likely due to Fitbit devices mistaking the periodic signal, which was being produced by the repetitive movements, for the cardiovascular cycle. Although walking resulted in heart rate measurements that were higher than the true heart rate, typing resulted in heart rate measurements that were lower than the true heart rate [18]. A study by Benedetto et al [15] assessed Fitbit Charge 2 heart rate accuracy during stationary biking and found that the device underestimated heart rate when compared to electrocardiography.

Some possible reasons for heart rate measurement error during motion include the device's sampling and interpolation methods, unstable device positioning, and variation in the pressure applied to the skin by the sensor [15,18,34]. Researchers should be aware of the impact of physical activity type (ie, motion intensity and periodicity) on Fitbit device heart rate measurement error.

The body positioning and fit of DHTs can also be sources of motion artifacts. For example, wrist-based Fitbit devices can misclassify nonambulatory arm movements as total body motion, which may result in the overestimation of physical activity and motion intensity [35]. This misclassification of physical activity may be worse if the device is not worn correctly. To address this challenge, the Fitbit device user manuals provide instructions for specific placement on the wrist to enable the acquisition of more reliable data [36].

Sources of Data Missingness in Fitbit Device Data

It can often be challenging to determine the minimum amount of data necessary to achieve a particular analysis goal when using DHT data. A systematic review by Chan et al [37] pointed to a common definition for a "valid day" of wearable data—at least 10 intermittent hours of data present within 1 day—and a "valid week" of data—at least 3 valid days during the week. It is important for researchers to note that most Fitbit devices need to be charged at least once per week for 1 to 2 hours at the time of writing, and the need to remove the device from the wrist to charge it results in at least some data missingness. More than the minimal necessary data missingness can occur in the event that the wearer forgets to put the device back on their wrist after charging it [35]. Such nonwear is an example of structured missingness, where a contiguous block of missing observations occurs when the device is not being worn. At scale, there may be observable nonwear patterns, such as times when people commonly remove their devices (eg, during sleep) [38].

In addition to nonwear, improper device wear can also result in data missingness. Improper wear, such as insufficient tightening of the wrist strap, can lead to the sensor orientation being askew or a loss of sensor-to-skin contact, which is required for high-fidelity optical measurements, such as PPG-based heart rate measurements [39]. Moreover, observations can be impacted by large motion artifacts, and such observations (eg, high accelerometry values) may be removed by the device firmware. This leads to missing values in the final data set. To explore the extent of such data removal, our team recently compared data missingness in optical heart rate and SpO_2 observations across multiple wearables [18,40]. We found that, for heart rate measurements, the Fitbit Charge 2 had the highest amount of missing data during both rest (18.7%) and physical activity (10.4%) when compared to other consumer-grade wearables [18]. Data missingness due to improper wear or firmware attempts to account for motion artifacts may be seen in the data set as random missingness, lacking structure and predictability. However, some wearers may be more prone to improper wear or high-intensity activity, which can lead to higher amounts of missing data in individual data sets. Other factors that can affect the presence or absence of Fitbit device data include the frequency of syncing the device with the smartphone app and poor device connectivity [41,42].

There is a taxonomy of mechanisms for missing data, including (1) data missing completely at random (MCAR), where missingness is unrelated to observed characteristics; (2) data missing at random (MAR), where missingness is related to observed characteristics; and (3) data missing not at random (MNAR), where missingness is related to unobserved characteristics. Different methods are required to best account for these three missingness mechanisms during data preprocessing; thus, identifying the type of missingness is an important step in DHT data analysis. In the case of Fitbit device data, observations that are MCAR may be due to nonsystematic device malfunctions, nonsystematic errors in data transfer, or sporadic improper device wear. Observations that are MAR may be the result of a particular device model missing a type of measurement capability (eg, a device that is known to not report heart rate measurements under high-intensity activity). An example of MNAR missingness might be nonwear during a bout of illness or due to a user having a poorly fitting device. In a free-living study, such as the AoURP, all three missingness mechanisms are likely to be present in the Fitbit device data and should be identified and appropriately addressed when possible by, for example, making assumptions about the reasons for the data missingness upon analyzing missingness patterns [43].

Mitigating Missingness and Noise in DHT Data

Accounting for Data Missingness

Avoiding data missingness is best done at the data collection stage. Prospective bring-your-own-device studies may increase wear time and improve device fit by incorporating reminders for users to wear the device, adding nonwear alerts for users or the study team, and educating users on fit and charging [44]. It should be noted that AoURP Fitbit device data collection is purely observational and, at this time, does not involve providing any alerts or interventions to improve Fitbit device wear habits.

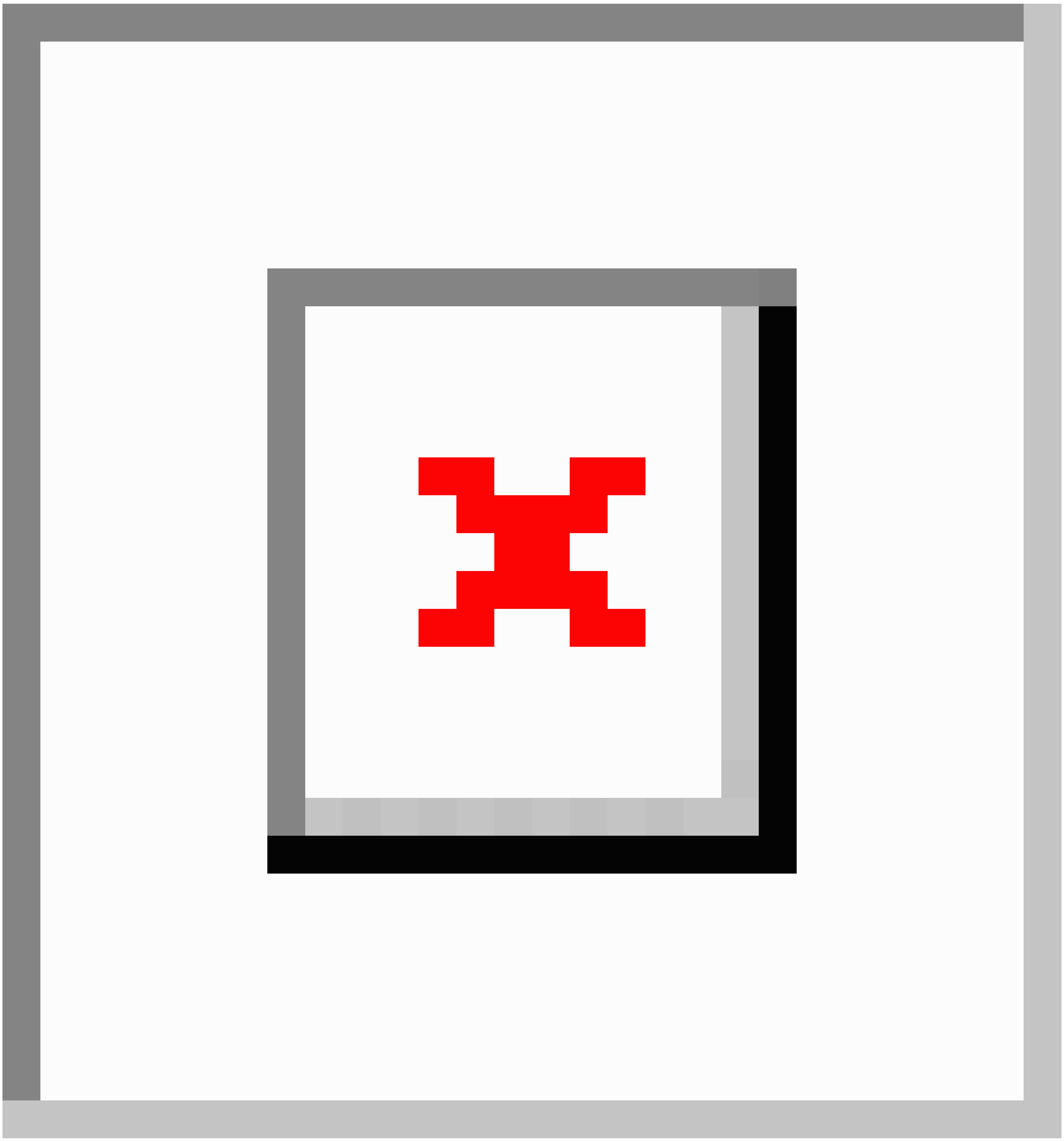
Some missingness in the data is inevitable. Accounting for data missingness begins by thoroughly identifying the reason for missingness and deciding upon the most appropriate strategy for mitigation. For AoURP researchers using Fitbit device data collected in free-living conditions, it is not always possible to distinguish MAR, MCAR, and MNAR missingness; therefore, they may need to make assumptions to decide how to proceed with mitigation [43]. In the context of Fitbit device data, we can distill a few practical solutions for addressing wear-related structured missingness [38].

The first step is to decide upon a definition for “wear” in the context of the data and the question at hand. Depending on what analysis or question is of interest, the definition of “wear” can drastically change. For example, an analysis involving sleep quality and staging may require the definition of “wear” to include common sleep hours (eg, 8 PM to 8 AM) or a minimum monitored sleep duration time, such as that reported by Fitbit devices, but other analyses may only require a daily wear level, which gives an idea of a participant’s activity and physiological state without requirements regarding wear during specific activities, such as sleep and exercise, as described in the *Sources of Data Missingness in Fitbit Device Data* section [37].

One way to calculate a daily wear level is to leverage minute-level heart rate data, as most consumer devices only collect these data when they are worn on the wrist. This would, for example, help to avoid including step count data that may have been collected when a device is in a purse. By dividing the total count of minute-level heart rate observations collected within a single day by the total number of minutes when such data were possible to collect (1440 min in 1 d), we can derive a reasonable estimate of the proportion of the day when the device was on the wrist.

Once established, the daily wear level threshold can be used for filtering out nonwear data and participants; the optimal threshold for filtering should be selected carefully to avoid unnecessary data loss (Figure 1A, Figure 1B). The optimal threshold is where data loss is minimized and there is adequate statistical power to draw conclusions from the analysis (Figure 1C).

Figure 1. Graphs A and B show the total counts of participants and days, respectively, that meet each wear threshold, with an optimal threshold of <0.4. Graph C shows the mean and median total days (per participant) with a wear level greater than or equal to each wear threshold, demonstrating an optimal threshold of between 0.3 and 0.4. The 95% CI (indicated in blue) was calculated as follows: $\mu \pm \sigma/\sqrt{n}$. The IQR (indicated in orange dashed lines) indicates the first and third quartile values. The controlled tier AoURP version 7 data set (C2022Q4R9) was used to generate Figure 1. AoURP: *All of Us* Research Program.



After removing nonwear data (ie, contiguous blocks of missing observations), it is important to identify other sources of data missingness and determine whether mitigation is best done by using imputation or by using the complete case method [43]. Sometimes, the decision can be made based on the extent of the missingness relative to the overall data volume needed for analysis, and at times, it may be deemed that the original analysis cannot be performed as planned due to insufficient data. As an example, a recent study on COVID-19 detection via Fitbit device data calculated the mean over 5-minute intervals of heart rate data; subsequently, any missing data over full 5-minute intervals were imputed by using the median heart rate value from a previously defined 14-day window [45]. Although it is common in wearable data analysis to use the mean and median values of heart rate data for imputation, new imputation methods for biomedical wearable data have also been developed to incorporate machine learning for improved imputation accuracy [46,47]. Although imputation can be beneficial because it preserves data, it should be noted that imputation is not always a good idea, particularly in cases with substantial missingness for which typical values cannot be established.

Accounting for Noise

Many methods exist to reduce noise in raw signal data (ie, sample-level or high-frequency signal data), particularly when the source of the noise is well characterized. Unfortunately, consumer devices typically do not give access to such high-frequency data, but their firmware and adaptive data collection methods are thought to include steps that account for skin tone–related errors and motion artifacts. Unfortunately, the public has no way of assessing how well these methods perform. Based on this, it is difficult for researchers to directly address errors resulting from skin tone or motion artifact errors. However, some general data cleaning strategies exist that may help to mitigate noise in the data, regardless of the source of the noise (Table 1). In this section, we discuss how to leverage changes from an individual's baseline, filtering during repeated activities, and *z* score normalization to improve the signal to noise ratio (SNR). We recommend using these techniques in combination with one another to best mitigate noise.

Table . Examples of noise mitigation methods for wearable data.

	Baseline comparisons	Sampling during periods of similar activity	*z* score normalization
Description	• Calculate the median value during a defined baseline period. • Calculate the Δ from the baseline for all other data.	• Establish specific wear times and use the "Activity Type" metric to filter an individual's Fitbit device data during similar time periods each day for comparable activity types. • Conduct analysis using these segmented data sets.	• Subtract the mean from each observation and divide by the SD.
Applicability	• Mitigate consistent measurement error (bias).	• Mitigate noise that is exacerbated under specific conditions.	• Mitigate short periods of noise.
Benefits	• Provides a "usual" picture of an individual [48].	• Assists in isolating confounding effects that may arise in different activity types and heart rate zones. • Standardizes the comparison of data across different individuals.	• Allows direct comparison of 2 observations originating from different segments of temporal data [49].
Limitations	• Need ample data to establish a baseline [48]. • Baselines can change over time.	• Recommended for large sets of longitudinal data to make accurate comparisons.	• Abstraction of units and range may make it difficult to interpret data.

On an individual level, the comparison of observations to a reliable baseline can be helpful for determining changes in biosignals over time while reducing the influence of both skin tone and motion artifacts. Reliable baselines can be established by first summarizing an individual's measurements during periods of sleep or inactivity or before a perturbation. The determination of which time period to use to establish a baseline is study dependent. Depending on the timescale of the analysis, it is also useful to consider a sliding window approach, wherein new baselines are established during predefined time periods to account for baseline changes over time. The median value of the biosignal serves as a useful baseline value because it is less susceptible to noise and outliers compared to other statistical summary metrics and provides a way to amplify the SNR during the next steps of the analysis. The establishment of and comparisons to reliable baselines have been performed in multiple studies [50-53]. One limitation of this approach is that substantial monitoring time may be needed to establish a reliable baseline for an individual due to inherent biological and behavioral variability and the effects of external factors that may be difficult to control for (eg, seasonality, circadian rhythms, weekdays vs weekends, etc) [48]. It should also be noted that comparisons to a reliable baseline would not improve the SNR in scenarios where there is a compound effect of the source of noise and the conditions of measurement [54]. For example, skin tone may only increase measurement error for certain heart rate zones (eg, high heart rate) or under circumstances of high motion. In such cases, removing data collected under certain conditions that exacerbate measurement error may be the most appropriate approach.

Another way to handle the challenge of confounding sources of measurement error is to only compare segments of data that are measured under the same conditions (eg, similar movement types and heart rate zones) [15,18,20,24,55]. This technique allows researchers to further isolate confounding sources of measurement error that may be exacerbated during different activities. First, one must define specific wear times of interest, such as wear during specific times of the day, which helps account for circadian variability. Second, when available, researchers should use the "Activity Type" provided by Fitbit devices to segment heart rate data into comparable sections. Researchers can also leverage this activity information to anticipate activities for which heart rate data may be less accurate.

In circumstances where there are short periods of incorrectly reported heart rates or step counts (eg, during high-intensity motion), simple normalization methods, such as *z* score normalization and minimum-maximum normalization, are the most useful [56,57]. Minimum-maximum normalization is useful when extreme outliers are not present in the data, especially when the data have a fixed possible range. *z* score normalization is particularly useful because it centers and scales the Fitbit device data to a mean of 0 and an SD of 1. *z* score normalization helps to reduce the comparatively higher impact of outliers within shorter data segments because it leverages information from longer segments (ie, the mean and SD) for normalization. Once normalized, the data can be compared across wearable data types and participants.

Future Directions

As the AoURP continues efforts to provide wearable data to researchers and expand the scope of the Fitbit device data made available on the Researcher Workbench, there are several future directions to be considered. Although the ideas presented herein are tailored to Fitbit device data originating from the bring-your-own-device facet of AoURP, Fitbit device data are now actively being collected from other studies, and these data may one day be integrated into the Researcher Workbench [3]. Each additional study may have unique characteristics, including the target population, which may play a role in the overall quality of the data. For example, whether a Fitbit device was provided to participants or whether they were using an existing device they purchased may be a factor in a participant's comfort level with using the device properly and regularly. Understanding the nuances and potential variations in data quality arising from different study protocols and data sources within the AoURP ecosystem necessitates further research. Investigating how specific study designs, participant demographics, and data collection protocols within the AoURP may influence the overall quality of the collected data will be crucial for researchers seeking to derive meaningful insights and improve the designs of future studies that implement DHTs.

Although Fitbit device model data are not currently available on the Researcher Workbench, it is worth considering how differing device models, software, and firmware may affect the data collected. At the time of writing, the underlying PurePulse PPG technology is the same across all Fitbit LLC heart rate tracking devices [14]. The largest differences in accelerometer-derived data have been observed between Fitbit LLC's early torso clip-on trackers and the newer wrist-based devices [20]. Additional research is needed to investigate whether there are any substantial differences in accelerometry performance across Fitbit LLC wrist-based models. With regard to data derived from heart rate and accelerometry, such as sleep tracking data, prior to Fitbit LLC's release of heart rate tracking devices in 2014 [58], Fitbit LLC's early accelerometry-only devices estimated sleep metrics based on movement alone. Only Fitbit devices with heart rate tracking will include sleep staging, wake heart rate, and sleep-time heart rate [59]. Identifying whether sleep staging metrics are available for a particular individual may be a convenient way to identify the broad type of Fitbit device that was worn. The future incorporation of other contextual information, such as environmental factors, user behaviors, and device models, will enhance the ability to detect and mitigate noise, improve overall data quality, and provide a more comprehensive understanding of an individual's health.

Conclusion

The development and validation of Fitbit device–derived digital biomarkers offer the potential for remote and continuous measurement of physiological data. Such digital biomarkers can help inform medical decisions and predict disease states [7]. The wide adoption of DHTs by both consumers and programs like the AoURP make DHTs a great source of data for researchers. Researchers can use various analytical, statistical, and machine learning approaches to further develop DHT data into digital biomarkers [60-63]. Like with any technology, there are inherent limitations and sources of error that stakeholders (eg, researchers using DHT data in their analyses) should be aware of. We encourage the *All of Us* community to use data processing techniques that address noise and missingness to reduce problems downstream in the data analysis. Although we focused on heart rate and motion data in this work, the error mitigation methods described are applicable to other forms of wearable data, including sleep data. For example, changes in total sleep time and sleep stages can be compared against baselines over time. Researchers should consider their study goals and expected outcomes when determining which data cleaning strategies are the most salient to their goals.

Acknowledgments

This work was funded by the Data and Research Center (5U2COD023196-05). The *All of Us* Research Program would not be possible without the partnership of its participants. The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: regional medical centers (1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA: AOD21037, AOD22003, AOD16037, AOD21041), federally qualified health centers (HHSN 263201600085U), Data and Research Center (5 U2C OD023196), Biobank (1 U24 OD023121), The Participant Center (U24 OD023176), Participant Technology Systems Center (1 U24 OD023163), communications and engagement partners (3 OT2 OD023205; 3 OT2 OD023206), and community partners (1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276).

Data Availability

To ensure the privacy of participants, the *All of Us* Research Program data used for this study are available to approved researchers following registration, completion of ethics training, and attestation of a data use agreement through the *All of Us* Research Workbench platform.

Conflicts of Interest

None declared.

References

1. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
2. All of Us Research Program expands data collection efforts with Fitbit. All of Us Research Program. 2019 Jan 16. URL: allofus.nih.gov/news-events/announcements/all-us-research-program-expands-data-collection-efforts-fitbit [accessed 2023-07-09]
3. Through ‘All of Us’ program, Scripps Research launches wearable technology study to accelerate precision medicine. Scripps Research. 2021 Feb 24. URL: www.scripps.edu/news-and-events/press-room/2021/20210224-aou-fitbit-study.html [accessed 2023-07-09]
4. Researcher workbench. All of Us Research Hub. URL: www.researchallofus.org/data-tools/workbench/ [accessed 2023-07-09]
5. Master H, Kouame A, Hollis H, Marginean K, Rodriguez K. 2022Q4R9 v7 data characterization report. All of Us Research Program. 2023. URL: support.researchallofus.org/hc/en-us/articles/14558858196628-2022Q4R9-v7-Data-Characterization-Report [accessed 2023-08-12]
6. Fitbit data. All of Us Research Hub. 2023. URL: databrowser.researchallofus.org/fitbit [accessed 2023-07-09]
7. Motahari-Nezhad H, Fgaier M, Abid MM, Péntek M, Gulácsi L, Zrubka Z. Digital biomarker–based studies scoping review of systematic reviews. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35722. [doi: 10.2196/35722] [Medline: 36279171]
8. Curry D. Fitbit revenue and usage statistics (2023). Business of Apps. 2023. URL: www.businessofapps.com/data/fitbit-statistics/ [accessed 2023-07-09]
9. Wearables market share companies 2023. Statista. 2023. URL: www.statista.com/statistics/435944/quarterly-wearables-shipments-worldwide-market-share-by-vendor/ [accessed 2023-07-09]
10. Holko M, Litwin TR, Munoz F, Theisz KI, Salgin L, Jenks NP, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022 Apr 25;5(1):53. [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]
11. General info and specifications. Fitbit Sense User Manual. 2023. URL: help.fitbit.com/manuals/sense/Content/manuals/html/General%20Info%20and%20Specifications.htm [accessed 2023-07-09]
12. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med 2020 Apr 14;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
13. Woolley SI, Collins T, Mitchell J, Fredericks D. Investigation of wearable health tracker version updates. BMJ Health Care Inform 2019 Oct;26(1):e100083. [doi: 10.1136/bmjhci-2019-100083] [Medline: 31597642]
14. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
15. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018 Feb 28;13(2):e0192691. [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
16. Pollreisz D, TaheriNejad N. Detection and removal of motion artifacts in PPG signals. Mobile Networks and Applications 2022;27:728-738. [doi: 10.1007/s11036-019-01323-6]
17. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR wireless heart rate monitor: validation study conducted under free-living conditions. JMIR Mhealth Uhealth 2017 Oct 20;5(10):e157. [doi: 10.2196/mhealth.8233] [Medline: 29055881]
18. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020 Feb 10;3(1):18. [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
19. How does my Fitbit device calculate my daily activity? Fitbit Help. URL: help.fitbit.com/articles/en_US/Help_article/1141.htm [accessed 2023-07-09]
20. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 9;6(8):e10527. [doi: 10.2196/10527] [Medline: 30093371]
21. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159. [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

22. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
23. Jung HC, Kang M, Lee NH, Jeon S, Lee S. Impact of placement of Fitbit HR under laboratory and free-living conditions. Sustainability 2020;12(16):6306. [doi: 10.3390/su12166306]
24. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016 Sep 2;11(9):e0161224. [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
25. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile device accuracy for step counting across age groups. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e88. [doi: 10.2196/mhealth.7870] [Medline: 28659255]
26. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019 Jun 19;7(6):e13084. [doi: 10.2196/13084] [Medline: 31219048]
27. Colvonen PJ. Response to: investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2021 Feb 26;4(1):38. [doi: 10.1038/s41746-021-00408-5] [Medline: 33637822]
28. Bent B, Enache OM, Goldstein B, Kibbe W, Dunn JP. Reply: matters arising 'Investigating sources of inaccuracy in wearable optical heart rate sensors'. NPJ Digit Med 2021 Feb 26;4(1):39. [doi: 10.1038/s41746-021-00409-4] [Medline: 33637842]
29. Lee J, Matsumura K, Yamakoshi K, Rolfe P, Tanaka S, Yamakoshi T. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:1724-1727. [doi: 10.1109/EMBC.2013.6609852] [Medline: 24110039]
30. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3. [doi: 10.3390/jpm7020003] [Medline: 28538708]
31. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478. [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
32. US patent for multi-channel photoplethysmography sensor patent (patent # 11,633,117). Justia Patents. 2019 Oct 3. URL: patents.justia.com/patent/11633117 [accessed 2023-07-09]
33. Li M, Kim YT. Design of a wireless sensor system with the algorithms of heart rate and agility index for athlete evaluation. Sensors (Basel) 2017 Oct 17;17(10):2373. [doi: 10.3390/s17102373] [Medline: 29039763]
34. Weiler DT, Villajuan SO, Edkins L, Cleary S, Saleem JJ. Wearable heart rate monitor technology accuracy in research: a comparative study between PPG and ECG technology. Proc Hum Factors Ergon Soc Annu Meet 2017;61(1):1292-1296. [doi: 10.1177/1541931213601804]
35. Wright SP, Brown TSH, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 1;312(3):R358-R367. [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
36. Düking P, Fuss FK, Holmberg HC, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102. [doi: 10.2196/mhealth.9341] [Medline: 29712629]
37. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022 Apr;160:104696. [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
38. Mitra R, McGough SF, Chakraborti T, Holmes C, Copping R, Hagenbuch N, et al. Learning from data with structured missingness. Nat Mach Intell 2023;5(1):13-23. [doi: 10.1038/s42256-022-00596-z]
39. Yamamoto S, editor. Human Interface and the Management of Information: Information and Interaction Design, 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I: Springer; 2013. [doi: 10.1007/978-3-642-39209-2]
40. Jiang Y, Spies C, Magin J, Bhosai SJ, Snyder L, Dunn J. Investigating the accuracy of blood oxygen saturation measurements in common consumer smartwatches. PLOS Digit Health 2023 Jul 12;2(7):e0000296. [doi: 10.1371/journal.pdig.0000296] [Medline: 37437005]
41. Constantinou V, Felber AE, Chan JL. Applicability of consumer activity monitor data in marathon events: an exploratory study. J Med Eng Technol 2017 Oct;41(7):534-540. [doi: 10.1080/03091902.2017.1366560] [Medline: 28954563]
42. Cleland I, Donnelly MP, Nugent CD, Hallberg J, Espinilla M, Garcia-Constantino M. Collection of a diverse, realistic and annotated dataset for wearable activity recognition. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops: Institute of Electrical and Electronics Engineers; 2018:555-560. [doi: 10.1109/PERCOMW.2018.8480322]
43. Di J, Demanuele C, Kettermann A, Karahanoglu FI, Cappelleri JC, Potter A, et al. Considerations to address missing data when deriving clinical trial endpoints from digital health technologies. Contemp Clin Trials 2022 Feb;113:106661. [doi: 10.1016/j.cct.2021.106661] [Medline: 34954098]

44. Demanuele C, Lokker C, Jhaveri K, Georgiev P, Sezgin E, Geoghegan C, et al. Considerations for conducting bring your own "device" (BYOD) clinical studies. Digit Biomark 2022 Jul 4;6(2):47-60. [doi: 10.1159/000525080] [Medline: 35949223]
45. Liu S, Han J, Puyal EL, Kontaxis S, Sun S, Locatelli P, et al. Fitbeat: COVID-19 estimation based on wristband heart rate using a contrastive convolutional auto-encoder. Pattern Recognit 2022 Mar;123:108403. [doi: 10.1016/j.patcog.2021.108403] [Medline: 34720200]
46. Feng T, Narayanan S. Imputing missing data in large-scale multivariate biomedical wearable recordings using bidirectional recurrent neural networks with temporal activation regularization. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:2529-2534. [doi: 10.1109/EMBC.2019.8856966] [Medline: 31946412]
47. Lin S, Wu X, Martinez G, Chawla NV. Filling missing values on wearable-sensory time series data. In: Proceedings of the 2020 SIAM International Conference on Data Mining (SDM): Society for Industrial and Applied Mathematics; 2020:46-54. [doi: 10.1137/1.9781611976236]
48. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
49. Standard score - understanding Z-scores and how to use them in calculations. Laerd Statistics. URL: statistics.laerd.com/statistical-guides/standard-score.php [accessed 2023-07-09]
50. Shandhi MMH, Cho PJ, Roghanizad AR, Singh K, Wang W, Enache OM, et al. A method for intelligent allocation of diagnostic testing by leveraging data from commercial wearable devices: a case study on COVID-19. NPJ Digit Med 2022 Sep 1;5(1):130. [doi: 10.1038/s41746-022-00672-z] [Medline: 36050372]
51. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34. [doi: 10.2196/mhealth.7043] [Medline: 28302596]
52. Viboud C, Santillana M. Fitbit-informed influenza forecasts. Lancet Digit Health 2020 Feb;2(2):e54-e55. [doi: 10.1016/S2589-7500(19)30241-9] [Medline: 33334559]
53. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. Nat Med 2022 Nov;28(11):2301-2308. [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
54. Louie A, Feiner JR, Bickler PE, Rhodes L, Bernstein M, Lucero J. Four types of pulse oximeters accurately detect hypoxia during low perfusion and motion. Anesthesiology 2018 Mar;128(3):520-530. [doi: 10.1097/ALN.0000000000002002] [Medline: 29200008]
55. Stein PK, Kleiger RE. Insights from the study of heart rate variability. Annu Rev Med 1999;50:249-261. [doi: 10.1146/annurev.med.50.1.249] [Medline: 10073276]
56. Böttcher S, Vieluf S, Bruno E, Joseph B, Epitashvili N, Biondi A, et al. Data quality evaluation in wearable monitoring. Sci Rep 2022 Dec 10;12(1):21412. [doi: 10.1038/s41598-022-25949-x] [Medline: 36496546]
57. Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recognit 2005 Dec;38(12):2270-2285. [doi: 10.1016/j.patcog.2005.01.012]
58. Burns M. Fitbit's latest activity trackers feature heart monitoring, smartwatch functions. TechCrunch. 2014 Oct 27. URL: techcrunch.com/2014/10/27/fitbits-latest-activity-trackers-feature-heartheart-monitoring-smartwatch-functions/ [accessed 2023-07-11]
59. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273. [doi: 10.2196/16273] [Medline: 31778122]
60. Husom EJ, Dautov R, Videsjorden AN, Gonidis F, Papatzelos S, Malamas N. Machine learning for fatigue detection using Fitbit fitness trackers. In: Capelli C, Verhagen E, Pezarat-Correia P, Vilas-Boas J, Cabri J, editors. Proceedings of the 10th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2022): SciTePress; 2022:41-52. [doi: 10.5220/0011527500003321]
61. Fuller D, Anaraki JR, Simango B, Rayner M, Dorani F, Bozorgi A, et al. Predicting lying, sitting, walking and running using Apple Watch and Fitbit data. BMJ Open Sport Exerc Med 2021 Apr 8;7(1):e001004. [doi: 10.1136/bmjsem-2020-001004] [Medline: 33907628]
62. Gavhane A, Kokkula G, Pandya I, Devadkar K. Prediction of heart disease using machine learning. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA): Institute of Electrical and Electronics Engineers; 2018:1275-1278. [doi: 10.1109/ICECA.2018.8474922]
63. Faust L, Purta R, Hachen D, Striegel A, Poellabauer C, Lizardo O, et al. Exploring compliance: observations from a large scale Fitbit study. In: SocialSens'17: Proceedings of the 2nd International Workshop on Social Sensing: Association for Computing Machinery; 2017:55-60. [doi: 10.1145/3055601.3055608]

Abbreviations

AoURP: *All of Us* Research Program
BYOD: Bring-Your-Own-Device

DHT: digital health technology
MAR: missing at random
MCAR: missing completely at random
MNAR: missing not at random
PPG: photoplethysmography
SNR: signal to noise ratio
SpO_2: saturation of peripheral oxygen

Edited by L Buis; submitted 16.12.22; peer-reviewed by A Angelucci, H Motahari-Nezhad, V Ukachukwu, W Zhou; revised version received 14.08.23; accepted 08.09.23; published 03.11.23.

Please cite as:
Lederer L, Breton A, Jeong H, Master H, Roghanizad AR, Dunn J
The Importance of Data Quality Control in Using Fitbit Device Data From the All of Us Research Program
JMIR Mhealth Uhealth 2023;11:e45103
URL: https://mhealth.jmir.org/2023/1/e45103
doi:10.2196/45103

Viewpoint

Mental Health in Urban Environments: Uncovering the Black Box of Person-Place Interactions Requires Interdisciplinary Approaches

Martina Kanning[1], Prof Dr; Li Yi[2], Phd; Chih-Hsiang Yang[3], PhD; Christina Niermann[4], PhD; Stefan Fina[5], Prof Dr

[1]Department of Sport Science, University of Konstanz, Konstanz, Germany

[2]Department of Nutrition, Harvard TH Chan School of Public Health, Boston, MA, United States

[3]Department of Exercise Science and TecHealth Center, University of South Carolina, Columbia, SC, United States

[4]Institute of Interdisciplinary Exercise Science and Sports Medicine, Medical School Hamburg, Hamburg, Germany

[5]Faculty of Architecture and Civil Engineering, University of Applied Sciences Augsburg, Augsburg, Germany

Corresponding Author:
Martina Kanning, Prof Dr
Department of Sport Science
University of Konstanz
Univeristätsstraße 10
Konstanz, 78464
Germany
Phone: 49 7531 88 3154
Fax: 49 7531 88 3154
Email: martina.kanning@uni-konstanz.de

Abstract

Living in urban environments affects individuals' mental health through different pathways. For instance, physical activity and social participation are seen as mediators. However, aiming to understand underlying mechanisms, it is necessary to consider that the individual is interacting with its environment. In this regard, this viewpoint discusses how urban health research benefits from integration of socioecological and interdisciplinary perspectives, combined with innovative ambulatory data assessments that enable researchers to integrate different data sources. It is stated that neither focusing on the objective and accurate assessment of the environment (from the perspective of spatial sciences) nor focusing on subjectively measured individual variables (from the public health as well as a psychosocial perspective) alone is suitable to further develop the field. Addressing person-place interactions requires an interdisciplinary view on the level of theory (eg, which variables should be focused on?), assessment methods (eg, combination of time-varying objective and subjective measures), as well as data analysis and interpretation. Firstly, this viewpoint gives an overview on previous findings addressing the relationship of environmental characteristics to physical activity and mental health outcomes. We emphasize the need for approaches that allow us to appropriately assess the real-time interaction between a person and a specific environment and examine within-subject associations. This requires the assessment of environmental features, the spatial-temporal behavior of the individual, and the subjective experiences of the situation together with other individual factors, such as momentary affective states. Therefore, we finally focused on triggered study designs as an innovative ambulatory data assessment approach that allows us to capture real-time data in predefined situations (eg, while walking through a specific urban area).

(JMIR Mhealth Uhealth 2023;11:e41345) doi:10.2196/41345

KEYWORDS

physical activity; urban health; ambulatory assessment; environment; mental health; real-time data; within-subject association

Introduction

Globally, 76% of people live in cities, and a growing number of people are expected to move into urban surroundings within the next two decades [1]. Although urban settings are frequently associated with locational advantages (eg, proximity to job and educational opportunities, cultural diversity, as well as service and infrastructure provision), they are also shown by empirical evidence to increase risks for psychological stress and mental disorders among their residents [2]. Even though mental health has complex determinants, theoretical assumptions about urban health and empirical evidence suggest that increasing physical activity levels and social interactions improve mental health (eg, well-being, quality of life, and satisfaction with life) in urban residents [3,4]. Recommendations for creating health-promoting urban environments include the vision of active lifestyles and opportunities for urban residents to participate in social life supported by suitable infrastructure (ie, walkable neighborhoods and people-oriented urban designs). For instance, there is extensive research in the field of walkability in urban areas and its association with health. However, we still have only preliminary insights into people's reactions to specific environmental features (eg, greenness and noise); we also have limited insight as to for whom and under which conditions such experiences result in more physical activity (eg, active mobility), social interactions, and improved mental health. For these reciprocal associations between the person and its environment, we use the term "person-place interactions," which former studies were not able to address adequately due to methodological constraints. This shortcoming can be overcome with innovative study designs that allow us to address within-subject relations, and therefore, allow for studying the underlying mechanisms of people's health behavior and mental health after they have been exposed to different environmental features [5]. In addition, such study designs provide the opportunity to draw from existing and newly emerging data sources and data fusion techniques of different disciplines (eg, spatial science, sport and movement science, and psychology) to get a deeper understanding of person-place interactions.

Why Do We Need Information About Person-Place Interactions to Create Health-Promoting Urban Environments?

According to socioecological approaches, individuals interact with their physical and natural environment as well as the social neighborhood setting, which affects health-related behaviors such as physical activity [6]. According to Stokols [7], this interaction can be described as "cycles of mutual influences"—the environmental features of a local neighborhood are associated with urban residents' behavior and health. Reciprocally, individuals live and act within these settings and engage with environmental features in a "more or less" health-enhancing way. For instance, on the one hand, design features of the built environment as well as opportunities to get active in social settings can stimulate physical activity (eg, creating attractive stairs in urban environments). On the other hand, a person who is highly motivated to improve activity levels, experiences and perceives the environment differently than a person who is less motivated, and therefore, acts differently based on these subjective experiences and perceptions (eg, using such stairs not only for stair climbing but also for a workout or to do parkour). Furthermore, according to socioecological approaches, individuals' behavior is affected by more than just the individual level (eg, motivation, self-efficacy, habits, and personal physiological constitution) and the perceived environment (eg, attractive stairs, perceptions of urban green or blue, and noise) but also by the sociocultural factors, factors arising from the built and the social environment, as well as policy factors [8]. Therefore, learning more about person-place interactions from an interdisciplinary perspective—integrating knowledge from spatial science, psychology, sport science, transport systems, politics, and sociology—is a precondition for creating health-promoting environments. For example, an urban health policy to add cycling lanes to promote physical activity levels may be less effective, if the fit between these environmental features (eg, cycling lanes) and the target groups' needs, preferences, social-cognitive constructs (eg, attitudes), and sociodemographic backgrounds (eg, age and proportion of bike owners) are not considered. As Stokols [9] emphasized, this fit serves as an important predictor of health and well-being.

Evidence About the Associations Between the Environment, Physical Activity, and Mental Health

Considering current studies, key findings confirm that environmental features are associated with physical activity and health. An analysis of previous systematic reviews and meta-studies [10] summarizes associations between built environmental features, dietary intake, physical activity, and obesity. More than half of the included reviews focused on physical activity (n=46) and reported consistent evidence about the positive associations between walkability and physical activity (supported by 83% of the reviews), followed by positive associations between access to recreational facilities, shops and services, and parks or trails and physical activity (supported by 63% to 70% of the reviews). Another systematic review of longitudinal studies (N=36) about the effects of the built environment on adults' physical activity came to similar conclusions: new infrastructures for walking, cycling, and using public transport increase overall physical activity [11]. Further reviews, especially in the past few years, support such results for older adults [12] and children [13].

In terms of mental health impacts, an overview of systematic reviews assessed the association between the built environment and different mental health indicators (eg, well-being, depression, and stress) [14]. The authors included 11 reviews and reported insufficient and heterogeneous evidence for health-enhancing effects of the environment, with a critically low methodological study quality of 80% of the included reviews. Another meta-narrative review synthesizes the impacts of urban green space on different indicators of mental health from 38 intervention studies [15]. The results were discussed in an international World Health Organization expert panel workshop, concluding that urban green space interventions are often multifaceted but can generally be categorized in 4 groups: park-based, greenways and trails, urban greening (eg, street trees), and green built interventions (eg, green roofs). Most studies in that meta-narrative review were designed as natural experiments, and the findings showed strong evidence for park-based as well as greenway and trail interventions to promote health and well-being through increased park use and physical activity [15].

Limitations of the Empirical Evidence When Analyzing Person-Place Interactions

The aforementioned reviews provide evidence about the relationships between specific environmental features and mental health or physical activity. However, they fall short of increasing our knowledge of the time-varying associations within subjects regarding how urban residents react to specific environmental features and under which conditions such experiences result in more physical activity and improved mental health. One reason for this is that these reviews focused mainly on environmental characteristics, such as accessibility or the amount or quality of greenness, and how these characteristics moderate the relationship between the environment and physical activity or mental health. They do not provide evidence about individuals' momentary perception, experience, and subsequent behavioral, cognitive, or affective states; nor do they show how these states are related to mental health.

Kwan [16] already criticized in 2009 that spatial research about associations between environmental features and physical activity or health mainly used a "place-based" instead of a "person-based" approach and operationalized environment exposure mostly by focusing on spatial units, such as census tracts, buffer zones, or postal codes. Such a "place-based" design neglects that individuals move around and do not stay in their "home spatial unit" during their daily activities (eg, workplace, school, and leisure activities) [17].

More than 10 years later, Zhang and colleagues [18] stated that there are still only a few studies investigating the association between environment and health from the perspective of the spatial-temporal behavior of the individual. According to the results of their survey with 1003 Chinese adults, there are significant differences between environmental exposures of individuals based on home buffer zones (ie, place-based) compared to time-weighted activity travel buffers (ie, person-based) [18].

A currently published scoping review [19] also stated that person-based approaches still are in their infancy. The review is about methodological approaches to measure the spatial contexts used in socioecological physical activity research, and the included studies have been mostly published within the last 7 years. In sum, person-based spatial methods have been used rarely; only 2% (10/412) of the included studies used activity spaces, and similarly, only 2% (8/412) of the studies buffered multiple points to capture the environment. Almost a third of the studies (118/412) used place-based approaches (eg, with administrative units) as an objective approach.

Furthermore, place-based approaches do not take into consideration that individuals have different lifestyles, psychosocial characteristics, and daily routines and may react differently to influences of similar environmental features. For instance, even persons living within the same building would perceive environmental exposure of their neighborhood differently [20]. That refers to the "uncertain geographic context problem" [21], and it also highlights the importance of interdisciplinary efforts by psychologists, sport scientists, geographers, and computer scientists [7,22]. Nevertheless, empirical evidence of interdisciplinary studies integrating assessment methods of spatial science and urban planning as well as social and health sciences is still lacking [23]. Further, these approaches do not allow detailed analysis; for example, to what extent social interaction could moderate the associations between the environment and mental health or for whom the quality of greenness may be relevant. They also neglect that the environmental exposure is not only directly associated with mental health but also via affecting health-related behaviors, such as physical activity. Thus, mediators or moderators of the associations between the environment and physical activity and health have hardly been examined so far [10]; a framework to explore relationships between place and mental health by combining GPS, Geographic Information System (GIS), and accelerometer data is available in a previous study [24].

Ambulatory Assessment Approaches Are Suitable to Analyze Person-Place Interactions

To advance our understanding of person-place interactions of urban residents in everyday life, we need more studies that collect intensive longitudinal data, which facilitates the estimation of time-varying associations between environmental features, individuals' behavior, as well as their momentary experiences. Ambulatory assessments are suitable approaches for addressing such within-subject relations because they allow us to monitor physical activity (eg, via accelerometry), physiological function (eg, heart rate or electrodermal activity), and environmental parameters (eg, via geolocation tracking) in real time.

In 2018, Chaix [23] published an overview of different wearable sensors and devices to capture the environment (eg, air pollution and the number of mobile phones nearby), the behavior (eg, physical activity and GPS receivers), and individuals' physiology (eg, heart rate and electrodermal activity). He recommended integrating different sensors to generate knowledge of healthy places and situations. Such an approach allows us to assess the duration, sequences, and accumulation of different environmental exposures; it also provides rich research possibilities to assess in situ changes in mental health according to different environments or environmental features [25].

A current example of combining different sensors is the study by Marquet and colleagues [26], which combined accelerometers to assess physical activity and GPS data that was linked to spatial data sets on walkability and greenness. They found that persons with high walkability and greenness in their activity spaces had higher levels of moderate-to-vigorous physical activity. In addition, a recent study [27] combined sensors that measure black carbon concentration with a sensor that assessed galvanic skin response (as a proxy measure of stress) and a GPS device and found that increases in black carbon are related to higher skin responses (indicating higher stress levels) during active travel. Green space and a good active travel infrastructure are associated with lower skin responses while walking or cycling.

To deepen our understanding of how different individuals react to specific environmental features, it is crucial to assess the above-mentioned environmental parameters and physiological functions using sensors in an objective way; however, it is also relevant to assess subjective experiences, preferably at that moment when an association between a person and a place is assumed. It is possible to schedule e-diaries throughout the day (eg, ecological momentary assessment) to assess different psychosocial constructs (eg, momentary affective states, momentary experiences of social interactions, and momentary motivation) via self-reported measures [28,29].

Ambulatory assessment approaches have already been applied in spatial research. Perchoux et al [30,31] introduce activity spaces as an individualized measure for environmental exposure. It considers individuals' daily mobility patterns, including major spatial-temporal cluster movements between home and different daily locations, and characterizes its temporal structure (ie, frequency, regularity, and duration). Further, to match person- and place-based data, some research groups combined the assessment of activity spaces with multiple self-reports per day via e-diaries (eg, feelings, emotions, and evaluations of their environment) resulting in geographically explicit ecological momentary assessments (GEMAs) [5,20]. GEMA studies implement innovative study designs that use mobile geographic location technologies to capture participants' activity space, which can then be used to assess the dynamic environmental exposure via GIS. e-Diaries allow us to capture subjective experiences in situ, and these data can be linked to participants' current position in time and space. However, GEMA struggles with the disadvantages of "time-based" assessments because prompts to answer the web-based questionnaires were usually triggered at random time intervals. With such a sampling scheme, self-reports during rare events (eg, being physically active in the neighborhood) are likely to be missed and could therefore hardly be used for analyzing time-varying associations. For instance, a GEMA study [32] assessed how urban green space is associated with stress in adolescents living in urban surroundings. Outdoor behaviors were assessed via GPS-enabled mobile phones. To capture momentary experiences of stress, participants received randomly 3-6 text messages throughout the day, including a link to a web-based questionnaire. However, 72% of the web-based questionnaires have been filled in at home and not during outdoor behaviors. To ask participants to report retrospectively about their feelings, experiences, and thoughts in response to specific situations is likely to increase recall bias, for instance [29]. Furthermore, spatial accuracy of GEMA is an issue and should be taken into account when analyzing and interpreting the data in urban settings when the GPS device signals are likely to be interrupted, such as in streets with dense tree covers [33].

Methodological Improvements and Future Directions of Research

Answers to research questions, such as the following, are crucial to inform initiatives aiming to create health-promoting urban environments: "How does mental health and physical activity vary due to the momentary exposure to specific physical (eg, streetscape greenery and noise) or social environmental features (eg, crime or places enabling social interactions)?" and "How are these time-varying associations moderated by personal factors, which could be time-invariant, such as lifestyle, attitudes, socioeconomic status, obesity, and gender, as well as dynamic such as momentary feelings, experiences, motivations, and thoughts?"

To address these questions, we need data of within-subject relations during predefined situations in everyday life in which a contextual effect is assumed (eg, being physically active within a neighborhood). The above-mentioned GEMA approach is an appropriate design to provide first answers. However, to capture data during the "right" situations, it could be extended by (1) assessing physical activity directly via accelerometers and (2) using triggered assessments. Triggered assessments allow capturing data in predefined situations and have already been applied in other fields, such as examining time-sensitive associations between physical activity and affective states [34] or the time-sensitive assessment of contextual factors during episodes of prolonged sedentary bouts [35]. A recent example of a triggered e-diary regarding outdoor activities was presented in a study protocol with older residents of Paris [36]. The study combined GEMA with a GPS receiver and used this novel methodology to initiate e-diaries when participants were outdoors. Another study used momentary physical activity levels (assessed via accelerometry) and locations by mobile phone positioning services (eg, GPS and transmission tower) to identify outdoor activities. A trigger algorithm was used to start an e-diary whenever movement acceleration exceeds a certain threshold and participant's locations were identified as outside the home [37]. The study included 46 middle-aged adults and showed that momentary affective states varied significantly due to different social (intensity of social interaction) and physical (amount of greenness) environments. The accuracy of the walking trigger has been examined in a previous study [38].

Furthermore, activity data could be integrated into a GIS to combine information of the physical environment of the activity spaces with movement data. Advanced GISs work with time-enabled spatial analysis functions to track movements with so-called "event-based feature classes." The challenge is to provide data on the actual exposure for the time of measurement. In this context, live sensor networks are preferable over archived data but only starting to become available, for example, in Smart City Sensor Observation Networks [39,40]. Through combining different data (eg, subjective experiences via self-reports and physical activity levels and physiological functions via specific sensors) in outdoor situations with exposures to different environmental features, we would be able to investigate the associations between specific uses of the environment (eg, walking, social interactions, and doing sports) and momentary experiences (eg, reduced stress and better feelings) and how the characteristics of the environment, the living conditions, or psychological factors moderate these associations.

Despite its promising future perspectives, these kinds of person-based spatial approaches lead to several methodological challenges concerning data processing, the linking of spatial and contextual exposures to individuals, special analytical and statistical methods, and ethical aspects of participants' privacy and security [41,42]. The solution to these challenges calls for the assembly of an interdisciplinary research team, which itself might also be challenging. However, this approach would enable us to take a broader perspective on this phenomenon and get closer to draw a "bigger picture" of person-place interactions during the everyday life of urban residents.

Conclusions

Urbanization with its advantages and disadvantages concerning health is on the rise. This viewpoint paper highlights the importance of gaining knowledge regarding the effect of urban environments on people's mental health by considering socioecological and interdisciplinary perspectives in combination with triggered ambulatory data assessments. It is crucial to assess time-varying associations to investigate person-place interactions between the individual and physical, social, and contextual features. Progress in technology and methodological advances enables researchers to study in more detail how people react to specific environmental features and which situational or personal factors may moderate these associations. Lastly, combining data and knowledge of different disciplines would deepen our understanding about the person-place interactions, which is crucial to create health-promoting urban environments.

Conflicts of Interest

None declared.

References

1. European Commission. Atlas of the Human Planet 2019.: Publications Office; 2020. URL: https://ghsl.jrc.ec.europa.eu/documents/Atlas_of_the_Human_Planet_2019_executivesummary.pdf?t=1576061612 [accessed 2023-04-06]
2. Vassos E, Pedersen C, Murray R, Collier DA, Lewis CM. Meta-analysis of the association of urbanicity with schizophrenia. Schizophr Bull 2012 Nov;38(6):1118-1123 [FREE Full text] [doi: 10.1093/schbul/sbs096] [Medline: 23015685]
3. Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. Annu Rev Public Health 2014 Mar 18;35(1):207-228. [doi: 10.1146/annurev-publhealth-032013-182443] [Medline: 24387090]
4. Lachowycz K, Jones A. Towards a better understanding of the relationship between greenspace and health: Development of a theoretical framework. Landsc Urban Plan 2013 Oct;118:62-69. [doi: 10.1016/j.landurbplan.2012.10.012]
5. Chaix B. How daily environments and situations shape behaviors and health: Momentary studies of mobile sensing and smartphone survey data. Health Place 2020 Jan;61:102241. [doi: 10.1016/j.healthplace.2019.102241] [Medline: 31784331]
6. Golden SD, Earp JAL. Social ecological approaches to individuals and their contexts: twenty years of health education and behavior health promotion interventions. Health Educ Behav 2012 Jun 20;39(3):364-372. [doi: 10.1177/1090198111418634] [Medline: 22267868]
7. Stokols D. Establishing and maintaining healthy environments: Toward a social ecology of health promotion. Am Psychol 1992;47(1):6-22. [doi: 10.1037/0003-066x.47.1.6]
8. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. Annu Rev Public Health 2006 Apr;27(1):297-322. [doi: 10.1146/annurev.publhealth.27.021405.102100] [Medline: 16533119]
9. Stokols D. Translating social ecological theory into guidelines for community health promotion. Am J Health Promot 1996 Aug 26;10(4):282-298. [doi: 10.4278/0890-1171-10.4.282] [Medline: 10159709]
10. Dixon B, Ugwoaba U, Brockmann A, Ross KM. Associations between the built environment and dietary intake, physical activity, and obesity: a scoping review of reviews. Obes Rev 2021 Apr;22(4):e13171 [FREE Full text] [doi: 10.1111/obr.13171] [Medline: 33369097]
11. Tcymbal A, Demetriou Y, Kelso A, Wolbring L, Wunsch K, Wäsche H, et al. Effects of the built environment on physical activity: a systematic review of longitudinal studies taking sex/gender into account. Environ Health Prev Med 2020 Nov 27;25(1):75 [FREE Full text] [doi: 10.1186/s12199-020-00915-z] [Medline: 33246405]
12. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E, Council on EnvironmentPhysical Activity (CEPA) – Older Adults working group. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 07;14(1):103 [FREE Full text] [doi: 10.1186/s12966-017-0558-z] [Medline: 28784183]
13. Davison K, Lawson CT. Do attributes in the physical environment influence children's physical activity? A review of the literature. Int J Behav Nutr Phys Act 2006 Jul 27;3(1):19 [FREE Full text] [doi: 10.1186/1479-5868-3-19] [Medline: 16872543]
14. Núñez-González S, Delgado-Ron JA, Gault C, Lara-Vinueza A, Calle-Celi D, Porreca R, et al. Overview of "systematic reviews" of the built environment's effects on mental health. J Environ Public Health 2020 Mar 19;2020:9523127-9523110 [FREE Full text] [doi: 10.1155/2020/9523127] [Medline: 32256618]

15. Hunter R, Cleland C, Cleary A, Droomers M, Wheeler B, Sinnett D, et al. Environmental, health, wellbeing, social and equity effects of urban green space interventions: a meta-narrative evidence synthesis. Environ Int 2019 Sep;130:104923 [FREE Full text] [doi: 10.1016/j.envint.2019.104923] [Medline: 31228780]
16. Kwan M. From place-based to people-based exposure measures. Soc Sci Med 2009 Nov;69(9):1311-1313. [doi: 10.1016/j.socscimed.2009.07.013] [Medline: 19665828]
17. Yi L, Wilson J, Mason T, Habre R, Wang S, Dunton GF. Methodologies for assessing contextual exposure to the built environment in physical activity studies: A systematic review. Health Place 2019 Nov;60:102226 [FREE Full text] [doi: 10.1016/j.healthplace.2019.102226] [Medline: 31797771]
18. Zhang L, Zhou S, Kwan M, Shen M. Assessing individual environmental exposure derived from the spatiotemporal behavior context and its impacts on mental health. Health Place 2021 Sep;71:102655. [doi: 10.1016/j.healthplace.2021.102655] [Medline: 34482159]
19. Rinne T, Kajosaari A, Söderholm M, Berg P, Pesola AJ, Smith M, et al. Delineating the geographic context of physical activities: A systematic search and scoping review of the methodological approaches used in social ecological research over two decades. Health Place 2022 Jan;73:102737 [FREE Full text] [doi: 10.1016/j.healthplace.2021.102737] [Medline: 34952474]
20. Kirchner TR, Shiffman S. Spatio-temporal determinants of mental health and well-being: advances in geographically-explicit ecological momentary assessment (GEMA). Soc Psychiatry Psychiatr Epidemiol 2016 Dec;51(9):1211-1223 [FREE Full text] [doi: 10.1007/s00127-016-1277-5] [Medline: 27558710]
21. Kwan M. The uncertain geographic context problem. Ann Assoc Am Geogr 2012 Sep;102(5):958-968. [doi: 10.1080/00045608.2012.687349]
22. John J, Haug V, Thiel A. Physical activity behavior from a transdisciplinary biopsychosocial perspective: a scoping review. Sports Med Open 2020 Oct 17;6(1):49 [FREE Full text] [doi: 10.1186/s40798-020-00279-2] [Medline: 33068179]
23. Chaix B. Mobile sensing in environmental health and neighborhood research. Annu Rev Public Health 2018 Apr 01;39(1):367-384 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-013731] [Medline: 29608869]
24. Jankowska MM, Schipperijn J, Kerr J. A framework for using GPS data in physical activity and sedentary behavior studies. Exerc Sport Sci Rev 2015 Jan;43(1):48-56 [FREE Full text] [doi: 10.1249/JES.0000000000000035] [Medline: 25390297]
25. Helbich M. Toward dynamic urban environmental exposure assessments in mental health research. Environ Res 2018 Feb;161:129-135 [FREE Full text] [doi: 10.1016/j.envres.2017.11.006] [Medline: 29136521]
26. Marquet O, Hirsch JA, Kerr J, Jankowska MM, Mitchell J, Hart JE, et al. GPS-based activity space exposure to greenness and walkability is associated with increased accelerometer-based physical activity. Environ Int 2022 Jul;165:107317 [FREE Full text] [doi: 10.1016/j.envint.2022.107317] [Medline: 35660954]
27. Yang X, Orjuela JP, McCoy E, Vich G, Anaya-Boig E, Avila-Palencia I, et al. The impact of black carbon (BC) on mode-specific galvanic skin response (GSR) as a measure of stress in urban environments. Environ Res 2022 Nov;214(Pt 4):114083 [FREE Full text] [doi: 10.1016/j.envres.2022.114083] [Medline: 35995220]
28. Reichert M, Giurgiu M, Koch ED, Wieland LM, Lautenbach S, Neubauer AB, et al. Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychol Sport Exerc 2020 Sep;50:101742 [FREE Full text] [doi: 10.1016/j.psychsport.2020.101742] [Medline: 32831643]
29. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013 Mar;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
30. Perchoux C, Chaix B, Cummins S, Kestens Y. Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. Health Place 2013 May;21:86-93. [doi: 10.1016/j.healthplace.2013.01.005] [Medline: 23454664]
31. Perchoux C, Chaix B, Kestens Y. Activity spaces in place and health research: novel exposure measures, data collection tools, and designs. Health Place 2019 Jul;58:102130. [doi: 10.1016/j.healthplace.2019.05.008] [Medline: 31377063]
32. Mennis J, Mason M, Ambrus A. Urban greenspace is associated with reduced psychological stress among adolescents: a Geographic Ecological Momentary Assessment (GEMA) analysis of activity space. Landsc Urban Plan 2018 Jun;174:1-9 [FREE Full text] [doi: 10.1016/j.landurbplan.2018.02.008] [Medline: 29576668]
33. Mennis J, Mason M, Ambrus A, Way T, Henry K. The spatial accuracy of geographic ecological momentary assessment (GEMA): Error and bias due to subject and environmental characteristics. Drug Alcohol Depend 2017 Sep 01;178:188-193. [doi: 10.1016/j.drugalcdep.2017.05.019] [Medline: 28654871]
34. Ebner-Priemer UW, Koudela S, Mutz G, Kanning M. Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. Front Psychol 2012;3:596 [FREE Full text] [doi: 10.3389/fpsyg.2012.00596] [Medline: 23355828]
35. Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M. Accuracy of sedentary behavior-triggered ecological momentary assessment for collecting contextual information: development and feasibility study. JMIR Mhealth Uhealth 2020 Sep 15;8(9):e17852 [FREE Full text] [doi: 10.2196/17852] [Medline: 32930668]
36. Fernandes A, Van Lenthe FJ, Vallée J, Sueur C, Chaix B. Linking physical and social environments with mental health in old age: a multisensor approach for continuous real-life ecological and emotional assessment. J Epidemiol Community Health 2021 May 04;75(5):477-483 [FREE Full text] [doi: 10.1136/jech-2020-214274] [Medline: 33148684]

37. Bollenbach L, Schmitz J, Niermann C, Kanning M. How do people feel while walking in the city? Using walking-triggered e-diaries to investigate the association of social interaction and environmental greenness during everyday life walking. Front Psychol 2022 Sep 26;13:970336 [FREE Full text] [doi: 10.3389/fpsyg.2022.970336] [Medline: 36225697]
38. Kanning M, Bollenbach L, Schmitz J, Niermann C, Fina S. Analyzing person-place interactions during walking episodes: innovative ambulatory assessment approach of walking-triggered e-diaries. JMIR Form Res 2022 Nov 25;6(11):e39322 [FREE Full text] [doi: 10.2196/39322] [Medline: 36427231]
39. Garcia Alvarez M, Morales J, Kraak M. Integration and exploitation of sensor data in smart cities through event-driven applications. Sensors (Basel) 2019 Mar 19;19(6):1372 [FREE Full text] [doi: 10.3390/s19061372] [Medline: 30893843]
40. Yang J, Jankowska M. Contextualizing space and time for GeoAI JITAIs (Just-in-Time Adaptive Interventions). In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery. 2019 Presented at: GeoAI 2019; Nov 5; Chicago, IL p. 66-68. [doi: 10.1145/3356471.3365239]
41. Hu H, Liu X, Zheng Y, He X, Hart J, James P, et al. Methodological challenges in spatial and contextual exposome-health studies. Crit Rev Environ Sci Technol 2022 Jul 04;53(7):827-846. [doi: 10.1080/10643389.2022.2093595]
42. James P, Jankowska M, Marx C, Hart JE, Berrigan D, Kerr J, et al. "Spatial Energetics": integrating data from GPS, accelerometry, and GIS to address obesity and inactivity. Am J Prev Med 2016 Nov;51(5):792-800 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.006] [Medline: 27528538]

Abbreviations

GEMA: geographically explicit ecological momentary assessment
GIS: geographic information system

Edited by L Buis; submitted 22.07.22; peer-reviewed by D Saleh, S Balula; comments to author 28.10.22; revised version received 20.12.22; accepted 06.03.23; published 11.05.23.

Please cite as:
Kanning M, Yi L, Yang CH, Niermann C, Fina S
Mental Health in Urban Environments: Uncovering the Black Box of Person-Place Interactions Requires Interdisciplinary Approaches
JMIR Mhealth Uhealth 2023;11:e41345
URL: https://mhealth.jmir.org/2023/1/e41345
doi:10.2196/41345
PMID:37166963

Digital Health Interventions to Enhance Tuberculosis Treatment Adherence: Scoping Review

Sol Lee[1,*], MPH; Vasuki Rajaguru[2,*], PhD; Joon Sang Baek[3], PhD; Jaeyong Shin[4,5], MD, MPH, PhD; Youngmok Park[5,6], MD

1

2

3

4

5

6

[*]these authors contributed equally

Corresponding Author:
Youngmok Park, MD

Abstract

Background: Digital health technologies are widely used for disease management, with their computing platforms, software, and sensors being used for health care. These technologies are developed to manage chronic diseases and infectious bacterial diseases, including tuberculosis (TB).

Objective: This study aims to comprehensively review the literature on the use of digital health interventions (DHIs) for enhancing TB treatment adherence and identify major strategies for their adoption.

Methods: We conducted a literature search in the PubMed, Cochrane Library, Ovid Embase, and Scopus databases for relevant studies published between January 2012 and March 2022. Studies that focused on web-based or mobile phone–based interventions, medication adherence, digital health, randomized controlled trials, digital interventions, or mobile health and ubiquitous health technology for TB treatment and related health outcomes were included.

Results: We identified 27 relevant studies and classified them according to the intervention method, a significant difference in treatment success, and health outcomes. The following interventions were emphasized: SMS text messaging interventions (8/27, 30%), medicine reminders (6/27, 22%), and web-based direct observation therapy (9/27, 33%). Digital health technology significantly promoted disease management among individuals and health care professionals. However, only a few studies addressed 2-way communication therapies, such as interactive SMS text messaging and feedback systems.

Conclusions: This scoping review classified studies on DHIs for patients with TB and demonstrated their potential for the self-management of TB. DHIs are still being developed, and evidence on the impact of digital technologies on enhancing TB treatment adherence remains limited. However, it is necessary to encourage patients' participation in TB treatment and self-management through bidirectional communication. We emphasize the importance of developing a communication system.

(JMIR mHealth and uHealth 2023;11:e49741) doi:10.2196/49741

KEYWORDS

tuberculosis; patient compliance; digital health; medication adherence; text messaging; mobile apps; application; medication; text; scoping review; disease management; chronic disease; communication; feedback; self-management; PRISMA

Introduction

Until the COVID-19 pandemic, tuberculosis (TB) was the leading cause of death from a single infectious disease, affecting approximately 10.6 million people in 2021 [1]. TB can be cured with appropriate medications; however, treatment adherence is affected by the complexity, tolerability, and long duration of the available regimens. Since low adherence increases the risk of poor treatment outcomes, several interventions have been attempted to enhance TB medication adherence [2].

Digital health interventions (DHIs) are promising for patient-centered care, as they allow for the remote monitoring of patients and can be used to conveniently remind patients to take their medications. Numerous studies have addressed how to enhance medication adherence during treatment by using mobile technologies, such as SMS text messaging [3], directly observed therapy (DOT) [3-5], video calls, phone call reminders [5,6], and web-based reports [3-7]. Studies have reported satisfaction [6-8], accuracy [6-8], acceptable uptake [5,7,8], improved drug adherence [3-5,7,9], higher rates of treatment success [5,7,8], and user acceptance [7-10] with regard to DHIs in TB management.

This review aims to summarize the existing literature on DHIs for TB treatment adherence, classify DHI techniques, identify the different types of interventions and their effects on treatment effectiveness, and evaluate adherence and health outcomes in TB treatment. This study reports on treatment outcomes, self-care management, follow-up, and the value of mobile-based communication activities that aim to improve TB treatment adherence.

Methods

We followed Arksey and O'Malley's [11] 5-stage scoping review framework, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [12], and the Joanna Briggs Institute protocol [13].

Identifying the Relevant Studies

We conducted a literature search in the PubMed, Cochrane Library, Ovid Embase, and Scopus databases for relevant studies published between January 2012 and March 2022. A comprehensive search strategy was developed to identify relevant studies, which included but was not confined to the following search string: *("Tuberculosis" OR "TB" OR "Tuberculosis infection") AND ("RCT" OR "Randomized controlled trial" OR "Experimental study") AND ("Behavior therapy" OR "Cognitive behavioral treatment" OR "Digital intervention" OR "Digital therapeutics" OR "App-based" OR "Web-based" OR "mHealth" OR "uHealth") AND ("treatment adherence" OR "medication adherence" OR "selfcare" OR "Management" OR "Persistence" OR "Compliance")*. The search terms and strategies are presented in Multimedia Appendix 1.

Eligibility and Exclusion Criteria

We included articles that met the following criteria: (1) published in peer-reviewed journals, (2) included TB treatment adherence and health outcomes as part of the study design, (3) written in English, (4) had full text available, and (5) published between January 2012 and March 2022. Studies were excluded if they were published before 2011 or did not focus on DHIs for TB. Reviews, case studies, reports, letters, conference proceedings, and abstract-only articles were also excluded.

Study Selection and Data Synthesis

Duplicates were eliminated from each database and recorded in the first stage. The second stage involved reviewing study titles and abstracts to ensure that articles were research studies that focused on digital health technology as a main intervention tool to improve the treatment adherence of patients with TB. The full texts of the articles were scrutinized in the last stage to verify whether they satisfied the key requirements.

Data were extracted by 1 reviewer (SL), and 2 independent reviewers (VR and YP) charted the data on different characteristics, including authors, publication year, country, study design, target population, number of participants, type of DHI, duration, follow-up, outcome measures, and major findings.

The retrieved data suggested that the core attributes of digital intervention strategies fell under the following three domains, which were based on the DHIs found in the selected articles: sending reminders via SMS text messages, monitoring progress, and tracking follow-ups for the self-management of TB treatment outcomes.

Quality Assessment and Risk of Bias

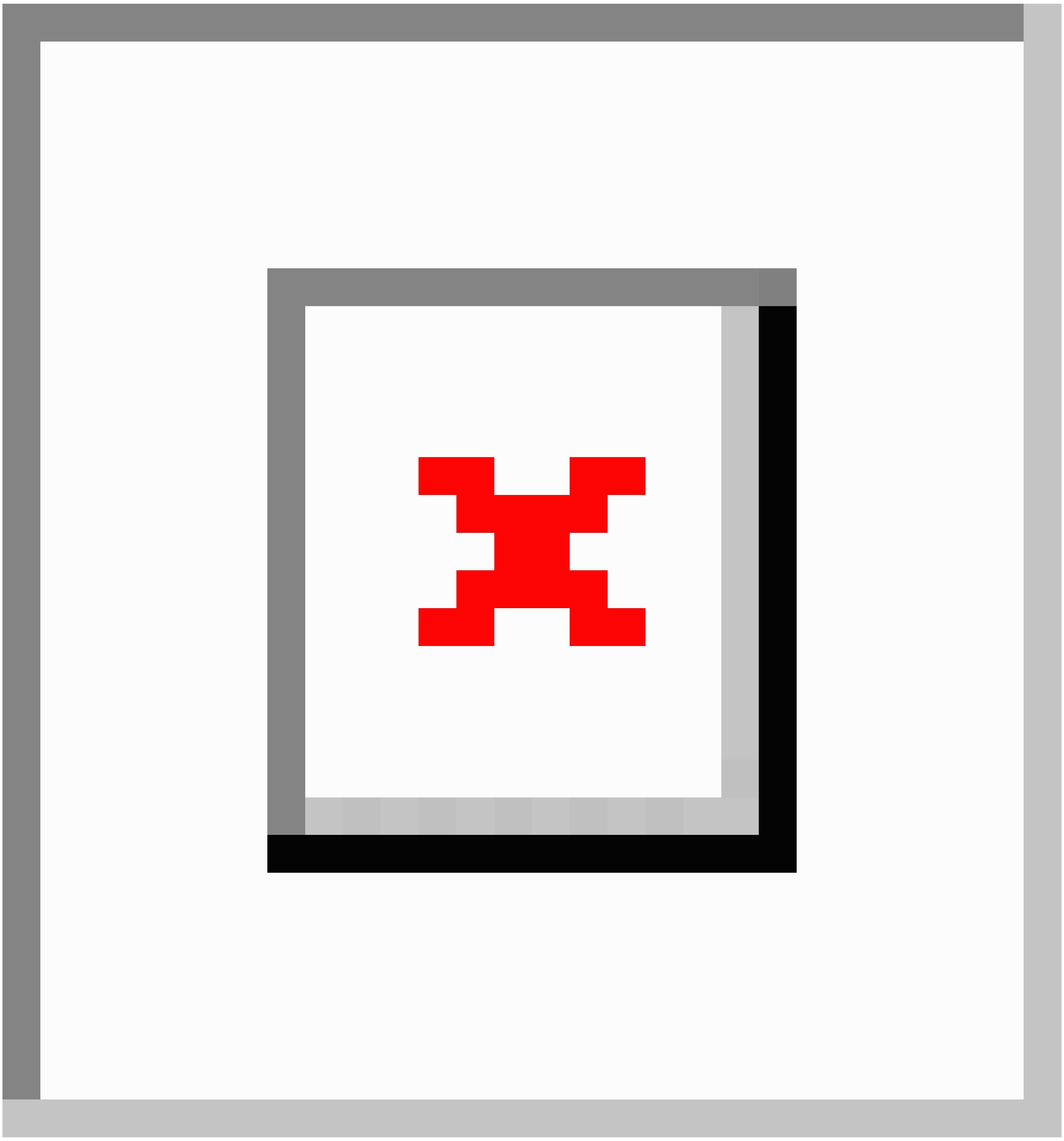
Two independent reviewers (SL and YP) evaluated the risk of bias as part of the quality assessment, using the Cochrane Collaboration's tool for assessing the risk of bias (RoB 2 [Risk of Bias 2]; version: August 9, 2019) [14]. The risk of bias was assessed based on 5 domains, and bias scores were assigned ("low risk," "some concern," or "high risk").

Results

Search Results

The literature search retrieved 305 articles; 72 duplicates were excluded, and 172 did not meet the inclusion criteria, based on the title and abstract review. As a result, 61 articles were screened for the full-text review, and 34 were excluded owing to implications regarding the exclusion criteria and unavailability of full texts. Ultimately, 27 studies were finalized for the data synthesis (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for selection of articles. RCT: randomized controlled trial.



Characteristics of the Selected Articles

Given the novelty of digital health technology in TB treatment, the number of publications was observed to have increased since 2018. A total of 27 articles [15-41] were selected; their characteristics are described in Multimedia Appendix 2. Most of the studies (19/27, 70%) were published in or after 2019 [21,22,24-31,33-41].

With regard to the study designs, 17 studies were randomized controlled trials (RCTs) [15-17,19-21,24-31,34,37,38], 8 were RCT protocols [22,23,32,33,35,36,39,40], and 2 were quasi-experimental studies [18,41]. Further, 13 studies were published in low- and middle-income countries (LMICs), including countries in Africa [18,20,22,23,28,33,35,36,40] and Asia [17,19,25,41]. Digital health technology for TB is actively used in LMICs due to the high prevalence of TB (Multimedia Appendix 3). The combined study population was aged >18 years and included participants who were diagnosed with TB or were taking TB medication. The average number of participants was 400.

Types of DHIs

Table 1 and Figure 2 present the most common technologies used in DHIs, including the duration, frequency, and outcomes of interventions. The commonest DHIs were SMS text messages and reminder messages (8/27, 30%) [15-22], DOT (9/27, 33%) [23-31], medication event reminder monitors (MERMs; 6/27, 22%) [32-37], and mobile apps (4/27, 15%) [38-41]. DOT-based DHIs included video observation therapy (VOT) [25], electronic DOT (e-DOT) [31], and wearable bracelet self-DOT [23]. Some studies evaluated a mix of interventions, including mobile app–based video observations [26,29,30,39], a web intervention [24], WhatsApp (Meta Platforms Inc) [38], and WeChat (Tencent Holdings Ltd) [39]. MERMs [32-37] were also used to determine the feasibility of a web-based follow-up [36] and a mobile-based (ie, evriMED1000 [Wisepill Technologies]) follow-up with phone call [34] reminders to enhance treatment adherence.

Table . Description of digital health technology tuberculosis (TB) interventions and related outcomes.

Study	Intervention	Main outcome	Secondary outcomes	Duration	Frequency
Bediang et al [15]	SMS text messaging	Treatment success	Treatment adherence, multidrug resistance, and satisfaction	6 mo	Daily
van der Kop et al [16]	SMS text messaging	Treatment success	Treatment adherence and treatment completion	9 mo	Weekly
Mohammed et al [17]	SMS text messaging	Treatment success	Treatment adherence and physical health measures	6 mo	Daily
Hermans et al [18]	SMS text messaging	Risk of LFU[a] in the first 2 mo of treatment	Treatment success, completion, adherence, satisfaction, and knowledge	2 mo	Other[b]
Farooqi et al [19]	SMS text messaging	Treatment default	TB treatment results according to the WHO[c,d]	2 mo	Daily
Bediang et al [20]	SMS text messaging	Treatment success	Self-reported adherence regarding attending appointments and satisfaction	6 mo	Daily
Moriarty et al [21]	SMS text messaging	TB treatment results according to the WHO[c]	Smoking cessation, reduction in alcohol use, and treatment adherence	6 mo	Twice weekly
Sahile et al [22]	SMS text messaging	Treatment adherence	ACTG[e], VAS[f], and clinic appointment attendance	2 mo	Daily
Huang et al [23]	e-DOT[g]	TB treatment results according to the WHO[c]	Treatment adherence, MGLS[h], knowledge, and quality of life	6 mo	Daily
Browne et al [24]	e-DOT	Positive detection accuracy	Treatment adherence	Other[i]	Daily
Holzman et al [25]	e-DOT	Treatment adherence	Proportion of all prescribed treatment	Other[i]	Daily
Story et al [26]	e-DOT	Treatment adherence	Treatment outcomes and health-related quality of life	6 mo	Daily
Khachadourian et al [27]	e-DOT	Treatment success	Treatment adherence, depressive symptoms, quality of life, and social support as nonclinical outcomes	4-5 mo	Daily
Crowder et al [28]	e-DOT	Treatment adherence	Reduced risk of LFU and cost-effectiveness	14 mo	Daily
Ravenscroft et al [29]	e-DOT	Treatment adherence	Treatment success at 12 mo	4 mo	Daily
Doltu et al [30]	e-DOT	Treatment adherence	Living conditions, health insurance before TB, previous treatment history, and mode of intensive phase	3 mo	Daily
Burzynski et al [31]	e-DOT	Completed doses and percentage differences between electronic vs in-person DOT[j]	Proportion of medication doses, patient adherence, and quality of care	Other[k]	Daily

Study	Intervention	Main outcome	Secondary outcomes	Duration	Frequency
Lewis et al [32]	MERM[l]	TB treatment results according to the WHO[c]	Adherence outcomes and cost-effectiveness outcomes	6 mo	Daily
Manyazewal et al [33]	MERM	Treatment adherence and sputum conversion	Adverse treatment outcomes, cost-effectiveness, and usability	15 d	Daily
Ratchakit-Nedsuwan et al [34]	MERM	Treatment success	Treatment adherence and patients' experiences	6 mo	Daily
Maraba et al [35]	MERM	Treatment adherence	Treatment success, acceptability of the intervention, and cost-effectiveness	18 mo	Daily
Tadesse et al [36]	MERM	Composite unfavorable outcome: treatment failure or death	Longitudinal technology engagement and fidelity to the intervention	6 mo	Daily
Acosta et al [37]	MERM	Treatment success	Treatment adherence, clinical failure, and LFU	4 mo	Daily
NoorHaslinda and Juni [38]	mHealth[m,n]	Treatment success and treatment adherence	N/A[o]	6 mo	Daily
Wei et al [39]	mHealth[n]	Rate of poor adherence	TB treatment results according to the WHO	6 mo	Daily
Byonanebye et al [40]	mHealth[n]	Treatment success	Treatment success, acceptability of the intervention, and cost-effectiveness	6 mo	Daily
Santra et al [41]	mHealth[n]	Treatment adherence and MGLS	N/A	Other[p]	Daily

[a]LFU: loss to follow-up.

[b]Compliance notifications (2, 7, and 11 d after the most recent appointment), appointment notifications (every 2 wk), and educational quizzes (3, 6, 9, and 12 d after the most recent appointment).

[c]Cured, treatment completed, treatment failed, died, lost to follow-up, not evaluated, or treatment success.

[d]WHO: World Health Organization.

[e]ACTG: AIDS Clinical Trial Group adherence questionnaire.

[f]VAS: visual analog scale.

[g]e-DOT: electronic directly observed therapy.

[h]MGLS: Morisky, Green, and Levine Adherence Scale.

[i]Until TB treatment completion.

[j]DOT: directly observed therapy.

[k]Completed 20 medication doses using 1 DOT method, then switched methods for another 20 doses.

[l]MERM: medication event reminder monitor.

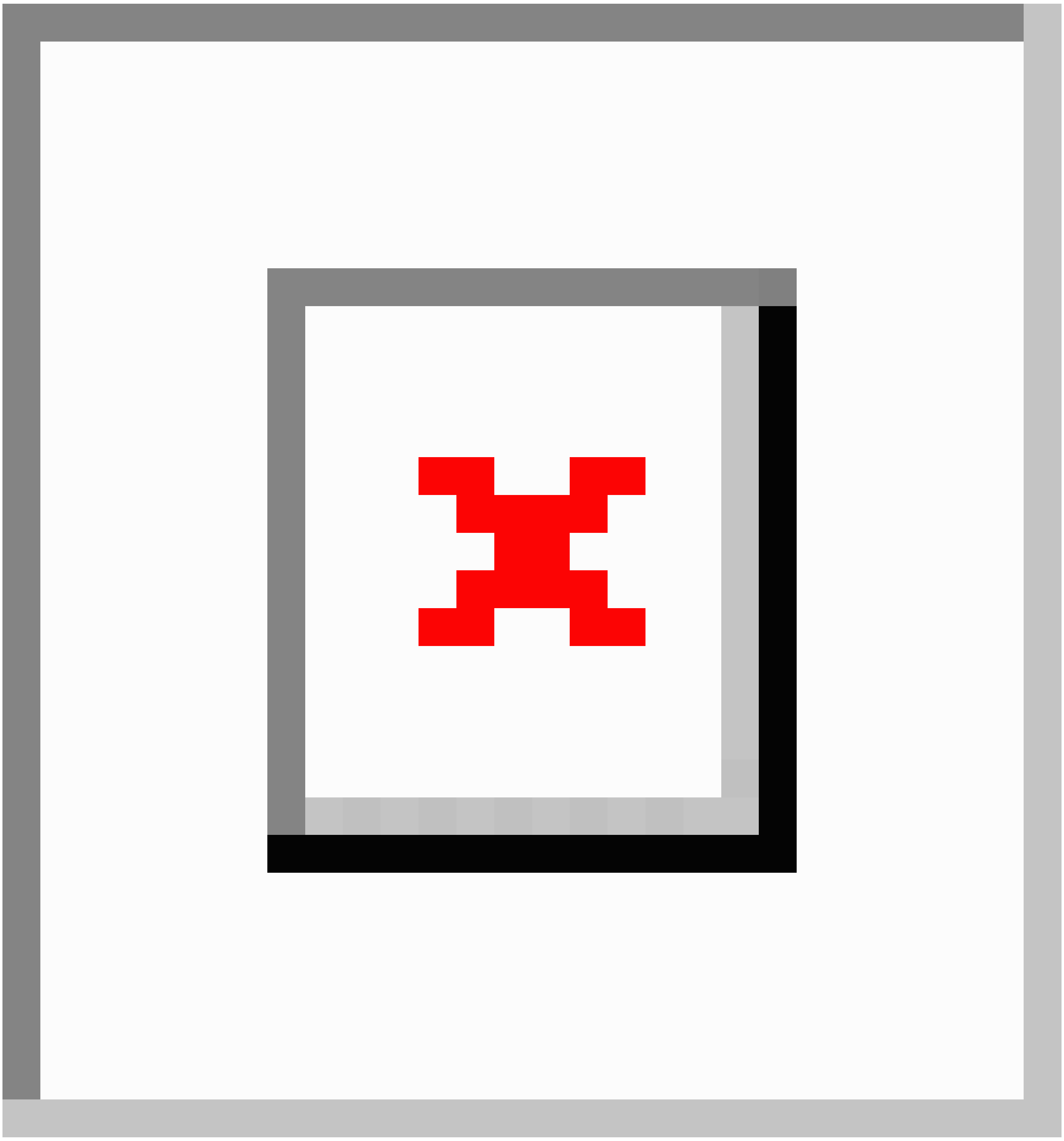
[m]mHealth: mobile health.

[n]Smartphone mobile app.

[o]N/A: not applicable.

[p]DOT for a minimum period of 30 d and a maximum of 90 d.

Figure 2. Types of digital health interventions and the number of articles published by year. e-DOT: electronic directly observed therapy; MERM: medication event reminder monitor; mHealth: mobile health.



Components of the DHIs and Outcomes

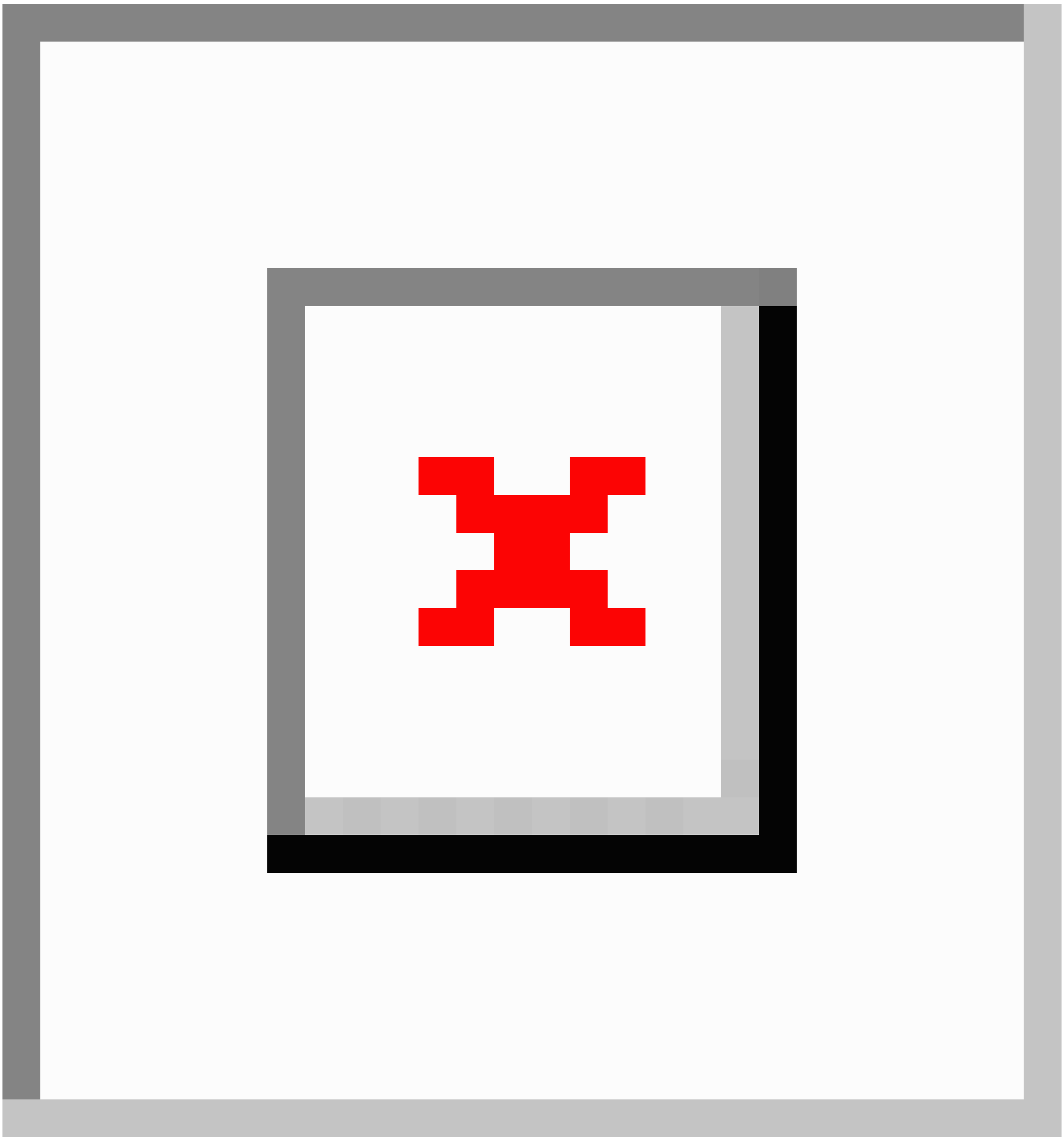
Table 2 presents the components of DHIs that were derived from the primary and secondary outcomes of the selected articles, including (1) sending reminders for treatment adherence via reinforcement SMS text messages [15-22], (2) monitoring treatment adherence by using digital technology [23-31], and (3) tracking treatment adherence through the use of mobile apps and mobile health (mHealth) technology [38-41] via treatment adherence [42] and modified behavior adherence [43] models. Figure 3 presents a modified adherence model.

Table . Distribution of digital health interventions (DHIs) and related interventions (N=27).

Components and DHIs		Articles, n (%)	References
Reminding			
	SMS text messaging	8 (30)	[15-22]
Monitoring			
	DOT[a] (e-DOT[b], VOT[c], and WOT[d])	9 (33)	[23-31]
	MERM[e]	6 (22)	[32-37]
Tracking			
	Mobile app and mHealth[f]	4 (15)	[38-41]

[a]DOT: directly observed therapy.
[b]e-DOT: electronic directly observed therapy.
[c]VOT: video observation therapy.
[d]WOT: wireless observation therapy.
[e]MERM: medication event reminder monitor.
[f]mHealth: mobile health.

Figure 3. Adherence to tuberculosis treatment is a repeated and ongoing self-management behavior. In this figure, *reminding* refers to reminding patients to take medications as prescribed (ie, correct dose, frequency, and time), *monitoring* refers to using digital health technology (eg, an app) to check whether patients are taking their medication at the prescribed frequency over the initial period, and *tracking* refers to following patients over time to determine whether they taking medications as prescribed [43].



Quality Assessment of the Selected Articles

A risk of bias assessment was performed to assess the quality of the selected articles. Only 8 of the 27 articles used an RCT design [17,20,24,26,27,29,31,37]. The risk of bias results are shown in Multimedia Appendix 4 [17,20,24,26,27,29,31,37] and Multimedia Appendix 5.

Discussion

Principal Results

This review aimed to identify DHIs related to TB treatment and management. We retrieved the relevant articles from electronic databases by using standard search terms and identified 27 articles published between 2012 and 2022. DHIs for improving treatment adherence were categorized as DHIs for sending reminders [15-22], DHIs for monitoring [23-31], and DHIs for tracking [38-41]. We identified various types of DHIs, including SMS text messaging [15-22], DOT [23-31], MERMs [32-37], and mobile apps [38-41], which improved the effectiveness of self-management, treatment adherence, and the prevention of TB in clinical and community settings.

A total of 19 studies focused on different types of interventions for reminding patients about treatment adherence and included outcomes such as medication adherence [16,21-23,29,33,35,37,41], self-reported survey satisfaction [17,20,22], and appointment attendance [20]. Treatment adherence was primarily accomplished through daily reminder SMS text messages [15-22] and phone calls [22,27,28,41] that requested confirmation of adherence. Furthermore, additional reminders were sent to patients for encouragement or motivation [15,20] if they did not respond within a given time period [16,17,23,28]. Studies also reported sending compliance reminders through daily quizzes [18]; sending reinforcement SMS text messages twice weekly for 12 weeks [21]; and sending system reminders or additional messages to remind patients about the time of medication use [17], confirm daily doses [28], notify patients about a consultation service for their upcoming monthly visits [32], encourage the use of an app [21], and promote self-satisfaction [17,20,22,27]. Rather than demonstrate treatment efficacy, SMS text messaging–based reminder interventions increased patient satisfaction [17,19]. SMS text messaging–based digital technology supports and helps patients and health care professionals to enhance health practices and clinical outcomes. An interactive reminder, such as an SMS text message or video conversation, should be developed according to the required medical monitoring process and incorporated into clinical practice.

Numerous studies have examined the use of DOT to monitor treatment adherence, including 99DOTS [28], VOT [4,29,39], asynchronous VOT [30], wireless observation therapy [24], and e-DOT [40]. DOT also includes treatment regimen monitoring interventions that are based on technology, such as wearable devices [23], mHealth apps [29], and wireless devices [24]. We identified 8 articles that reported e-DOT interventions for TB treatment adherence. Prior studies reported that participants preferred e-DOT over traditional therapy for supporting daily TB medication use during the long-term phase of TB treatment [24,27,29,30]. e-DOT should be tested in areas with a high risk of TB contraction, as e-DOT could greatly enhance the development of programs for treating the disease in LMICs. In addition, VOT interventions for new TB cases were used in combination with a mobile app [26], WeChat (for education and knowledge) [39], and treatment follow-up (with a maximum follow-up interval of 6 months). Story et al [26] reported that VOT resulted in an 80% medication adherence rate in 2 months when compared to DOT, and Ravenscroft et al [29] reported that VOT resulted in about a 45% decrease in nonadherence, which was statistically significant. Further, smartphone-enabled video surveillance of TB therapy has been proven successful and has many advantages over conventional DOT. Wade et al [44] found that VOT increased the proportion of observed treatment doses when compared to DOT; however, the effect on the treatment adherence rate was not statistically significant. Thus, audio- and video-based DHIs may be useful in reducing attrition and improving treatment adherence and health outcomes in acute care settings.

In this review, 4 RCT protocols for MERM-related monitoring interventions were also included [32,33,35,36] to obtain data on the methodological pattern of treatment adherence. Most MERMs are designed to ensure drug compliance, such as evriMED500 [32,33] or evriMED1000 [35,36]. Maraba et al [35] developed an MERM for the daily monitoring of patients and children with drug-susceptible TB during a 6- to 12-month follow-up. Additionally, Ratchakit-Nedsuwan et al [34] conducted a clinical trial of an MERM for patients with pulmonary TB for approximately 6 months; a total of 54 doses were delivered over 70 days, and the adherence rate was approximately 90%. Further, Acosta et al [37] reported that an MERM was significantly more effective than DOT. Hence, we suggest that further RCTs using MERM-based digital intervention strategies should be conducted to enhance TB treatment adherence and clinical outcomes. Since most outcomes were self-reported, additional trials are recommended to determine the accuracy of MERM system–based adherence rates.

Tracking and guiding patients remain important for the follow-up of treatment adherence in a therapeutic context. We found that 4 smartphone-, mHealth-, and mobile app–based digital devices were used to evaluate TB treatment adherence [34,40,41] and acceptability [38]. Patients with pulmonary TB who received intervention through the WhatsApp TB@Clicks module (an mHealth-based DHI) were approximately 4.1 times more likely to have favorable treatment results than a control group [38]. Another DHI for daily drug tracking resulted in drug adherence rates increasing from 85.5% to 96.4% over time [41], and a health-related VOT resulted in decreased nonadherence rates within 4 days [29]. Some apps were combined with a mobile-based pillbox system for a second consultation, resulting in satisfaction and confidence among patients [34]. These outcomes must be incorporated into future clinical trial designs that adopt trustworthy quantitative methods to determine the relative contribution of each digital health technology component.

This review's findings revealed that DHIs encouraged self-management among patients with TB and empowered them to participate in collaborative discussions during consultations. However, we found that studies on real-time, conversation-based digital technology are lacking; such technology could improve treatment adherence and foster positive health outcomes in various clinical settings. Due to the rapid development of artificial intelligence technologies, including digital tool kits and generative artificial intelligence, 2-way communication–based chatbots in TB treatment may lead to improved self-management in patients with TB.

Limitations

This review had some limitations. First, our review included studies that focused on treatment outcome–based interventions rather than health care delivery. Therefore, we did not focus on other details, such as TB prevalence, costs, or health insurance. Second, this study focused on the effects of commonly used DHIs on TB treatment outcomes in clinical and community settings. Further studies should determine how DHIs vary between the two contexts and how they interact with multidomain therapies. Third, this study did not specifically describe treatment adherence and self-management. There are no clear differences between the accurate meaning and measurement of treatment adherence in a clinical trial setting and those of self-management in a clinical or community context, and few studies have attempted to provide answers [45-47]. Fourth, many of the included studies (13/27, 48%) were conducted in LMICs because of the high prevalence of TB cases, even though high-income nations have a considerable number of studies. This could be attributed to our study's selection criteria, such as our criterion for language. Therefore, additional studies are required to identify DHIs across the entire TB care continuum.

Conclusions

This study examined 27 studies published between 2012 and 2022 and selected the most recent articles. The following three domains were identified from the selected studies: reminding, monitoring, and tracking. The preponderance of treatment adherence was reinforced by mHealth strategies, such as the use of SMS text messaging, mobile apps, mHealth technology, and MERMs. Our findings have implications for TB-related digital health research, which frequently fails to adequately address patients with TB. To preserve treatment adherence and self-care management, patients should have access to real-time, conversation-based interventions (dialogue or communication between patients and health care professionals), such as mobile- or app-based chats, regardless of the restrictions imposed by the COVID-19 pandemic. This scoping review study was conducted before our ongoing chatbot project, which focuses on a mixed methods study on chatbot communication for the treatment adherence of patients with TB. Thus, we emphasize the importance of developing a communication system. DHIs provide several advantages, including improved patient engagement, availability, and accessibility, in addition to lower workloads for practitioners. These results should be considered in the context of national TB control programs and policies to establish a strategy for sustaining TB control and health outcomes. We propose that these developments can significantly improve TB treatment adherence through global collaboration and investment.

Acknowledgments

This study was supported by the Severance Hospital Research Fund for Clinical Excellence Grant (grant number: C-2022-0017).

Authors' Contributions

YP, JSB, and JS conceived and designed this study and were responsible for the methodology. SL and VR conducted the data extraction. SL, VR, and YP conducted the formal analysis. SL and VR wrote the manuscript. All authors contributed to manuscript revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File, 18 KB - mhealth_v11i1e49741_app1.docx]

Multimedia Appendix 2
Characteristics of the selected articles.
[DOCX File, 44 KB - mhealth_v11i1e49741_app2.docx]

Multimedia Appendix 3
Number of articles published by continent (Africa: Cameroon, Ethiopia, South Africa, and Uganda; Asia: China, India, Malaysia, Pakistan, and Thailand; North America and South America: United States, Peru, and Canada; Europe: United Kingdom and Moldova).
[PNG File, 27 KB - mhealth_v11i1e49741_app3.png]

Multimedia Appendix 4
Quality assessment and risk of bias based on the five RoB 2 (Risk of Bias 2) domains. Domain 1: randomization process; domain 2: deviations from intended interventions; domain 3: missing outcome data; domain 4: measurement of the outcome; domain 5: selection of the reported result; domain 6: overall.
[PNG File, 80 KB - mhealth_v11i1e49741_app4.png]

Multimedia Appendix 5

Quality assessment and risk of bias, by intention-to-treat percentage, based on the five RoB 2 (Risk of Bias 2) domains. Domain 1: randomization process; domain 2: deviations from intended interventions; domain 3: missing outcome data; domain 4: measurement of the outcome; domain 5: selection of the reported result; domain 6: Overall.
[PNG File, 19 KB - mhealth_v11i1e49741_app5.png]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File, 49 KB - mhealth_v11i1e49741_app6.docx]

References

1. Global tuberculosis report 2022. World Health Organization. 2022. URL: www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 [accessed 2023-11-10]
2. Alipanah N, Jarlsberg L, Miller C, et al. Adherence interventions and outcomes of tuberculosis treatment: a systematic review and meta-analysis of trials and observational studies. PLoS Med 2018 Jul 3;15(7):e1002595. [doi: 10.1371/journal.pmed.1002595] [Medline: 29969463]
3. Ngwatu BK, Nsengiyumva NP, Oxlade O, et al. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan 11;51(1):1701596. [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]
4. Story A, Garfein RS, Hayward A, et al. Monitoring therapy compliance of tuberculosis patients by using video-enabled electronic devices. Emerg Infect Dis 2016 Mar;22(3):538-540. [doi: 10.3201/eid2203.151620] [Medline: 26891363]
5. Ridho A, Alfian SD, van Boven JFM, et al. Digital health technologies to improve medication adherence and treatment outcomes in patients with tuberculosis: systematic review of randomized controlled trials. J Med Internet Res 2022 Feb 23;24(2):e33062. [doi: 10.2196/33062] [Medline: 35195534]
6. Lee Y, Raviglione MC, Flahault A. Use of digital technology to enhance tuberculosis control: scoping review. J Med Internet Res 2020 Feb 13;22(2):e15727. [doi: 10.2196/15727] [Medline: 32053111]
7. Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016 May;20(5):588-593. [doi: 10.5588/ijtld.15.0738] [Medline: 27084810]
8. Thomas BE, Kumar JV, Chiranjeevi M, et al. Evaluation of the accuracy of 99DOTS, a novel cellphone-based strategy for monitoring adherence to tuberculosis medications: comparison of digital adherence data with urine isoniazid testing. Clin Infect Dis 2020 Dec 3;71(9):e513-e516. [doi: 10.1093/cid/ciaa333] [Medline: 32221550]
9. Wang N, Shewade HD, Thekkur P, et al. Electronic medication monitor for people with tuberculosis: implementation experience from thirty counties in China. PLoS One 2020 Apr 29;15(4):e0232337. [doi: 10.1371/journal.pone.0232337] [Medline: 32348351]
10. Fiseha D, Demissie M. Assessment of directly observed therapy (DOT) following tuberculosis regimen change in Addis Ababa, Ethiopia: a qualitative study. BMC Infect Dis 2015 Sep 30;15:405. [doi: 10.1186/s12879-015-1142-2] [Medline: 26423277]
11. Westphaln KK, Regoeczi W, Masotya M, et al. From Arksey and O'Malley and beyond: customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021 May 7;8:101375. [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
12. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700. [doi: 10.1136/bmj.b2700] [Medline: 19622552]
13. Peters M, Godfrey C, McInerney P, Soares C, Khalil H, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews: The Joanna Briggs Institute; 2015.
14. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
15. Bediang G, Stoll B, Elia N, et al. SMS reminders to improve the tuberculosis cure rate in developing countries (TB-SMS Cameroon): a protocol of a randomised control study. Trials 2014 Jan 24;15:35. [doi: 10.1186/1745-6215-15-35] [Medline: 24460827]
16. van der Kop ML, Memetovic J, Patel A, et al. The effect of weekly text-message communication on treatment completion among patients with latent tuberculosis infection: study protocol for a randomised controlled trial (WelTel LTBI). BMJ Open 2014 Apr 9;4(4):e004362. [doi: 10.1136/bmjopen-2013-004362] [Medline: 24719431]
17. Mohammed S, Glennerster R, Khan AJ. Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. PLoS One 2016 Nov 1;11(11):e0162944. [doi: 10.1371/journal.pone.0162944] [Medline: 27802283]
18. Hermans SM, Elbireer S, Tibakabikoba H, Hoefman BJ, Manabe YC. Text messaging to decrease tuberculosis treatment attrition in TB-HIV coinfection in Uganda. Patient Prefer Adherence 2017 Aug 31;11:1479-1487. [doi: 10.2147/PPA.S135540] [Medline: 28919720]

19. Farooqi RJ, Ashraf S, Zaman M. The role of mobile SMS-reminders in improving drugs compliance in patients receiving anti-TB treatment from DOTS program. Journal of Postgraduate Medical Institute 2017 May 12;31(2):156-162 [FREE Full text]
20. Bediang G, Stoll B, Elia N, Abena JL, Geissbuhler A. SMS reminders to improve adherence and cure of tuberculosis patients in Cameroon (TB-SMS Cameroon): a randomised controlled trial. BMC Public Health 2018 May 2;18(1):583. [doi: 10.1186/s12889-018-5502-x] [Medline: 29720146]
21. Moriarty AS, Louwagie GM, Mdege ND, et al. ImPROving TB outcomes by modifying LIFE-style behaviours through a brief motivational intervention followed by short text messages (ProLife): study protocol for a randomised controlled trial. Trials 2019 Jul 26;20(1):457. [doi: 10.1186/s13063-019-3551-9] [Medline: 31349850]
22. Sahile Z, Perimal-Lewis L, Arbon P, Maeder AJ. Protocol of a parallel group randomized control trial (RCT) for mobile-assisted medication adherence support (Ma-MAS) intervention among tuberculosis patients. PLoS One 2021 Dec 31;16(12):e0261758. [doi: 10.1371/journal.pone.0261758] [Medline: 34972128]
23. Huang R, Ren G, Hu J. Bracelet- and self-directed observational therapy for control of tuberculosis: study protocol for a cluster randomized controlled trial. Trials 2017 Jul 4;18(1):286. [doi: 10.1186/s13063-017-1996-2] [Medline: 28673323]
24. Browne SH, Umlauf A, Tucker AJ, et al. Wirelessly observed therapy compared to directly observed therapy to confirm and support tuberculosis treatment adherence: a randomized controlled trial. PLoS Med 2019 Oct 4;16(10):e1002891. [doi: 10.1371/journal.pmed.1002891] [Medline: 31584944]
25. Holzman SB, Atre S, Sahasrabudhe T, et al. Use of smartphone-based video directly observed therapy (vDOT) in tuberculosis care: single-arm, prospective feasibility study. JMIR Form Res 2019 Aug 27;3(3):e13411. [doi: 10.2196/13411] [Medline: 31456581]
26. Story A, Aldridge RW, Smith CM, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224. [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
27. Khachadourian V, Truzyan N, Harutyunyan A, et al. People–centred care versus clinic–based DOT for continuation phase TB treatment in Armenia: a cluster randomized trial. BMC Pulm Med 2020 Apr 25;20(1):105. [doi: 10.1186/s12890-020-1141-y] [Medline: 32334553]
28. Crowder R, Kityamuwesi A, Kiwanuka N, et al. Study protocol and implementation details for a pragmatic, stepped-wedge cluster randomised trial of a digital adherence technology to facilitate tuberculosis treatment completion. BMJ Open 2020 Nov 27;10(11):e039895. [doi: 10.1136/bmjopen-2020-039895] [Medline: 33247012]
29. Ravenscroft L, Kettle S, Persian R, et al. Video-observed therapy and medication adherence for tuberculosis patients: randomised controlled trial in Moldova. Eur Respir J 2020 Aug 6;56(2):2000493. [doi: 10.1183/13993003.00493-2020] [Medline: 32381495]
30. Doltu S, Ciobanu A, Sereda Y, et al. Short and long-term outcomes of video observed treatment in tuberculosis patients, the Republic of Moldova. J Infect Dev Ctries 2021 Sep 29;15(9.1):17S-24S. [doi: 10.3855/jidc.14601] [Medline: 34609956]
31. Burzynski J, Mangan JM, Lam CK, et al. In-person vs electronic directly observed therapy for tuberculosis treatment adherence: a randomized noninferiority trial. JAMA Netw Open 2022 Jan 4;5(1):e2144210. [doi: 10.1001/jamanetworkopen.2021.44210] [Medline: 35050357]
32. Lewis JJ, Liu X, Zhang Z, et al. Evaluation of a medication monitor-based treatment strategy for drug-sensitive tuberculosis patients in China: study protocol for a cluster randomised controlled trial. Trials 2018 Jul 25;19(1):398. [doi: 10.1186/s13063-018-2650-3] [Medline: 30045757]
33. Manyazewal T, Woldeamanuel Y, Holland DP, Fekadu A, Blumberg HM, Marconi VC. Electronic pillbox-enabled self-administered therapy versus standard directly observed therapy for tuberculosis medication adherence and treatment outcomes in Ethiopia (SELFTB): protocol for a multicenter randomized controlled trial. Trials 2020 May 5;21(1):383. [doi: 10.1186/s13063-020-04324-z] [Medline: 32370774]
34. Ratchakit-Nedsuwan R, Nedsuwan S, Sawadna V, et al. Ensuring tuberculosis treatment adherence with a mobile-based CARE-call system in Thailand: a pilot study. Infect Dis (Lond) 2020 Feb;52(2):121-129. [doi: 10.1080/23744235.2019.1688862] [Medline: 31710250]
35. Maraba N, Orrell C, Chetty-Makkan CM, et al. Evaluation of adherence monitoring system using evriMED with a differentiated response compared to standard of care among drug-sensitive TB patients in three provinces in South Africa: a protocol for a cluster randomised control trial. Trials 2021 Jun 9;22(1):389. [doi: 10.1186/s13063-021-05337-y] [Medline: 34108036]
36. Tadesse AW, Mohammed Z, Foster N, et al. Evaluation of implementation and effectiveness of digital adherence technology with differentiated care to support tuberculosis treatment adherence and improve treatment outcomes in Ethiopia: a study protocol for a cluster randomised trial. BMC Infect Dis 2021 Nov 10;21(1):1149. [doi: 10.1186/s12879-021-06833-x] [Medline: 34758737]
37. Acosta J, Flores P, Alarcón M, Grande-Ortiz M, Moreno-Exebio L, Puyen ZM. A randomised controlled trial to evaluate a medication monitoring system for TB treatment. Int J Tuberc Lung Dis 2022 Jan 1;26(1):44-49. [doi: 10.5588/ijtld.21.0373] [Medline: 34969428]

38. NoorHaslinda I, Juni MH. Effectiveness of health education module delivered through WhatsApp to enhance treatment adherence and successful outcome of tuberculosis in Seremban District, Negeri Sembilan, Malaysia. International Journal of Public Health and Clinical Sciences 2019;6(4):145-159. [doi: 10.32827/ijphcs.6.4.145]
39. Wei X, Hicks JP, Pasang P, et al. Protocol for a randomised controlled trial to evaluate the effectiveness of improving tuberculosis patients' treatment adherence via electronic monitors and an app versus usual care in Tibet. Trials 2019 May 16;20(1):273. [doi: 10.1186/s13063-019-3364-x] [Medline: 31097023]
40. Byonanebye DM, Mackline H, Sekaggya-Wiltshire C, et al. Impact of a mobile phone-based interactive voice response software on tuberculosis treatment outcomes in Uganda (CFL-TB): a protocol for a randomized controlled trial. Trials 2021 Jun 13;22(1):391. [doi: 10.1186/s13063-021-05352-z] [Medline: 34120649]
41. Santra S, Garg S, Basu S, Sharma N, Singh MM, Khanna A. The effect of a mhealth intervention on anti-tuberculosis medication adherence in Delhi, India: a quasi-experimental study. Indian J Public Health 2021;65(1):34-38. [doi: 10.4103/ijph.IJPH_879_20] [Medline: 33753687]
42. Julius RJ, Novitsky MAJ, Dubin WR. Medication adherence: a review of the literature and implications for clinical practice. J Psychiatr Pract 2009 Jan;15(1):34-44. [doi: 10.1097/01.pra.0000344917.43780.77] [Medline: 19182563]
43. Adherence & behavioral science. Acare HCP Global. URL: acarepro.abbott.com/adherence-behavioral-science/ [accessed 2023-11-10]
44. Wade VA, Karnon J, Eliott JA, Hiller JE. Home videophones improve direct observation in tuberculosis treatment: a mixed methods evaluation. PLoS One 2012;7(11):e50155. [doi: 10.1371/journal.pone.0050155] [Medline: 23226243]
45. Howren MB, Gonzalez JS. Treatment adherence and illness self-management: introduction to the special issue. J Behav Med 2016 Dec;39(6):931-934. [doi: 10.1007/s10865-016-9804-0] [Medline: 27766482]
46. Li J, Pu J, Liu J, et al. Determinants of self-management behaviors among pulmonary tuberculosis patients: a path analysis. Infect Dis Poverty 2021 Jul 30;10(1):103. [doi: 10.1186/s40249-021-00888-3] [Medline: 34330337]
47. Vernon A, Fielding K, Savic R, Dodd L, Nahid P. The importance of adherence in tuberculosis treatment clinical trials and its relevance in explanatory and pragmatic trials. PLoS Med 2019 Dec 10;16(12):e1002884. [doi: 10.1371/journal.pmed.1002884] [Medline: 31821323]

Abbreviations

DHI: digital health intervention
DOT: directly observed therapy
e-DOT: electronic directly observed therapy
LMIC: low- and middle-income country
MERM: medication event reminder monitor
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: Risk of Bias 2
TB: tuberculosis
VOT: video observation therapy

Edited by L Buis; submitted 21.06.23; peer-reviewed by B Sim, L Perimal-Lewis; revised version received 05.10.23; accepted 27.10.23; published 04.12.23.

Please cite as:
Lee S, Rajaguru V, Baek JS, Shin J, Park Y
Digital Health Interventions to Enhance Tuberculosis Treatment Adherence: Scoping Review
JMIR mHealth and uHealth 2023;11:e49741
URL: https://mhealth.jmir.org/2023/1/e49741
doi:10.2196/49741

The Importance of Activating Factors in Physical Activity Interventions for Older Adults Using Information and Communication Technologies: Systematic Review

Ellen Bentlage, MSc; John Jnr Nyamadi, BSc; Rosemary Dubbeldam, PhD

Department of Movement Science, Institute of Sport and Exercise Sciences, University of Münster, Münster, Germany

Corresponding Author:
Ellen Bentlage, MSc

Related Article:

https://mhealth.jmir.org/2024/1/e55486

Abstract

Background: In an aging population, it is important to activate older adults in taking care of their own health. Increasing physical activity is one way to avoid or lessen age-related physical and mental impairments. Interest in the use of information and communication technology (ICT) tools to promote physical activity among older adults is growing considerably. Such tools are suitable for communicating activation factors—skills, knowledge, and motivation—by integrating a variety of behavior change techniques (BCTs) to enhance physical activity. Although activation factors have been incorporated into physical activity interventions using ICT, little is known about the actual integration methods used in such interventions or about the effects of activation factors on influencing behavior change.

Objective: The first aim of this study was to identify which of the activation factors were covered in physical activity–promoting ICT interventions for older adults and which BCTs were used to address them. The second objective was to classify the user interaction interfaces and delivery modes that were used to promote these activation factors.

Methods: The search engines of PubMed, Web of Science, and ScienceDirect were used to search for and identify articles examining the effectiveness of ICT interventions for promoting physical activity in older adults. References and related data were selected, extracted, and reviewed independently by 2 reviewers. The risk of bias was assessed, and any conflict was addressed by a third separate reviewer. Selected articles included older adults aged ≥55 years without pre-existing medical diseases and other physical or mental conditions that could hinder movement.

Results: In total, 368 records were retrieved, and 13 studies met all inclusion criteria. Articles differed in terms of themes, timescales, user interaction interfaces, and outcome measures; therefore, a quantitative data synthesis was not feasible. Motivation was the most promoted activation factor among all trials (33 times). An app and a smartwatch were used in the majority of intervention groups (7/20, 35%) for tracking physical activity and receiving personalized feedback based on the individual goals. Skills (25 times) and knowledge (17 times) were the next most commonly addressed activation factors. Face-to-face interaction was the most used approach to targeting users' skills, including providing instructions on how to perform a behavior and exchanging knowledge via education on the health consequences of insufficient physical activity. Overall, integrating all 3 activation factors and using multiple user interaction interfaces with a variety of delivery modes proved the most effective in improving physical activity.

Conclusions: This study highlights commonly used BCTs and preferred modes of their delivery. So far, only a limited number of available BCTs (21/102, 21%) have been integrated. Considering their effectiveness, a larger variety of BCTs that address skills, knowledge, and motivation should be exploited in future ICT interventions.

(JMIR Mhealth Uhealth 2023;11:e42968) doi:10.2196/42968

KEYWORDS

older adults; information and communication technology; healthy aging; activation factors; skills; knowledge; motivation; behavior change techniques; physical activity

Introduction

Consequences of Aging

The World Health Organization projects a 34% increase in the global population of 1 billion older adults by 2030, showing a demographic trend toward an older population [1]. Several pathologies, such as pulmonary disease, neurodegenerative disorders, and cardiovascular disease, share aging as their dominant pathogenesis risk factor, though these pathologies can be positively influenced by physical activity [2-4]. Hence, managing the health of the older population is important [5].

Activation in Health Care Management

Individuals themselves can successfully be activated in their health management if they are equipped with activating factors. Activation of the individual leads to behavioral changes and consequently enhances, for example, physical activity levels [6]. Activating people involves making the individuals believe in their participatory role and fostering their confidence through improving their skills and abilities regarding their well-being. It also involves creating awareness by communicating knowledge on the necessity to act. Furthermore, it includes motivating individuals to take action to maintain and improve their health outcomes [7,8]. Evidence proves that activated individuals are independently able to better control their health and have better health outcomes [9]. As a result, activated individuals develop confidence in self–health care management [10].

Activation Using Information and Communication Technology

First, information and communication technology (ICT) interfaces communicate activation factors effectively to the user. Such activation factors can be described by using the behavior change technique (BCT) taxonomy from Michie et al [11]. This taxonomy includes 16 categories for promoting skills, knowledge, or motivation; these categories include shaping knowledge, comparing behavior, natural consequences, comparison of outcomes, goals and planning, and feedback and monitoring. Further, considering the possibilities of ICT tools, Dugas and colleagues [12] added 2 more categories (ie, personalization and gamification), including 9 BCTs in total. A review by Aldawood et al [13] also pointed out that ICT interfaces in health interventions offer various BCTs that indeed raise awareness of health and promote more self-awareness among people. Second, ICT interventions that include activation factors are effective for improving health. Such interventions include, among others, remote coaching or monitoring, automated feedback, and increased accessibility to credible health information [14,15]. Providing feedback based on detected behavior patterns, when paired with reminders to be active, leads to improvements in physical activity behavior, better health-related knowledge, and increased motivation [16]. Additionally, a web-based intervention [17] and an intervention using wearable activity trackers connected with a smartphone app [18] both showed improvements in participants' health skills, knowledge, and motivation for developing and maintaining positive health-related practices. Lastly, entertainment, such as exergames, can be used to teach health-related skills, provide feedback, and constantly motivate the user [19].

Although activation factors have been incorporated into physical activity interventions using ICT, little is known about the actual integration methods used in such interventions or about the effects of activation factors on influencing behavior change. McGarrigle and Todd [20] stated that ICT interventions incorporating BCTs may be more effective in promoting physical activity than those interventions that do not focus on such techniques. Accordingly, Dugas and colleagues [12] performed a systematic review on health behavior interventions within mobile health apps and reported on the integrated BCTs and how they influenced outcomes.

Study Aims

This systematic review, as a primary objective, aimed to identify which of the activation factors—skills, knowledge, and motivation—were covered in ICT interventions that promote the physical activity of older adults and to report the incorporated BCTs. The secondary objective was to classify the user interaction interfaces and delivery modes that were used to promote the activation factors.

Methods

Information Sources, Databases, and Searching Process

The search engines of PubMed, Web of Science, and ScienceDirect were used to search the MEDLINE, Web of Science Core Collection, and ScienceDirect databases, respectively, for peer-reviewed publications that were published until February 28, 2022. The search strategy was customized for each selected database according to their filtering options (eg, for PubMed, a combination of Medical Subject Headings and other index terms was used).

The final search string included *("mhealth" OR "telemedicine" OR "mobile application") AND ("older adults") AND ("activation" OR "physical activity" OR "self-healthcare" OR "healthy ageing")*. The results that were generated by using the abovementioned search strategy in all databases were uploaded to Rayyan (Rayyan Systems Inc) [21] for the cleaning and selection process. First, the titles and abstracts of identified studies were independently screened by 2 reviewers to select relevant studies. Second, the full texts of potentially relevant studies were obtained and independently reviewed. A third assessor, who was not part of the previous screening of articles, decided on the inclusion or exclusion of the articles in cases of conflicts.

Selection of Studies and Data Extraction

Guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement were used for the reporting of this systematic review [22]. The three databases and the reference lists of the included articles were searched and evaluated based on a set of inclusion and exclusion criteria.

Articles that met the following criteria were included: (1) a target group with an average age of ≥55 years; (2) participants without pre-existing chronic medical diseases (eg, diabetes) or mental and physical impairments (eg, repetitive falling); (3) articles that reported about digital physical activity intervention(s); (4) articles that assessed physical activity effectiveness through objective methods (eg, pedometer), subjective methods (eg, questionnaires), or a combination of objective and subjective methods; and (5) methods were embedded in one of the following designs: randomized controlled trial, quasi-experimental, clinical, or feasibility study designs.

Articles that met at least one of the following exclusion criteria were excluded: (1) articles not written in English; (2) no access to the full text; (3) a target group with an average age of <55 years; (4) a focus on older adults with pre-existing chronic medical diseases or in acute rehabilitation scenarios (eg, patients in rehabilitation after stroke); (5) articles that did not include a digital physical activity intervention; (6) articles that did not report about intervention effectiveness for physical activity; and (7) nonempirical research (eg, editorials and commentary papers).

After the selection of relevant articles, the data were extracted, and quality was assessed. The data extracted from each selected study included the author, year of publication, study design, sample size, population, intervention, technology, timescales, outcome measures, and main findings. The delivery modes of the user interaction interfaces were classified (1) as a personalized exercise introduction session at the beginning of the intervention, (2) as digital (calls, text messages, apps, the web, smartwatches, and activators), (3) as traditional (face-to-face modes and printed materials), or (4) as digital and traditional (hybrid).

The supervision of the delivery modes for the intervention groups was defined by Denton et al [23], as follows: (1) supervised (ie, physical activity is undertaken in the presence of a health care professional or qualified fitness instructor, either virtually or in person, to ensure safety and or correct technique), (2) facilitated (ie, physical activity is undertaken without the presence of a health care professional or qualified fitness instructor but with scheduled meetings or check-ins between sessions to monitor progress and provide support [virtually or in person]), or (3) unsupervised (ie, physical activity is undertaken without the presence of a health care professional or qualified fitness instructor; no support or progress tracking appointments are scheduled).

We identified and classified the used activation factors and corresponding user interfaces. The definitions of the activation factors and their classification are described in Table 1. The classification items were based on the BCT taxonomy with 16 categories developed by Michie and colleagues [11]. Additionally, the two categories suggested by Dugas and colleagues [12]—personalization and gamification—were incorporated as categories 17 and 18.

Table . Activation factors and behavior change technique taxonomy.

	Activation factor		
	Skills (ability)	Knowledge (awareness)	Motivation (triggers)
Definition	• A person will be equipped with skills via instructions or tips on the correct performance of the behavior. Observing others succeeding in performing the targeted activity is another strategy [24]. With increased ability, self-confidence will be fostered [9].	• A person will be educated by providing them with knowledge on the benefits of sufficient physical activity as well as the consequences of insufficient physical activity [25]. The awareness that it is necessary to be active will become present, and intentions will be formed [9].	• A person will be triggered to have high motivation [7,8] by gaining personalized values [26] through self-monitoring [25], monitoring by another person, monitoring via a technical device, and encouraging feedback [27].
Behavior change taxonomy categories[a,b]	• Category 4: Shaping knowledge • Category 6: Comparison of behavior • Category 8: Repetition and substitution • Category 13: Identity • Category 15: Self-belief	• Category 5: Natural consequences • Category 9: Comparison of outcomes • Category 11: Regulation	• Category 1: Goals and planning • Category 2: Feedback and monitoring • Category 3: Social support • Category 7: Associations • Category 10: Reward and thread • Category 12: Antecedents • Category 14: Scheduled consequences • Category 16: Covert learning • Category 17: Personalization • Category 18: Gamification

[a]Behavior change taxonomy categories 1 to 16 by Michie et al [11].

[b]Behavior change taxonomy categories 17 and 18 by Dugas et al [12].

Quality Assessment

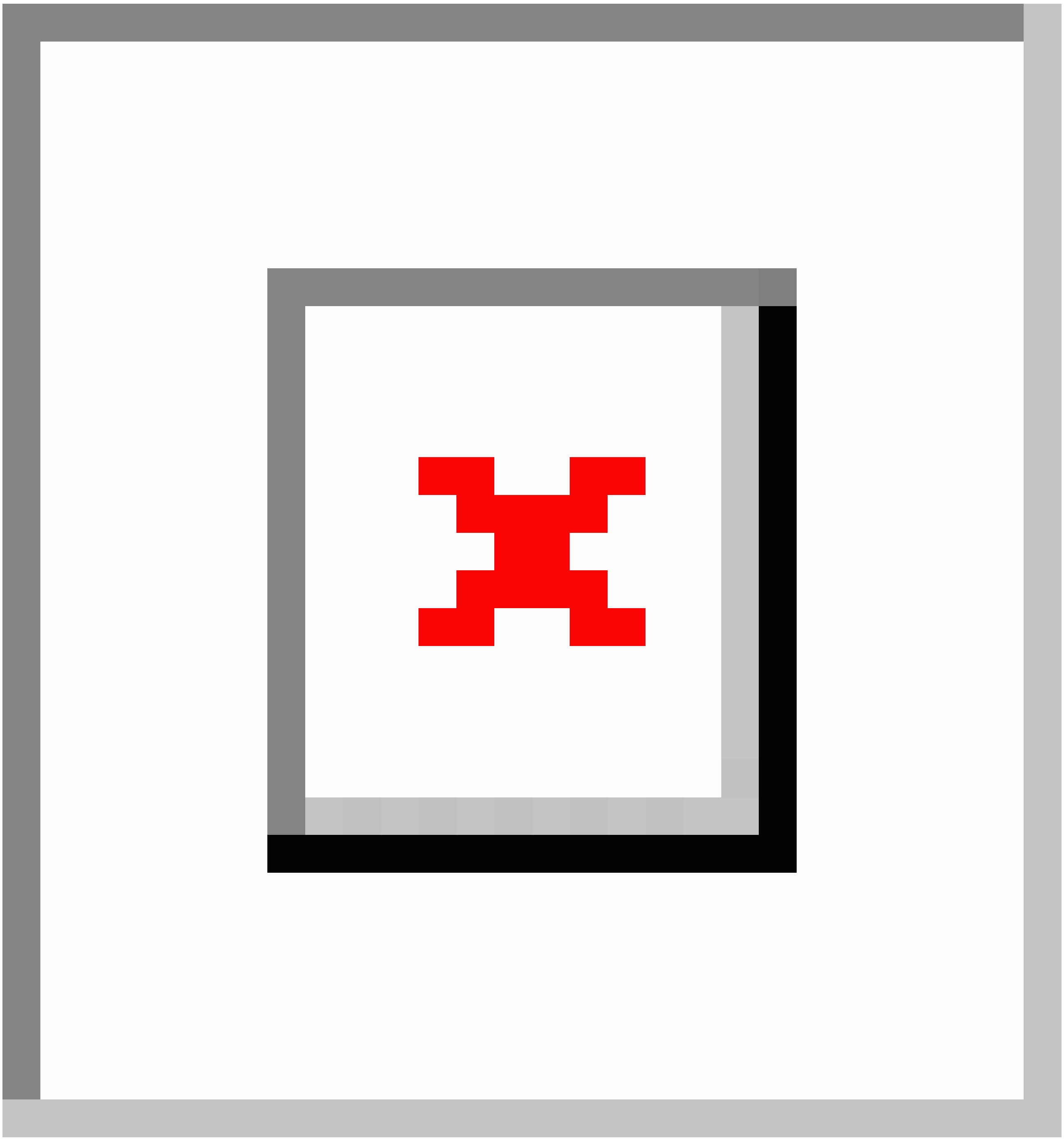
To assess the risk of bias, the Cochrane Collaboration tool Risk of Bias 2 (ROB 2), which focuses on different aspects of trial design (ie, conducting and reporting), was applied [28,29]. According to the tool, a study was classified as having a low risk of bias when it scored low on all 3 domains, a moderate risk of bias when 2 of the 3 domains were scored low, and a high risk of bias when 1 or no main domain was scored low. A meta-analysis was not feasible due to the selected studies being different with regard to the types of interventions.

Results

Studies Identified Through the Searching Process

After searching the databases, 353 abstracts were identified. Further, 15 additional abstracts were identified by searching the reference lists of the initially identified articles. After the removal of duplicates and the screening of the involved abstracts, the full texts of 46 articles were assessed for eligibility. At the end, 13 articles met the inclusion criteria and were therefore included in this systematic review. Details on the study selection process are presented in Figure 1.

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



Study Characteristics

Study Design

Of the 13 included articles, 6 were randomized controlled trials, and 8 used a quasi-experimental design. The 13 articles included a total sample of 1622 participants. Male and female participants were involved in all articles; however, their exact composition was not reported in two of them [30,31]. The mean age of the study participants ranged from 63 to 80 years. The duration of the intervention ranged from 2 weeks to 24 months. The study characteristics are presented in Multimedia Appendix 1 [30-42].

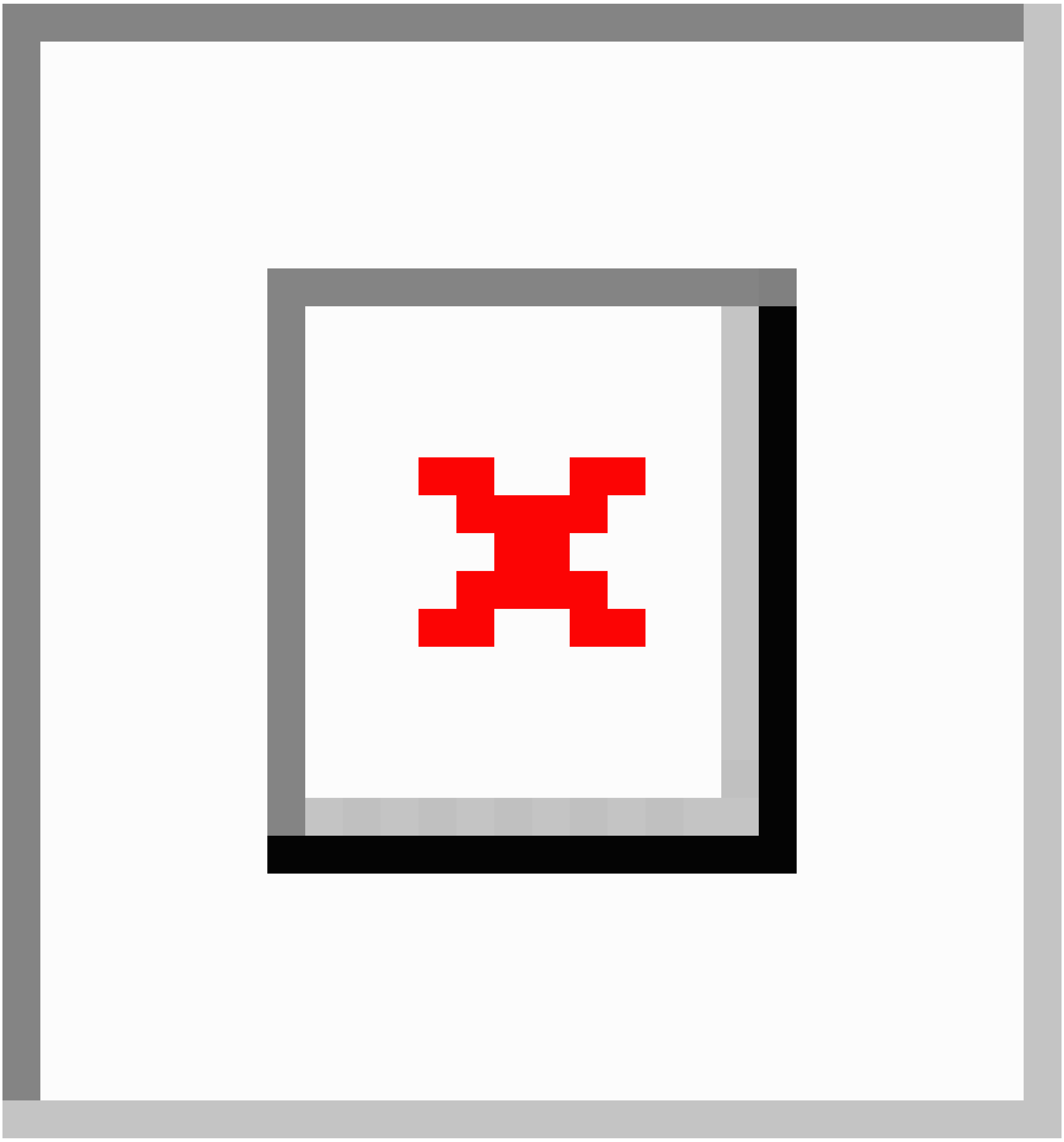
Physical Activity Assessment

To be able to analyze the effectiveness of an intervention, subjective or objective outcome measures for physical activity are necessary. In 4 articles (with 6 intervention groups), only subjective methods were used to measure physical activity (ie, questionnaires like the International Physical Activity Questionnaire). In 4 of the intervention groups, a significant beneficial effect of the digital or hybrid intervention was reported when compared to a control group [30-33]. In 8 articles (with 9 intervention groups), physical activity was measured by using only an objective assessment (ie, pedometer and accelerometer). In 3 of these intervention groups, a significant positive effect of the digital or traditional intervention on physical activity values was reported [31,34,35]. In the remaining intervention group, both objective assessments and subjective assessments were integrated, which showed significant effects on improving physical activity [36].

Activation Factors

Figure 2 presents the integrated activation factors, as well as the corresponding interfaces and delivery modes used for the user interaction in each article and intervention group. Data on the BCTs used in the intervention were extracted for each article, and the number of BCTs for each factor of activation was determined. An overview of integrated BCTs and those that were not used is presented in Multimedia Appendix 2.

Figure 2. User interaction interfaces and delivery modes for delivering activation factors [30,31,32,33,34,35,36,37,38,39,40,41,42]. Orange blocks show intervention groups with significant beneficial findings regarding physical activity. Grey blocks highlight the user interfaces used along with the integrated activation factors. A high resolution version of the figure can be found in Multimedia Appendix 3. D: digital; FA: facilitated; H: hybrid; K: knowledge; M: motivation; OB: objective; PE: personalized exercise introduction; S: skills; SU: subjective; SUP: supervised; T: traditional; U: unsupervised.



Delivering Activation Factors

For all of the interventions, motivation was the most promoted activation factor (33 times), followed by skills (25 times) and knowledge (17 times). All 3 activation factors were identified in 12 of 20 interventions. Interventions that included all 3 activation factors were shown to be more successful in increasing or maintaining physical activity levels (articles: 7/12, 58%) compared to interventions that limited themselves to 2 or fewer activation factors (articles: 3/8, 38%).

BCTs that promoted skills, knowledge, and motivation were addressed in 19, 15, and 17 of the 20 intervention groups, respectively. Effective outcomes regarding physical activity were reported while promoting skills in 47% (9/19) of trials, while promoting knowledge in 53% (8/15) of trials, and while promoting motivation in 53% (9/17) of trials.

To promote skills, instruction on how to optimally perform a behavior was most commonly delivered in the form of training prescriptions and exercise tips (eg, key movements and where the exercises should be felt) or through descriptions using pictures (16 intervention groups). The next most common BCT for promoting skills was face-to-face demonstration or demonstration through tutorial videos (8 intervention groups), followed by the provision of information on how to integrate exercise into daily activities (habit formation; 3 intervention groups) and practice (2 intervention groups). Lastly, participants were informed about antecedents of the behavior (1 intervention group). Of the 23 available BCTs [11,12] that target skills, the aforementioned 6 (23%) were used among all included articles.

The most used strategy for changing physical activity behavior by promoting knowledge was the integration of information about the health consequences that arise if physical activity is not sufficiently part of daily behaviors (13 intervention groups). Besides that, information from a credible source (7 intervention groups) and educative information, such as the salience of consequences (6 intervention groups) and the pros and cons of physical activity (1 intervention group), were also part of the interventions. Of the 13 available BCTs [11,12] that target knowledge, the aforementioned 4 (31%) were used among all included articles.

With regard to targeting motivation, the self-monitoring of physical activity (12 intervention groups) and personalized feedback (8 intervention groups), which was often combined with individual goal setting (7 intervention groups) and social support (8 intervention groups), were the most used BCTs. Others were prompts and cues (4 intervention groups), adjustment of intervention content to the performance (4 intervention groups), action planning (2 intervention groups) and problem-solving discussions for finding ways to overcome barriers (1 intervention group). Of the 66 available BCTs [11,12] that target motivation, 11 were used among all included articles.

Altogether, of the 102 potential BCTs [11,12], 21 (21%) were integrated in all articles.

User Interaction Interfaces and Delivery Modes

Various digital user interaction interfaces were used. The delivery modes of the user interaction interfaces varied, including personalized exercise introduction sessions at the beginning of the intervention, digital modes, traditional modes, and hybrids of digital and traditional modes. Further, 2 interventions were supervised, 12 were unsupervised, and 6 were facilitated.

To address skills, face-to-face interaction was commonly preferred (14 intervention groups). Other interaction interfaces included the web (5 intervention groups), smartwatches (2 intervention groups), apps (2 intervention groups), and printed materials (2 intervention groups). To cover knowledge, face-to-face interaction was often used (8 intervention groups), followed by the web (5 intervention groups), printed materials (3 intervention groups), and apps (1 intervention group). Calls, text messages, and activators were not used to target skills or knowledge. To target motivation, all user interaction interfaces, except printed materials, were used. Apps (7 intervention groups), smartwatches (7 intervention groups), and the web (6 intervention groups) were the most commonly integrated interaction interfaces, followed by face-to-face interaction (4 intervention groups), text messages (4 intervention groups), calls (3 intervention groups), and activators (2 intervention groups). The forms of feedback therefore differed, including visual, audio, and tactile feedback. The combination of digital and traditional user interaction modes resulted in significant positive physical activity outcomes.

Altogether, of 20 intervention groups, 14 used personal interactions, and the remaining 6 used solely nonpersonal interactions. The effectiveness of the physical activity of the nonpersonal interaction groups (intervention groups: 4/6, 67% effective) differed from the results of the personal interaction groups (intervention groups: 6/14, 43% effective).

Only Mouton and Cloes [30] reported on the delivered intervention dose, that is, how many attempts were made to contact each participant, how many attempts were received, and how many attempts were acted on by the participants. During their 3-month web-based interventions, the web-based intervention group visited the website, on average, 18 (SD 14) times, and the combined intervention group visited, on average, 39 (SD 21) times; these were much lower than the intended number of visits (potentially 90 visits).

Risk of Bias

Table 2 presents the results of the risk of bias assessment for the 13 included articles. The risk of bias for most of the included studies (11/13, 85%) was rated as *moderate*. Only 2 of the articles showed an overall high risk of bias [30,34]. In the article by Muntaner-Mas et al [34], a convenience sample method was used, which does not exactly randomize the sample, and the allocation of participants was not concealed, thereby introducing a selection bias. In the article by Mouton and Cloes [30], 28% of the participants dropped out of the study, thereby creating a bias and affecting the effectiveness of the study. In cases where the domains of the ROB 2 were not described in the included article, the risk of bias was rated as *unclear*. In 3 articles, the blinding of participants and personnel was not possible (performance bias) [30,34,37]. Furthermore, in 3 articles, the blinding of the outcome assessment was not possible [30,33,38] because subjective methods or a combination of subjective and objective methods was used to measure physical activity (detection bias). Additionally, 2 articles showed a high risk of attrition bias (incomplete outcome data) [30,31]. Other possible biases included small sample sizes [36,39], self-selection bias, baseline differences between study groups, and short intervention periods [35,36,39,40].

Table . Cochrane collaboration tool for assessing risk of bias [28,29].

Author, year (study design)	Selection bias		Detection bias		Attrition bias	Reporting bias	
	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Summary of risk of bias
Albergoni et al [39], 2020 (pilot study)	?[a]	?	+[b]	?	+	+	±[c]
Brew-Sam et al [32], 2022 (quantitative study)	?	?	?	?	+	+	±
Brickwood et al [41], 2021 (RCT[d])	+	+	?	?	+	+	±
Compernolle et al [40], 2020 (mixed methods study)	?	+	?	?	+	+	±
Compernolle et al [35], 2021 (mixed methods study)	?	+	?	?	+	+	±
Li et al [36], 2020 (pilot feasibility study)	?	?	?	?	+	+	±
Mansson et al [42], 2020 (feasibility study)	+	?	+	?	?	+	±
Mouton and Cloes [30], 2015 (parallel group RCT)	+	+	?	–[e]	–	?	–
Müller et al [38], 2016 (2-arm RCT)	+	+	+	–	+	?	±
Muntaner-Mas et al [34], 2017 (pilot 3-group CTS[f])	–	–	?	+	?	?	–
Rowley et al [31], 2019 (RCT)	+	+	+	?	–	+	±
Thompson et al [37], 2014 (RCT)	+	+	?	+	+	?	±
Van Dyck et al [33], 2016 (RCT)	+	+	+	–	+	?	±

[a]Unclear risk of bias.

[b]Low risk of bias.

[c]Moderate risk of bias.

[d]RCT: randomized controlled trial.

[e]High risk of bias.

[f]CTS: clinical trial study.

Discussion

Key Findings

This systematic review, as a primary objective, aimed to identify which of the three activation factors—skills, knowledge, and motivation—were promoted in ICT interventions for older adults and which corresponding BCTs were applied. The secondary objective was to classify the user interaction interfaces and delivery modes that promoted these activation factors. A summary of the main findings from the 13 included articles are presented in Textbox 1.

Textbox 1. Summary of the main findings.

All activation factors

- Integrating all 3 activation factors proved most effective for increasing physical activity levels

Skills

- Promoted in 19 of 20 interventions
- Six BCTs were used; main behavior change techniques (BCTs) were the instruction of optimal behavior performance and the demonstration of behavior

Knowledge

- Promoted in 15 of 20 interventions
- Four BCTs were used; main BCTs were providing information about health consequences, information from credible sources, and information about the salience of consequences

Motivation

- Promoted in 17 of 20 interventions
- Eleven BCTs were used; main BCTs were self-monitoring, feedback on behavior, social support, and goal setting

User interaction interfaces

- A mixture of interfaces raised the chances of addressing all activation factors, which resulted in effective interventions

Digital delivery modes

- Commonly used to target motivation
- Allowed for the automatic detection of activity patterns and personalized feedback

Traditional delivery modes

- Commonly used to target skills and knowledge
- Only successful if they supported digital user interaction interfaces

Subjective and Objective Outcome Measures

In this review, articles with subjective measures reported more positive outcomes than those using objective measures alone. In agreement with literature, articles that used only subjective methods to measure physical activity showed a positive outcome of the intervention, including improved physical activity levels [29]. This however could have possibly been due to self-reporting bias, whereby participants assess their own improvement over the intervention period [43,44]. Nevertheless, in the article that used a combination of subjective and objective factors, the authors also reported that their intervention was effective in improving or maintaining physical activity levels [36].

Delivering Activation Factors

Although a large variety of BCTs were integrated to promote activation factors, our findings show that for each activation factor, only a limited number of BCTs from Michie et al [11] and Dugas et al [12] were exploited. Preissner et al [45] highlighted the relevance of multiple behavioral determinants (ie, social cognitive, habitual, automatic, postintentional, and planning processes) to physical activity intention in older adults. Additionally, adult users themselves have identified a variety of BCTs to fulfill their needs and preferences for engagement with physical activity apps [46]. Our findings coincide with previous reports and indicate that BCTs that integrated a combination of all 3 activation factors indeed showed the most beneficial physical activity effects. This was also reported in a review that analyzed internet-based interventions promoting health behavioral change [47].

Improving skills results in better outcomes of interventions [48]. The most used BCTs for promoting this activation factor in the interventions of the selected studies were instructions on how to perform a behavior and demonstration of a behavior, which have proven effective in improving physical activity levels [49,50]. The integration of some other likely effective BCTs was not reported, or these BCTs were not used (eg, behavioral experiments, habit reversal, graded tasks, or the identification of self as a role model). In other studies, habit reversal [51,52] and graded tasks [50] were shown to be effective in reducing sedentary behavior and improving physical activity. These can as well be integrated into interventions using ICT tools, to increase the chance of effectiveness.

To increase knowledge, 4 BCTs that focused on information about health consequences, information from credible sources, and information about the salience of consequences were integrated. In the literature, providing information about health consequences is likewise often used to target the user's knowledge and has proved effective in reducing sedentary behavior and improving physical activity [49,53]. Other strategies (eg, providing information about social, environmental, or emotional consequences) are also available and can improve effectiveness, as seen in a systematic review with a meta-analysis [54].

To target participants' motivation, 11 BCTs, such as self-monitoring, feedback on behavior, social support, adjusting intervention content to the performance and goal setting, were used. Gamification was not used to motivate participants, even though studies show that it can also be effective as a mode of motivation [12]. Prompts and cues were used in some interventions, but considering their effectiveness in mobile health apps [12], they should be incorporated more frequently in future interventions. Other BCTs, such as self-rewards [50] and reduced rewards [55], were also not used.

User Interfaces and Delivery Modes

Of the 20 interventions, most (n=14, 70%) still used a personal interaction. Of course, personal interaction enables the addressing of users' issues and gaps in skills and knowledge (eg, raising questions and receiving a demonstration of the physical activity behavior). However, based on our findings, interventions with nonpersonal interaction can be just as effective as or even more effective than interventions with personal interactions in promoting physical activity among the population of older adults. Personal interactions are time and cost intensive, and our findings suggest that solely integrating nonpersonal interactions into physical activity interventions could be a good alternative option. However, a combination of various interfaces increases the chance of addressing all activation factors, as shown by op den Akker et al [56]. Evita—a mobile, 3D virtual fitness trainer—guides users through explanations and demonstrations of how to perform exercises. An external sensor, which is attached to a belt on the patient's hip, measures physical activity behavior. Feedback messages (eg, "you have taken more rest – take a nice walk") are also sent to the users. The advantage of digital devices is the possibility of personalized and regular feedback [57,58]. For a hybrid intervention with a mixture of traditional (face-to-face) delivery modes targeting skills and digital (web) delivery modes targeting knowledge and motivation, positive physical activity effects were detected [30]. However, for some individuals, participating with others can still be a source of inspiration to improve or maintain physical activity levels [59].

Issues With Digital Interfaces

We also reported digital interventions without significant beneficial effects on physical activity. Our results indicate that interventions involving the unsupervised usage of digital technology may be challenging for participants because of, for example, problems with log-ins, as well as access to or an abundance of information regarding physical activity (eg, recommendations, success stories, tips for exercise, goal setting, physical activity diaries, tools to measure physical activity, and local physical activity opportunities) [30]. A focus group with 46 older adults affirmed this argument; one of the main results of the focus group was that a web-based physical activity program was preferred to be simple; be not cluttered; and include personalized advice, reminders to check-in, and the ability to review goals [60]. An initial instruction and a helpline are useful for necessary technological support [61].

Limitations and Recommendations

The database search was restricted to studies published only in English. Additionally, a quantitative data synthesis (ie, meta-analysis) was not feasible because the included studies were too dissimilar in terms of the intervention content, duration, assessment of outcome measures, follow-up, and comparator groups. Future articles should consider a standard method of assessing outcome measures (possibly a combination of both subjective methods and objective methods), such as using trackers for activity levels, the Behavioral Regulation in Exercise Questionnaire for covering knowledge and motivation levels [62], and the eHealth Usability Benchmarking Instrument for measuring usability (retention and acceptability) [63]. These would make the results more homogeneous for meta-analyses. Further, using the ROB 2 tool was difficult because some of its criteria could not be accurately applied to public health interventions (eg, blinding of study personnel or participants), or the information needed to determine the categories for risk of bias was not provided or was unclear in the publication. Lastly, during our research, we noticed that some studies did not report or sufficiently describe which BCTs were included.

Conclusions

Motivation was the most promoted activation factor in the ICT interventions. However, integrating BCTs that promote all activation factors resulted in better effects in improving physical activity compared to the effects of using only 1 or 2 activation factors. Although a broad variety of BCTs were used in the articles, they were limited to about 21% (21/102) of available BCTs. Hence, many more BCTs could be exploited in future interventions. Integrating multiple interaction interfaces (eg, interfaces for delivering the intervention program and tracking one's own activity to guarantee regular and personalized feedback) was shown to be the most effective in promoting physical activity. Time-consuming and costly personal interactions are not crucial for increasing physical activity in the older population, though they are effective in supporting digital interactions. At present, study outcomes are too diverse, which hinders intervention comparisons. To make the effects of interventions comparable, future studies should report both objective measures and subjective measures.

Authors' Contributions

EB, JJN, and RD conceptualized this study. EB and JJN were responsible for screening (phase 1). RD was responsible for postscreening (phase 2: conflicts). EB and JJN wrote and prepared the original draft. RD reviewed and edited the manuscript. JJN was responsible for reference management. EB was responsible for project administration and journal submission. RD supervised this study. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study characteristics.
[DOCX File, 67 KB - mhealth_v11i1e42968_app1.docx]

Multimedia Appendix 2
Behavior change technique overview.
[DOCX File, 21 KB - mhealth_v11i1e42968_app2.docx]

Multimedia Appendix 3
High resolution version of Figure 2. User interaction interfaces and delivery modes for delivering activation factors [30-42] Orange blocks show intervention groups withsignificant beneficial findings regarding physical activity. Grey blocks highlight the user interfaces used along with the integrated activation factors. D: digital; FA: facilitated; H: hybrid; K: knowledge; M: motivation; OB: objective; PE: personalized exercise introduction; S: skills; SU: subjective; SUP: supervised; T: traditional; U: unsupervised.
[XLSX File, 43 KB - mhealth_v11i1e42968_app3.xlsx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 97 KB - mhealth_v11i1e42968_app4.pdf]

References

1. UN decade of healthy ageing: plan of action. World Health Organization. 2020. URL: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf [accessed 2023-09-06]
2. Juan SMA, Adlard PA. Ageing and cognition. Subcell Biochem 2019;91:107-122. [doi: 10.1007/978-981-13-3681-2_5] [Medline: 30888651]
3. Costantino S, Paneni F, Cosentino F. Ageing, metabolism and cardiovascular disease. J Physiol 2016 Apr 15;594(8):2061-2073. [doi: 10.1113/JP270538] [Medline: 26391109]
4. Husted KLS, Brink-Kjær A, Fogelstrøm M, Hulst P, Bleibach A, Henneberg K, et al. A model for estimating biological age from physiological biomarkers of healthy aging: cross-sectional study. JMIR Aging 2022 May 10;5(2):e35696. [doi: 10.2196/35696] [Medline: 35536617]
5. Taylor D. Physical activity is medicine for older adults. Postgrad Med J 2014 Jan;90(1059):26-32. [doi: 10.1136/postgradmedj-2012-131366] [Medline: 24255119]
6. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222. [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
7. Blakemore A, Hann M, Howells K, Panagioti M, Sidaway M, Reeves D, et al. Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. BMC Health Serv Res 2016 Oct 18;16(1):582. [doi: 10.1186/s12913-016-1843-2] [Medline: 27756341]
8. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026. [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
9. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
10. Chen J, Mullins CD, Novak P, Thomas SB. Personalized strategies to activate and empower patients in health care and reduce health disparities. Health Educ Behav 2016 Feb;43(1):25-34. [doi: 10.1177/1090198115579415] [Medline: 25845376]

11. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (V1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
12. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020 Feb 20;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
13. Aldawood H, Alabadi M, Alharbi O, Skinner G. A contemporary review of raising health awareness using ICT for application in the Cyber security domain. Presented at: 2019 International Conference in Engineering Applications (ICEA); July 8-11, 2019; Sao Miguel, Portugal p. 1-8. [doi: 10.1109/CEAP.2019.8883454]
14. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198. [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
15. Jimison HB, Hagler S, Kurillo G, Bajcsy R, Pavel M. Remote health coaching for interactive exercise with older adults in a home environment. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5485-5488. [doi: 10.1109/EMBC.2015.7319633] [Medline: 26737533]
16. Knight E, Stuckey MI, Petrella RJ. Health promotion through primary care: enhancing self-management with activity prescription and mHealth. Phys Sportsmed 2014 Sep;42(3):90-99. [doi: 10.3810/psm.2014.09.2080] [Medline: 25295771]
17. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116. [doi: 10.2196/jmir.5429] [Medline: 27334418]
18. Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: qualitative study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e22488. [doi: 10.2196/22488] [Medline: 33464216]
19. Chtourou H, Trabelsi K, H'mida C, Boukhris O, Glenn JM, Brach M, et al. Staying physically active during the quarantine and self-isolation period for controlling and mitigating the COVID-19 pandemic: a systematic overview of the literature. Front Psychol 2020 Aug 19;11:1708. [doi: 10.3389/fpsyg.2020.01708] [Medline: 33013497]
20. McGarrigle L, Todd C. Promotion of physical activity in older people using mHealth and eHealth technologies: rapid review of reviews. J Med Internet Res 2020 Dec 29;22(12):e22201. [doi: 10.2196/22201] [Medline: 33372894]
21. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021 Mar 29;18(3):e1003583. [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
23. Denton F, Power S, Waddell A, Birkett S, Duncan M, Harwood A, et al. Is it really home-based? A commentary on the necessity for accurate definitions across exercise and physical activity programmes. Int J Environ Res Public Health 2021 Sep 1;18(17):9244. [doi: 10.3390/ijerph18179244] [Medline: 34501833]
24. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
25. DiClemente CC, Graydon MM. Changing behavior using the transtheoretical model. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, United Kingdom: Cambridge University Press; 2020:136-149. [doi: 10.1017/9781108677318]
26. Meixner C, Baumann H, Fenger A, Wollesen B. Gamification in health apps to increase physical activity within families. Presented at: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); October 21-23, 2019; Barcelona, Spain p. 15-20. [doi: 10.1109/WiMOB.2019.8923332]
27. Warner LM, French DP. Self-efficacy interventions. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, United Kingdom: Cambridge University Press; 2020:461-478. [doi: 10.1017/9781108677318]
28. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
29. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
30. Mouton A, Cloes M. Efficacy of a web-based, center-based or combined physical activity intervention among older adults. Health Educ Res 2015 Jun;30(3):422-435. [doi: 10.1093/her/cyv012] [Medline: 25772974]
31. Rowley TW, Lenz EK, Swartz AM, Miller NE, Maeda H, Strath SJ. Efficacy of an individually tailored, internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]
32. Brew-Sam N, Chib A, Torres AYF, Ng JXJ, Wong YTJ, Sze-G Y. An integrated mHealth campaign to reduce the risk of falling for older adults. J Appl Gerontol 2022 May;41(5):1336-1347. [doi: 10.1177/07334648211062877] [Medline: 35085043]

33. Van Dyck D, Plaete J, Cardon G, Crombez G, De Bourdeaudhuij I. Effectiveness of the self-regulation eHealth intervention 'MyPlan1.0.' on physical activity levels of recently retired Belgian adults: a randomized controlled trial. Health Educ Res 2016 Oct;31(5):653-664. [doi: 10.1093/her/cyw036] [Medline: 27422898]
34. Muntaner-Mas A, Vidal-Conti J, Borràs PA, Ortega FB, Palou P. Effects of a Whatsapp-delivered physical activity intervention to enhance health-related physical fitness components and cardiovascular disease risk factors in older adults. J Sports Med Phys Fitness 2017;57(1-2):90-102. [doi: 10.23736/S0022-4707.16.05918-1] [Medline: 26364690]
35. Compernolle S, Van Dyck D, Cardon G, Brondeel R. Exploring breaks in sedentary behavior of older adults immediately after receiving personalized haptic feedback: intervention study. JMIR Mhealth Uhealth 2021 May 10;9(5):e26387. [doi: 10.2196/26387] [Medline: 33970109]
36. Li J, Hodgson N, Lyons MM, Chen KC, Yu F, Gooneratne NS. A personalized behavioral intervention implementing mHealth technologies for older adults: a pilot feasibility study. Geriatr Nurs 2020;41(3):313-319. [doi: 10.1016/j.gerinurse.2019.11.011] [Medline: 31810730]
37. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]
38. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 7;18(1):e5. [doi: 10.2196/jmir.5235] [Medline: 26742999]
39. Albergoni A, Hettinga FJ, Stut W, Sartor F. Factors influencing walking and exercise adherence in healthy older adults using monitoring and interfacing technology: preliminary evidence. Int J Environ Res Public Health 2020 Aug 24;17(17):6142. [doi: 10.3390/ijerph17176142] [Medline: 32846988]
40. Compernolle S, Cardon G, van der Ploeg HP, Van Nassau F, De Bourdeaudhuij I, Jelsma JJ, et al. Engagement, acceptability, usability, and preliminary efficacy of a self-monitoring mobile health intervention to reduce sedentary behavior in Belgian older adults: mixed methods study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e18653. [doi: 10.2196/18653] [Medline: 33118951]
41. Brickwood KJ, Ahuja KDK, Watson G, O'Brien JA, Williams AD. Effects of activity tracker use with health professional support or telephone counseling on maintenance of physical activity and health outcomes in older adults: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 5;9(1):e18686. [doi: 10.2196/18686] [Medline: 33399541]
42. Mansson L, Lundin-Olsson L, Skelton DA, Janols R, Lindgren H, Rosendahl E, et al. Older adults' preferences for, adherence to and experiences of two self-management falls prevention home exercise programmes: a comparison between a digital programme and a paper booklet. BMC Geriatr 2020 Jun 15;20(1):209. [doi: 10.1186/s12877-020-01592-x] [Medline: 32539711]
43. Fadnes LT, Taube A, Tylleskär T. How to identify information bias due to self-reporting in epidemiological research. The Internet Journal of Epidemiology 2008;7(2). [doi: 10.5580/1818]
44. Chaabene H, Prieske O, Herz M, Moran J, Höhne J, Kliegl R, et al. Home-based exercise programmes improve physical fitness of healthy older adults: a PRISMA-compliant systematic review and meta-analysis with relevance for COVID-19. Ageing Res Rev 2021 May;67:101265. [doi: 10.1016/j.arr.2021.101265] [Medline: 33571702]
45. Preissner CE, Charles K, Knäuper B, Kaushal N. Predicting decisional determinants of physical activity among older adults: an integrated behavior approach. J Aging Health 2022;34(4-5):569-580. [doi: 10.1177/08982643211049079] [Medline: 34657497]
46. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 8;7(2):e11636. [doi: 10.2196/11636] [Medline: 30735143]
47. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
48. Brach JS, Vanswearingen JM. Interventions to improve walking in older adults. Curr Transl Geriatr Exp Gerontol Rep 2013 Dec;2(4):230-238. [doi: 10.1007/s13670-013-0059-0] [Medline: 24319641]
49. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]
50. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 1;9(1):147-157. [doi: 10.1093/tbm/iby010] [Medline: 29506209]
51. Patterson K, Davey R, Keegan R, Kunstler B, Woodward A, Freene N. Behaviour change techniques in cardiovascular disease smartphone apps to improve physical activity and sedentary behaviour: systematic review and meta-regression. Int J Behav Nutr Phys Act 2022 Jul 7;19(1):81. [doi: 10.1186/s12966-022-01319-8] [Medline: 35799263]
52. Meyer-Schwickerath C, Morawietz C, Baumann FT, Huber G, Wiskemann J. Efficacy of face-to-face behavior change counseling interventions on physical activity behavior in cancer survivors - a systematic review and meta-analysis. Disabil Rehabil 2022 Sep;44(19):5386-5401. [doi: 10.1080/09638288.2021.1938247] [Medline: 34261403]

53. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192. [doi: 10.2196/jmir.3469] [Medline: 25131661]
54. Tang MY, Smith DM, Mc Sharry J, Hann M, French DP. Behavior change techniques associated with changes in postintervention and maintained changes in self-efficacy for physical activity: a systematic review with meta-analysis. Ann Behav Med 2019 Aug 16;53(9):801-815. [doi: 10.1093/abm/kay090] [Medline: 30534971]
55. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2020 Nov;54(21):1259-1268. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
56. op den Akker H, Moualed LS, Jones VM, Hermens HJ. A self-learning personalized feedback agent for motivating physical activity. Presented at: ISABEL '11: International Symposium on Applied Sciences in Biomedical and Communication Technologies; October 26-29, 2011; Barcelona, Spain. [doi: 10.1145/2093698.2093845]
57. Bao L, Intille SS. Activity recognition from user-Annotated acceleration data. In: Ferscha A, Mattern F, editors. Pervasive Computing: Springer; 2004:1-17 ISBN: 3540218351.
58. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71. [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
59. Patterson M, McDonough MH, Hewson J, Culos-Reed SN, Bennett E. Social support and body image in group physical activity programs for older women. J Sport Exerc Psychol 2022 Jul 23;44(5):335-343. [doi: 10.1123/jsep.2021-0237] [Medline: 35894880]
60. Alley SJ, Samra P, Rebar AL, Schoeppe S, Parkinson L, Power D, et al. A focus group study of older adults' perceptions and preferences towards web-based physical activity interventions. Inform Health Soc Care 2020 Sep;45(3):273-281. [doi: 10.1080/17538157.2019.1656210] [Medline: 31690152]
61. Dekker-van Weering M, Jansen-Kosterink S, Frazer S, Vollenbroek-Hutten M. User experience, actual use, and effectiveness of an information communication technology-supported home exercise program for pre-frail older adults. Front Med (Lausanne) 2017 Nov 27;4:208. [doi: 10.3389/fmed.2017.00208] [Medline: 29250523]
62. Castro PC, Romano LB, Frohlich D, Lorenzi LJ, Campos LB, Paixão A, et al. Tailoring digital apps to support active ageing in a low income community. PLoS One 2020 Dec 10;15(12):e0242192. [doi: 10.1371/journal.pone.0242192] [Medline: 33301455]
63. Broekhuis M, van Velsen L. Improving usability benchmarking for the eHealth domain: the development of the eHealth UsaBility Benchmarking instrument (HUBBI). PLoS One 2022 Feb 17;17(2):e0262036. [doi: 10.1371/journal.pone.0262036] [Medline: 35176033]

Abbreviations

BCT: behavior change technique
ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROB 2: Risk of Bias 2

Edited by L Buis; submitted 16.12.22; peer-reviewed by M Katapodi, M Metcalf; revised version received 14.06.23; accepted 04.08.23; published 24.10.23.

Please cite as:
Bentlage E, Nyamadi JJ, Dubbeldam R
The Importance of Activating Factors in Physical Activity Interventions for Older Adults Using Information and Communication Technologies: Systematic Review
JMIR Mhealth Uhealth 2023;11:e42968
URL: https://mhealth.jmir.org/2023/1/e42968
doi:10.2196/42968

Efficacy of mHealth Interventions for Improving the Pain and Disability of Individuals With Chronic Low Back Pain: Systematic Review and Meta-Analysis

Bruna de Melo Santana[1,*], MSc; Julia Raffin Moura[1,*], BSc; Aline Martins de Toledo[1,*], PhD; Thomaz Nogueira Burke[2,*], PhD; Livia Fernandes Probst[3,*], PhD; Fernanda Pasinato[1,*], PhD; Rodrigo Luiz Carregaro[1,*], PhD

1

2

3

[*]all authors contributed equally

Corresponding Author:
Rodrigo Luiz Carregaro, PhD

Abstract

Background: Low back pain is one of the main causes of disability worldwide. Individuals with chronic conditions have been widely affected by the COVID-19 pandemic. In this context, mobile health (mHealth) has become popular, mostly due to the widespread use of smartphones. Despite the considerable number of apps for low back pain available in app stores, the effectiveness of these technologies is not established, and there is a lack of evidence regarding the effectiveness of the isolated use of mobile apps in the self-management of low back pain.

Objective: We summarized the evidence on the effectiveness of mHealth interventions on pain and disability for individuals with chronic low back pain.

Methods: We conducted a systematic review and meta-analysis comparing mHealth to usual care or no intervention. The search terms used were related to low back pain and mHealth. Only randomized controlled trials were included. The primary outcomes were pain intensity and disability, and the secondary outcome was quality of life. Searches were carried out in the following databases, without date or language restriction: PubMed, Scopus, Embase, Physiotherapy Evidence Database (PEDro), the Cochrane Library, and OpenGrey, in addition to article references. The risk of bias was analyzed using the PEDro scale. Data were summarized descriptively and through meta-analysis (pain intensity and disability). In the meta-analysis, eligible studies were combined while considering clinical and methodological homogeneity. The certainty of evidence was assessed using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) criteria.

Results: A total of 5 randomized controlled trials were included, totaling 894 participants (447 allocated to the mHealth group and 445 to the usual care group), and they had similar methodological structure and interventions. Follow-up ranged from 6 weeks to 12 months. The studies did not demonstrate significant differences for pain intensity (mean difference −0.86, 95% CI −2.29 to 0.58; *P*=.15) and disability (standardized mean difference −0.24, 95% CI −0.69 to 0.20; *P*=.14) when comparing mHealth and usual care. All studies showed biases, with emphasis on nonconcealed allocation and nonblinding of the outcome evaluator. The certainty of evidence was rated as low for the analyzed outcomes.

Conclusions: mHealth alone was no more effective than usual care or no treatment in improving pain intensity and disability in individuals with low back pain. Due to the biases found and the low certainty of evidence, the evidence remains inconclusive, and future quality clinical trials are needed.

Trial Registration: PROSPERO CRD42022338759; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=338759

(JMIR Mhealth Uhealth 2023;11:e48204) doi:10.2196/48204

KEYWORDS

physiotherapy; back pain; mobile technology; efficacy; disability; chronic condition; chronic; effectiveness; self-management; systematic review; pain; meta-analysis; treatment; mHealth; mobile health

Introduction

Low back pain is one of the main causes of years lived with disability in all people aged ≥18 years in the world [1] and causes serious socioeconomic problems due to its high health care costs in several countries [2-4]. For example, the annual costs for this health condition have been estimated to be approximately US $200 billion in the United States, including direct health care spending and indirect costs due to productivity losses and reduced quality of life [5]. In Brazil, between 2012 and 2016, the societal costs (treatment and productivity losses) arising from low back pain were estimated to be US $2.2 billion [6].

Low back pain is recognized for its high prevalence in all age groups ≥18 years. Globally, the prevalence of this condition was estimated to be at 7.5% in 2017, representing approximately 577 million people worldwide [7]. It is worth noting that people with low back pain are frequent users of health and social care services, which causes high expenses [6,8,9]. Thus, currently, one of the great challenges is to use effective strategies to manage this condition and avoid unnecessary expenses [9]. In this context, self-management of low back pain is recommended by international clinical guidelines [10,11]. This strategy involves care programs that facilitate the management and monitoring of the health condition itself, to enable the individual to manage symptoms as well as lifestyle changes [12-14]. It is recommended that self-management includes exercise and psychotherapy to limit the use of drugs and surgical procedures in clinical practice [11,15,16].

In the past decades, there has been a growth in the use of technological resources as a means for health promotion [17,18]. One of the main resources is mobile health (mHealth), which uses mobile and wireless technologies (eg, mobile phones, patient monitoring devices, and virtual assistants) [19,20]. One of the main advantages of mHealth is easy access and usability, as well as applicability in monitoring a health condition [21]. In addition, mHealth can encourage self-management actions; provide greater speed and practicality in the delivery of information; and promote adherence to treatment and other care, including for individuals with low back pain [22-24].

Despite the considerable number of apps for low back pain available in app stores, the effectiveness of these technologies is not established, and most are of low quality [25,26]. Notwithstanding, recent systematic reviews [5,27] have demonstrated positive results using eHealth (eg, the delivery of health resources via traditional internet and interventions with computer access) in the context of self-management of low back pain while considering different outcomes, such as pain and disability. Regarding mHealth, Chen et al [28] demonstrated that this modality combined with usual care (eg, SMS text messages, telephone calls, real-time monitoring, exercises, and counseling) improved the pain intensity and disability of individuals with low back pain. However, the review had limitations, including searches being restricted to the English language and possible selection biases (eg, there was no registration of the protocol, and the authors did not present a list of excluded studies during the full-text reading). Additionally, the review did not analyze the certainty of evidence nor discussed the impacts of the risk of bias of the included studies. Thus, there is a lack of evidence regarding the effectiveness of the isolated use of mobile apps, without interaction with therapists, in the self-management of low back pain.

Accordingly, this study aimed to synthesize updated data focusing on studies that investigated the use of apps for mobile devices (ie, smartphone back pain apps) as the only form of intervention for people with low back pain, without interaction with therapists. This aspect is relevant, given the impact of the COVID-19 pandemic and the subsequent increase in the number of apps available and the use of remote treatments [18,29]. Thus, the aim of this study was to investigate the effectiveness of mHealth interventions in improving the pain intensity and disability of individuals with chronic low back pain, compared to no intervention or usual health care strategies.

Methods

Overview

This systematic review is reported according to the recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [30]. The protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42022338759).

Eligibility Criteria

Randomized controlled trials were eligible if they met the inclusion criteria, as defined in Table 1 according to the Population, Intervention, Comparators, and Outcomes question.

The search did not restrict the year or language of publication of the studies. Studies that investigated individuals with specific low back pain and studies that used apps with interference or contact with the therapist during the intervention period were excluded.

Table . Eligibility criteria for the study according to the PICO[a] question.

PICO question item	Inclusion criteria
Population	• Economically active adult population (aged 18-59 y) with nonspecific low back pain for more than 3 mo
Intervention	• mHealth[b] technology [27]
Comparators	• No intervention or usual care (eg, maintenance of medical and pharmacological care or counseling regarding physical activity and exercise prescription) [31]
Outcomes	• Primary outcomes: pain intensity and disability • Secondary outcome: quality of life
Study design	• Randomized controlled trials

[a]PICO: Population, Intervention, Comparators, and Outcomes.

[b]mHealth: mobile health.

Information Sources

Systematic searches were performed in the following databases, with no restriction on publication date: MEDLINE (via PubMed), Scopus, Embase, Physiotherapy Evidence Database (PEDro), and the Cochrane Library, in addition to gray literature (via OpenGrey [32]). The references of the included studies were also screened, and the entire search process took place between December 13 and 26, 2022.

Search Strategy

Search strategies were composed of controlled vocabulary terms and words, according to each database. Terms referring to the investigated condition (low back pain) were combined with terms referring to the intervention of interest (mHealth). No search filters were used for study design, and the search was individually adapted for each database (Multimedia Appendix 1). The search strategy was validated by an experienced librarian.

Screening Process

The studies retrieved in the search were uploaded to Rayyan software (Rayyan Systems Inc) [33]. After confirming and deleting the duplicates, 2 reviewers independently performed the screening by title and abstract. Any disagreement between the reviewers at this stage resulted in the inclusion of the study in the full-text reading stage. Authors of registered protocols were contacted to confirm the publication of data. The second selection phase was carried out by the same reviewers independently, taking into account the eligibility criteria. Any disagreements were resolved through discussion and consensus.

Data Collection Process

The data extraction process of the included studies was performed by 2 reviewers independently; they used a previously prepared and standardized form for this review.

Data List

The information extracted included the sample size, the intervention type of the experimental and control group, the duration of the intervention, the outcomes, sources of funding, and the declaration of conflicts of interest.

Assessment of the Risk of Bias

The risk of bias in the included studies was assessed using the PEDro scale [34]. This step was performed by 2 independent assessors, with subsequent consensus. The PEDro scale contains 11 criteria to be considered from the study analysis, and each item is equivalent to 1 point in the total score of the scale. The final score ranges from 0 to 10, and the first item (eligibility) must be disregarded in the score.

Effect Measures

The following effect measures were extracted from the included studies: means and SDs for pain intensity, disability, and quality-of-life outcomes.

Synthesis Methods

For the meta-analysis, the primary outcomes were considered. To combine the results, the eligible studies were analyzed while considering the clinical and methodological homogeneity and the follow-up period of the intervention. Mean differences and 95% CIs were used as an effect measure for the pain intensity outcome. For disability, standardized mean differences and 95% CIs were calculated, grouped with Hedges correction, in view of the differences in the scales of the disability instruments adopted in the studies (eg, differences in scales and direction of effects). The values were multiplied by −1 to restore effect direction [35].

The random effects model with Knapp-Hartung adjustment [36] was used in the calculation of both outcomes. Heterogeneity was assessed by visual analysis of the similarity of point estimates and overlapping of CIs and using the χ^2 test and I^2 measure. The results were considered heterogeneous when the I^2 values were >50% and P<.10 for χ^2 values [35]. Meta-analyses were performed using the SPSS software (version 29.0; IBM Corp).

Assessment of Publication Bias

We planned to perform publication bias analysis if there were more than 10 studies included in the same comparison, by visual inspection of the funnel plot and the Egger statistical test.

Assessing the Certainty of Evidence

Certainty in the final set of evidence was assessed using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) criteria. The 5 items of the GRADE criteria were analyzed: methodological limitations (risk of bias), inconsistency, indirectness, imprecision, and publication bias. Each of these criteria has items to be judged through a qualitative assessment of the evidence for each analyzed outcome, allowing the classification of confidence in the estimate of effects as high, moderate, low, or very low, thus making it possible to reduce or increase the level of evidence [37]. In this evaluation, pain intensity and disability were considered critical outcomes. This evaluation was performed in GRADEpro software (McMaster University and Evidence Prime Inc).

Adverse Events and Adherence

We extracted information pertaining to the number of adverse events and intervention adherence in the included studies. Adverse effects were defined as unintended responses that occur during or after an intervention but are not necessarily caused by a causal relationship to the trial intervention. An adverse event was defined as an event for which the causal relation between the intervention and the event is at least a reasonable possibility [38].

Results

Study Selection

A total of 1824 publications relevant to the review were identified. After the exclusion of duplicates and selection by title and abstract, 18 were considered eligible for full-text reading, according to the inclusion and exclusion criteria. Subsequently, 5 publications [22,39-42] were included after the full-text reading (Figure 1). The 13 excluded studies during the full-text phase are described in Multimedia Appendix 2 with the exclusion justifications.

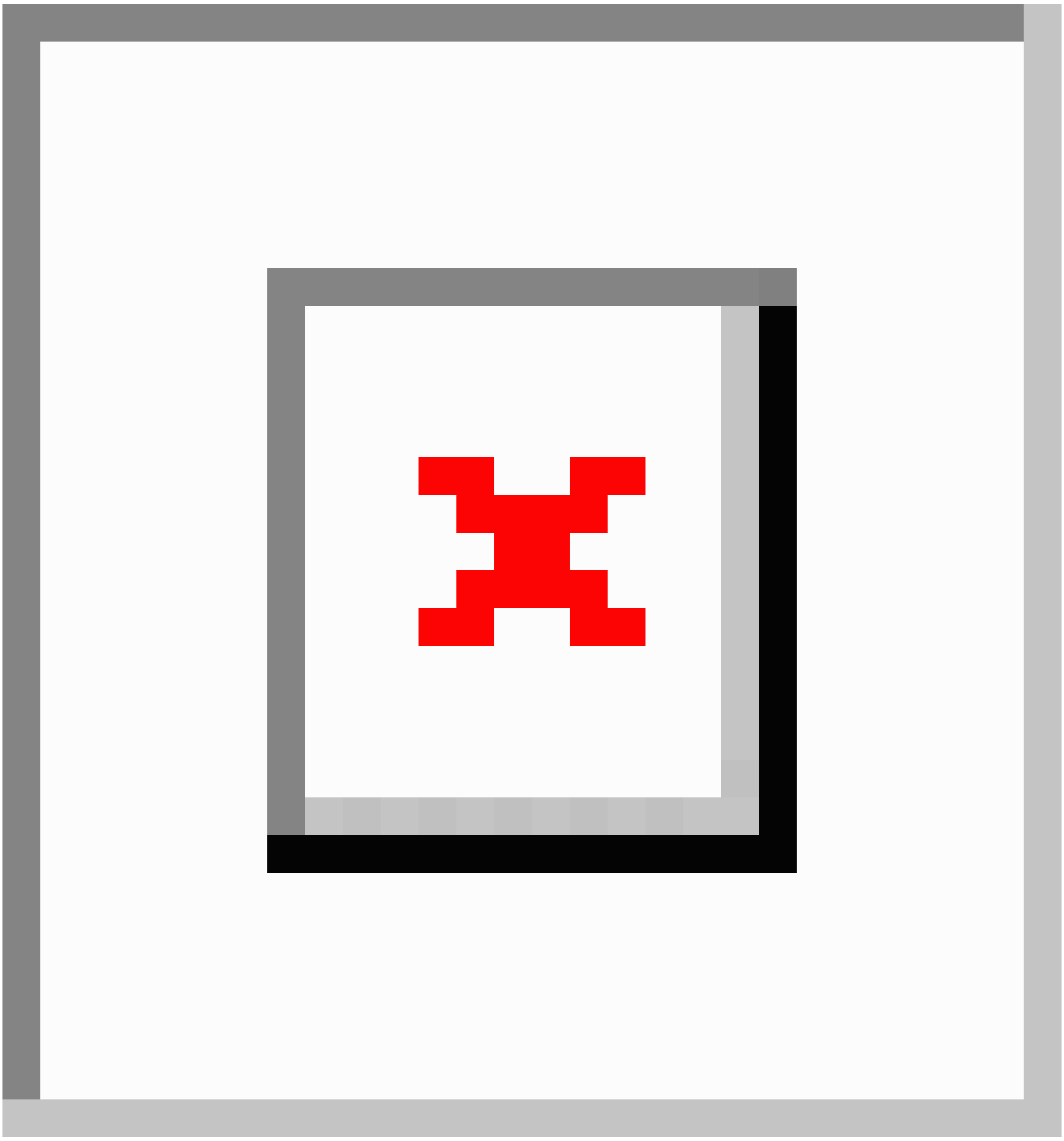


Figure 1. Flowchart of the screening and selection of studies. PEDro: Physiotherapy Evidence Database.

Characteristics and Results of Individual Studies

The included studies had a total of 894 participants (447 allocated to the mHealth group and 445 to the usual care group) and similar methodological structure and interventions. Follow-up ranged from 6 weeks to 12 months, and the studies evaluated the pain intensity and disability outcomes. The characteristics of the included studies and the findings are shown in Table 2. The studies were conducted in Jordan, India, Denmark and Norway, and Germany.

Table . Characteristics of the studies included in the review.

Characteristics	Included studies (reference and published year)				
	Almhdawi et al [39], 2020	Chhabra et al [22], 2018	Sandal et al [40], 2021	Toelle et al [41], 2019	Weise et al [42], 2022
Study design	Pilot RCT[a] with follow-up at 6 wks	RCT with follow-up at 12 wks	RCT with follow-up at 9 mo	RCT with follow-up at 12 wks	RCT with follow-up at 12 wks
Protocol number (record)	NCT03994458	Not reported	NCT03798288	DRKS00016329	DRKS00022781
Country	Jordan	India	Denmark and Norway	Germany	Germany
Study period	January to August 2019	Beginning September 2016; no information on the end date	March to December 2019	August 2017 to October 2018	August 2020 to April 2021
Population	Office workers for more than 5 y between 30 and 55 y of age, with low back pain for more than 12 wk	Individuals over 18 y of age, with persistent chronic low back pain for more than 12 wk with or without the presence of radicular symptoms	Participants aged 18 y or older, with nonspecific low back pain for more than 8 wk	Participants with nonspecific low back pain between 18 and 65 y of age[b], with continuous pain for more than 6 wk	Participants >18 y of age, with nonspecific low back pain
Participants, n	39 (20 intervention and 19 control)	93 (45 intervention and 48 control)	461 (232 intervention and 229 control)	86 (42 intervention and 44 control)	215 (108 intervention and 107 control)
Analysis	Per protocol	Intention to treat	Intention to treat	Per protocol	Intention to treat
Intervention	"Relieve my back" provides general advice, instructions, and stretching and strengthening exercises for lower back and abdominal muscles. Four phone notifications (sound and vibration with a pop-up window) were used to notify participants on taking breaks for walking, correct posture reminders, and exercise reminders.	Snapcare app+written prescription aimed to motivate, promote, and guide participants to increase their level of physical activity and adherence to exercise, including lumbar and aerobic exercises.	selfBACK app+usual care provides individualized weekly self-management recommendations for 3 key components: physical activity (number of steps), strength and flexibility exercises, and daily education messages. In addition, the app provides general information about low back pain and access to various tools (goal setting, mindfulness audios, pain relief exercises, and sleep reminders)	Kaia app involves 3 therapy modules: specific education for back pain, physical therapy or physical exercise, and mindfulness and relaxation techniques.	ViViRA provides a self-directed home exercise program using the principles of movement therapy and functional regional interdependence, plus daily reminders displayed as a notification.
Comparator	Placebo app (nutrition advice posts, along with 4 notifications: sound and vibration, along with a pop-up window with instructions) containing nutritional information unrelated to low back pain	Prescription drugs and their dosages and physical activity	Medical treatment and instruction to manage the condition according to clinician advice	Individual face-to-face sessions of standard physiotherapy once a wk (physical exercises and manual therapy). Encouragement to perform the physiotherapeutic exercises at home and to an active lifestyle. Weekly emails with a brief motivating message and links to medically oriented websites providing web-based resources for patient education about pathophysiology, diagnoses, treatment, and self-management in low back pain.	Physical exercises lasting 15 to 25 min, guided by a certified physiotherapist

Characteristics	Included studies (reference and published year)				
	Almhdawi et al [39], 2020	Chhabra et al [22], 2018	Sandal et al [40], 2021	Toelle et al [41], 2019	Weise et al [42], 2022
Duration of intervention	6 wk	12 wk	6 wk	6 wk	12 wk

Characteristics	Included studies (reference and published year)				
	Almhdawi et al [39], 2020	Chhabra et al [22], 2018	Sandal et al [40], 2021	Toelle et al [41], 2019	Weise et al [42], 2022
Outcomes	• Pain intensity: VAS[c] • Disability: ODI[d] • Quality of life: 12-item Short-Form Health Survey • Sleep quality: Pittsburgh Sleep Quality Index • Physical activity level: International Physical Activity Questionnaire	• Pain intensity: NPRS[e] • Disability: MODI[f]	• Disability: RMDQ[g] • Pain intensity: NRS[h] • Confidence in the ability to cope despite pain: Pain Self-Efficacy Questionnaire • Fear-avoidance: Fear-Avoidance Beliefs Questionnaire, physical activity subscale • Cognitive and emotional representations of disease: Brief Illness Perception Questionnaire • Quality of life: EuroQol-5 Dimension questionnaire and EuroQol VAS • Level of physical activity during leisure time: Saltin-Grimby Physical Activity Level • General improvement: Global Perceived Effect Scale	• Pain intensity: NRS • Functional measures: HFAQ[i] • Behavioral measures: GCPS[j] • Quality of life: VR-12[k]	• Pain intensity: VNRS[l]
Results	At the 6-wk assessment, pain intensity showed a significant reduction in the app group (mean 2.30, SD 2.13) compared to the control group (mean 5, SD 1.97; *P*<.001). There was also a significant reduction in disability in the app group (mean 20.25, SD 13.47) compared to the control group (mean 30.63, SD 10.63; *P*=.002). Regarding quality of life, there was a significant change in the physical component of the app group (mean 79.95, SD 16.09) compared to the control group (mean 62.26, SD 19.76; *P*=.001), a trend that was not followed by the mental component (*P*=.68).	Regarding pain intensity, no significant differences (*P*>.05) were found between the groups over time. As for disability, the scores at baseline were significantly different between the groups: mean 52.1 (SD 14.4) for the app group and mean 20.2 (SD 17.8) for the control group (*P*=.03). Nevertheless, after 12 wk of intervention, the app group (mean 41.4, SD 18.8) registered a significant improvement in disability compared to the control group (mean 29.9, SD 20.1; *P*=.001).		Both groups reported a reduction in pain intensity over time, but the app group reported a significantly lower pain intensity (mean 2.70, SD 1.51) at 12 wk compared to the control group (mean 3.40, SD 1.63; *P*=.02). Regarding disability and quality of life, no significant differences (*P*>.05) were observed between the groups, although both showed improvement over time.	Pain intensity showed a significant reduction in favor of mHealth[m] at all times (2, 6, and 12 wk; *P*<.001; Cohen *d*>0.8).

Characteristics	Included studies (reference and published year)				
	Almhdawi et al [39], 2020	Chhabra et al [22], 2018	Sandal et al [40], 2021	Toelle et al [41], 2019	Weise et al [42], 2022
			Pain intensity showed a reduction in the app group (mean 3.3, SD 2.2) compared to the control group (mean 3.9, SD 2.4; *P*=.001) at the 3-mo assessment, and this effect was maintained at the 9-mo assessment. Disability showed a significant improvement at 3 mo for the app group (mean 6.7, SD 4.7) compared to the control group (mean 7.4, SD 5.4; *P*=.03). This effect was maintained at 9 mo, but in an attenuated form: mean 6.0 (SD 5.3) for the app group and mean 6.9 (SD 5.6) for the control group. Quality of life showed no significant difference (*P*>.05) between groups at the 3- and 9-mo assessments.		
Funding sources	Jordan University of Science and Technology and Erasmus + Program of the European Union	Snapcare Technologies Pvt. Lt	European Union Horizon 2020 research and innovation programme	Kaia Health Software GmbH, Munich, Germany	ViViRA Health Lab GmbH

Characteristics	Included studies (reference and published year)				
	Almhdawi et al [39], 2020	Chhabra et al [22], 2018	Sandal et al [40], 2021	Toelle et al [41], 2019	Weise et al [42], 2022
Conflicts of interest	None declared.	None declared.	"Dr Kjaer reported receiving personal fees from UCL University College outside the submitted work. No other disclosures were reported."	None declared.	"HW, BZ, MB, DS, and KW were responsible for devising the study design and overseeing the study and data analysis. They are researchers, clinicians, and statisticians who are independent of ViViRA Health Lab GmbH. They received salaries (BZ, MB, and DS) or honoraria (HW and KW) for their involvement in the study. BS and LB are employed by ViViRA Health Lab GmbH."

[a]RCT: randomized controlled trial.

[b]Although the authors have considered participants outside the age range of our inclusion criteria (ie, participants up to 65 y of age), we decided to include it because we observed that few participants aged >59 years were included.

[c]VAS: Visual Analogue Scale.

[d]ODI: Oswestry Disability Index.

[e]NPRS: Numeric Pain Rating Scale.

[f]MODI: Modified Oswestry Disability Index.

[g]RMDQ: Roland-Morris Disability Questionnaire.

[h]NRS: Numeric Rating Scale.

[i]HFAQ: Hannover Functional Ability Questionnaire.

[j]GCPS: Graded Chronic Pain Scale.

[k]VR-12: Veterans RAND 12-Item Health Survey.

[l]VNRS: Verbal Numerical Rating Scale.

[m]mHealth: mobile health.

Based on data extraction, a summary of the results of the included studies was performed (Multimedia Appendix 3 [22,39-42]), containing the means and SDs for the pain intensity, disability, and quality-of-life outcomes. Overall, the studies reported benefits of mHealth in pain intensity, disability, and quality of life.

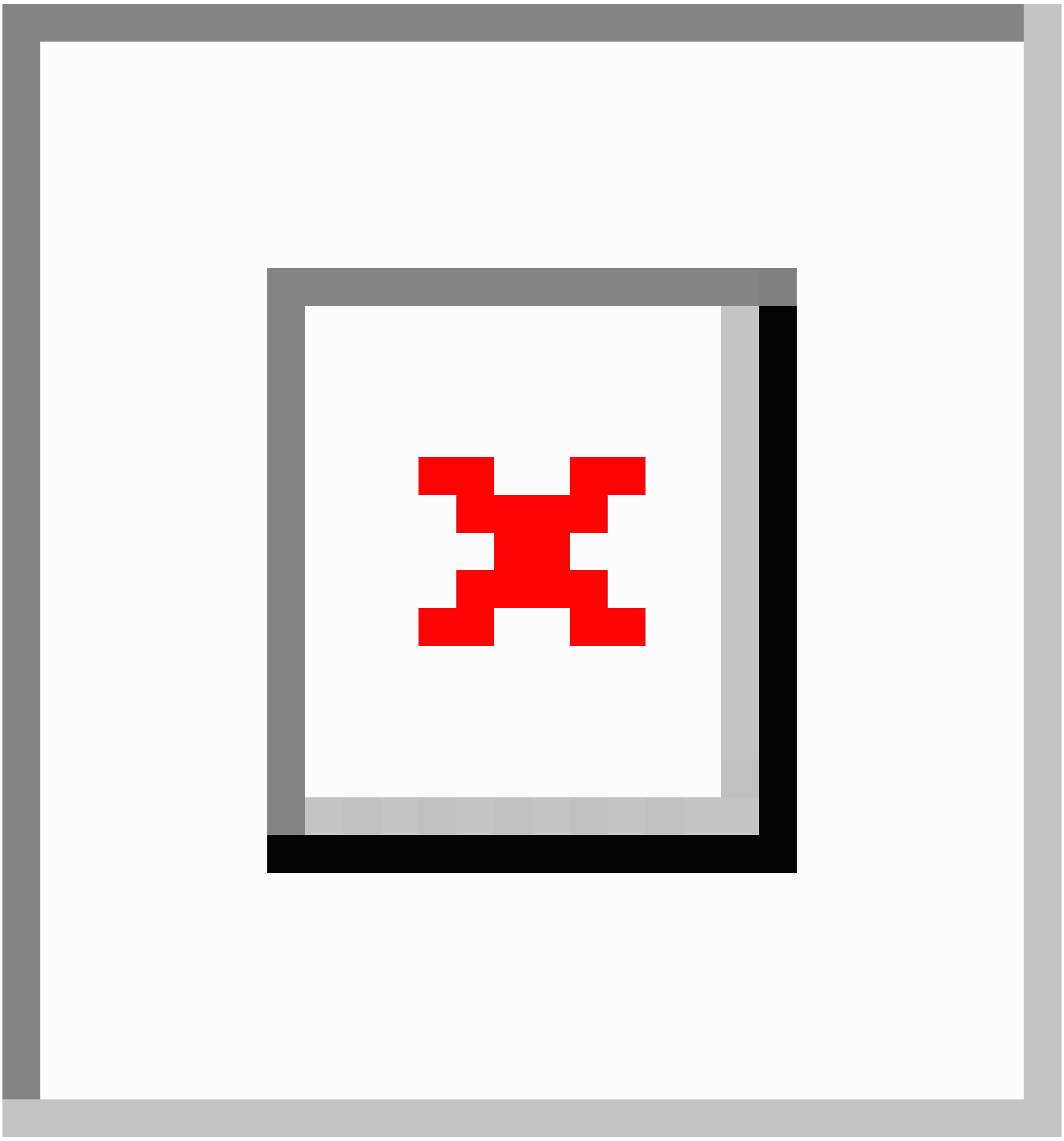
Synthesis Results

A meta-analysis was carried out for the pain intensity and disability outcomes, consisting of 4 of the 5 included studies [22,40-42] that adopted a follow-up of 12 weeks.

Pain Intensity

Of the studies included in the meta-analysis, 3 studies [22,40,41] used the Numeric Pain Rating Scale [43] and 1 study [42] used the Verbal Numerical Rating Scale to assess pain intensity. Both scales assess and rate pain from 0 to 10 points, where 0 represents the absence of pain and 10 represents intense pain [44,45]. The effects were classified as low-quality evidence (Figure 2 [22,40-42]).

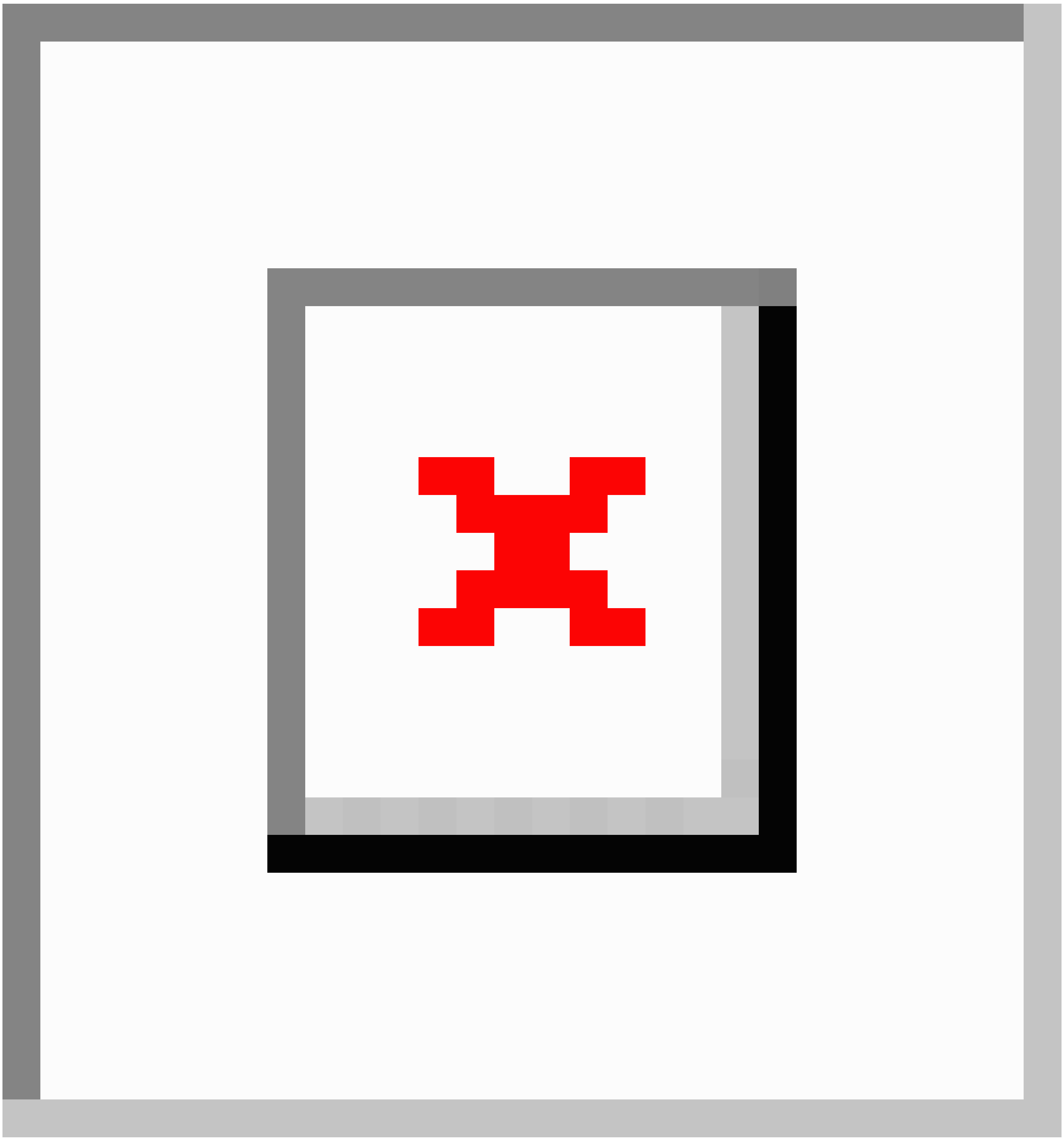
Figure 2. Forest plot of pain intensity (app: mobile health; control: usual care).



Disability

The Modified Oswestry Disability Index [22], Roland-Morris Disability Questionnaire [40] and Hanover Functional Ability Questionnaire [41] were used to assess disability. The effects were classified as low-quality evidence, and no significant differences were found between mHealth compared to usual care (*P*=.14; Figure 3 [22,40,41]).

Figure 3. Forest plot for disability (app: mobile health; control: usual care).



Risk of Bias of the Included Studies

The assessment of the risk of bias in the included studies is presented in Table 3. In all, 4 studies were classified with a final score of 7 and 1 study was classified with a score of 5. Overall, the nonblinding of participants and outcome assessors were common biases. It is worth noting that none of the included studies adopted the blinding of therapists.

Table . Risk of bias of included studies using the Physiotherapy Evidence Database (PEDro) scale.

Studies	PEDro scale items									
	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	Total score
Toelle et al [41]	Y[k]	N[l]	Y	N	N	N	Y	N	Y	Y	5
Chhabra et al [22]	Y	Y	N	N	N	N	Y	Y	Y	Y	6
Sandal et al [40]	Y	Y	Y	N	N	N	Y	Y	Y	Y	7
Almhdawi et al [39]	Y	N	Y	Y	N	Y	Y	N	Y	Y	7
Weise et al [42]	Y	Y	Y	N	N	N	Y	Y	Y	Y	7

[a]1: Participants were randomly distributed.
[b]2: Concealed allocation.
[c]3: Initially, the groups were similar.
[d]4: All participants were blinded.
[e]5: All therapists administered the therapy blindly.
[f]6: All evaluators measured the results blindly.
[g]7: Measurement of key outcomes were obtained in more than 85% of participants.
[h]8: All participants received the treatment as allocated, or the analysis was done by intention to treat.
[i]9: The results of the comparisons were described in at least 1 key result.
[j]10: The study presents both measures of accuracy and variability.
[k]Y: Yes, item met.
[l]N: No, item not met.

Publication Bias

It was not possible to perform publication bias analysis through visual inspection of the funnel plot and the Egger statistical test since only 5 studies were included. However, we consider the probability of publication bias to be low, since the searches were sensitive and gray literature was also consulted.

Certainty of Evidence

The certainty of evidence of mHealth effects was rated as low quality for both outcomes (pain intensity and disability). Details of the evidence profile are presented in Multimedia Appendices 4 and 5.

Adverse Events and Adherence

Only 2 studies [41,42] reported nonserious adverse events; however, there was no clear definition pertaining to the occurrence and severity. Weise et al [42] reported several nonserious adverse events and nonserious adverse reactions not requiring the interruption of the intervention (eg, nausea, pain increase, and transient muscle cramp). Moreover, Toelle et al [41] reported 1 participant in the mHealth group being diagnosed with a lumbar disk herniation; however, this event was deemed not related to the intervention.

Adherence to mHealth interventions was monitored by different methods and definitions. For instance, some studies defined adherence as the number of complete active days of app use [41,42], average time using the app [40], or the number of plans for self-management using the app during the first 12 weeks after randomization [39]. Participants receiving mHealth interventions had a higher adherence compared to the control group (ie, placebo app) [39]. The authors reported that participants in the mHealth group had, on average, 6 times higher daily use of the app than the control group participants. In addition, Toelle et al [41] estimated that participants used the app, on average, for 35 days within the 12 weeks of follow-up, and Sandal et al [40] demonstrated an adherence of 78% of app use. Although adherence was not associated with symptom improvements, 1 study highlighted a higher frequency of app use when pain severity was higher [42].

Discussion

Our systematic review synthesized recent evidence on the use of mHealth technology in the management of individuals with low back pain. We found 5 studies totaling 894 patients, which reported positive effects on improving pain intensity and disability. However, we found a low certainty of evidence in favor of mHealth, and our meta-analyses showed no significant differences between mHealth versus usual care or no intervention (pain intensity: P=.15 and disability: P=.14). There were no reports of serious adverse events.

Even though our review demonstrated no significant differences between mHealth versus usual care or no intervention, the adoption of mHealth provided some beneficial effects in reducing pain intensity in people with low back pain. The combined effect of the included studies was approximately 0.9 (95% CI –2.29 to 0.58) points of improvement, demonstrating that a portion of the participants benefited, specifically those who had a score above 2 points [44]. Likewise, we found no significant differences in reducing disability, which was associated with a small effect size of 0.24 (95% CI –0.69 to 0.20) in favor of mHealth. However, the study by Almhdawi et al [39] investigated the use of a mobile app in office workers with low back pain and observed an effect size of Cohen d=1.08, which was considered large. It is worth noting that, despite the use of effect size measures in meta-analyses composed of standardized means, this interpretation is still considered conflicting [46]. In this context, previous studies have shown that the minimal clinically important difference in disability for low back pain is at least a 30% reduction in the score of the scales [47,48], and the findings of our study were below this threshold. Interestingly, Zheng et al [49] demonstrated that exercise combined with self-management training delivered via mHealth presents a faster improvement in disability when compared to exercise alone via mHealth. Thus, considering the findings of these previous studies, it is possible to assume that mHealth provides, to some extent, clinically relevant effects for the management of low back pain.

We found that the quality of life of the participants improved after the use of mHealth; however, this difference was not significant compared to usual care. Among the 3 studies that investigated quality of life, Sandal et al [40] and Toelle et al [41] found no differences between mHealth and usual care. These findings corroborate those of Schlicker et al [50], which also showed no significant differences between mHealth and usual care. The study by Zheng et al [51] investigated the effects of exercise delivered via mHealth, with and without a health education process, and found significant improvements in the physiological functional aspects of quality of life in both groups. Likewise, Almhdawi et al [39] found a large effect size in favor of mHealth (Cohen d=1.18), specifically for the improvement of the physical component of quality of life, but did not find improvement in the mental component. These results indicate that the effects of mHealth on quality of life are still conflicting. The quality of life is influenced by cultural, physical, and social aspects, which makes it difficult to compare the results considering different contexts [52]. In addition, the improvement in the quality of life is more related to the improvement of disability than to pain intensity [53], and in our study, disability presented a small effect size, which may reflect the nonsignificant difference found for the quality of life.

A recent review [21] carried out a qualitative synthesis of the evidence on the perceptions and experiences of health professionals regarding the use of mHealth. The findings showed advantages arising from mHealth, such as the optimization of tasks, the speed of delivery of information, and the possibility of monitoring these patients remotely and recording data about their routine. Other studies [54,55] have shown that the satisfaction of patients who used digital interventions is similar to those who receive face-to-face care, with emphasis on the ease of use, efficiency in communication, and low cost, in addition to technology overcoming distance barriers. Another advantage is the fact that the technologies are based on active interventions, which focus on physical exercise and self-management—strategies that are considered effective to treat patients with musculoskeletal conditions [56]. Thus, mHealth can be a valuable tool for symptom control in patients with chronic low back pain. Nevertheless, factors such as adherence and the individual's ability to manage their symptoms can have a determining effect on the clinical relevance of the results. In this sense, it is suggested that strategies that favor adherence and self-efficacy should be included in the service packages delivered by mHealth. Therefore, individualized strategies are recommended, given that the use of technological resources can be a positive factor for better adherence to treatment [57].

All included studies in our review showed some methodological biases. None of the 5 included studies blinded the therapists and 4 did not blind the patients [22,40-42]. It is noteworthy that 2 studies did not adopt concealed allocation [39,41]; 4 studies did not adopt the blinding of outcome assessors [22,40-42]; and in 1 study [22], the scores at baseline were significantly different between groups. The occurrence of biases is relevant because they may overestimate or underestimate the effect of interventions [58,59]. Concealed allocation refers to how participants are allocated to groups, and an inadequate allocation increases the estimate of effect size and can generate a difference in the investigator's approach to participants, causing selection bias [60,61]. Studies that adopted an adequate blinding process showed less predisposition to findings that favored a given intervention [62]. Thus, inadequate blinding is a factor associated with biases and can alter the conduct of participants and researchers, who can change their behavior [63]. However, it is not always possible to blind therapists and participants, mainly due to the characteristics of interventions in certain areas (eg, the adoption of exercise and booklets) [64]. Two studies [39,41] did not perform the analysis of participants according to allocation; in these cases, participants who do not comply with the initial protocol are not considered, resulting in the loss of the benefits of randomization. This fact increases the risk of selection bias and the probability that changes are attributable to external factors or confounding variables [65].

Our review has the following strengths. Initially, we highlight the fact that we investigated the isolated effect of mHealth compared to usual care or no intervention in people with low back pain. This aspect reduced the risk of heterogeneity regarding the intervention and divergences in interpretations [66], which is contrary to previous studies [5,27,28]. Moreover, we took steps to minimize bias, such as including a minimum of 2 reviewers to independently assess the studies for inclusion and carry out the data extraction. In addition, 2 other independent reviewers carried out the risk-of-bias and certainty-of-evidence assessments. Furthermore, a comprehensive search strategy was adopted, comprising the major databases without language or date restrictions.

As a limitation, our review included a small number of studies due to the eligibility criteria, which favored the inclusion of a clinically homogeneous intervention. A second limitation was differences in the target audience of the included studies. The most heterogeneous study [39] carried out the research in a specific environment (ie, office), whereas the other studies included individuals from the general population. A third limitation concerns the biases present in the included studies, mainly the absence of concealed allocation and nonblinding of outcome assessors, which limited our conclusions. We also observed high heterogeneity in the pain intensity meta-analysis, which might be influenced by aspects related to the design and intervention characteristics of the included studies. For instance, the study of Weise et al [42] adopted a pragmatic trial, and the authors highlighted that the staff maintained close contact with the enrolled participants. Hence, this aspect might have influenced their intervention effects compared to the other trials [22,40,41]. Finally, owing to the small number of included studies, we have not performed sensitivity analyses.

Our review demonstrated no significant differences between mHealth interventions versus no intervention or usual care, neither on pain intensity and disability nor on quality of life. Notwithstanding, our findings suggest positive clinical effects from the use of mHealth in individuals with low back pain compared to no intervention or usual care. Owing to the biases found, the evidence remains inconclusive, and future high-quality clinical trials are warranted.

Acknowledgments

The authors thank Mr Francisco Rafael Amorim dos Santos, the librarian at the Campus UnB Ceilândia, for validating the search strategy and Mrs Renata Monique for her support as one of the independent reviewers in the title and abstract screening. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article. This work was supported by Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF; process 00193-00000758/2021-24); Universidade de Brasília; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES) Finance Code 001.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies adopted.
[DOCX File, 16 KB - mhealth_v11i1e48204_app1.docx]

Multimedia Appendix 2
List of excluded studies, with reasons for exclusion after full-text reading.
[DOCX File, 17 KB - mhealth_v11i1e48204_app2.docx]

Multimedia Appendix 3
Data related to the outcomes (mean and SD) during the intervention period of the studies included in the review.
[DOCX File, 21 KB - mhealth_v11i1e48204_app3.docx]

Multimedia Appendix 4
Summary of findings table (Grading of Recommendations, Assessment, Development, and Evaluations).
[DOCX File, 17 KB - mhealth_v11i1e48204_app4.docx]

Multimedia Appendix 5
Result of the assessment of the certainty of evidence for the primary outcomes (pain intensity and disability).
[DOCX File, 17 KB - mhealth_v11i1e48204_app5.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 116 KB - mhealth_v11i1e48204_app6.pdf]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Olafsson G, Jonsson E, Fritzell P, Hägg O, Borgström F. Cost of low back pain: results from a national register study in Sweden. Eur Spine J 2018 Nov;27(11):2875-2881. [doi: 10.1007/s00586-018-5742-6] [Medline: 30155730]
3. Geurts JW, Willems PC, Kallewaard JW, van Kleef M, Dirksen C. The impact of chronic discogenic low back pain: costs and patients' burden. Pain Res Manag 2018;2018:4696180. [doi: 10.1155/2018/4696180] [Medline: 30364097]

4. Coombs DM, Machado GC, Richards B, Wilson R, Chan J, Storey H, et al. Healthcare costs due to low back pain in the emergency department and inpatient setting in Sydney, Australia. Lancet Reg Health West Pac 2021 Feb;7:100089. [doi: 10.1016/j.lanwpc.2020.100089] [Medline: 34327417]
5. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179. [doi: 10.2196/jmir.7290] [Medline: 28550009]
6. Carregaro RL, Tottoli CR, Rodrigues DDS, Bosmans JE, da Silva EN, van Tulder M. Low back pain should be considered a health and research priority in Brazil: lost productivity and healthcare costs between 2012 to 2016. PLoS One 2020;15(4):e0230902. [doi: 10.1371/journal.pone.0230902] [Medline: 32236113]
7. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299. [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
8. Becker A, Held H, Redaelli M, Strauch K, Chenot JF, Leonhardt C, et al. Low back pain in primary care: costs of care and prediction of future health care utilization. Spine (Phila Pa 1976) 2010 Aug 15;35(18):1714-1720. [doi: 10.1097/brs.0b013e3181cd656f] [Medline: 21374895]
9. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet 2018 Jun 9;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
10. Delitto A, George SZ, Van L, Whitman JM, Sowa G, Shekelle P, et al. Low back pain: clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. J Orthop Sports Phys Ther 2012 Apr 1;42(4):A1-A57. [doi: 10.2519/jospt.2012.42.4.A1]
11. Bernstein IA, Malik Q, Carville S, Ward S. Low back pain and sciatica: summary of NICE guidance. BMJ 2017 Jan 6;356:i6748. [doi: 10.1136/bmj.i6748] [Medline: 28062522]
12. May S. Self-management of chronic low back pain and osteoarthritis. Nat Rev Rheumatol 2010 Apr;6(4):199-209. [doi: 10.1038/nrrheum.2010.26] [Medline: 20357789]
13. Oliveira VC, Ferreira PH, Maher CG, Pinto RZ, Refshauge KM, Ferreira ML. Effectiveness of self-management of low back pain: systematic review with meta-analysis. Arthritis Care Res (Hoboken) 2012 Nov;64(11):1739-1748. [doi: 10.1002/acr.21737] [Medline: 22623349]
14. Kongsted A, Ris I, Kjaer P, Hartvigsen J. Self-management at the core of back pain care: 10 key points for clinicians. Braz J Phys Ther 2021;25(4):396-406. [doi: 10.1016/j.bjpt.2021.05.002] [Medline: 34116904]
15. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 9;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
16. George SZ, Fritz JM, Silfies SP, Schneider MJ, Beneciuk JM, Lentz TA, et al. Interventions for the management of acute and chronic low back pain: revision 2021: clinical practice guidelines linked to the International Classification of Functioning, Disability and Health from the Academy of Orthopaedic Physical Therapy of the American Physical Therapy Association. J Orthop Sports Phys Ther 2021 Nov;51(11):CPG1-CPG60. [doi: 10.2519/jospt.2021.0304]
17. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71. [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
18. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747. [doi: 10.3390/ijerph19073747] [Medline: 35409431]
19. The MAPS toolkit: mHealth assessment and planning for scale. World Health Organization. 2015 Jan 1. URL: www.who.int/publications/i/item/9789241509510 [accessed 2023-10-10]
20. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2023-10-10]
21. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942. [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
22. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
23. Hasenöhrl T, Windschnurer T, Dorotka R, Ambrozy C, Crevenna R. Prescription of individual therapeutic exercises via smartphone app for patients suffering from non-specific back pain: a qualitative feasibility and quantitative pilot study. Wien Klin Wochenschr 2020 Mar;132(5-6):115-123. [doi: 10.1007/s00508-020-01616-x] [Medline: 32060724]
24. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682. [doi: 10.2196/39682] [Medline: 36018713]
25. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]

26. Carvalho C, Prando BC, Dantas LO, Serrão P. Mobile health technologies for the management of spine disorders: a systematic review of mHealth applications in Brazil. Musculoskelet Sci Pract 2022 Aug;60:102562. [doi: 10.1016/j.msksp.2022.102562] [Medline: 35413592]
27. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
28. Chen M, Wu T, Lv M, Chen C, Fang Z, Zeng Z, et al. Efficacy of mobile health in patients with low back pain: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e26095. [doi: 10.2196/26095] [Medline: 34114965]
29. Negreiro M, European Parliamentary Research Service. The rise of digital health technologies during the pandemic. European Parliament. 2021 Apr 14. URL: www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)690548 [accessed 2023-10-10]
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89. [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
31. Kamper SJ, Logan G, Copsey B, Thompson J, Machado GC, Abdel-Shaheed C, et al. What is usual care for low back pain? a systematic review of health care provided to patients with low back pain in family practice and emergency departments. Pain 2020 Apr;161(4):694-702. [doi: 10.1097/j.pain.0000000000001751] [Medline: 31738226]
32. Grey literature database. OpenGrey. URL: opengrey.eu/ [accessed 2023-10-10]
33. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
34. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [Medline: 12882612]
35. Deeks JJ, Higgins JPT, Altman DG, Cochrane Statistical Methods Group. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons; 2022.
36. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14:25. [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
37. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
38. Peryer G, Golder S, Junqueira DR, Vohra S, Loke YK, Cochrane Adverse Effects Methods Group. Chapter 19: Adverse effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions: John Wiley & Sons; 2022.
39. Almhdawi KA, Obeidat DS, Kanaan SF, Oteir AO, Mansour ZM, Alrabbaei H. Efficacy of an innovative smartphone application for office workers with chronic non-specific low back pain: a pilot randomized controlled trial. Clin Rehabil 2020 Oct;34(10):1282-1291. [doi: 10.1177/0269215520937757] [Medline: 32602362]
40. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Jensen JSD, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021 Oct 1;181(10):1288-1296. [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]
41. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34. [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
42. Weise H, Zenner B, Schmiedchen B, Benning L, Bulitta M, Schmitz D, et al. The effect of an app-based home exercise program on self-reported pain intensity in unspecific and degenerative back pain: pragmatic open-label randomized controlled trial. J Med Internet Res 2022 Oct 28;24(10):e41899. [doi: 10.2196/41899] [Medline: 36215327]
43. Numeric Pain Rating Scale. Physiopedia. URL: www.physio-pedia.com/Numeric_Pain_Rating_Scale [accessed 2023-10-10]
44. Modarresi S, Lukacs MJ, Ghodrati M, Salim S, MacDermid JC, Walton DM. A systematic review and synthesis of psychometric properties of the Numeric Pain Rating Scale and the Visual Analog Scale for use in people with neck pain. Clin J Pain 2022 Oct 26;38(2):132-148. [doi: 10.1097/AJP.0000000000000999]
45. Suzuki H, Aono S, Inoue S, Imajo Y, Nishida N, Funaba M, et al. Clinically significant changes in pain along the Pain Intensity Numerical Rating Scale in patients with chronic low back pain. PLoS One 2020;15(3):e0229228. [doi: 10.1371/journal.pone.0229228] [Medline: 32126108]
46. Bakker A, Cai J, English L, Kaiser G, Mesa V, Van Dooren W. Beyond small, medium, or large: points of consideration when interpreting effect sizes. Educ Stud Math 2019 Sep;102(1):1-8. [doi: 10.1007/s10649-019-09908-4]
47. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol 2006 Jan;59(1):45-52. [doi: 10.1016/j.jclinepi.2005.03.018] [Medline: 16360560]

48. Ostelo R, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008 Jan 1;33(1):90-94. [doi: 10.1097/BRS.0b013e31815e3a10] [Medline: 18165753]
49. Zheng F, Zheng Y, Liu S, Yang J, Xiao W, Xiao W, et al. The effect of m-health-based core stability exercise combined with self-compassion training for patients with nonspecific chronic low back pain: a randomized controlled pilot study. Pain Ther 2022 Jun;11(2):511-528. [doi: 10.1007/s40122-022-00358-0] [Medline: 35133634]
50. Schlicker S, Baumeister H, Buntrock C, Sander L, Paganini S, Lin J, et al. A web- and mobile-based intervention for comorbid, recurrent depression in patients with chronic back pain on sick leave (Get.Back): pilot randomized controlled trial on feasibility, user satisfaction, and effectiveness. JMIR Ment Health 2020 Apr 15;7(4):e16398. [doi: 10.2196/16398] [Medline: 32293577]
51. Zheng F, Liu S, Zhang S, Yu Q, Lo WLA, Li T, et al. Does m-health-based exercise (guidance plus education) improve efficacy in patients with chronic low-back pain? a preliminary report on the intervention's significance. Trials 2022 Mar 3;23(1):190. [doi: 10.1186/s13063-022-06116-z] [Medline: 35241140]
52. Pucci G, Rech CR, Fermino RC, Reis RS. Association between physical activity and quality of life in adults. Rev Saude Publica 2012 Feb;46(1):166-179. [doi: 10.1590/s0034-89102012000100021] [Medline: 22249758]
53. Tagliaferri SD, Miller CT, Owen PJ, Mitchell UH, Brisby H, Fitzgibbon B, et al. Domains of chronic low back pain and assessing treatment effectiveness: a clinical perspective. Pain Pract 2020 Feb;20(2):211-225. [doi: 10.1111/papr.12846] [Medline: 31610090]
54. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 3;7(8):e016242. [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
55. Eannucci EF, Hazel K, Grundstein MJ, Nguyen JT, Gallegro J. Patient satisfaction for telehealth physical therapy services was comparable to that of in-person services during the COVID-19 pandemic. HSS J 2020 Nov;16(Suppl 1):10-16. [doi: 10.1007/s11420-020-09800-5] [Medline: 33100933]
56. Beresford L, Norwood T. Can physical therapy deliver clinically meaningful improvements in pain and function through a mobile app? an observational retrospective study. Arch Rehabil Res Clin Transl 2022 Jun;4(2):100186. [doi: 10.1016/j.arrct.2022.100186] [Medline: 35756979]
57. Agnew JMR, Hanratty CE, McVeigh JG, Nugent C, Kerr DP. An investigation into the use of mHealth in musculoskeletal physiotherapy: scoping review. JMIR Rehabil Assist Technol 2022 Mar 11;9(1):e33609. [doi: 10.2196/33609] [Medline: 35275089]
58. Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006 Mar 15;163(6):493-501. [doi: 10.1093/aje/kwj069] [Medline: 16443796]
59. Kamper SJ. Risk of bias and study quality assessment: linking evidence to practice. J Orthop Sports Phys Ther 2020 May;50(5):277-279. [doi: 10.2519/jospt.2020.0702] [Medline: 32354317]
60. Nunan D, Heneghan C, Spencer EA. Catalogue of bias: allocation bias. BMJ Evid Based Med 2018 Feb;23(1):20-21. [doi: 10.1136/ebmed-2017-110882] [Medline: 29367320]
61. Schulz KF, Chalmers I, Altman DG, Grimes DA, Moher D, Hayes RJ. 'Allocation concealment': the evolution and adoption of a methodological term. J R Soc Med 2018 Jun;111(6):216-224. [doi: 10.1177/0141076818776604] [Medline: 29877772]
62. Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. JAMA 2010 Aug 18;304(7):793-794. [doi: 10.1001/jama.2010.1161] [Medline: 20716744]
63. Buehler AM, Cavalcanti AB, Suzumura EA, Carballo MT, Berwanger O. How to assess intensive care randomized trials. Rev Bras Ter Intensiva 2009 Jun;21(2):219-225. [doi: 10.1590/S0103-507X2009000200016] [Medline: 25303354]
64. Shiwa SR, Costa LOP, Moser ADL, Aguiar IC, de Oliveira LVF. PEDro: the physiotherapy evidence database. Fisioterapia em Movimento 2011 Sep;24(3):523-533. [doi: 10.1590/S0103-51502011000300017]
65. Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. Phys Ther Rev 2009 Feb;14(1):36-49. [doi: 10.1179/174328809X405928]
66. Tabacof L, Baker TS, Durbin JR, Desai V, Zeng Q, Sahasrabudhe A, et al. Telehealth treatment for nonspecific low back pain: a review of the current state in mobile health. PM R 2022 Sep;14(9):1086-1098. [doi: 10.1002/pmrj.12738] [Medline: 34786870]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
mHealth: mobile health
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews

Edited by L Buis; submitted 15.04.23; peer-reviewed by AS Ferreira, C Didyk; revised version received 14.07.23; accepted 21.07.23; published 02.11.23.

Please cite as:
de Melo Santana B, Raffin Moura J, Martins de Toledo A, Burke TN, Fernandes Probst L, Pasinato F, Luiz Carregaro R
Efficacy of mHealth Interventions for Improving the Pain and Disability of Individuals With Chronic Low Back Pain: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2023;11:e48204
URL: https://mhealth.jmir.org/2023/1/e48204
doi:10.2196/48204

Mobile Phone–Based Interventions for Smoking Cessation Among Young People: Systematic Review and Meta-Analysis

Xinmei Zhou[1], MSc; Xiaowen Wei[1,2], MSc; Anqi Cheng[1], PhD; Zhao Liu[1], PhD; Zheng Su[1], PhD; Jinxuan Li[1,2], MSc; Rui Qin[1,3], MSc; Liang Zhao[1], MSc; Ying Xie[1,4], BA; Zhenxiao Huang[1,4], MSc; Xin Xia[1,5], BA; Yi Liu[1,5], BA; Qingqing Song[1,2], BA; Dan Xiao[1], PhD; Chen Wang[6,7], PhD

1
2
3
4
5
6
7

Corresponding Author:
Dan Xiao, PhD

Abstract

Background: Mobile phone–based cessation interventions have emerged as a promising alternative for smoking cessation, while evidence of the efficacy of mobile phone–based smoking cessation programs among young people is mixed.

Objective: This study aimed to determine the efficacy of mobile phone–based interventions compared to usual practice or assessment-only controls on smoking cessation in young people.

Methods: In this systematic review and meta-analysis, we searched Cochrane Library, Embase, PubMed, and Web of Science on March 8, 2023. We included randomized controlled trials that examined the efficacy of mobile phone–based interventions on smoking cessation in young people (age ≤30 years). The risk of bias was assessed with Cochrane Risk of Bias 2.

Results: A total of 13 eligible studies, comprising 27,240 participants, were included in this analysis. The age range of the participants was between 16 and 30 years. Nine studies were SMS text messaging interventions, and 4 studies were app-based interventions. The duration of the smoking cessation intervention varied from 5 days to 6 months. The included studies were conducted in the following countries: the United States, China, Sweden, Canada, Switzerland, and Thailand. The meta-analysis revealed that SMS text messaging interventions significantly improved continuous abstinence rates compared to inactive control conditions (risk ratio [RR] 1.51, 95% CI 1.24-1.84). The subgroup analysis showed pooled RRs of 1.90 (95% CI 1.29-2.81), 1.64 (95% CI 1.23-2.18), and 1.35 (95% CI 1.04-1.76) for continuous abstinence at the 1-, 3-, and 6- month follow-up, respectively. Pooling across 7 studies, SMS text messaging interventions showed efficacy in promoting 7-day point prevalence abstinence (PPA), with an RR of 1.83 (95% CI 1.34-2.48). The subgroup analysis demonstrated a significant impact at the 1- and 3-month follow-ups, with pooled RRs of 1.72 (95% CI 1.13-2.63) and 2.54 (95% CI 2.05-3.14), respectively, compared to inactive control conditions. However, at the 6-month follow-up, the efficacy of SMS text messaging interventions in promoting 7-day PPA was not statistically significant (RR 1.45, 95% CI 0.92-2.28). In contrast, app-based interventions did not show significant efficacy in promoting continuous abstinence or 7-day PPA. However, it is important to note that the evidence for app-based interventions was limited.

Conclusions: SMS text messaging–based smoking cessation interventions compared to inactive controls were associated with abstinence among young people and could be considered a viable option for smoking cessation in this population. More research is needed on smoking cessation apps, especially apps that target young people. Future research should focus on identifying the most effective mobile phone–based cessation approaches and on developing strategies to increase their uptake and intention.

Trial Registration: PROSPERO CRD42022318845; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=318845

(JMIR Mhealth Uhealth 2023;11:e48253) doi:10.2196/48253

KEYWORDS

smoking cessation; young people; mobile health; text messaging; mHealth; PRISMA

Introduction

Background

Tobacco use is one of the leading risk factors for premature morbidity and mortality worldwide [1]. Smoking among young people is of particular concern. Despite the well-documented health risks associated with tobacco use, many young people continue to smoke or experiment with smoking. The prevalence of smoking among young people is especially troubling, as this age group is in the midst of crucial physical and psychological development. The harmful effects of smoking at this stage of life can have lifelong consequences, including increased risk of chronic disease, impaired cognitive function [2], and reduced quality of life [3]. In 2019, an estimated 155 million (95% uncertainty interval 150-160) people aged 15-24 years worldwide were tobacco smokers, with a prevalence of 20.1% in males and 4.95% in females [4]. Quitting before the age of 30 years can prevent more than 97% of the excess mortality caused by continued smoking [5]. Given the serious health risks associated with smoking, quitting is critical for young people.

While traditional cessation methods such as pharmacotherapy [6] and behavioral counseling [7] can be effective, their widespread implementation at a population level faces barriers [8]. Mobile phone–based smoking cessation interventions have emerged as a promising alternative to assist with smoking cessation [9-12]. Phone interventions are a cost-effective use of health care resources [13]. These interventions can provide personalized interactive support that is tailored to individual needs and characteristics, irrespective of location and time constraints [14,15], making them a valuable approach for smoking cessation in this demographic. Furthermore, young individuals are more open to novel and innovative approaches [16]. According to the International Telecommunications Union, approximately 66% of the global population had internet access in 2022 [17]. While previous research has suggested that SMS text message–based smoking cessation interventions were more effective than minimal smoking cessation support in the general population [18], evidence of the efficacy in young people remained inconclusive [19,20]. As countries work toward achieving the goal of reducing the prevalence of tobacco use [21], timely data on the efficacy of mobile phone–based smoking cessation programs among young people are necessary to guide effective policy and planning.

Objective

To the best of our knowledge, there are no meta-analyses supporting the efficacy of mobile phone–based smoking cessation interventions among young people. The aim of this meta-analysis was to determine the efficacy of mobile phone–based smoking cessation interventions, excluding pharmacological treatment, in helping young smokers to quit.

Methods

We adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews of interventions. We used a prespecified protocol registered with PROSPERO (CRD42022318845).

Search Strategy and Selection Criteria

We included randomized controlled trials (RCTs) with young smokers (30 years or younger) who wanted to quit. Included trials had to be clearly focused on smoking cessation using SMS text messaging or a smoking cessation app without pharmacotherapy, compared to a control intervention. Trials that had a focus on pregnant women were not eligible for inclusion.

Data Extraction

Studies were assessed for inclusion if they reported cigarette smoking cessation as the primary outcome. Self-reported abstinence from cigarette smoking and biochemically validated measures of abstinence were used to define smoking cessation.

Data extracted from each study included the study location, study design, population, inclusion criteria, exclusion criteria, follow-up period, details of the intervention group, details of the control group, definition of smoking cessation, number of participants, and smoking cessation rates. Wherever possible, an intention-to-treat analysis was used.

The following electronic bibliographic databases were last searched in March 2023: Cochrane Tobacco Addiction Group's Specialised Register (Source: PubMed, Embase, ClinicalTrials.gov, and the ICTRP), Embase, PubMed, and Web of Science. The search strategies used in the Cochrane Library, PubMed, Embase, and Web of Science are listed in Multimedia Appendix 1. The database literature search was restricted to the English language and studies on humans. The search terms were text messaging, phone-based, smartphone, app, mobile health, sms, txt, young, student, adolescent, and smoking cessation. Both abstracts and full manuscripts were considered.

Statistical Analysis

Authors XZ, XW, and AC independently confirmed study eligibility. Authors JL, YX, ZH, XX, YL, QS, and XZ extracted data, which were then checked by a second author (RQ, LZ, or AC). Two authors (XZ, AC, ZL, or ZS) independently assessed quality using the Cochrane Risk of Bias tool. All differences were resolved by discussion.

We used random-effects meta-analysis to analyze pooled outcome data among smokers who used SMS text messaging or an app compared with a control. Binary outcomes were estimated using risk ratios (RRs) and 95% CIs, with priority given to intention-to-treat data when available. For smoking cessation, meta-analyses were conducted for continuous abstinence and 7-day point prevalence abstinence (PPA). For 2 studies [22,23], data specifically for individuals 30 years and younger were extracted from the original data set and reanalyzed. The 7-day PPA at the 1-month follow-up for one study was derived from the third figure in Chulasai et al [24]. Heterogeneity between studies was assessed using the I^2 statistic. A subgroup analysis of the length of follow-up was also performed. Additionally, we performed a sensitivity analysis by removing the studies with a high risk of bias. All analyses were performed using R 4.2.0 (R Foundation for Statistical Computing) and Revman 5.4 (The Cochrane Collaboration).

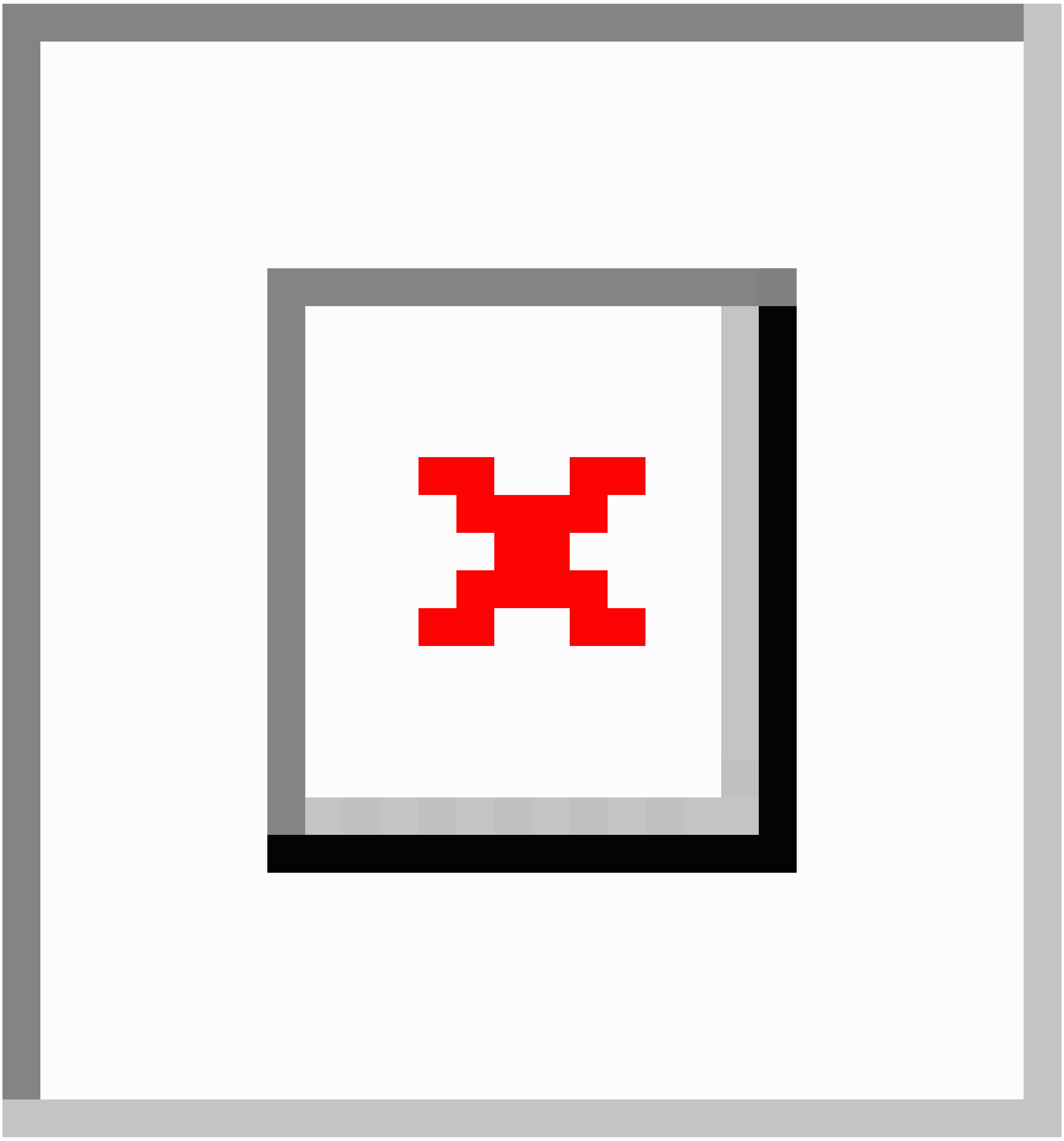
We did not perform funnel plot asymmetry because no outcome had more than 10 studies in the meta-analysis [25].

The risk of bias for studies was assessed using the Cochrane Risk of Bias 2 tool [26]. Studies were considered to be at high risk in the domain of missing outcome data if the overall loss to follow-up was more than 50% or if there was a difference in follow-up rates of more than 20% between study arms.

Results

We identified 1046 full-text trial reports or titles and abstracts (Figure 1) and identified 13 RCTs [10,15,19,20,22-27,33-33] for inclusion in the final review. The complete process is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of evidence search and selection.



General Characteristics of the Selected Studies

Details of the eligible studies are presented in Table 1. A total of 13 eligible studies, comprising 27,240 participants, were included in this analysis. The age range of the participants was mainly between 16 and 30 years. Of the 13 included studies, 2 were cluster RCTs [29,32] and 11 were individual RCTs. There were 5 studies in the United States [15,28,31,33,34], 2 studies in Sweden [10,20], 2 studies in China [22,32], and 1 study each in Canada [19], Switzerland [29], and Thailand [24]. One study was conducted online [23], in which registered users of the Smoke Free app were the study participants, regardless of location. Measures of current smoking varied between studies: 4 studies [10,20,29,32] included daily smokers and weekly smokers, 2 studies [19,22] included only daily smokers, and Palmer et al [31] used a definition related to daily smoking (vaping nicotine at least 25 days/month). Three studies [15,24,28] included participants who smoked monthly or more, and Ybarra et al [34] included participants who smoked 24 cigarettes or more per week (at least 4 per day on at least 6 days in a week). One study used the definition of having smoked 100 cigarettes in life and now smoking every day or some days [33]. The remaining study [23] did not report the definition of smoker. Of the 13 studies, 5 recruited from vocational schools [29,32], high schools [20], or colleges [10,24]. The remaining studies recruited smokers from the community, health care facilities, online, or a combination of sources.

Table . Characteristics of included studies.

Study	Participants, n	Age (years)	Intervention	Control	Outcome
Baskerville et al [19], 2018, Canada, RCT[a]	1599	19-29	Crush the Crave Application	On the Road to Quitting–Self Help: The control group received an evidence-based self-help guide known as “On the Road to Quitting” that has been developed by Health Canada for young adult smokers.	Continuous self-reported abstinence at 6 months; self-reported 30-day PPA[b] from smoking at 3 and 6 months, operationalized as not having smoked any cigarettes, even a puff, or used other tobacco in the last 30 days; self-reported 7-day PPA at 3 and 6 months.
Bendtsen et al [20], 2021, Sweden, RCT	535	High school students, median 17 (IQR 16-18)	The intervention starts with a 1-week motivational phase. After 1 week the core intervention begins and runs for 12 weeks. Participants receive up to 4 texts per day during the first few weeks, and then the number of messages per day decreases.	Self-help materials: The control group could use the website of the national quit line or contact their high school’s health service for more help	Prolonged abstinence and point prevalence of smoking cessation at the 3- and 6-month follow-ups. Prolonged abstinence, following the Russel standard, was defined at the 3-month follow-up as not smoking more than 5 cigarettes in the past 8 weeks. At 6 months, the definition was altered to not smoking more than 5 cigarettes in the past 5 months. This outcome thus measures abstinence from the start of the 12-week smoking cessation program, allowing for a 4-week grace period. Point prevalence of smoking cessation, a recommended outcome by the Society for Research on Nicotine and Tobacco, was defined as not having smoked a single cigarette in the past 4 weeks. This question measures current behavior and, thus, was the same at both the 3- and 6-month follow-ups.
Chulasai et al [24], 2022, Thailand, RCT	273	18-24	The Quit with US app integrated with smoking cessation counseling from pharmacists	Smoking cessation counseling from pharmacists	The primary smoking abstinence outcome was the 7-day point prevalence at the 12-week follow-up, as recommended for smoking abstinence measures. The outcome was defined as a self-report of the previous 7 consecutive days of continuous abstinence from smoking plus an exhaled CO concentration level of ≤6 ppm.

Study	Participants, n	Age (years)	Intervention	Control	Outcome
Crane et al [23], 2018, online, RCT	18,400	18-30	Full version of the smoke-free app	Reduced version that did not include the missions	The primary outcome measure was self-reported continuous abstinence up to the 12-week follow-up.
Graham et al [28], 2021, United States, RCT	2588	18-24	This is Quitting (texting)	Assessment only control	The primary outcome was self-reported 30-day PPA at 7 months analyzed under an intention-to-treat analysis, which counted nonresponders as vaping. Secondary outcomes were 7-day PPA under the intention-to-treat analysis and retention weighted complete case analysis of 30-day and 7-day PPA.
Haug et al [29], 2013, Switzerland, cluster RCT	755	Vocational school students, mean 18.2 (SD 2.3)	Individualized texts to support smoking cessation (2 texts per week for a period of 3 months); possibility to register for a more intensive program providing strategies for smoking cessation around a self-defined quit date (2 texts per day for a period of 4 weeks)	Assessment only control group	7-day point prevalence smoking abstinence at the 6-month follow-up (ie, not having smoked a puff within the past 7 days preceding the follow-up) and 4-week point prevalence smoking abstinence were assessed
Liao et al [22], 2018, China, RCT	344	18-30	Mobile phone–based texts (3-5 messages per day); interventions (Happy Quit) for smoking cessation for the 12-week and 24-week follow-up	No cessation message intervention: Texts every week thanking them for being in the study and texts for assessment	Biochemically verified continuous smoking abstinence at 24 weeks; self-reported 7-day PPA (ie, not even a puff of smoke, for the last 7 days) at 1, 4, 8, 12, 16, 20, and 24 weeks; self-reported continuous abstinence at 4, 12, and 24 weeks.
Mays et al [15], 2021, United States, RCT	349	18-30	6-week mobile messaging intervention	Assessments only	Self-reported cessation was assessed at 6 weeks, 3 months, and 6 months.
Müssener et al [10], 2016, Sweden, RCT	1590	Mainly between 21 and 30	Texts: Those in group 1 received motivational messages (the intervention) 5 times a day for 3 days before their stated quit day and then continue to receive 3-5 motivational texts per day for week 1, 2-4 messages per day for the next 2-4 weeks, and then 10 messages per week for the remaining 8 weeks.	Texts unrelated to quitting	At the 4-month follow-up, 8 weeks of prolonged abstinence (having smoked ≤5 cigarettes during this time); self-reported 4-week PPA (not having smoked a single cigarette); self-reported 7-day PPA (defined as not smoking any cigarettes in the past 7 days)

Study	Participants, n	Age (years)	Intervention	Control	Outcome
Palmer et al [31], 2022, United States, RCT	27	17-21	Contingency management was delivered through DynamiCare Health's smartphone app for 4 weeks, in which financial incentives were delivered contingent on abstinent cotinine samples after the quit day until the end of treatment.	Assessment only: Control participants earned incentives for submitting cotinine, regardless of abstinence.	Abstinence at 1-month follow-up
Shi et al [32], 2013, China, cluster RCT	179	16-19	Tailored information via mobile phone texts for 12 weeks	A self-help pamphlet about smoking cessation	Self-reported 7- and 30-day abstinence at 12-week follow-up
Villanti et al [33], 2022, United States, RCT	437	18-30	12-week tailored text smoking-cessation program with a companion web-based intervention	Referral to online quit resources	Self-reported 30- and 7-day PPA at 12-week follow-up
Ybarra et al [34], 2013, United States, RCT	164	18-25	SMS USA, a 6-week smoking cessation intervention. Intervention participants received texts daily pre- and postquitting. Everyone receives messages 14 days prior to the quit day and through the day after the quit day. Then, participants are ”pathed“ to particular messages based upon their self-reported smoking status on day 2 and day 7 post quit. Those who are successful at quitting receive messages aimed at relapse prevention, whereas those who have slipped receive messages aimed at getting the person to recommit to quitting and trying again. Other name: Stop My Smoking USA	No intervention: Attention-matched control; messages aimed at improving one's sleep and increasing one's fitness, along with general messages about the most well-known health dangers of smoking. Messages sent on the same schedule as the intervention group.	Three-month continuous abstinence; smoking ≤5 cigarettes since the quit day 4 weeks post quit; 7-day PPA at 4 weeks

[a]RCT: randomized controlled trial.

[b]PPA: point prevalence abstinence.

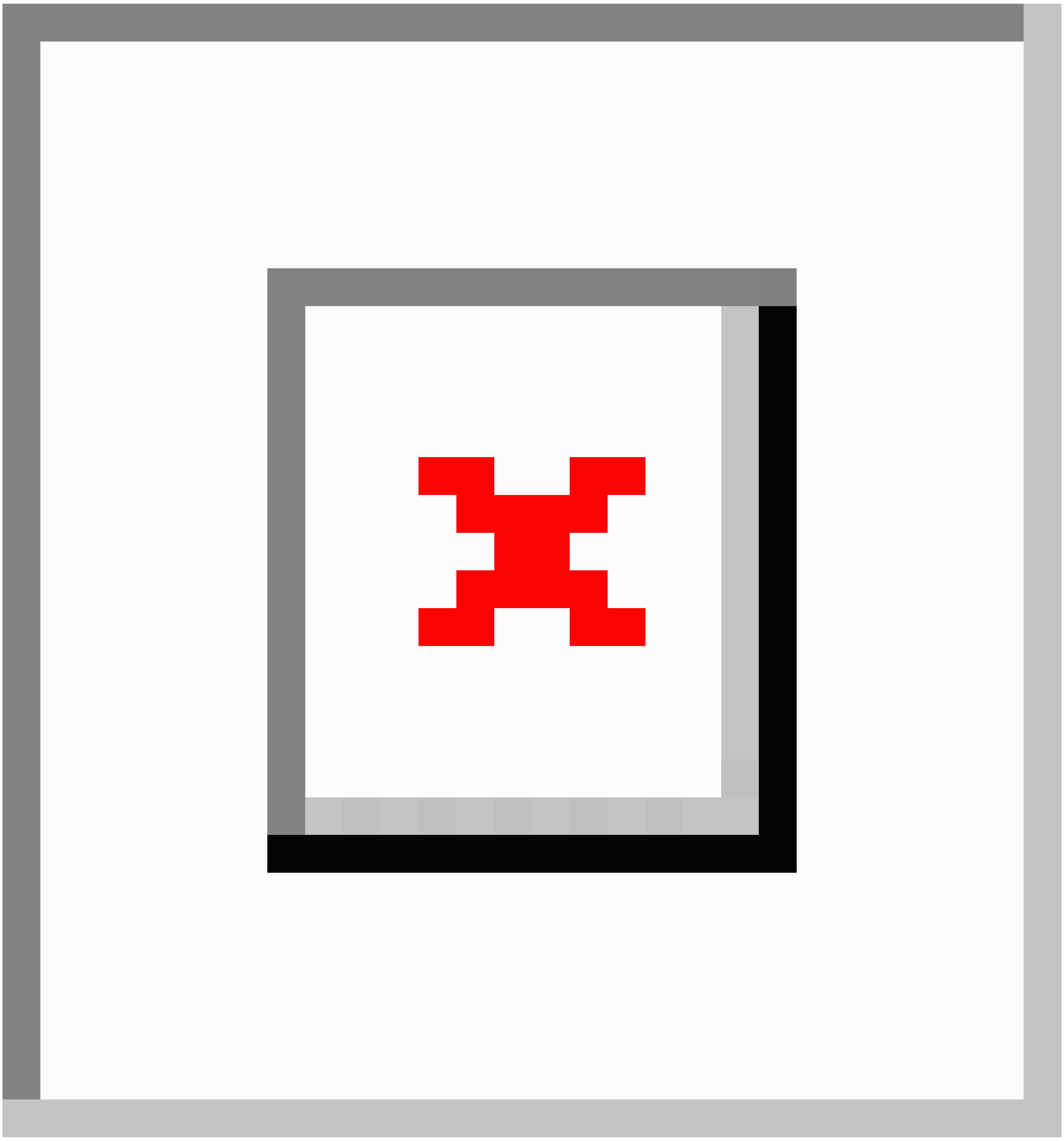
Most studies [10,15,19,22,24,31,33,34] provided incentives to participants in the form of financial rewards for follow-up, with the highest incentive being US $310 [31]. One study offered an alternative incentive in the form of a lottery draw at the end of the study instead of a monetary incentive. Nine of the selected studies provided a tailored intervention to participants [15,19,22,24,28,29,32-34]. These interventions were tailored to the user's age, stage of quitting, smoking history, stage of readiness to quit, demographics, etc.

Nine studies used SMS text messaging–based interventions [10,15,20,22,28,29,32-34], and 4 studies used app-based interventions [19,23,24,31]. The duration of the smoking cessation intervention varied from 5 days to 6 months, with 8 studies lasting ≥12 weeks [10,19,20,22,24,29,32,33], and the remaining studies lasting between 5 days and 6 weeks [15,23,28,31,34]. Most studies used an inactive control, such as an assessment-only control group [15,28,29,31], SMS text messages that were unrelated to quitting [10,22,34], or self-help cessation materials [19,20,32,33]. One study [23] provided a reduced version app that did not include the mission. One study [24] provided the control with smoking cessation counseling from a pharmacist. The majority of studies used self-reported abstinence, without biochemical verification. Four studies [22,24,31,33] used biochemically verified abstinence, such as salivary cotinine test [31,33], urine cotinine test [22], and exhaled CO concentration for verification [24].

Risk of Bias

The risk of bias assessments for individual studies is shown in Figure 2. The majority of studies reported methods of randomization and allocation concealment that were judged to be of low risk for the randomization process. The main source of some concerns was the measurement of the outcome, as these studies used self-reported smoking cessation rates without biochemical validation, and the intervention could not be blinded to participants due to its inherent characteristics. We judged 3 studies [19,23,32] to be at a high risk of bias for missing outcome data because more than 50% of participants were lost to follow-up or the difference in follow-up rates between study arms was more than 20%. The remaining studies were judged to be at a low risk in the domain of missing outcome data. Overall, 2 studies were at a low risk of bias (judged at low risk for all domains) [22,24], 3 were at high risk (judged to be at high risk in at least one domain) [19,23,32], and the remaining studies were of some concern.

Figure 2. Risk of bias summary: reviewers' judgments about each risk of bias item for each included study [10,15,19,20,22-28,29,31,34-34].



Continuous Abstinence

The result of continuous abstinence is illustrated in Figure 3. Combining data from 5 studies using a random-effects meta-analysis, a significant improvement in continuous abstinence rates was observed with SMS text messaging interventions, with an RR of 1.51 (95% CI 1.24-1.84) compared with inactive control conditions (assessment only, non–quit-related SMS text messages, or self-help materials).

For continuous abstinence, no high-risk study was identified when comparing the SMS text message intervention with the inactive control. The subgroup analysis (Figure 4) showed a pooled RR of 1.90 (95% CI 1.29-2.81) for continuous abstinence at the 1-month follow-up, with no significant heterogeneity observed among the included studies. At the 3-month follow-up, the pooled RR for SMS text messaging interventions versus an inactive control was 1.64 (95% CI 1.23-2.18), with an I^2 value of 50.4% (P=.089). At the 6-month follow-up, the SMS text messaging intervention yielded similar results as the 3-month follow-up for continuous abstinence (RR 1.35, 95% CI 1.04-1.76), with no significant heterogeneity observed (I^2=0.0%).

Figure 3. Random-effects meta-analysis for SMS text messaging compared to an inactive control on continuous abstinence [10,15,20,22,34]. MH: Mantel-Haenszel; RR: risk ratio.

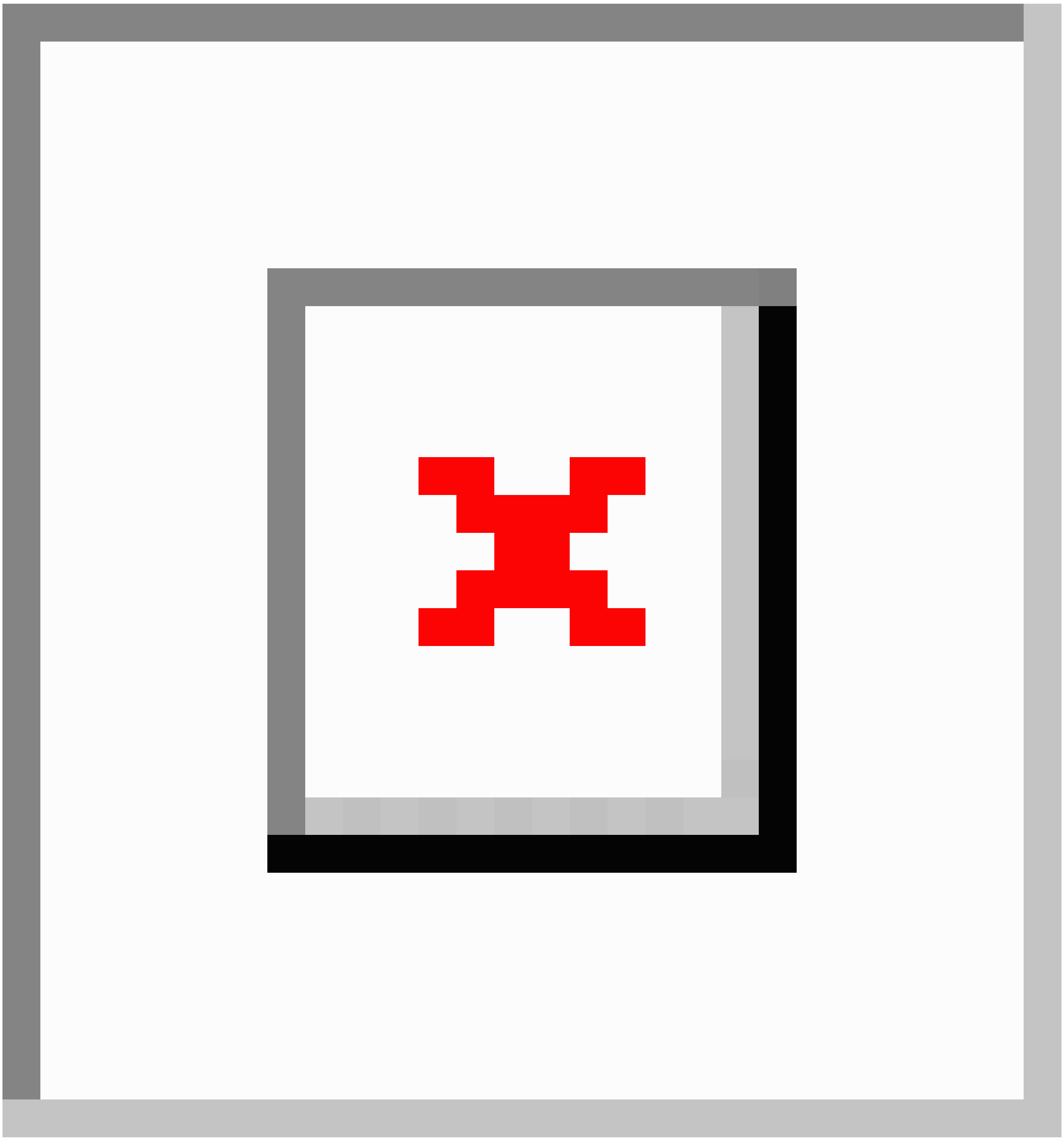
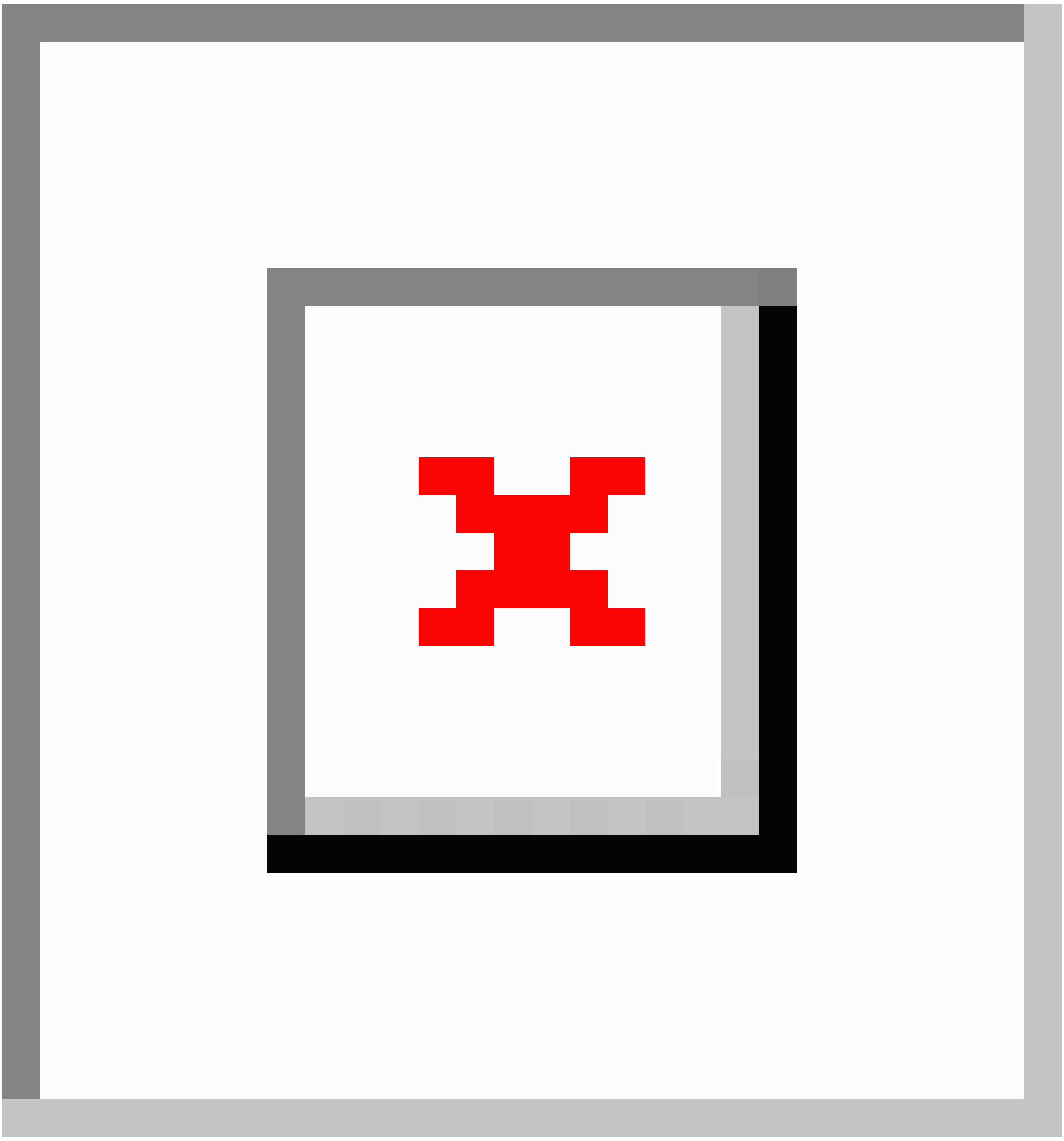
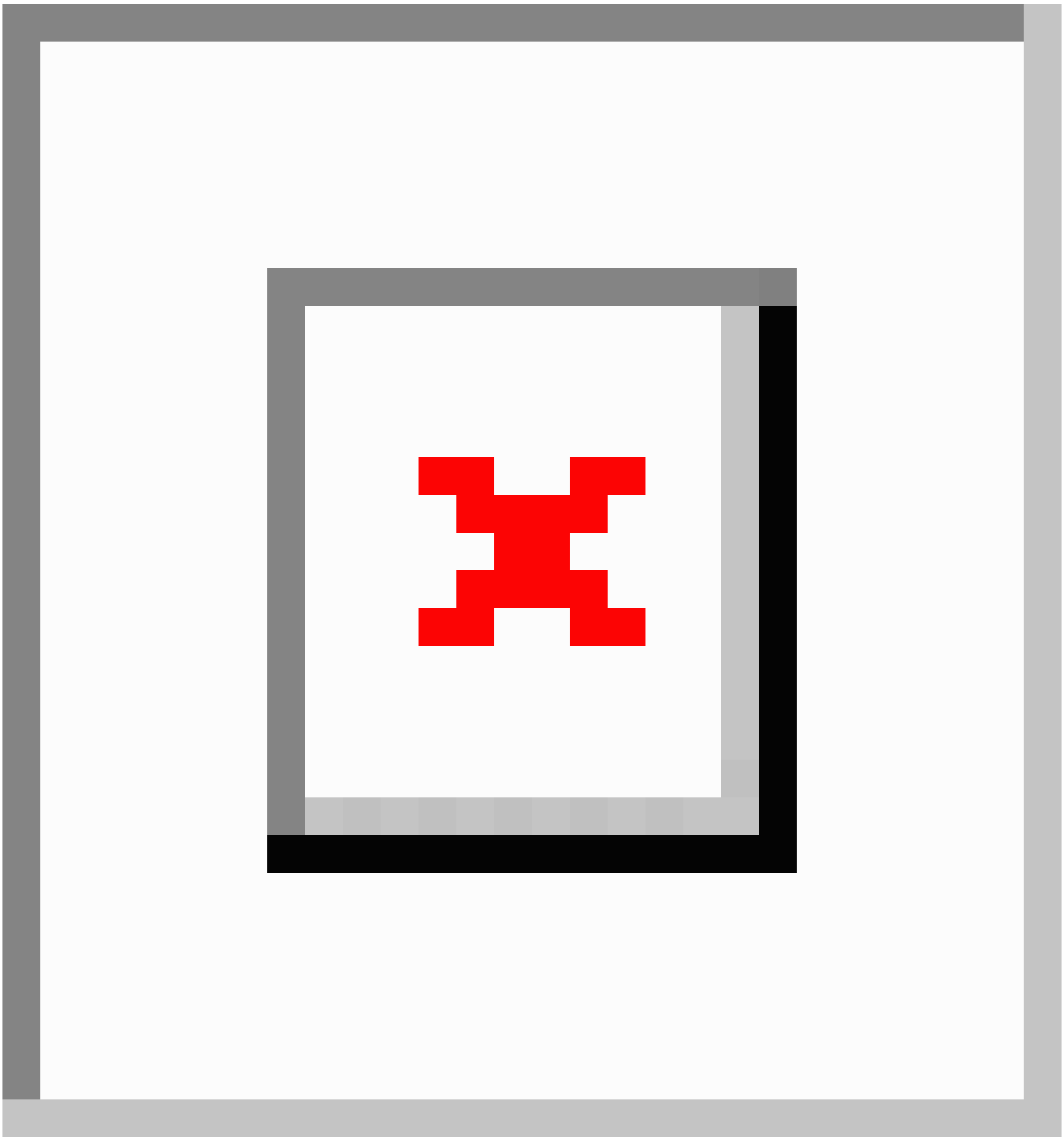


Figure 4. Subgroup analysis: random-effects meta-analysis for SMS text messaging compared to an inactive control on continuous abstinence at the 1-, 3-, and 6-month follow-ups [10,15,20,22,34]. MH: Mantel-Haenszel; RR: risk ratio.



Only 2 studies provided data on the RR of continuous abstinence in the comparison between app-based interventions and controls (Figure 5). However, these studies yielded conflicting results. It is worth noting that both studies included in this analysis were subject to a high risk of bias due to missing outcome data.

Figure 5. Random-effects meta-analysis for apps compared to a control on continuous abstinence [19,23]. MH: Mantel-Haenszel; RR: risk ratio.



7-Day PPA

In terms of 7-day PPA, as illustrated in Figure 6, the meta-analysis of 7 studies showed an RR of 1.83 (95% CI 1.34-2.48), with an I^2 value of 87% (*P*<.001). A sensitivity analysis excluding high-risk studies showed consistent results (RR 1.84, 95% CI 1.32-2.57; *k*=6), indicating the stability of the results. The subgroup analysis (Figure 7) showed that SMS text messaging interventions had a significant impact at the 1- and 3-month follow-ups, with pooled RRs of 1.72 (95% CI 1.13-2.63) and 2.54 (95% CI 2.05-3.14), respectively, compared with inactive control conditions. When pooling across 3 studies, SMS text messaging interventions showed nonsignificant efficacy in promoting 7-day PPA at the 6-month follow-up (RR 1.45, 95% CI 0.92-2.28).

Figure 6. Random-effects meta-analysis for SMS text messaging compared to an inactive control on 7-day point prevalence abstinence [10,22,28,29,32-34]. MH: Mantel-Haenszel; RR: risk ratio.

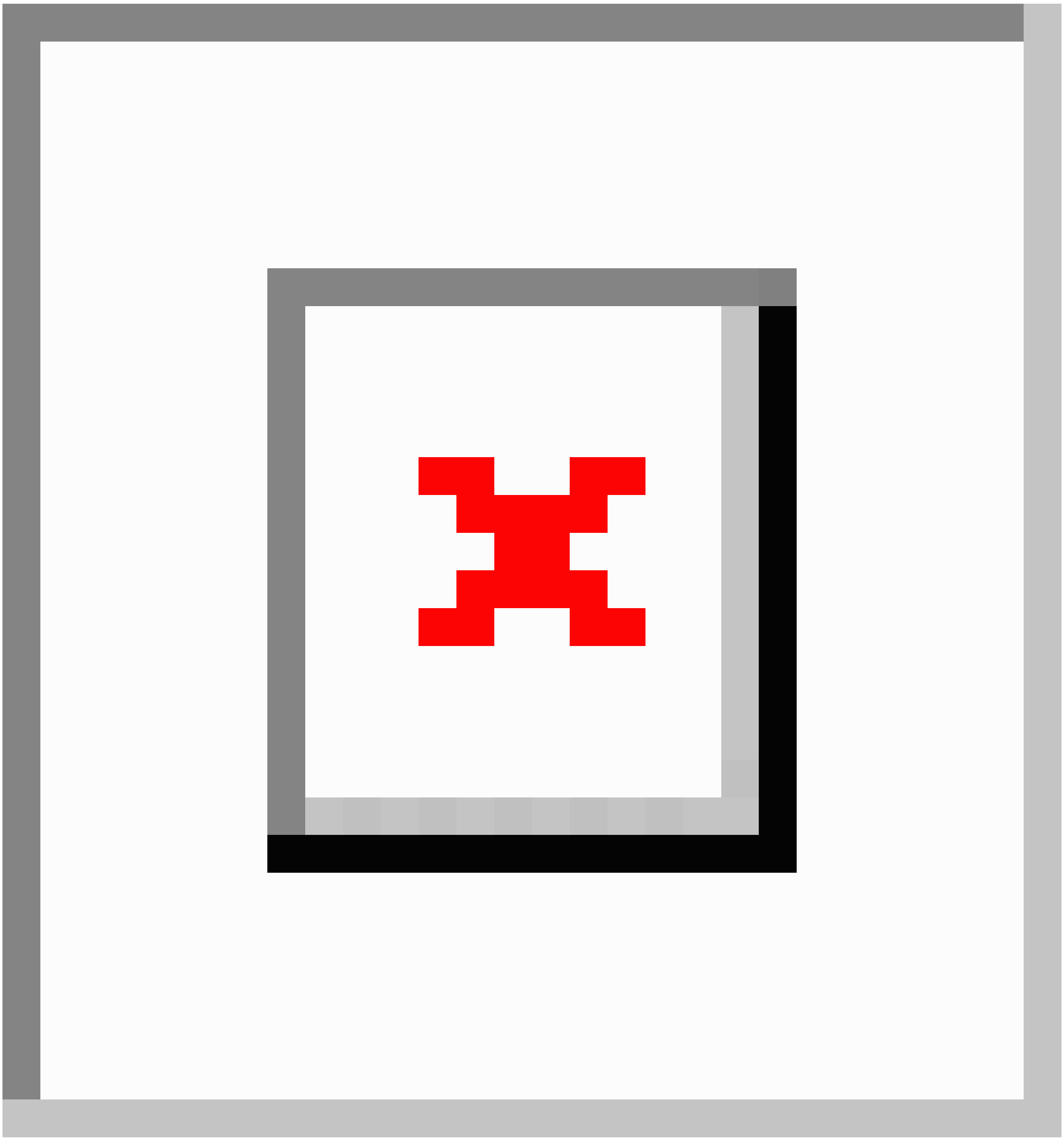
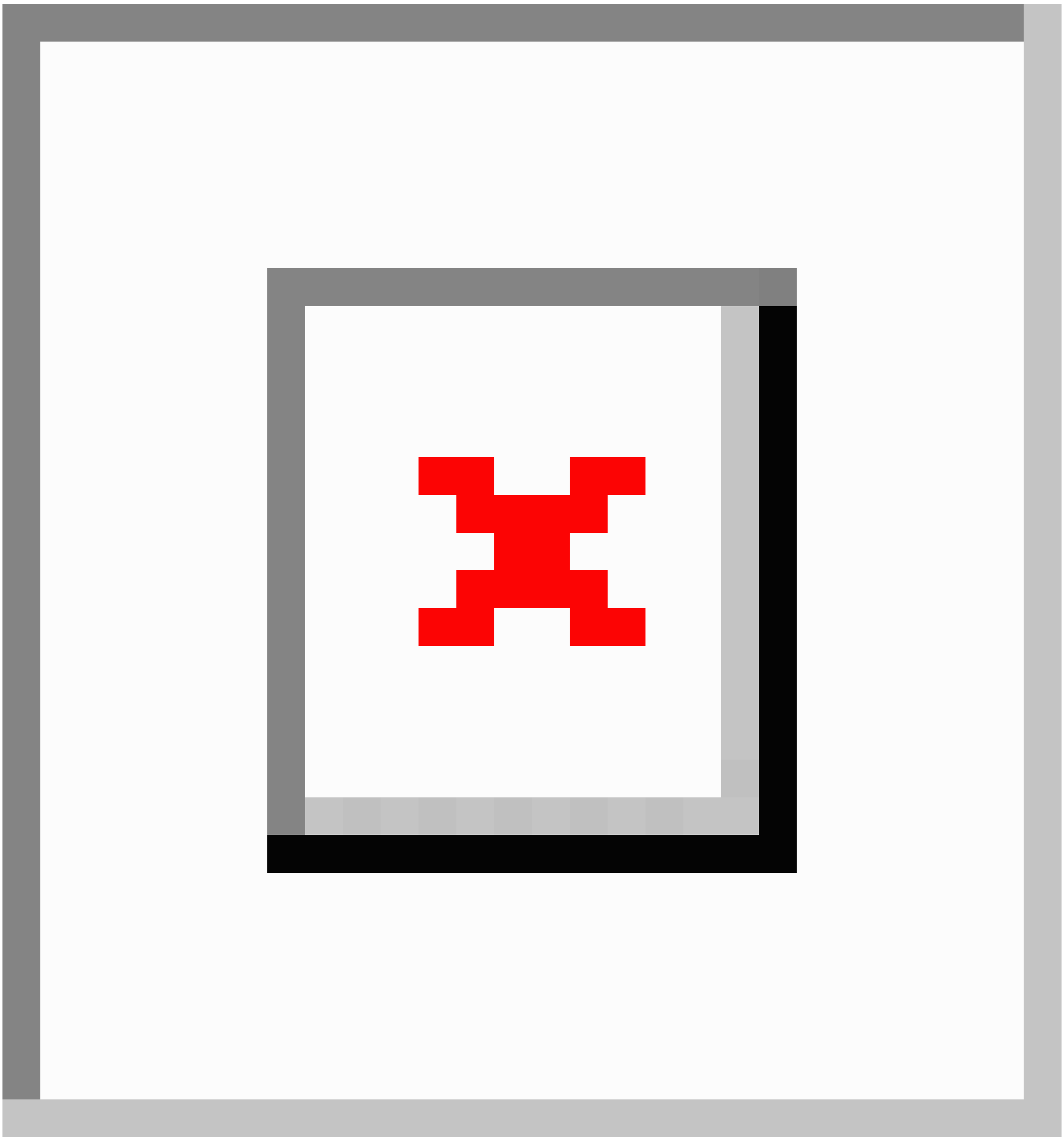
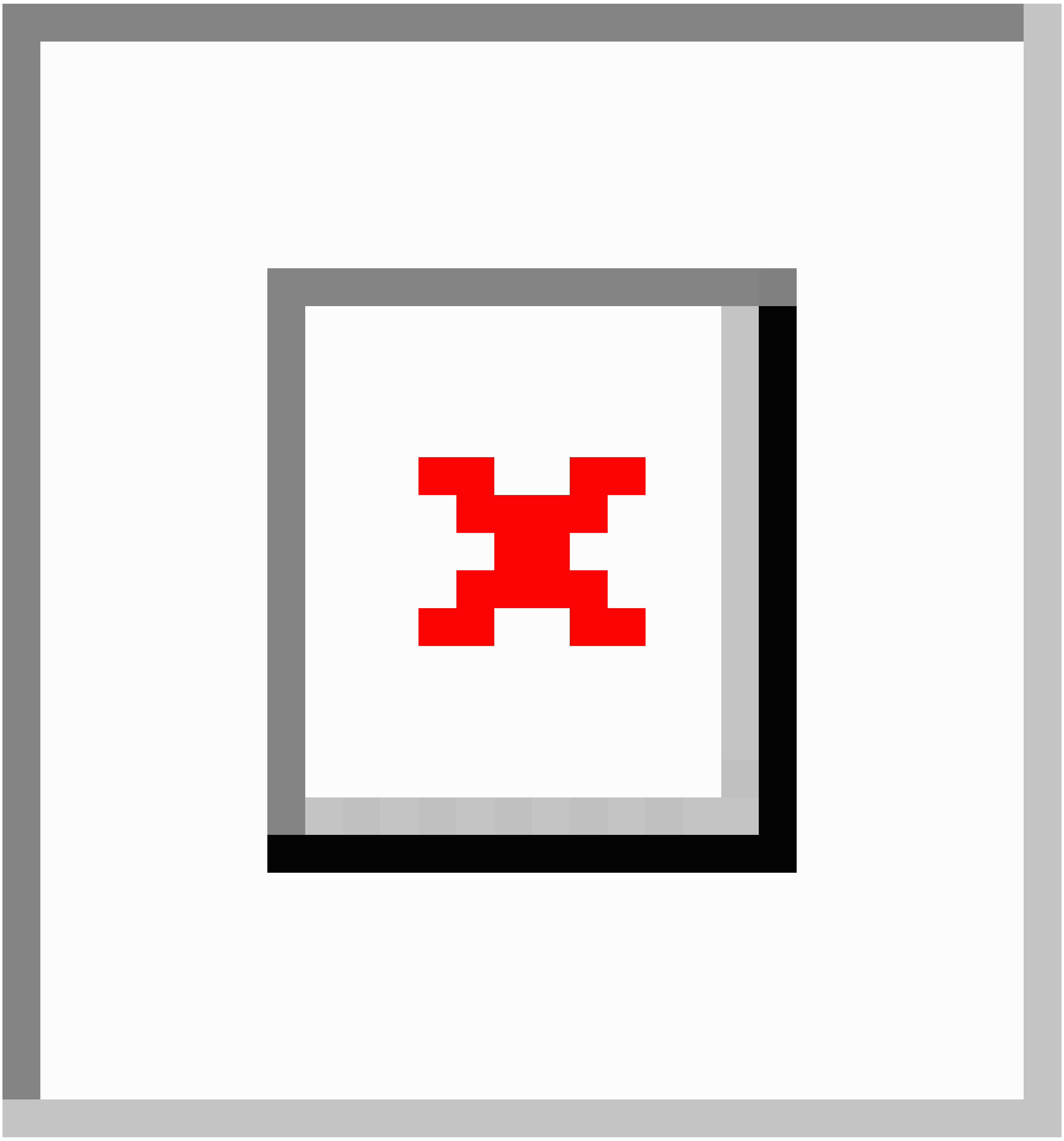


Figure 7. Subgroup analysis: random-effects meta-analysis for SMS text messaging compared to an inactive control on the 7-day point prevalence abstinence at the 1-, 3-, and 6-month follow-ups [10,22,28,29,32-34]. MH: Mantel-Haenszel; RR: risk ratio.



Pooling across 3 studies, app-based interventions showed no significant efficacy in promoting 7-day PPA (RR 1.27, 95% CI 0.69-2.34), indicating a lack of substantial impact (Figure 8). Notably, a high level of heterogeneity was observed among the included studies, with an I^2 value of 91% (P<.001), suggesting significant variation in the results. A sensitivity analysis was conducted by removing the study identified as having a high risk of bias. After the exclusion, the remaining studies showed a pooled RR of 1.86 (95% CI 1.41-2.46), indicating a relatively higher effect size in favor of app-based interventions. Importantly, the removal of the high-risk study resulted in a substantial decrease in heterogeneity, from 91% to 0%.

Figure 8. Random-effects meta-analysis for apps compared to a control on 7-day point prevalence abstinence [19,24,31]. MH: Mantel-Haenszel; RR: risk ratio.



Discussion

Principal Results

This study aimed to synthesize the published literature on the efficacy of mobile phone–based interventions for smoking cessation among young people. Our findings suggest that SMS text messaging interventions could be effective for smoking cessation among young individuals, whereas the evidence for app-based interventions is inconclusive. The sensitivity analysis showed stable results for SMS text messaging interventions, but conflicting results for the app-based study.

Comparison With Prior Work

Previous reviews of mobile phone–based interventions have varied considerably in terms of population characteristics [35-39] and geographical limitations [40,41]. To our knowledge, there is currently no meta-analysis supporting the efficacy of mobile phone–based interventions among young people. Our review fills this gap by providing evidence that SMS text messaging approaches to smoking cessation are robust among young people. Due to the nontemporal and nonspatial nature of mobile phone–based interventions, they can reach a wider audience and serve as a good adjunct to smoking cessation interventions for this population.

The meta-analysis showed that SMS text messaging smoking cessation interventions were effective for continuous abstinence among young people, which is in line with a previous study on general smokers [18]. The RR value at 6 months of continuous abstinence was slightly lower than that for general smokers (RR 1.54, 95% CI 1.19-2.00) [18]. This discrepancy may be attributed to the interventions lacking specific tailoring to address the unique characteristics and needs of young individuals. Smoking cessation was related to motivation to quit, which can differ across age groups [42]. A nationwide study conducted in the United States revealed the common reasons for smoking cessation in young adults aged 18-34 years; the two most popular reasons were physical fitness (64%) and the cost of tobacco (64%). More than half of current smokers also identified "encouragement from friend or relative" (55.2%) and "info about health hazards" (59.7%) as reasons for quitting smoking [43]. Despite most of the included studies in our analysis allowing for customization of the quit date, the interventions' content may not have been specifically tailored for young people. Developing more targeted cessation interventions that take into account young people's unique motivations and use patterns is crucial. This approach may help promote a positive attitude toward quitting smoking [44]. Furthermore, the efficacy of smoking cessation interventions among young people was observed to be slightly higher at 3 months compared to 1 month, as indicated by the 7-day PPA. This finding aligns with the cycle of withdrawal response, with the most intense withdrawal response occurring in the weeks when smokers first attempt to quit [45].

Less than one-third of smokers use cessation medications or behavioral counseling to support quit attempts [46]. Young smokers are more reluctant to seek treatment for smoking cessation than older smokers [47]. Pharmacotherapy and counseling often require face-to-face contact and the presence of a health care provider, which can be time-consuming and may not be covered by insurance. Additionally, adverse drug reactions to medication can be a barrier to quitting [48]. Young people are more likely to use mobile phones and novel technology in their daily lives [16]. Mobile phone–based interventions can be delivered remotely and offer a more cost-effective, discreet, convenient, and accessible alternative. Therefore, mobile phone–based smoking cessation interventions are a promising alternative for young people.

In recent years, the development of technology has led to the increasing popularity of mobile phone apps designed for health management [49,50]. The included studies were published between 2012-2022 for the SMS text messaging–based studies and between 2018-2022 for the app-based studies, indicating a relatively recent focus on app-based interventions. Mobile phone apps can provide more interactive and personalized features than SMS text messaging, such as tracking progress, setting goals, and sending notifications [23]. Apps can also offer real-time support and a wealth of resources, including educational materials, coping strategies, and social support networks [24]. However, the existing literature on app-based interventions exhibits significant heterogeneity and conflicting results. The lack of standardized protocols and guidelines for app-based smoking cessation interventions further exacerbates the challenge of generating consistent evidence. Therefore, there is a need for further research on smoking cessation apps targeted toward young people. In particular, future studies should explore the optimal features and design of smoking cessation apps.

Despite its many advantages, high dropout [51] and nonadherence [52] remain significant limitations. Incentives were offered in more than half of the included trials, and these studies generally had lower dropout rates than those without incentives. The majority of included studies with incentives had loss to follow-up rates ranging from 6% to 24%, while studies without incentives had rates between 26% and 92%. We note that the study by Crane et al [23], which was conducted in a real-world setting, reported the highest rate of missing data. This finding was in line with previous research highlighting the challenge of retaining participants in real-world studies [53]. A review showed that the personalization of content, app design, reminder form, and personal support help to improve adherence, but research on factors that influence adherence to mobile health apps remains limited [52]. Increasing people's engagement and retention is important as a key consideration for such interventions on a large scale in the real world. Future research should focus on identifying the most effective design, personalized content, features, and support mechanisms to increase their uptake and intention.

Limitations

Several limitations must be considered when interpreting the findings of this review. First, few studies examined app-based interventions, which limited the amount of available evidence for analysis and interpretation. Furthermore, considerable heterogeneity was observed among app-based studies. This scarcity of data highlights the need for further research in this area to improve the understanding of the efficacy of app-based interventions for smoking cessation. Second, the studies included in this review were largely conducted in high-income countries, limiting the generalizability of the findings to other settings. Lastly, a significant number of studies relied on self-reported abstinence without biochemical validation, raising some concerns in the domain of outcome measurement. In light of these limitations, future research should aim to address these issues and ensure the production of high-quality evidence to guide mobile phone–based smoking cessation interventions.

Conclusions

Our meta-analysis provides evidence that SMS text messaging smoking cessation interventions are effective among young people. There is a need for further research on smoking cessation apps, especially those targeted at young people. Future research should also focus on identifying the most effective mobile phone–based cessation approaches and on developing strategies to increase their uptake and intention.

Acknowledgments

This work was supported by the National High Level Hospital Clinical Research Funding (2022-NHLHCRF-LX-01-0304). The sponsor of the study had no role in the study design, collection, analysis, interpretation, or writing of this report.

Authors' Contributions

All authors were involved in the planning of the study, literature review, interpretation of the findings, and manuscript preparation. DX and XZ conceived and designed the study. DX supervised the study. DX and XZ drafted the report. XZ and ZS did the statistical analysis. All authors contributed to the data collection, analysis, and interpretation. All authors revised the report and approved the final version before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies of PubMed, Web of Science, Embase, and the Cochrane Library.
[DOCX File, 27 KB - mhealth_v11i1e48253_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist
[PDF File, 99 KB - mhealth_v11i1e48253_app2.pdf]

References

1. Reitsma MB, Fullman N, Ng M, Salama JS, Abajobir A, Abate KH. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the global burden of disease study 2015. Lancet 2017 May 13;389(10082). [doi: 10.1016/S0140-6736(17)30819-X] [Medline: 28390697]
2. Campos MW, Serebrisky D, Castaldelli-Maia JM. Smoking and cognition. Curr Drug Abuse Rev 2016;9(2):76-79. [doi: 10.2174/1874473709666160803101633] [Medline: 27492358]
3. Tian J, Venn AJ, Blizzard L, Patton GC, Dwyer T, Gall SL. Smoking status and health-related quality of life: a longitudinal study in young adults. Qual Life Res 2016 Mar;25(3):669-685. [doi: 10.1007/s11136-015-1112-6] [Medline: 26310284]
4. Reitsma MB, Flor LS, Mullany EC, Gupta V, Hay SI, Gakidou E. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990-2019. Lancet Public Health 2021 Jul;6(7):e472-e481. [doi: 10.1016/S2468-2667(21)00102-X] [Medline: 34051921]
5. Pirie K, Peto R, Reeves GK, Green J, Beral V, Million Women Study Collaborators. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. Lancet 2013 Jan 12;381(9861):133-141. [doi: 10.1016/S0140-6736(12)61720-6] [Medline: 23107252]
6. Shang X, Guo K, E F, Deng X, Wang Y, Wang Z, et al. Pharmacological interventions on smoking cessation: a systematic review and network meta-analysis. Front Pharmacol 2022 Oct 24;13:1012433. [doi: 10.3389/fphar.2022.1012433] [Medline: 36353488]
7. Hartmann-Boyce J, Livingstone-Banks J, Ordóñez-Mena JM, Fanshawe TR, Lindson N, Freeman SC, et al. Behavioural interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2021 Jan 4;1:CD013229. [doi: 10.1002/14651858.CD013229.pub2] [Medline: 33411338]
8. Adams JM. Smoking cessation-progress, barriers, and new opportunities: the surgeon general’s report on smoking cessation. JAMA 2020 Jun 23;323(24):2470-2471. [doi: 10.1001/jama.2020.6647] [Medline: 32432659]
9. Graham AL, Papandonatos GD, Jacobs MA, Amato MS, Cha S, Cohn AM, et al. Optimizing text messages to promote engagement with internet smoking cessation treatment: results from a factorial screening experiment. J Med Internet Res 2020 Apr 2;22(4):e17734. [doi: 10.2196/17734] [Medline: 32238338]
10. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328. [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
11. Wang MP, Luk TT, Wu Y, Li WH, Cheung DY, Kwong AC, et al. Chat-based instant messaging support integrated with brief interventions for smoking cessation: a community-based, pragmatic, cluster-randomised controlled trial. Lancet Digit Health 2019 Aug;1(4):e183-e192. [doi: 10.1016/S2589-7500(19)30082-2] [Medline: 33323188]
12. Tang J, Liao Y. Effectiveness of a text messaging-based smoking cessation intervention (Happy Quit) in China: a randomised controlled trial. Lancet 2018 Oct;392:S2. [doi: 10.1016/S0140-6736(18)32635-7] [Medline: 30562352]
13. Daly AT, Deshmukh AA, Vidrine DJ, Prokhorov AV, Frank SG, Tahay PD, et al. Cost-effectiveness analysis of smoking cessation interventions using cell phones in a low-income population. Tob Control 2019 Jan;28(1):88-94. [doi: 10.1136/tobaccocontrol-2017-054229] [Medline: 29886411]

14. Coleman T, Clark M, Welch C, Whitemore R, Leonardi-Bee J, Cooper S, et al. Effectiveness of offering tailored text message, self-help smoking cessation support to pregnant women who want information on stopping smoking: miQuit3 randomised controlled trial and meta-analysis. Addiction 2022 Apr;117(4):1079-1094. [doi: 10.1111/add.15715] [Medline: 34636086]
15. Mays D, Johnson AC, Phan L, Sanders C, Shoben A, Tercyak KP, et al. Tailored mobile messaging intervention for waterpipe tobacco cessation in young adults: a randomized trial. Am J Public Health 2021 Sep;111(9):1686-1695. [doi: 10.2105/AJPH.2021.306389] [Medline: 34436927]
16. Puigdomènech E, Trujillo-Gómez JM, Martín-Cantera C, Díaz-Gete L, Manzano-Montero M, Sánchez-Fondevila J, et al. Information and communication technologies for approaching smokers: a descriptive study in primary healthcare. BMC Public Health 2015;15:2. [doi: 10.1186/1471-2458-15-2] [Medline: 25971903]
17. ITU. Individuals using the internet. 2023. URL: www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2023-08-25]
18. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10(10):CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
19. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893. [doi: 10.2196/10893] [Medline: 30355563]
20. Bendtsen M, Bendtsen P, Müssener U. Six-month outcomes from the NEXit Junior trial of a text messaging smoking cessation intervention for high school students: randomized controlled trial with Bayesian analysis. JMIR Mhealth Uhealth 2021 Oct;9(10):e29913. [doi: 10.2196/29913] [Medline: 34673532]
21. World Health Organization. WHO framework convention on tobacco control. 2003. URL: apps.who.int/iris/handle/10665/42811 [accessed 2023-08-29]
22. Liao Y, Wu Q, Kelly BC, Zhang F, Tang YY, Wang Q, et al. Effectiveness of a text-messaging-based smoking cessation intervention ("Happy Quit") for smoking cessation in China: a randomized controlled trial. PLoS Med 2018 Dec 18;15(12):e1002713. [doi: 10.1371/journal.pmed.1002713] [Medline: 30562352]
23. Crane D, Ubhi HK, Brown J, West R. Relative effectiveness of a full versus reduced version of the 'smoke free' mobile application for smoking cessation: an exploratory randomised controlled trial. F1000Res 2018 Sep 21;7:1524. [doi: 10.12688/f1000research.16148.2] [Medline: 30728950]
24. Chulasai P, Chinwong D, Vientong P, Lertsinudom S, Kanjanarat P, Hall JJ, et al. Smartphone application for smoking cessation (quit with US): a randomized controlled trial among young adult light smokers in Thailand. Int J Environ Res Public Health 2022 Jul 6;19(14):8265. [doi: 10.3390/ijerph19148265] [Medline: 35886120]
25. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
26. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
27. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55. [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
28. Graham AL, Amato MS, Cha S, Jacobs MA, Bottcher MM, Papandonatos GD. Effectiveness of a vaping cessation text message program among young adult e-cigarette users: a randomized clinical trial. JAMA Intern Med 2021 Jul 1;181(7):923-930. [doi: 10.1001/jamainternmed.2021.1793] [Medline: 33999133]
29. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171. [doi: 10.2196/jmir.2636] [Medline: 23956024]
30. Mason M, Mennis J, Way T, Zaharakis N, Campbell LF, Benotsch EG, et al. Text message delivered peer network counseling for adolescent smokers: a randomized controlled trial. J Prim Prev 2016 Oct;37(5):403-420. [doi: 10.1007/s10935-016-0439-2] [Medline: 27388626]
31. Palmer AM, Tomko RL, Squeglia LM, Gray KM, Carpenter MJ, Smith TT, et al. A pilot feasibility study of a behavioral intervention for nicotine vaping cessation among young adults delivered via telehealth. Drug Alcohol Depend 2022 Mar 1;232:109311. [doi: 10.1016/j.drugalcdep.2022.109311] [Medline: 35123362]
32. Shi HJ, Jiang XX, Yu CY, Zhang Y. Use of mobile phone text messaging to deliver an individualized smoking behaviour intervention in Chinese adolescents. J Telemed Telecare 2013 Jul;19(5):282-287. [doi: 10.1177/1357633X13495489] [Medline: 24163238]
33. Villanti AC, Peasley-Miklus C, Cha S, Schulz J, Klemperer EM, LePine SE, et al. Tailored text message and web intervention for smoking cessation in U.S. socioeconomically-disadvantaged young adults: a randomized controlled trial. Prev Med 2022 Dec;165(Pt B):107209. [doi: 10.1016/j.ypmed.2022.107209] [Medline: 35995105]

34. Ybarra ML, Holtrop JS, Prescott TL, Rahbar MH, Strong D. Pilot RCT results of stop my smoking USA: a text messaging-based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399. [doi: 10.1093/ntr/nts339] [Medline: 23348969]
35. Fanshawe TR, Halliwell W, Lindson N, Aveyard P, Livingstone-Banks J, Hartmann-Boyce J. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2017 Nov 17;11(11):CD003289. [doi: 10.1002/14651858.CD003289.pub6] [Medline: 29148565]
36. Eppes EV, Augustyn M, Gross SM, Vernon P, Caulfield LE, Paige DM. Engagement with and acceptability of digital media platforms for use in improving health behaviors among vulnerable families: systematic review. J Med Internet Res 2023 Feb 3;25:e40934. [doi: 10.2196/40934] [Medline: 36735286]
37. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]
38. Sha L, Yang X, Deng R, Wang W, Tao Y, Cao H, et al. Automated digital interventions and smoking cessation: systematic review and meta-analysis relating efficiency to a psychological theory of intervention perspective. J Med Internet Res 2022 Nov 16;24(11):e38206. [doi: 10.2196/38206] [Medline: 36383408]
39. Vijayaraghavan M, Elser H, Frazer K, Lindson N, Apollonio D. Interventions to reduce tobacco use in people experiencing homelessness. Cochrane Database Syst Rev 2020 Dec 3;12(12):CD013413. [doi: 10.1002/14651858.CD013413.pub2] [Medline: 33284989]
40. Villanti AC, West JC, Klemperer EM, Graham AL, Mays D, Mermelstein RJ, et al. Smoking-cessation interventions for U.S young adults: updated systematic review. Am J Prev Med 2020 Jul;59(1):123-136. [doi: 10.1016/j.amepre.2020.01.021] [Medline: 32418800]
41. Krishnan N, Gu J, Abroms LC. Mobile phone-based messaging for tobacco cessation in low and middle-income countries: a systematic review. Addict Behav 2021 Feb;113:106676. [doi: 10.1016/j.addbeh.2020.106676] [Medline: 33038676]
42. Gallus S, Muttarak R, Franchi M, Pacifici R, Colombo P, Boffetta P, et al. Why do smokers quit? Eur J Cancer Prev 2013 Jan;22(1):96-101. [doi: 10.1097/CEJ.0b013e3283552da8] [Medline: 22644233]
43. Villanti AC, Bover Manderski MT, Gundersen DA, Steinberg MB, Delnevo CD. Reasons to quit and barriers to quitting smoking in US young adults. Fam Pract 2016 Apr;33(2):133-139. [doi: 10.1093/fampra/cmv103] [Medline: 26733658]
44. Wang T, Wang W, Liang J, Nuo M, Wen Q, Wei W, et al. Identifying major impact factors affecting the continuance intention of mHealth: a systematic review and multi-subgroup meta-analysis. NPJ Digit Med 2022 Sep 15;5(1):145. [doi: 10.1038/s41746-022-00692-9] [Medline: 36109594]
45. McLaughlin I, Dani JA, De Biasi M. Nicotine withdrawal. Curr Top Behav Neurosci 2015;24:99-123. [doi: 10.1007/978-3-319-13482-6_4] [Medline: 25638335]
46. US Department of Health and Human Services. Smoking cessation: a report of the surgeon general. URL: www.hhs.gov/sites/default/files/2020-cessation-sgr-full-report.pdf [accessed 2023-08-29]
47. Messer K, Trinidad DR, Al-Delaimy WK, Pierce JP. Smoking cessation rates in the United States: a comparison of young adult and older smokers. Am J Public Health 2008 Feb;98(2):317-322. [doi: 10.2105/AJPH.2007.112060] [Medline: 18172143]
48. Gray KM, Rubinstein ML, Prochaska JJ, DuBrava SJ, Holstein AR, Samuels L, et al. High-dose and low-dose varenicline for smoking cessation in adolescents: a randomised, placebo-controlled trial. Lancet Child Adolesc Health 2020 Nov;4(11):837-845. [doi: 10.1016/S2352-4642(20)30243-1] [Medline: 32979939]
49. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn’t work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
50. Barroso-Hurtado M, Suárez-Castro D, Martínez-Vispo C, Becoña E, López-Durán A. Smoking cessation apps: a systematic review of format, outcomes, and features. Int J Environ Res Public Health 2021 Nov 6;18(21):11664. [doi: 10.3390/ijerph182111664] [Medline: 34770178]
51. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
52. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
53. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014 Apr;16(4):e109. [doi: 10.2196/jmir.3084] [Medline: 24735567]

Abbreviations

MH: Mantel-Haenszel

PPA: point prevalence abstinence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio

Edited by L Buis; submitted 17.04.23; peer-reviewed by F Khordastan, H Guo, N Mungoli, N Li; revised version received 04.07.23; accepted 25.07.23; published 12.09.23.

Please cite as:
Zhou X, Wei X, Cheng A, Liu Z, Su Z, Li J, Qin R, Zhao L, Xie Y, Huang Z, Xia X, Liu Y, Song Q, Xiao D, Wang C
Mobile Phone–Based Interventions for Smoking Cessation Among Young People: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2023;11:e48253
URL: https://mhealth.jmir.org/2023/1/e48253
doi:10.2196/48253

Assessing the Effectiveness of mHealth Interventions for Diabetes and Hypertension Management in Africa: Systematic Review and Meta-Analysis

Pearl Aovare[1,*], MPH; Kasim Abdulai[2,*], PhD; Amos Laar[3,*], PhD; Eva L van der Linden[4,*], MD; Nicolaas Moens[5,*], PhD; Edo Richard[6,7,*], PhD; Eric P Moll van Charante[1,8,*], MD, PhD; Charles Agyemang[1,*], PhD

1

2

3

4

5

6

7

8

[*]all authors contributed equally

Corresponding Author:
Pearl Aovare, MPH

Abstract

Background: Mobile health (mHealth) interventions are effective in improving chronic disease management, mainly in high-income countries. However, less is known about the efficacy of mHealth interventions for the reduction of cardiovascular risk factors, including for hypertension and diabetes, which are rapidly increasing in low- and middle-income countries.

Objective: This study aimed to assess the efficacy of mHealth interventions for diabetes and hypertension management in Africa.

Methods: We searched PubMed, Cochrane Library, Google Scholar, African Journals Online, and Web of Science for relevant studies published from inception to July 2022. The main outcomes of interest were changes in hemoglobin A_{1c} (HbA_{1c}), systolic blood pressure, and diastolic blood pressure. The random or fixed effect model was used for the meta-analysis, and the I^2 statistic was used to gauge study heterogeneity. *Z* tests and *P* values were used to evaluate the effect of mHealth interventions on HbA_{1c} and blood pressure levels.

Results: This review included 7 studies (randomized controlled trials) with a total of 2249 participants. Two studies assessed the effect of mHealth on glycemic control, and 5 studies assessed the effect of mHealth on blood pressure control. The use of mHealth interventions was not associated with significant reductions in HbA_{1c} levels (weighted mean difference [WMD] 0.20, 95% CI −0.40 to 0.80; *P*=.51) among patients with diabetes and systolic blood pressure (WMD −1.39, 95% CI −4.46 to 1.68; *P*=.37) and diastolic blood pressure (WMD 0.36, 95% CI −1.37 to 2.05; *P*=.69) among patients with hypertension. After conducting sensitivity analyses using the leave-one-out method, the Kingue et al study had an impact on the intervention, resulting in a 2 mm Hg reduction in systolic blood pressure (WMD −2.22, 95% CI −3.94 to −0.60; *P*=.01) but was nonsignificant for diastolic blood pressure and HbA_{1c} levels after omitting the study.

Conclusions: Our review provided no conclusive evidence for the effectiveness of mHealth interventions in reducing blood pressure and glycemic control in Africa among persons with diabetes and hypertension. To confirm these findings, larger randomized controlled trials are required.

Trial Registration: PROSPERO CRD42021230642; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=230642

(JMIR Mhealth Uhealth 2023;11:e43742) doi:10.2196/43742

KEYWORDS

mobile health; interventions; diabetes; blood sugar; hypertension; management; effectiveness; chronic disease; Africa; blood pressure; glycemic; intervention; mHealth; efficiency; resource

Introduction

Managing chronic diseases often calls for a long-term care strategy [1]. Diabetes and hypertension remain two of the most common chronic conditions globally, resulting in the highest health care resource use and mortality [2-4]. Type 2 diabetes prevalence has become a substantial health issue, especially in African regions where type 2 diabetes is predicted to increase at the quickest rate (129%) in the world by 2045 [5-7]. Similarly, hypertension remains a major public health challenge among older adults in the African region, with an estimated pooled prevalence of 30.8% in Africa and between 30% and 31.1% in sub-Saharan Africa [8]. Poor blood pressure control among persons with hypertension is thought to involve intricate interactions between patients, health care providers, and socioeconomic variables [9]. Medication adherence has also been identified as one of the critical disease management issues, especially in enhancing life quality, health outcomes, and access to affordable health care worldwide [10,11].

Disease management programs using mobile health (mHealth) are promising emerging strategies to help patients self-manage their conditions (eg, measuring their blood pressure and sugar levels with remote professional support when needed [7,12]). mHealth is a medical and public health practice supported by portable electronic devices such as cell phones, patient monitoring devices, personal digital assistants, and other wireless gadgets [13]. This includes the use of phones and remote monitoring devices in health care and public health practice for communication, data collection, patient monitoring, and education, and to facilitate adherence to chronic disease management [14,15]. mHealth devices can improve service delivery and impact patient outcomes [15].

Previous studies in some low- and middle-income countries have assessed the application of mHealth as a tool to increase drug compliance in patients with a range of long-term illnesses, including diabetes, chronic obstructive pulmonary disease, and HIV infection [11,16]. Although several African countries are still in the pilot and development stages, an increasing number of mHealth apps have been put into use in clinical care settings [17]. The majority of small-scale pilot or feasibility mHealth intervention studies in Africa have been based on SMS text messaging systems to improve disease management [17,18].

Most individuals now possess mobile phones, and there are over 5.3 billion subscribers to mobile services worldwide, 67% of the world's population [19,20]. There will be 400 million more new mobile service customers between now and 2025, the majority of whom will come from Asia Pacific and sub-Saharan Africa, increasing the total number of subscribers to 5.7 billion (70% of the global population) [21]. There have been individual studies in Africa about mHealth interventions on disease management [22], although the data on the efficacy of mHealth in the management of diabetes and hypertension in Africa are limited and have not yet been systematically evaluated. Therefore, this systematic review assessed the effectiveness of mHealth interventions on blood pressure control among patients with hypertension and glycemic control among patients with diabetes in Africa. The findings of this paper will guide improvements to the adoption of mHealth for the management of diabetes and hypertension in African countries.

Methods

This systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23] (Multimedia Appendix 1). The protocol was registered in PROSPERO (CRD42021230642).

Search Strategy

PubMed, Cochrane Library, Google Scholar, African Journals Online, and Web of Science were searched for relevant studies published from inception to July 2022, with assistance from a clinical librarian. The full search strategies, common Medical Subject Headings (MeSH), and search terms used across databases are available in Table S1 in Multimedia Appendix 2. The reference lists of the included studies were hand-searched to identify additional relevant studies.

Study Selection

Two independent authors (PA and KA) manually assessed and screened studies for both the titles and abstracts as well as full-text articles using an Excel sheet (Microsoft Corporation). Disagreements were resolved by consensus with a third author (CA) as necessary. This was performed in three stages as follows. First, PA screened the titles of all papers to determine their relevance. KA performed a cross-check by screening 20% of the titles excluded by the first reviewer, and it was confirmed that none of the titles screened by the second reviewer met the inclusion criteria. Second, abstracts of the papers selected after the title screening stage were again screened by PA and KA following the same procedure as described in step one. Finally, the full texts of potentially relevant papers were retrieved and evaluated by PA and KA independently to ascertain their relevance and usefulness to the review. Disagreements were settled through dialogue with CA to reach an agreement. Duplicates were also identified using EndNote reference manager (version x9; Clarivate).

Inclusion Criteria

We included studies that met the following criteria: the patients had hypertension or diabetes and were 18 years or older; the patients had received treatment at a selected health care setting; the intervention included an mHealth component; the results included target values of hemoglobin A_{1c} (HbA_{1c}), systolic blood pressure, or diastolic blood pressure; the studies were randomized controlled trials (RCTs); the articles were written in English; and the studies were conducted in hospitals and primary health centers.

Exclusion Criteria

We excluded studies in which the full text was not available after attempts to contact the author, the research participants were pregnant women or a specific patient population (eg, patients with cancer), the results did not describe primary outcomes, the primary intervention did not use mHealth devices, or the articles were unpublished manuscripts or conference abstracts.

Risk of Bias Assessment

The quality of each study was assessed using a 28-point scoring system as adopted from the Downs and Black checklist [24]. The included studies focused on the following items for assessment: items 1 through 10 evaluated whether the information provided was adequate for the reader to make an objective assessment of the study's findings; items 11 through 13 evaluated external validity, which examined the extent to which study findings could be applied to the population from which the study participants were drawn; items 14 through 20 assessed possible bias, which focused on biases in the assessment of the intervention and the result; and items 21 through 26 assessed confounding, which focused on biases in the research participant selection. To determine if neutral research results may be the result of chance or insufficient power, item 27 evaluated the study's power (Table 1).

Table . Downs and Black quality assessment.

Studies	Information based on study findings score (questions 1-10)	External validity score (questions 11-13)	Potential bias score (questions 14-20)	Confounding score (questions 21-26)	Power of study score (question 27)	Total score (maximum score of 27)	Quality as per the cutoff described
Abaza and Marschollek [25], 2017	10	3	6	5	0	24	Good
Adjei et al [26], 2015	10	3	5	5	0	23	Good
Asante et al [27], 2020	10	3	4	5	0	22	Good
Bobrow et al [28], 2016	10	3	5	6	0	24	Good
Kingue et al [29], 2013	10	3	5	6	0	24	Good
Owolabi et al [30], 2019	10	3	6	5	0	24	Good
Sarfo et al [31], 2019	10	3	6	6	0	25	Good

Data Extraction

Two authors (PA and KA) independently extracted the following study characteristics from each included article using a tested extraction form: first author, year of publication, mean age, the country where the study was conducted, the participant (patient with diabetes or hypertension), mHealth location (primary care setting, hospital, clinic, etc), condition (diabetes/hypertension), sample size, mHealth intervention, study design, and outcome of the intervention.

Data Synthesis and Analysis

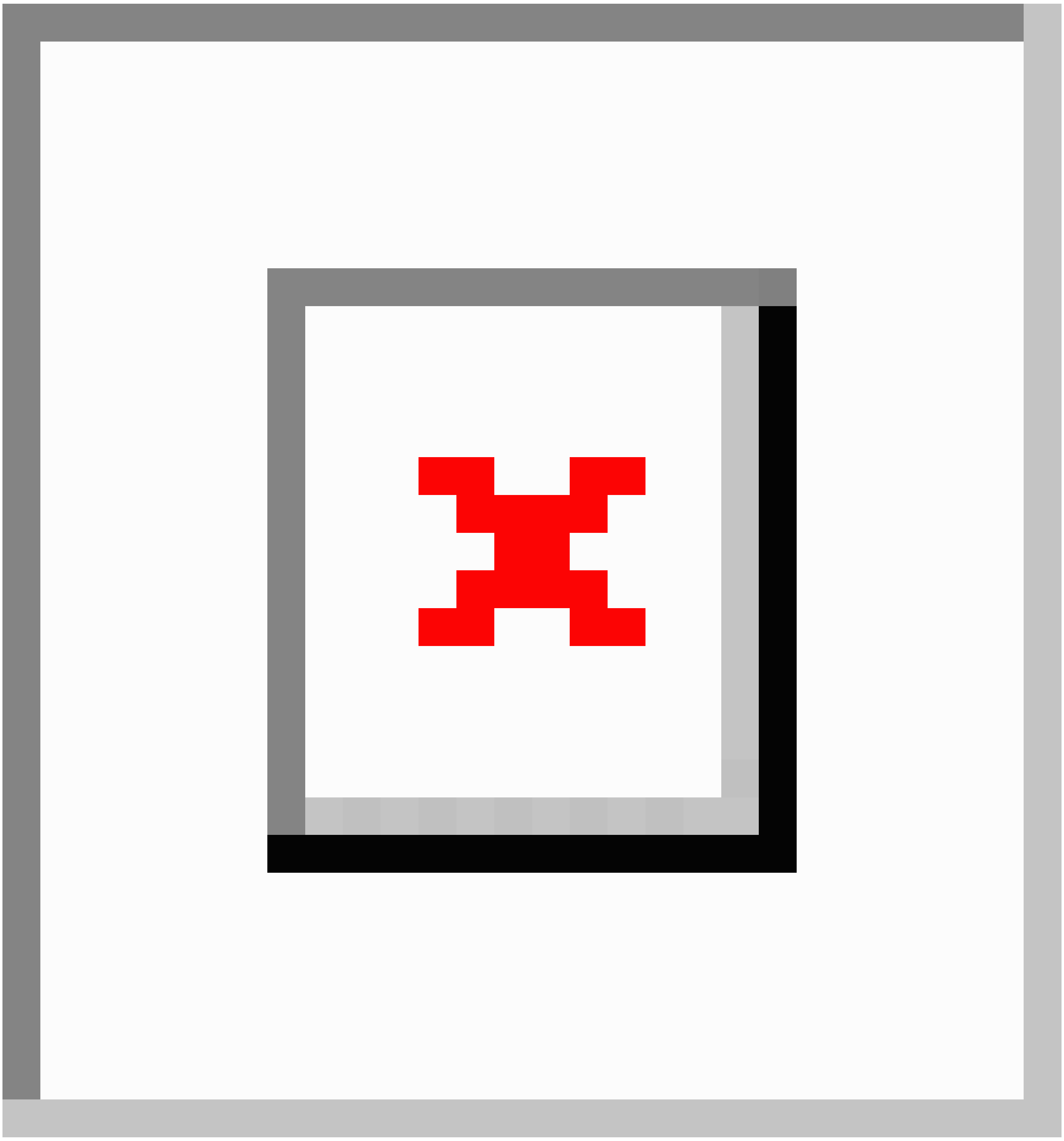
The data for the primary outcomes (HbA_{1c}, systolic blood pressure, and diastolic blood pressure) were analyzed separately using random or fixed effects models with a weighted mean difference (WMD) in ReviewManager (version 5.4; The Cochrane Collaboration) [32]. The I^2 statistic was calculated to measure the percentage of variation across trials due to heterogeneity, with values of <50% and ≥50% indicating low and high levels of heterogeneity, respectively. The WMD for blood pressure and HbA_{1c} between the intervention and control and *Z* tests were used to compare groups, and $P<.05$ was regarded as statistically significant. We checked publication bias subjectively by funnel plots and objectively by Begg and Egger tests using Stata version 16 (StataCorp). Begg and Egger tests with $P<.05$ were considered to have significant publication bias.

Results

Study Selection

We identified 2908 records from our search; authors screened 2880 titles and abstracts after duplicates were removed. In total, 7 studies (RCTs) [25-29,31] were considered eligible for inclusion. The reasons for excluding a study are provided in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of evidence search and selection.



Study Characteristics

The characteristics of the studies are shown in Tables 2 and 3. Of the 2249 participants, 1824 (54.1%) were female, the mean age ranged from 51.2 to 60.6 years, and the sample sizes ranged from 60 to 1372 participants. The studies were conducted at hospitals and primary health centers in the following countries: Ghana (n=3), South Africa (n=2), Egypt (n=1), and Cameroon (n=1). Five studies [26-29,31] reported the apps' effectiveness in controlling blood pressure among patients with hypertension (Table 4), while 2 studies [25,27] reported the effect on HbA_{1c} levels in patients with diabetes (Table 5). Of the 7 included studies, 2 articles had more than one intervention group with the same outcome measure [25,28].

Table . Study characteristics.

Studies	Study design	Country	Gender (male/female; %)	Condition	Type of patient	Patients who received treatment (%)	mHealth[a] study settings
Abaza and Marschollek [25], 2017	RCT[b]	Egypt	47/53	Diabetes	Adult with diabetes	—[c]	Clinic
Adjei et al [26], 2015	RCT	Ghana	44/56	Diabetes	Adult with diabetes	—	Health care center
Asante et al [27], 2020	RCT	Ghana	27/73	Diabetes	Adults with type 2 diabetes	66.3	Health center
Bobrow et al [28], 2016	RCT	South Africa	56/44	Hypertension	Adults with hypertension	50	Primary care clinic
Kingue et al [29], 2013	RCT	Cameroon	35/65	Hypertension	Young adults with hypertension	73.3	Clinic
Owolabi et al [30], 2019	RCT	South Africa	84/16	Diabetes	Adult with diabetes	75	Primary health care centers
Sarfo et al [31], 2019	RCT	Ghana	65/35	Hypertension	Adults with hypertension	13.3	Clinic

[a]mHealth: mobile health.

[b]RCT: randomized controlled trial.

[c]Not available.

Table . Study intervention and control description.

Studies	Intervention type (duration)	Intervention group	Control group
Abaza and Marschollek [25], 2017	SMS text messaging (3 mo)	Patients received daily messages and weekly reminders addressing various diabetes care categories.	The control group did not receive SMS text messages but received paper-based educational material.
Adjei et al [26], 2015	Electronic reminders (6 mo)	The intervention group was given electronic reminders for their clinical appointments, and their physicians were prompted for abnormal laboratory results [33] for 6 months.	Patients received only the usual care.
Asante et al [27], 2020	Mobile phone calls (3 mo)	The mobile phone call intervention was delivered by nurses in addition to care as usual over 12 weeks. The intervention group received up to 16 mobile phone calls (mean duration 12 min) from a diabetes specialist nurse in addition to their care as usual.	The control group received only care as usual.
Bobrow et al [28], 2016	SMS text messaging (6 and 12 mo)	SMS text messages were delivered automatically via an open source, web-based electronic medical record system. Texts were sent for 1 year from enrollment. Blood pressure measurements were collected from participants as they attended their routine clinic visits. The delivery of texts was automatically tracked, and if undelivered, a research assistant that was blinded to group allocation would contact the number of a friend or relative to obtain a new mobile phone number	The usual care group continued to receive care from the clinic and some form of written information about hypertension and healthy living, but no personalized SMS text messages were sent.
Kingue et al [29], 2013	Mobile phone calls (24 wk)	Interactive electronic communication were delivered between the patient and the provider or between multiple providers in either synchronous or asynchronous settings for the provision of health care services or consultation.	The control group only received routine treatment and care from the clinic.
Owolabi et al [30], 2019	SMS text messages (6 mo)	Participants in the intervention arm received daily educational SMS text messages on diabetes for 6 months. In addition, the intervention group received the text at an agreed time of the day, according to their needs, care plan, and goals.	The control groups proceeded with their usual care including all medical visits, tests, and diabetes support at the clinic.
Sarfo et al [31], 2019	SMS text messages (9 mo)	Patients received a Bluetooth blood pressure device and smartphone with an app for monitoring blood pressure measurements and medication intake under nurse guidance for 3 months. Participants also received motivational and support messages, advice on lifestyle behaviors like diets, physical activity, smoking cessation, and medication and appointment reminders.	The control arm received only the usual care.

Table . Study outcome for blood pressure.

Studies	Sample size, N	Age (years), mean (SD)	Intervention, mean (SD)		Control, mean (SD)	
			Systolic blood pressure (mm Hg)	Diastolic blood pressure (mm Hg)	Systolic blood pressure (mm Hg)	Diastolic blood pressure (mm Hg)
Adjei et al [26], 2015	200	47.6 (9.1)	122.9 (18.3)	71.3 (8.5)	124.8 (4.2)	72.3 (9.7)
Asante et al [27], 2020	60	55.1 (10.9)	134 (27.4)	85.2 (17)	150.9 (24.9)	87.3 (12.9)
Bobrow et al [28], 2016[a]	1372	54.3 (11.5)	132.7 (17.5)	—[b]	134.3 (17.3)	—
Bobrow et al [28], 2016[c]	1372	54.3 (11.5)	132.1 (16.6)	—	134.3 (17.3)	—
Kingue et al [29], 2013	268	59.9 (10.4)	169.2 (27.9)	100.4 (18.3)	160.8 (23.7)	95.2 (14.8)
Owolabi et al [30], 2019	216	60.6 (11.6)	144.3 (21.2)	82.3 (10.3)	146.3 (23.8)	82.8 (15.1)
Sarfo et al [31], 2019	60	55 (13)	141.3 (30.3)	91.4 (18.0)	146.3 (22.5)	89.6 (12.9)

[a]Interactive intervention group vs control.

[b]Not available.

[c]Information only intervention group vs control.

Table . Study outcome for hemoglobin A_{1c} (HbA_{1c}).

Studies	Sample size, N	Age (years), mean (SD)	Intervention HbA_{1c} (%), mean (SD)	Control HbA_{1c} (%), mean (SD)
Abaza and Marschollek [25], 2017 (baseline)	73	51.2 (8.7)	9.8 (2.5)	9.5 (2.8)
Abaza and Marschollek [25], 2017 (end point)	73	51.2 (8.7)	8.7 (2.0)	8.8 (2.4)
Asante et al [27], 2020	60	55.1 (10.9)	9.5 (2.0)	9.1 (1.7)

Meta-Analysis of the Effects on Primary Outcomes

A total of 7 studies, 5 for blood pressure [26-29,31] and 2 for HbA_{1c} [25,27], were included in the meta-analysis.

Systolic Blood Pressure

As shown in Figure 2, one study had more than one intervention group with the same outcome measured [28]; therefore, 7 interventions are shown in the forest plot of systolic blood pressure, and the estimated WMD of systolic blood pressure between intervention and control groups was not statistically significant at −1.39 mm Hg (95% CI −4.46 to 1.68; P=.37; I^2=61%). No significant publication bias was detected visually by the funnel plot (Figure 3) or statistically by Begg (P=.30) and Egger (P=.10) tests.

Figure 2. Forest plot of the difference in systolic blood pressure between the mHealth intervention group and control group in 6 studies [26-31]. Bobrow et al [28]: (A) interactive intervention group vs control; (B) information only intervention group vs control. mHealth: mobile health.

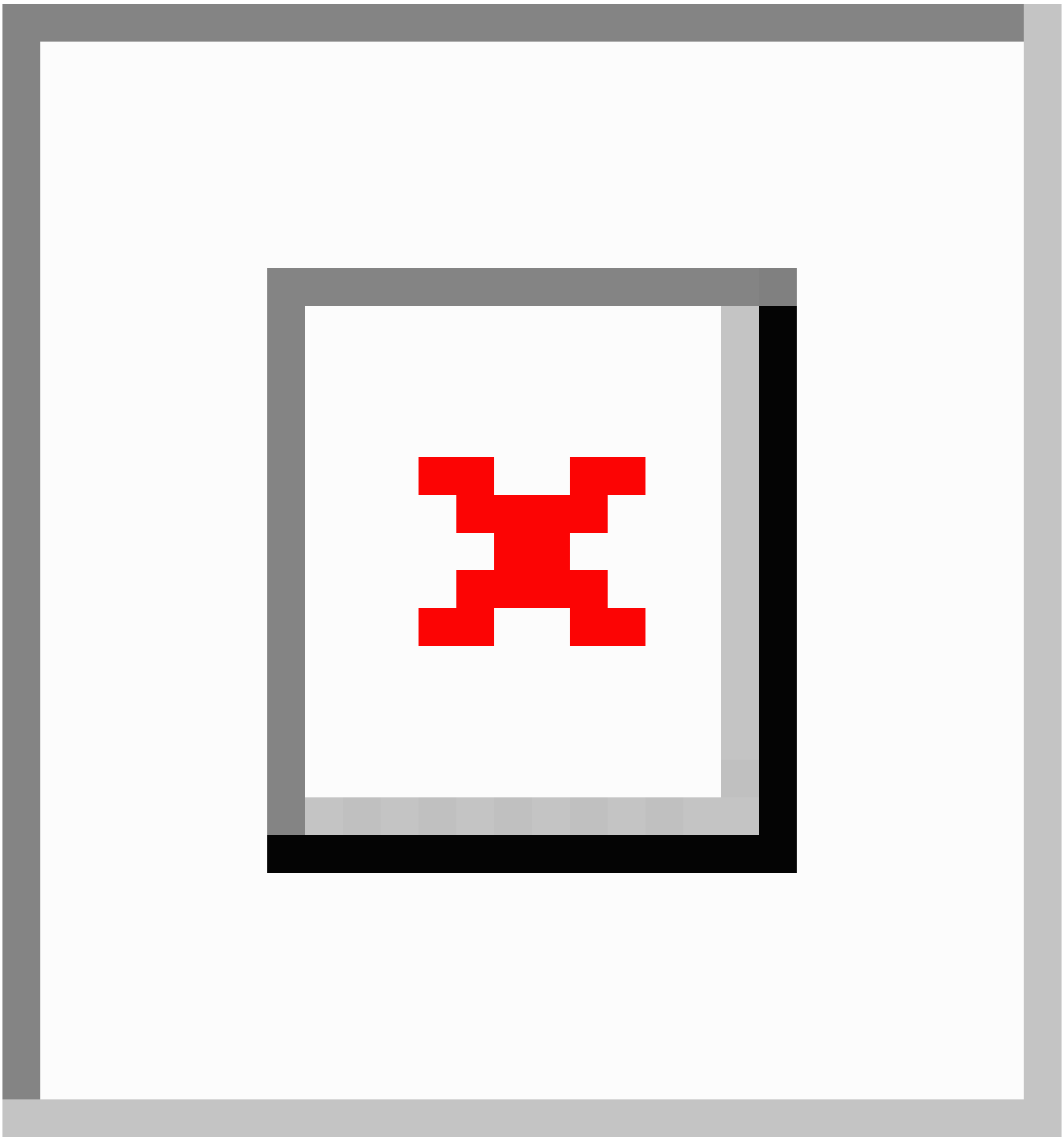
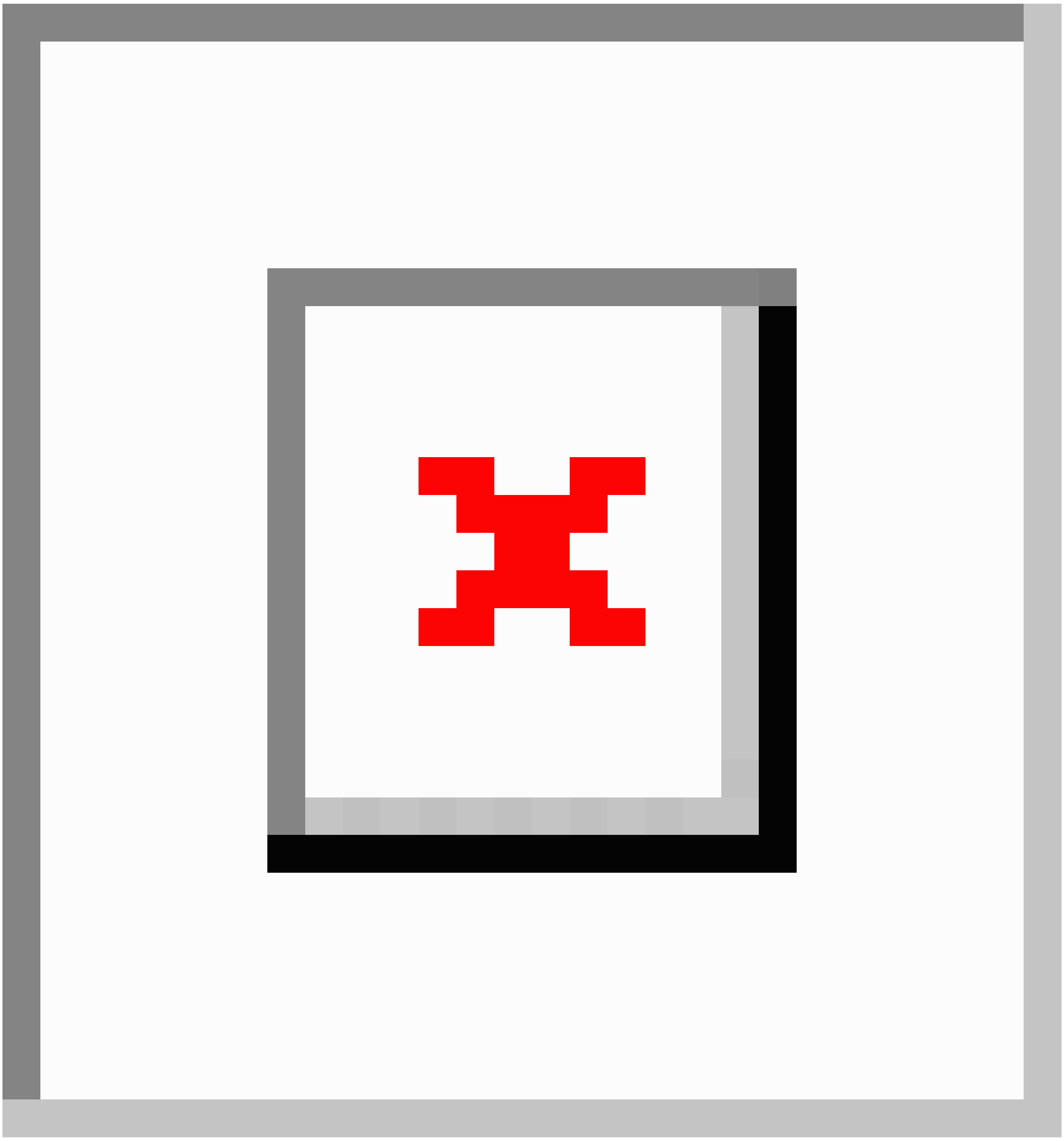


Figure 3. Funnel plot of the difference in systolic blood pressure between the mobile health intervention group and control group. diff.: difference.



Diastolic Blood Pressure

There was no statistically significant difference in diastolic blood pressure (0.36 mm Hg, 95% CI −1.37 to 2.08; *P*=.69; I^2=47%) between the intervention and control groups (Figure 4). No significant publication bias was detected visually by the funnel plot (Figure 5) or statistically by Begg (*P*=.65) and Egger (*P*=.81) tests.

Figure 4. Forest plot of the difference in diastolic blood pressure between the mHealth intervention group and control group in 5 studies [26,27,29-31]. mHealth: mobile health.

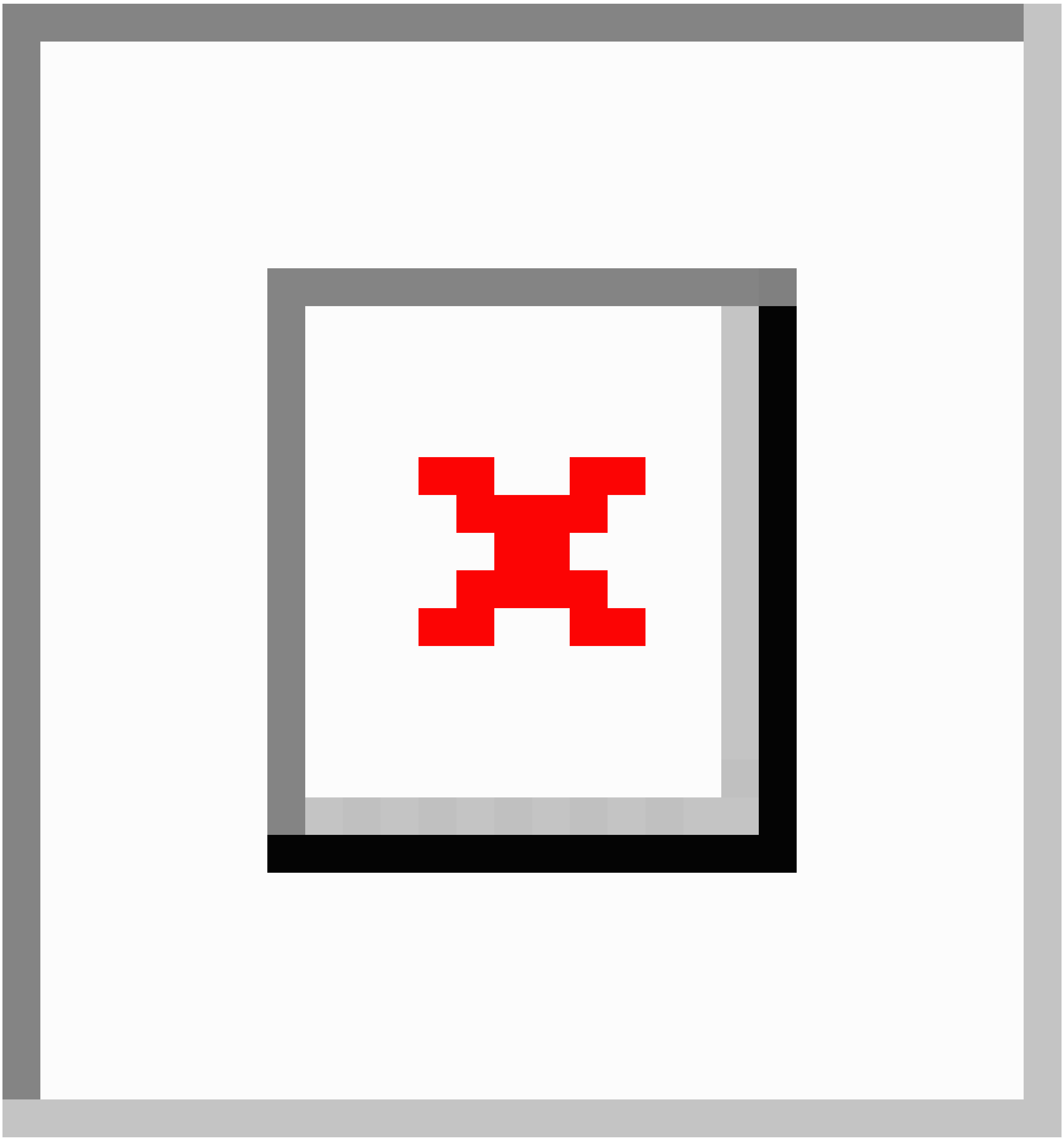
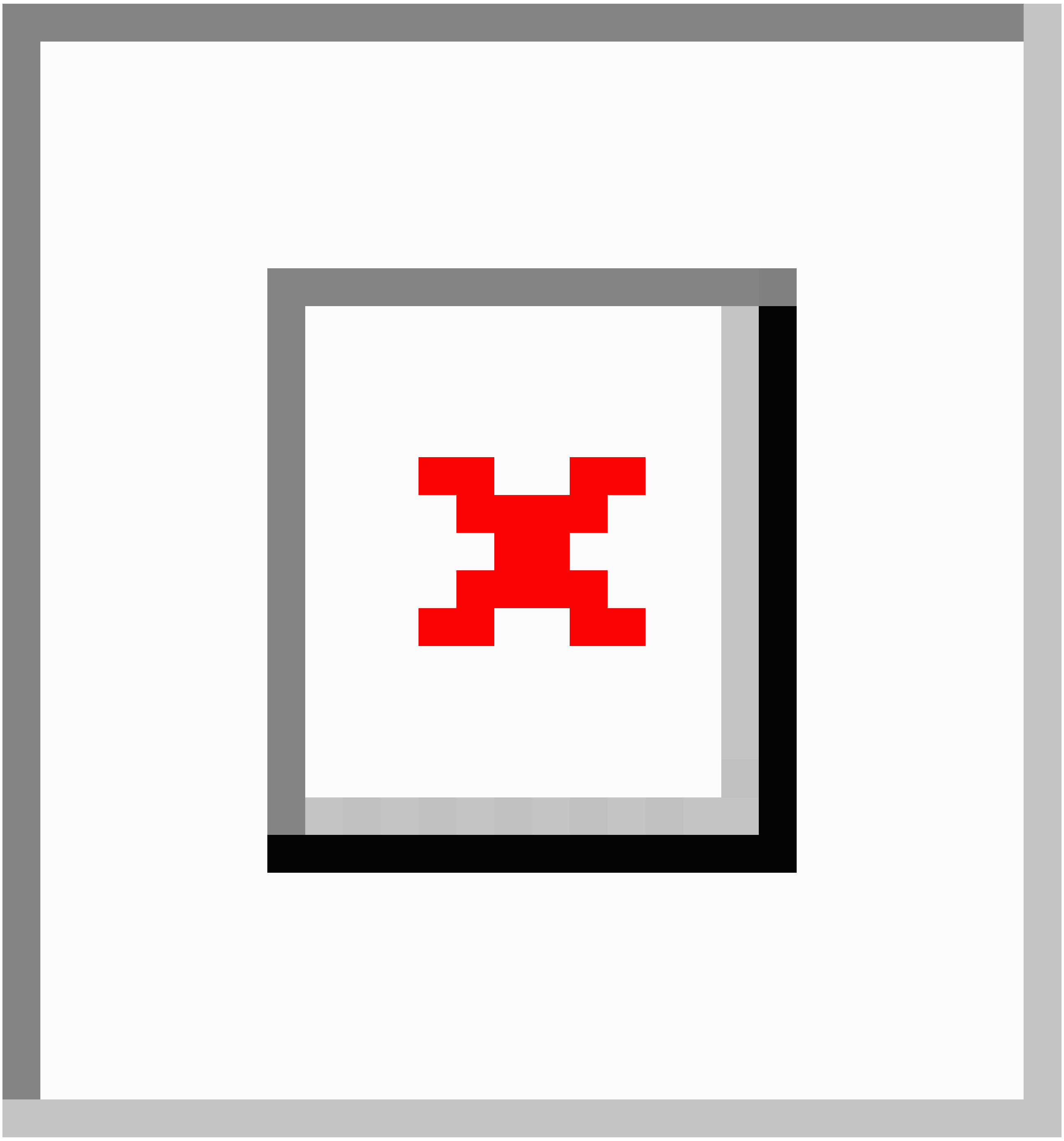


Figure 5. Funnel plot of the difference in diastolic blood pressure between the mobile health intervention group and control group. diff.: difference.



Hemoglobin A_{1c}

Additionally, as shown in Figure 6, one study had more than one intervention group with the same outcome measured [25]; hence, 3 interventions were shown in the forest plot of HbA_{1c}, and the mHealth intervention had no significant lowering effects on HbA_{1c} levels among patients with diabetes in the pooled meta-analysis at 0.20 mmol/mol (95% CI −0.40 to 0.80; *P*=.51; I^2=0%). No significant publication bias was detected visually by the funnel plot (Figure 7) or statistically by Begg (*P*=.96) and Egger (*P*=.10) tests.

Figure 6. Forest plot of the difference in hemoglobin A_{1c} between the mHealth intervention group and control group in 2 studies [25,27]. Abaza et al [25]: (A) baseline measurement; (B) end point measurement. mHealth: mobile health.

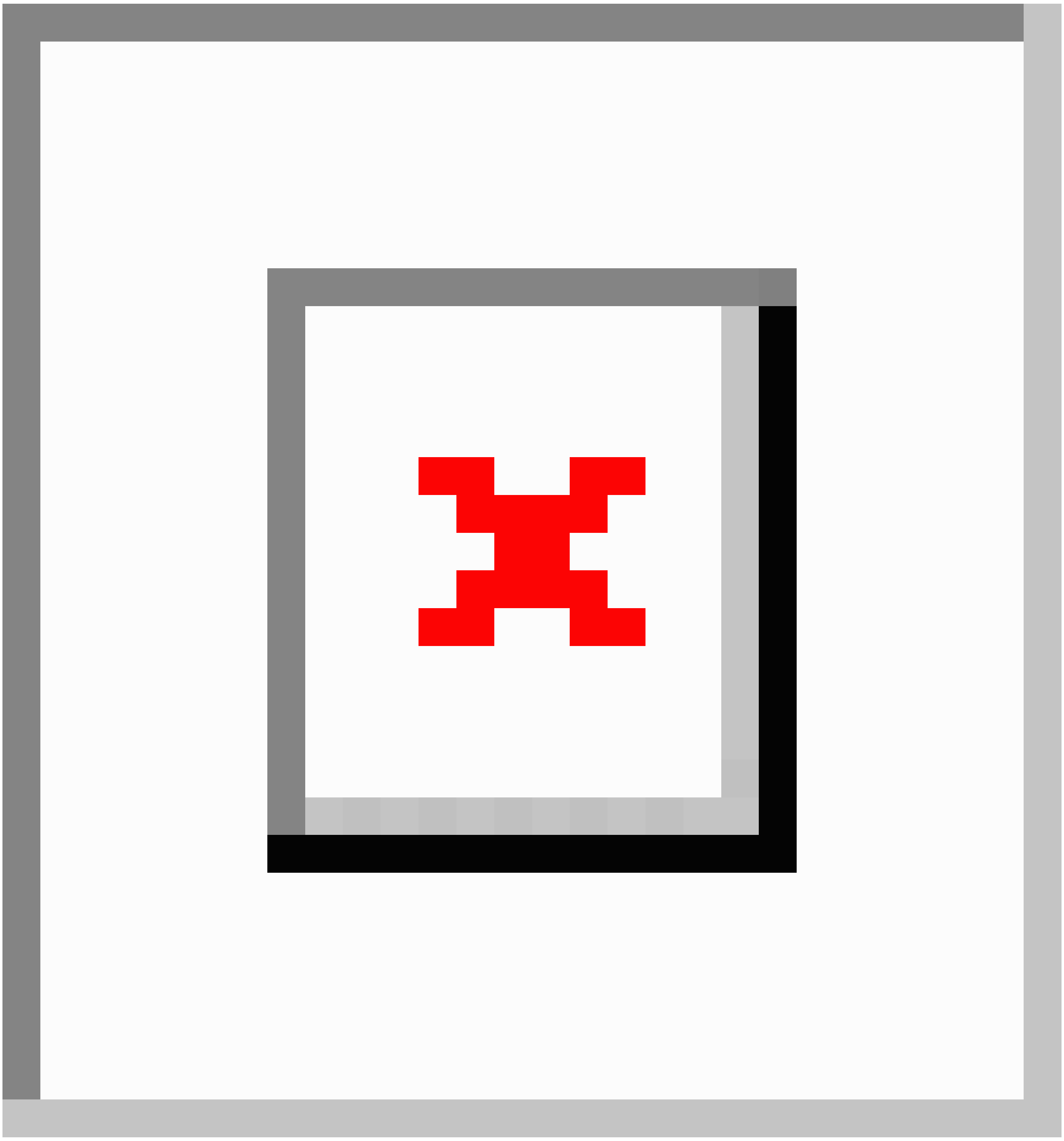
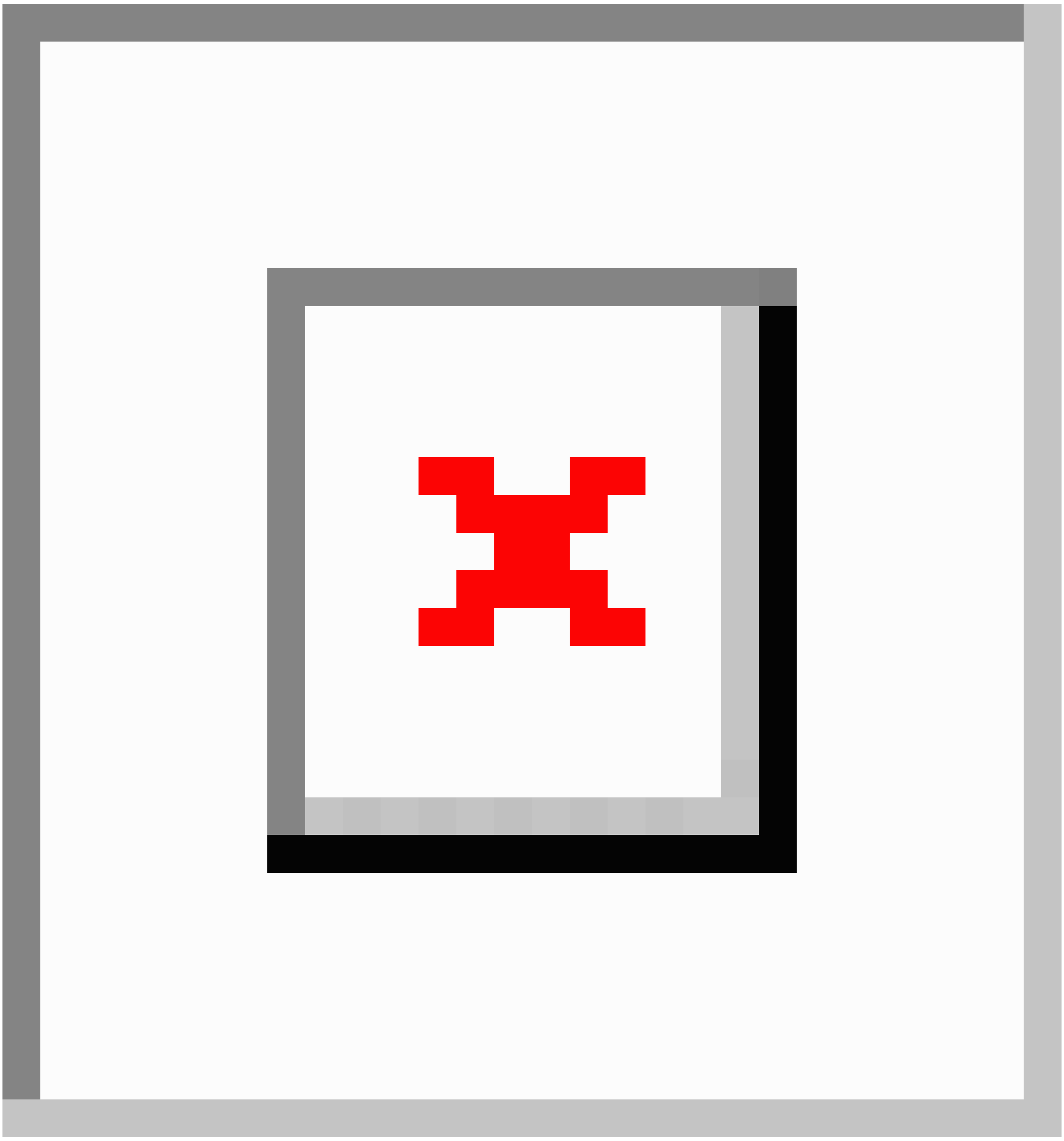


Figure 7. Forest plot of the difference in hemoglobin A_{1c} between the mobile health intervention group and control group. diff.: difference.



Sensitivity Analyses

Sensitivity analyses were conducted using the leave-one-out method. For systolic blood pressure, the Kingue et al [29] study had an impact on the WMD, with the pooled WMD being statistically significant after the exclusion of the Kingue et al [29] study (−2.22, 95% CI −3.94 to −0.60; *P*=.01; Figure 8). For diastolic blood pressure and HbA_{1c}, the exclusion of each of the studies rendered the WMD nonsignificant (Figures 9 and 10).

Figure 8. Leave-one-out forest plot for systolic blood pressure in 6 studies [26-31]. Bobrow et al [28]: (A) interactive intervention group vs control; (B) information only intervention group vs control.

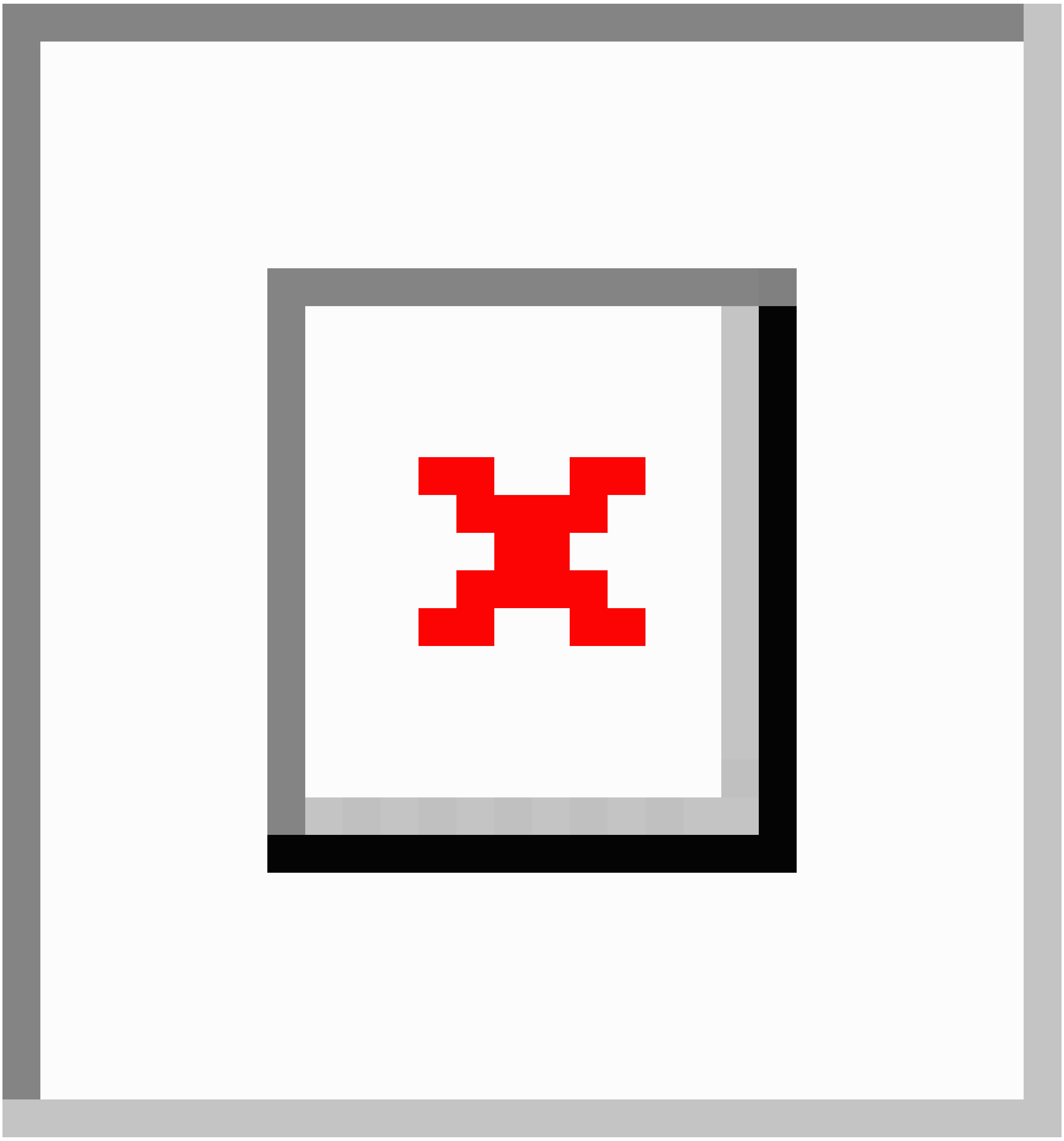


Figure 9. Leave-one-out forest plot for diastolic blood pressure in 5 studies [26,27,29-31].

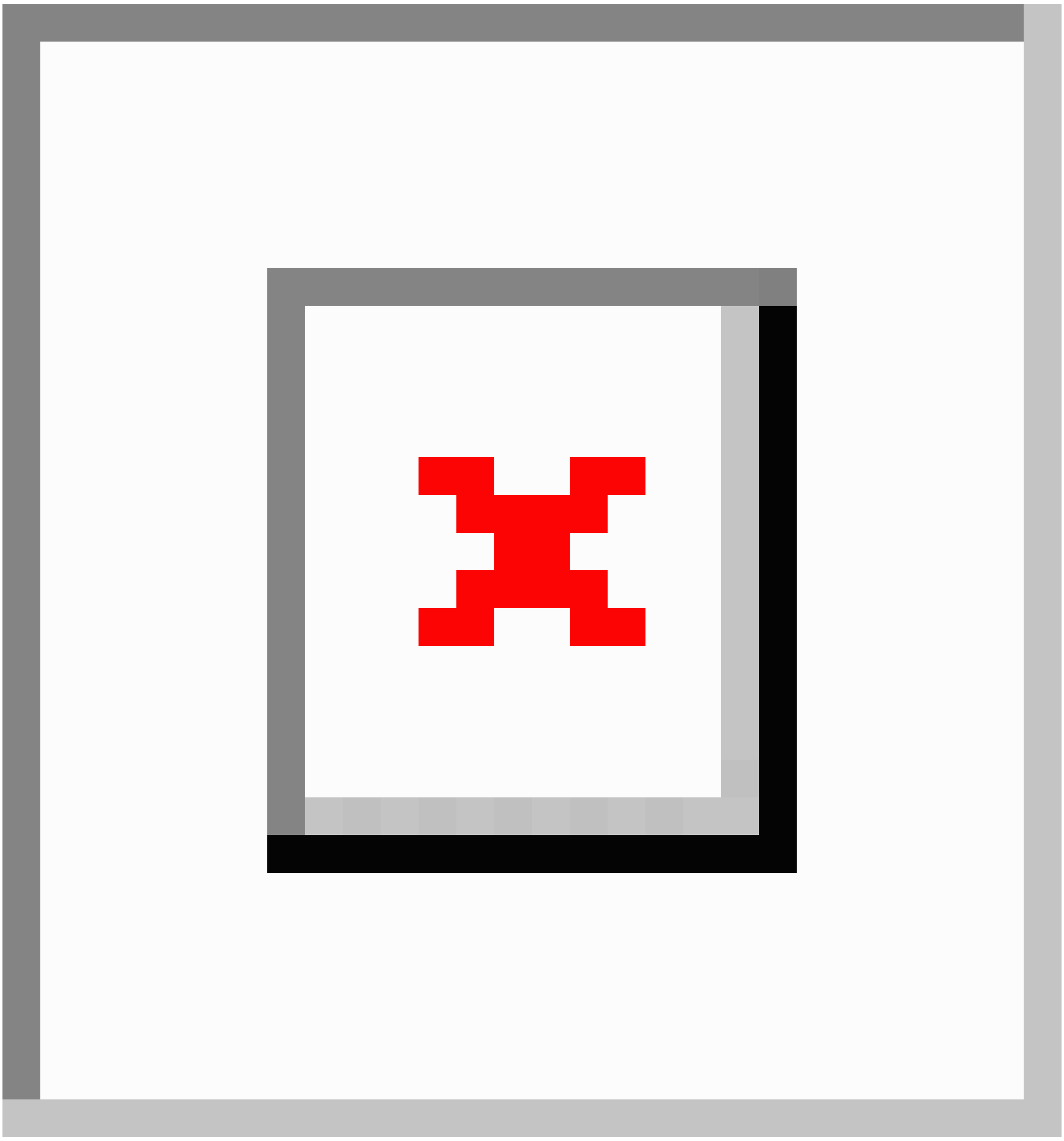
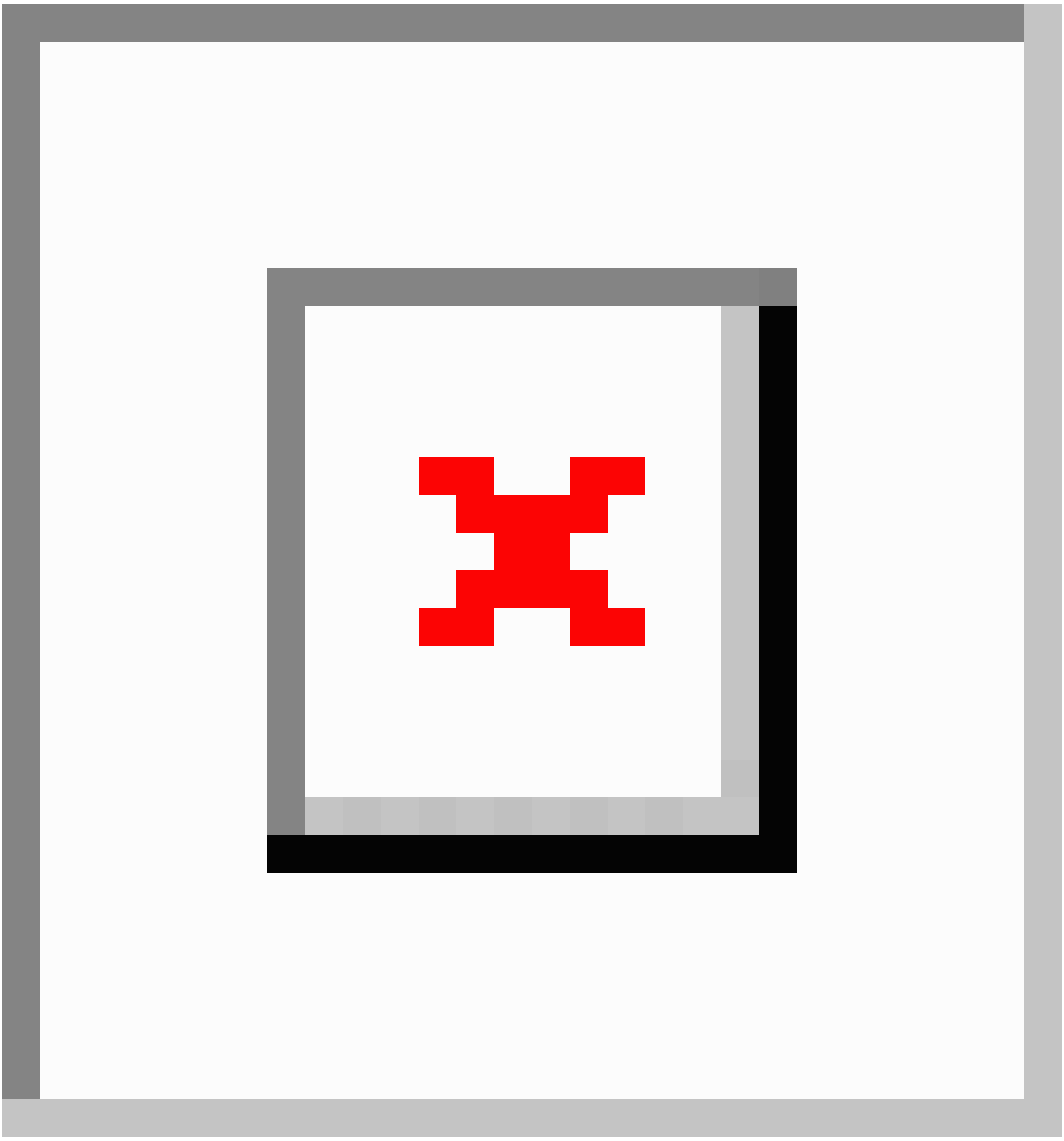


Figure 10. Leave-one-out forest plot for hemoglobin A_{1c} in 2 studies [25,27]. Abaza et al [25]: (A) baseline measurement; (B) end point measurement.



Discussion

Discussion of the Key Findings

This systematic review and meta-analysis identified 7 RCT studies that assessed the effectiveness of mHealth interventions on blood pressure and glycemic control among patients with hypertension and diabetes in Africa. In this review, the effectiveness of mHealth interventions on blood pressure and glycemic control among patients with hypertension and diabetes in Africa did not show conclusive evidence.

Systolic Blood Pressure Control

For systolic blood pressure, we observed a reduction after the mHealth intervention compared to usual care by an average of 1.39 mm Hg; however, it was not statistically significant. After conducting a leave-one-out analysis, a study by Kingue et al [29] had an impact on the WMD, and the exclusion of this study resulted in a pooled WMD reduction of 2.22 mm Hg in systolic blood pressure. This finding is consistent with previous RCT studies [20,34-36] that examined mHealth interventions on systolic blood pressure control and showed that mHealth interventions reduced systolic blood pressure by 10.4 mm Hg [20], 5.5 mm Hg [34], 3 mm Hg [35], and 3.9 mm Hg [36], respectively. In contrast, a study performed by Rubinstein et al [13] reported that the mHealth intervention did not reduce systolic blood pressure compared with usual care. This discrepancy could be explained by the relatively small sample number of studies included in this review. Another reason could be due to the different study populations, interventions, ages, and medications used.

Diastolic Blood Pressure Control

For diastolic blood pressure, our study observed no lowering effect of the mHealth intervention, which reduced by an average of 0.36 mm Hg, which is inconsistent with studies performed by Lu et al [36] and Zhang et al [20] who reported a reduction of 2.2 and 4.8 mm Hg, respectively, after the mHealth intervention compared to usual care. The disparity is that the previous studies [20,36] were conducted among patients with stroke and heart failure with a more complicated pathogenesis of hypertension, which might have resulted in the observed significant decrease in diastolic blood pressure control in this study. In patients with stroke, lower blood pressure might be achieved with strict treatment targets that also lead to a controlled condition. In previous studies, they noted a significant net reduction in body weight and intake of high-fat and high-sugar foods after the intervention [20,36]. Despite no significant findings on diastolic blood pressure control after mHealth interventions, the study by Rubinstein et al [13] reported that each 1 mm Hg decrease in diastolic blood pressure is associated with a 7% decrease in mortality from stroke and ischemic heart disease. Thus, the mHealth intervention may still be a measure worth considering for reducing blood pressure.

Glycemic Control

For glycemic control, the meta-analysis results showed no improvement after the mHealth interventions. Our study contradicts previous studies by Mao et al [35], Moattari et al [37], Kitsiou et al [4], and Huang et al [38] who found significant improvements in glycemic control following mHealth interventions among patients with diabetes. These studies have reported that patients and health care professionals who communicated by SMS text messages, telephone calls, and even electronic reminders or web servers reported greater improvement in HbA_{1c} outcomes compared with usual care [4,35,37,38]. Thus, patients with poorly controlled diabetes might benefit more from using mHealth, therefore more clinical trials are needed to confirm these findings. The Adjei and Marschollek [25] study in Ghana, despite not reporting HbA_{1c}, saw a substantial reduction in fasting plasma glucose of −1.6 mmol/L. Although evidence is scarce about the effect of mHealth interventions on the management of patients with diabetes, the difference could be attributed to the better care the system generates from the health providers. Another reason is that long-term interventions likely result in more significant changes in glycemic control than short-term mHealth interventions.

Strengths and Limitations

To the best of our knowledge, this review was the first that assessed the effectiveness of mHealth interventions in diabetes and hypertension management in Africa. Quality appraisal suggests that the quality of the included studies was good. Additionally, the included studies show no publication bias. However, there are limitations to acknowledge. Despite a thorough search, the number of included studies was relatively small, signifying that using mHealth interventions in Africa on patients with hypertension and diabetes remains an emerging area. This review may not be able to capture some significant effects due to the small samples in the included studies. Given the above limitations, future studies with larger samples are needed to validate our findings.

Conclusion

Our study showed no conclusive evidence on the effect of mHealth interventions on systolic blood pressure, diastolic blood pressure, or glycemic control. However, the sample sizes of the included studies were small; therefore, there is a need for larger RCT studies to confirm these findings.

Acknowledgments

Funding/sponsorship for the research was from the Africa eHealth Foundation.

Authors' Contributions

PA and CA designed the study. PA and KA carried out the literature search and data extraction. PA conducted the analysis and drafted the manuscript. CA and ELvdL provided critical inputs into the manuscript revision, and AL, NM, ER, and EPMvC read and approved the final manuscript. All authors had access to the data presented in this paper and accept the responsibility to submit it for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 1040 KB - mhealth_v11i1e43742_app1.pdf]

Multimedia Appendix 2
Supplementary keywords used in the search.
[PNG File, 50 KB - mhealth_v11i1e43742_app2.png]

References

1. Yasmin F, Banu B, Zakir SM, Sauerborn R, Ali L, Souares A. Positive influence of short message service and voice call interventions on adherence and health outcomes in case of chronic disease care: a systematic review. BMC Med Inform Decis Mak 2016 Apr 22;16:1-14. [doi: 10.1186/s12911-016-0286-3] [Medline: 27106263]
2. Tuglo LS. Prevalence and determinants of lower extremity amputations among type I and type II diabetic patients: a multicenter-based study. Int Wound J 2023 Apr;20(4):903-909. [doi: 10.1111/iwj.13935] [Medline: 36054437]
3. Lee JK, McCutcheon LRM, Fazel MT, Cooley JH, Slack MK. Assessment of interprofessional collaborative practices and outcomes in adults with diabetes and hypertension in primary care: a systematic review and meta-analysis. JAMA Netw Open 2021 1;4(2):e2036725. [doi: 10.1001/jamanetworkopen.2020.36725] [Medline: 33576817]
4. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017 Mar 1;12(3):e0173160. [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
5. Tuglo LS. Comparison of adiposity anthropometric indices and their associations with visceral fat levels determined by bioelectrical impedance analysis among diabetic patients. Sci Rep 2022 Oct 24;12(1):17824. [doi: 10.1038/s41598-022-22848-z] [Medline: 36280694]
6. Hardy ST, Loehr LR, Butler KR, Chakladar S, Chang PP, Folsom AR, et al. Reducing the blood pressure-related burden of cardiovascular disease: impact of achievable improvements in blood pressure prevention and control. J Am Heart Assoc 2015 Oct 27;4(10):e002276. [doi: 10.1161/JAHA.115.002276] [Medline: 26508742]
7. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
8. Bosu WK, Bosu DK. Prevalence, awareness and control of hypertension in Ghana: a systematic review and meta-analysis. PLoS One 2021 Mar 5;16(3):e0248137. [doi: 10.1371/journal.pone.0248137] [Medline: 33667277]
9. Boima V, Ademola AD, Odusola AO, Agyekum F, Nwafor CE, Cole H, et al. Factors associated with medication nonadherence among hypertensives in Ghana and Nigeria. Int J Hypertens 2015;2015:205716. [doi: 10.1155/2015/205716] [Medline: 26509081]
10. Cavero-Redondo I, Saz-Lara A, Sequí-Dominguez I, Gómez-Guijarro MD, Ruiz-Grao MC, Martinez-Vizcaino V, et al. Comparative effect of eHealth interventions on hypertension management-related outcomes: a network meta-analysis. Int J Nurs Stud 2021 Dec;124:104085. [doi: 10.1016/j.ijnurstu.2021.104085] [Medline: 34601205]
11. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
12. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014 Oct 27;7:25606. [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
13. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
14. Lee S, Cho YM, Kim SY. Mapping mHealth (mobile health) and mobile penetrations in sub-Saharan Africa for strategic regional collaboration in mHealth scale-up: an application of exploratory spatial data analysis. Global Health 2017 Aug 22;13(1):63. [doi: 10.1186/s12992-017-0286-9] [Medline: 28830540]
15. Zhai P, Hayat K, Ji W, Li Q, Shi L, Atif N, et al. Efficacy of text messaging and personal consultation by pharmacy students among adults with hypertension: randomized controlled trial. J Med Internet Res 2020 May 20;22(5):e16019. [doi: 10.2196/16019] [Medline: 32432556]
16. Rehman H, Kamal AK, Morris PB, Sayani S, Merchant AT, Virani SS. Mobile health (mHealth) technology for the management of hypertension and hyperlipidemia: slow start but loads of potential. Curr Atheroscler Rep 2017 Mar;19(3):12. [doi: 10.1007/s11883-017-0649-y] [Medline: 28210974]
17. Betjeman TJ, Soghoian SE, Foran MP. mHealth in sub-Saharan Africa. Int J Telemed Appl 2013;2013:482324. [doi: 10.1155/2013/482324] [Medline: 24369460]
18. Adler AJ, Laar A, Prieto-Merino D, Der RMM, Mangortey D, Dirks R, et al. Can a nurse-led community-based model of hypertension care improve hypertension control in Ghana? Results from the ComHIP cohort study. BMJ Open 2019 Apr 2;9(4):e026799. [doi: 10.1136/bmjopen-2018-026799] [Medline: 30944139]
19. Khoong EC, Olazo K, Rivadeneira NA, Thatipelli S, Barr-Walker J, Fontil V, et al. Mobile health strategies for blood pressure self-management in urban populations with digital barriers: systematic review and meta-analyses. NPJ Digit Med 2021 Jul 22;4(1):1-12. [doi: 10.1038/s41746-021-00486-5] [Medline: 34294852]
20. Zhang W, Cheng B, Zhu W, Huang X, Shen C. Effect of telemedicine on quality of care in patients with coexisting hypertension and diabetes: a systematic review and meta-analysis. Telemed J E Health 2021 Jun;27(6):603-614. [doi: 10.1089/tmj.2020.0122] [Medline: 32976084]

21. GSMA. The mobile economy 2022. 2022. URL: www.gsma.com/mobileeconomy/wp-content/uploads/2022/02/280222-The-Mobile-Economy-2022.pdf [accessed 2023-08-11]
22. Osei E, Agyei K, Tlou B, Mashamba-Thompson TP. Availability and use of mobile health technology for disease diagnosis and treatment support by health workers in the Ashanti region of Ghana: a cross-sectional survey. Diagnostics (Basel) 2021 Jul 9;11(7):1233. [doi: 10.3390/diagnostics11071233] [Medline: 34359316]
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
24. Freeman K, Connock M, Cummins E, Gurung T, Taylor-Phillips S, Court R, et al. Fluorouracil plasma monitoring: systematic review and economic evaluation of the My5-FU assay for guiding dose adjustment in patients receiving fluorouracil chemotherapy by continuous infusion. Health Technol Assess 2015 Nov;19(91):1-321. [doi: 10.3310/hta19910] [Medline: 26542268]
25. Abaza H, Marschollek M. SMS education for the promotion of diabetes self-management in low & middle income countries: a pilot randomized controlled trial in Egypt. BMC Public Health 2017 Dec 19;17(1):962. [doi: 10.1186/s12889-017-4973-5] [Medline: 29258499]
26. Adjei DN, Agyemang C, Dasah JB, Kuranchie P, Amoah AGB. The effect of electronic reminders on risk management among diabetic patients in low resourced settings. J Diabetes Complications 2015 Aug;29(6):818-821. [doi: 10.1016/j.jdiacomp.2015.05.008] [Medline: 26025699]
27. Asante E, Bam V, Diji AA, Lomotey AY, Owusu Boateng A, Sarfo-Kantanka O, et al. Pilot mobile phone intervention in promoting type 2 diabetes management in an urban area in Ghana: a randomized controlled trial. Diabetes Educ 2020 Oct;46(5):455-464. [doi: 10.1177/0145721720954070] [Medline: 32998649]
28. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu LM, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 9;133(6):592-600. [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
29. Kingue S, Angandji P, Menanga AP, Ashuntantang G, Sobngwi E, Dossou-Yovo RA, et al. Efficiency of an intervention package for arterial hypertension comprising telemanagement in a cameroonian rural setting: the TELEMED-CAM study. Pan Afr Med J 2013 Aug 29;15:153. [doi: 10.11604/pamj.2013.15.153.2655] [Medline: 24396559]
30. Owolabi EO, Goon DT, Ajayi AI. Efficacy, acceptability and feasibility of daily text-messaging in promoting glycaemic control and other clinical outcomes in a low-resource setting of South Africa: a randomised controlled trial. PLoS One 2019 Nov 27;14(11):e0224791. [doi: 10.1371/journal.pone.0224791] [Medline: 31774842]
31. Sarfo FS, Treiber F, Gebregziabher M, Adamu S, Nichols M, Singh A, et al. Phone-based intervention for blood pressure control among Ghanaian stroke survivors: a pilot randomized controlled trial. Int J Stroke 2019 Aug;14(6):630-638. [doi: 10.1177/1747493018816423] [Medline: 30465630]
32. Cochrane Training. RevMan. 2020. URL: training.cochrane.org/online-learning/core-software/revman [accessed 2023-08-11]
33. ScienceDirect. Abnormal laboratory result. URL: www.sciencedirect.com/topics/medicine-and-dentistry/abnormal-laboratory-result [accessed 2023-08-11]
34. Lv M, Wu T, Jiang S, Chen W, Zhang J. Effects of telemedicine and mHealth on systolic blood pressure management in stroke patients: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e24116. [doi: 10.2196/24116] [Medline: 34114961]
35. Mao Y, Lin W, Wen J, Chen G. Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis. BMJ Open Diabetes Res Care 2020 Sep;8(1):e001225. [doi: 10.1136/bmjdrc-2020-001225] [Medline: 32988849]
36. Lu X, Yang H, Xia X, Lu X, Lin J, Liu F, et al. Interactive mobile health intervention and blood pressure management in adults: a meta-analysis of randomized controlled trials. Hypertension 2019 Sep;74(3):697-704. [doi: 10.1161/HYPERTENSIONAHA.119.13273]
37. Moattari M, Hashemi M, Dabbaghmanesh MH. The impact of electronic education on metabolic control indicators in patients with diabetes who need insulin: a randomised clinical control trial. J Clin Nurs 2013 Jan;22(1-2):32-38. [doi: 10.1111/j.1365-2702.2012.04200.x] [Medline: 22905971]
38. Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J. A smartphone app to improve medication adherence in patients with type 2 diabetes in Asia: feasibility randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 12;7(9):e14914. [doi: 10.2196/14914] [Medline: 31516127]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
WMD: weighted mean difference

Edited by L Buis; submitted 23.10.22; peer-reviewed by E Ammenwerth, I Shubina; revised version received 28.03.23; accepted 09.06.23; published 29.08.23.

Please cite as:
Aovare P, Abdulai K, Laar A, van der Linden EL, Moens N, Richard E, Moll van Charante EP, Agyemang C
Assessing the Effectiveness of mHealth Interventions for Diabetes and Hypertension Management in Africa: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2023;11:e43742
URL: https://mhealth.jmir.org/2023/1/e43742
doi:10.2196/43742

Review

Behavior Change Effectiveness Using Nutrition Apps in People With Chronic Diseases: Scoping Review

Emily Salas-Groves[1], MSc, RDN, LD; Shannon Galyean[1], PhD, RDN, LD; Michelle Alcorn[2], PhD; Allison Childress[1], PhD, RDN, LD

[1]Department of Nutritional Sciences, Texas Tech University, Lubbock, TX, United States

[2]Department of Hospitality & Retail Management, Texas Tech University, Lubbock, TX, United States

Corresponding Author:
Shannon Galyean, PhD, RDN, LD
Department of Nutritional Sciences
Texas Tech University
1301 Akron Avenue
Lubbock, TX, 79409
United States
Phone: 1 8068342286
Email: shannon.galyean@ttu.edu

Abstract

Background: Cardiovascular disease, cancer, diabetes mellitus, and obesity are common chronic diseases, and their prevalence is reaching an epidemic level worldwide. As the impact of chronic diseases continues to increase, finding strategies to improve care, access to care, and patient empowerment becomes increasingly essential. Health care providers use mobile health (mHealth) to access clinical information, collaborate with care teams, communicate over long distances with patients, and facilitate real-time monitoring and interventions. However, these apps focus on improving general health care concerns, with limited apps focusing on specific chronic diseases and the nutrition involved in the disease state. Hence, available evidence on the effectiveness of mHealth apps toward behavior change to improve chronic disease outcomes is limited.

Objective: The objective of this scoping review was to provide an overview of behavior change effectiveness using mHealth nutrition interventions in people with chronic diseases (ie, cardiovascular disease, diabetes mellitus, cancer, and obesity). We further evaluated the behavior change techniques and theories or models used for behavior change, if any.

Methods: A scoping review was conducted through a systematic literature search in the MEDLINE, EBSCO, PubMed, ScienceDirect, and Scopus databases. Studies were excluded from the review if they did not involve an app or nutrition intervention, were written in a language other than English, were duplicates from other database searches, or were literature reviews. Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, the systematic review process included 4 steps: identification of records through the database search, screening of duplicate and excluded records, eligibility assessment of full-text records, and final analysis of included records.

Results: In total, 46 studies comprising 256,430 patients were included. There was diversity in the chronic disease state, study design, number of participants, in-app features, behavior change techniques, and behavior models used in the studies. In addition, our review found that less than half (19/46, 41%) of the studies based their nutrition apps on a behavioral theory or its constructs. Of the 46 studies, 11 (24%) measured maintenance of health behavior change, of which 7 (64%) sustained behavior change for approximately 6 to 12 months and 4 (36%) showed a decline in behavior change or discontinued app use.

Conclusions: The results suggest that mHealth apps involving nutrition can significantly improve health outcomes in people with chronic diseases. Tailoring nutrition apps to specific populations is recommended for effective behavior change and improvement of health outcomes. In addition, some studies (7/46, 15%) showed sustained health behavior change, and some (4/46, 9%) showed a decline in the use of nutrition apps. These results indicate a need for further investigation on the sustainability of the health behavior change effectiveness of disease-specific nutrition apps.

(JMIR Mhealth Uhealth 2023;11:e41235) doi:10.2196/41235

KEYWORDS
mobile apps; apps; mobile health; mHealth; eHealth; nutrition education; cancer; obesity; diabetes; cardiovascular disease; mobile phone

Introduction

Background

Cardiovascular disease (CVD), cancer, diabetes mellitus (DM), and obesity are common chronic diseases [1], and their prevalence is reaching a substantial epidemic level internationally [2]. Chronic diseases are defined by the Centers for Disease Control and Prevention broadly as “conditions that last one year or more and require ongoing medical attention or limit activities of daily living or both” [3]. Chronic diseases affect hospitalization, mortality rates, and people’s overall health and quality of life (QOL) [1]. For example, CVD remains the most prevalent cause of morbidity and mortality in high-income countries despite significant advances in treatment over the last 5 decades. Recent epidemiological data show that CVD mortality is no longer declining and is indeed rising in some communities [4], and hospitalization rates are universally increasing [5]. Furthermore, chronic conditions such as cancer, CVD, DM, and chronic respiratory diseases caused approximately 33.2 million deaths worldwide in 2019 [6]. The prevalence of obesity has increased in all World Health Organization regions since 2000, which affects other chronic conditions as it is a risk factor for the development of CVD, DM, and several cancers [6].

As the impact of chronic diseases continues to increase, finding strategies to improve care, access to care, and patient empowerment becomes increasingly essential. Therefore, mobile health (mHealth) technology is rapidly gaining popularity among health care providers and consumers [7]. mHealth technology is defined as mobile devices (ie, mobile phones or monitoring devices) intended to be worn, carried, or accessed by patients or health care providers to monitor health status or improve health outcomes [8]. Among mobile devices, the most used are smartphones, with more than three-quarters of Americans having one, and at least one-third of smartphone owners use a health app [7,9]. Health care providers use mHealth to access clinical information, collaborate with care teams, communicate over long distances with patients, and facilitate real-time monitoring and interventions. These apps provide an opportunity to increase health care access for vulnerable populations [10]. Patients use mHealth to track their health data, access their clinical records, and communicate with their providers [11].

Furthermore, a meta-analysis [12] reported promising results for mHealth interventions in the improvement of patient outcomes such as body measurements (ie, weight and waist circumference), metabolic and physiological measurements (ie, blood pressure and glucose levels), adherence to and safe use of medication, physical activity performance, meal management, and awareness of health conditions and treatment options. Mobile apps that provide tools intended to facilitate nutrition care via smartphone technologies provide patients with more autonomy, thus empowering them and offsetting patient disengagement [13]. Moreover, mobile app–based interventions effectively improve diet and diet-related health outcomes, and effect sizes are comparable with those of traditional nondigital interventions [14]. For example, mHealth interventions can help improve lifestyle behaviors related to CVD [15,16]. A recent meta-analysis [17] found that using mHealth interventions for CVD was associated with improved blood pressure and hospitalization rates. In addition, mHealth apps have emerged as supportive tools in managing cancer as they reduce the financial burden, provide access to information, and facilitate communication [18]. Different studies and meta-analyses of patients with cancer have shown the benefits of mHealth, which include reducing fatigue or pain [19,20], providing distance physical activity programs [21,22], the use of social networks to improve QOL [23], and monitoring of symptoms [24]. mHealth offers improved and cost-effective care to people with type 2 DM (T2DM) through improved diabetes self-management [25,26]. These apps seem to increase the perception of self-care by contributing to better knowledge among people with T2DM [10]. Individuals with DM also become more confident in dealing with their illness, primarily because of decreased fear resulting from a lack of information [27,28]. For obesity, a systematic review [29] has found that technology-based interventions can provide a moderately effective way of facilitating lifestyle changes and weight loss. New technologies could help reduce economic costs, improve accessibility and adherence to treatment, and increase patient motivation and weight control [30].

The use of mobile apps for mHealth and of general health apps is increasing rapidly. However, these apps focus on improving general health care concerns, with limited apps focusing on specific chronic diseases and their nutritional intervention. The available evidence on the effectiveness of mHealth apps toward behavior change to improve chronic disease outcomes is also limited. Systematic reviews to date regarding nutrition apps have focused on healthy participants or examined the effects of dietary apps on diet improvement but not on chronic diseases [31]. Most reviews were inconclusive, with the authors recommending further research in this area to demonstrate possible benefits [32-34]. From a clinical point of view, it is essential to know if an app designed for chronic diseases produces behavior changes to improve an individual’s chronic disease outcomes. From an app creator’s point of view, it is crucial to see what needs to be done when developing an app directed toward people with chronic diseases to enhance the app’s effectiveness on behavior change and produce positive outcomes. The goal of this review was to define behavioral theories associated with mHealth use, evaluate behavior change techniques (BCTs), and describe the behavior change effectiveness using mHealth nutrition interventions in people with chronic diseases (ie, CVD, DM, cancer, and obesity).

Models and Theories of Health Behavior and Clinical Interventions

Health care and self-management of chronic conditions require effort and commitment on the part of the patient to follow treatment plans. These treatments involve many behaviors, such as dietary intake, physical activity, and prescription drug use. [35]. Theories and models are used in treatment planning to understand and explain the health behavior of individuals and can help guide clinicians to develop interventions that increase the effectiveness of health behavior change. The role of behavioral theories and models in informing digital health interventions is important to support sustainable health behavior change. A brief explanation of some of the theories and models of behavior used in these studies may help with understanding their constructs [36]: (1) the transtheoretical model shows how individuals move through 6 stages of behavior change (ie, precontemplation, contemplation, preparation, action, maintenance, and termination), which can be used to support behavior change and self-efficacy for the cessation of unhealthy behaviors and encourage a healthy lifestyle; (2) Social Cognitive Theory emphasizes self-efficacy and focuses on behavior change and outcome expectations by mastering steps to behavior change, observing others who are successful, improving mood, and ultimately increasing self-efficacy for health behavior change; (3) the Health Belief Model is based on the concept of expectancy-value and constitutes an individual's belief that their health condition is serious, their actions will reduce their risk or illness, there is a benefit to taking action for their health condition, and they have the ability to take action for health behavior change (self-efficacy); and (4) the theory of planned behavior emphasizes that motivation that is directly influenced by ability (ie, an individual's self-efficacy or perceived control over outside factors) is key to making a health behavior change [36]. There are other theories that can be used as well. For health behavior change, there is a general acceptance of the Health Belief Model along with a focus on self-efficacy found in many behavioral models such as the Social Cognitive Theory [35]. In addition, a commonly used theory to guide lifestyle interventions is the transtheoretical model [36]. There are several theories identified in this review that are important to chronic disease research on health behavior.

Methods

Information Sources and Search Strategies

In April 2022, a systematic literature search was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) strategy for each of the following diseases: obesity, DM, CVD, and cancer [37]. The PRISMA method allows for the transparent selection and analysis of literature for inclusion. The databases used in the search were MEDLINE, EBSCO, PubMed, ScienceDirect, and Scopus. These databases were chosen based on their extensive health- and technology-related literature [38]. The search strategy was similar for all diseases. It included the following keywords searched among titles, keywords, and abstracts: (1) *mobile applications* or *apps* or *mobile apps* or *mHealth* or *eHealth* AND *nutritioneducation* AND *obesity*, (2) *mobile applications* or *apps* or *mobile apps* or *mHealth* or *eHealth* AND *nutritioneducation* AND *type 2 diabetes*, (3) *mobile applications* or *apps* or *mobile apps* or *mHealth* or *eHealth* AND *nutritioneducation* AND *cardiovascular disease*, and (4) *mobile applications* or *apps* or *mobile apps* or *mHealth* or *eHealth* AND *nutritioneducation* AND *cancer.*

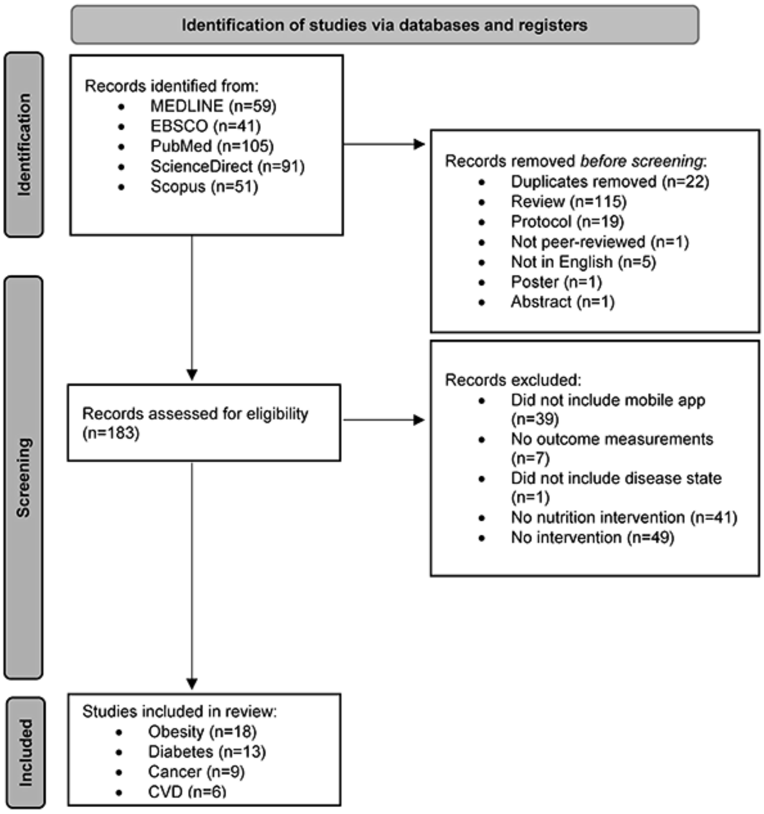
Eligibility Criteria

To summarize the evidence available on the topic, we included primary research studies involving an app and a nutrition intervention. Studies were excluded from the review if they did not involve an app or nutrition intervention, were written in a language other than English, were duplicates from other database searches, or were literature reviews.

Procedures

Following the PRISMA 2020 method, the systematic review process included four steps: (1) identification of records through the database search, (2) screening of duplicate and excluded records, (3) eligibility assessment of full-text records, and (4) final analysis of included records. This process can be seen in Figure 1. The results of the database searches were exported to Microsoft Excel (Microsoft Corp) for further analysis. Duplicates were removed. The remaining studies were moved to the eligibility phase, in which the authors assessed the eligibility of the articles based on the inclusion and exclusion criteria of the full text. The reason for exclusion was listed for each excluded article. The citation of the article, name of the mobile app used if available, type of intervention used, year of publication, population and sample details, study design details, purpose of the research, behavioral effectiveness, behavioral techniques, theory used, outcome measures, and results were recorded for all included studies. The authors independently reviewed the articles found in each search, and a second screening was performed by a different author. The process for screening was as follows: SG initially screened articles found in the obesity search, MA initially screened articles found in the DM search, ESG initially screened articles found in the cancer search, and AC initially screened articles found in the CVD search. SG performed a second screening of articles found in the DM search, MA performed a second screening of articles found in the obesity search, AC performed a second screening of articles found in the cancer search, and ESG performed a second screening of articles found in the CVD search.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [37]. CVD: cardiovascular disease.



Results

CVD Results

Demographic, Participant, and Study Design Details

There were 6 studies and mHealth app nutrition interventions comprising 451 patients included for CVD. Of the 6 studies, 4 (67%) randomized controlled trials, 1 (17%) qualitative descriptive design study, and 1 (17%) intervention evaluation were analyzed. The populations analyzed had diverse ethnicities, education levels, ages, and diagnoses. Furthermore, there were different diets or treatments among the studies, with variations in targeted behavioral domains and outcomes measured in the mHealth app nutrition interventions. App features for these studies included food tracking [8,39-41], physical activity tracking [8,41,42], nutrition and exercise knowledge and recommendations [8,40-43], scheduled reminders [39,40,42,43], clinician portals [41-43], connectivity to digital health devices [42], demonstration of the desired behavior [40,43], and challenges [41]. The aim of all studies (6/6, 100%) was to evaluate the effectiveness of mHealth in the CVD population. The included studies on nutrition apps for people with CVD can be found in Table 1.

Table 1. Included studies on cardiovascular disease (CVD; n=6).

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Eyles et al [39]	New Zealand	Randomized controlled trial	Evaluate whether the SaltSwitch app is effective in helping people with CVD select lower-salt food purchases	SaltSwitch—all-smartphone app; features: food tracking and reminders	None	A significant reduction in mean household purchases of salt was observed during the 4-week intervention phase. A total of 87% of the participants reported using the SaltSwitch app, and 75% reported finding the SaltSwitch app very easy to use.
Indraratna et al [42]	New Zealand	Randomized controlled trial	Investigate the feasibility, efficacy, and cost-effectiveness of a smartphone app–based model of care in patients discharged after ACS[a] or HF[b] admission	TeleClinical Care TCC—all-smartphone app; features: physical activity tracking, knowledge, reminders, connectivity to digital health devices, and clinician portal	None	The intervention was associated with a significant reduction in unplanned hospital readmissions, including cardiac readmissions, and higher rates of cardiac rehabilitation completion and medication adherence. The average usability rating for the app was 4.5/5.
Agher et al [8]	France	Survey	Design an mHealth[c] app, Prevent Connect, to assess its quality for evaluating patient behavior for 4 cardiovascular risk factors (unhealthy eating, sedentary lifestyle, and alcohol and tobacco consumption) and suggest personalized recommendations and mHealth interventions for risky behaviors	Prevent Connect—all-smartphone app; features: food and physical activity tracking and knowledge	None	The app was deemed of good quality, with a mean uMARS[d] quality score of 4 on a 5-point Likert scale. The functionality and information content of the app were particularly appreciated, with a mean uMARS score of >4. The esthetics and engagement of the app were appreciated (uMARS score of 3.7). A total of 80% (42/52) of the participants declared that the app helped them become aware of the importance of addressing health behavior, and 65% (34/52) said that the app helped motivate them to change lifestyle habits.
Schmaderer et al [43]	United States	Qualitative analysis	Explore the experience of using a self-management mHealth intervention in individuals with HF to inform a future mHealth intervention study	Play-It Health—all-smartphone app; features: knowledge, reminders, clinician portal, and demonstration of behavior	None	Participants reported an overall positive experience. The education provided during the study increased self-awareness and promoted self-management of their HF. The mHealth app supported patient empowerment, resulting in better HF management and improved quality of life.
Steinberg et al [40]	United States	Randomized controlled trial	Improve adherence to the DASH[e] diet among women with HTN[f] or pre-HTN	Nutritionix—all-smartphone app; features: food tracking, knowledge, reminders, and demonstration of behavior	Behavior change principles	Intervention participants had lower systolic and diastolic BP[g] compared with active comparator participants. Most intervention participants (23/29, 79%) said that they would recommend the DASH Cloud intervention to a friend or family member. However, only 34% (10/59) indicated that the feedback SMS text messages helped them reach their diet goals.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Choi et al [41]	United States	Randomized controlled trial	Discover whether dietary counseling supplied through a custom smartphone app results in better adherence to a Mediterranean diet in a non-Mediterranean population than the traditional standard-of-care counseling	Unknown name—all-smartphone app; features: food and physical activity tracking, knowledge, clinician portal, and challenges	None	There were no significant differences between EXP[h] and SOC[i] regarding BP, lipid parameters, HbA_{1c}, or C-reactive protein. EXP achieved a significantly greater weight loss on average of 3.3 lbs versus 3.1 lbs for SOC. Adherence to the Mediterranean diet increased significantly over time for both groups, but there was no significant difference between the groups. Similarly, there was no significant difference in diet satisfaction between EXP and SOC, although diet satisfaction increased significantly over time for both groups. The proportion of participants with high Mediterranean diet compliance increased significantly over time—from 18.4% to 57.1% for SOC and from 27.5% to 64.7% for EXP; however, there was no significant difference between the groups.

[a]ACS: acute coronary syndrome.

[b]HF: heart failure.

[c]mHealth: mobile health.

[d]uMARS: User Version of the Mobile App Rating Scale.

[e]DASH: Dietary Approaches to Stop Hypertension.

[f]HTN: hypertension.

[g]BP: blood pressure.

[h]EXP: experimental.

[i]SOC: standard of care.

Targeted Behavior and Outcome Measures

The most commonly measured outcome in the CVD studies was usability or engagement in 83% (5/6) of the studies, followed by diet intake or quality and metabolic and physiological measurements (ie, blood pressure and urinary sodium) in 67% (4/6) of the studies. In addition, treatment adherence (ie, diet, cardiac rehabilitation, and medication) was measured in 50% (3/6) of the studies. Other measured outcomes included physical activity, weight loss, dietary knowledge, sustainability, tobacco or alcohol consumption, hospital readmissions, cost-effectiveness, food purchases, and self-management in 17% (1/6) of the studies.

BCTs Used

The most commonly used BCT was motivation or encouragement, which was used in all reviewed studies (6/6, 100%). Knowledge or education and self-monitoring were used in 83% (5/6) of the studies. Both prompts or cues and feedback were used in 50% (3/6) and 33% (2/6) of the studies, respectively. The least common BCTs included graded tasks or challenges, demonstration of behavior, self-efficacy, and goal setting in 17% (1/6) of the studies reviewed.

Behavioral Theory or Model

In total, 17% (1/6) of the analyzed studies were based on behavioral theories or models. The others (5/6, 83%) did not mention a theory. The transtheoretical model, Social Cognitive Theory, the theory of planned behavior, the Health Belief Model, the precaution adoption model, and goal-setting theories were used as the basis for 17% (1/6) of the interventions in this review section. Evidence for informing digital technology interventions reveals that the Health Belief Model has been widely used for goal setting and lifestyle changes for reducing cardiovascular risk as it focuses on confidence in one's ability to take action [36].

Behavior Change Effectiveness

CVD-specific mHealth apps significantly improved the completion of cardiac rehabilitation, were significantly cost-effective, and resulted in substantial weight loss and less salt purchases in 17% (1/6) of the studies. Furthermore, significant engagement was observed in 67% (4/6) of the studies.

In addition, some studies (1/6, 17%) showed improvement in blood pressure and self-management and rated the app quality as good. However, linking BCTs and theoretical frameworks to behavior change and CVD health measures is challenging. Therefore, this area of study should be investigated further.

Cancer Results

Demographic, Participant, and Study Design Details

In total, 9 cancer-related studies comprising 645 patients were included. Of the 9 studies, 3 (33%) focused on survivors of breast cancer, 2 (22%) focused on breast cancer, 2 (22%) focused on esophageal cancer, 1 (11%) focused on patients with either gastric or colon cancer, and 1 (11%) focused on pancreatic cancer. Various study designs were used: 22% (2/9) prospective quasi-experimental studies, 22% (2/9) prospective pilot studies, 11% (1/9) interventional observation studies, 11% (1/9) nonrandomized 2-group controlled design studies, 11% (1/9) randomized controlled trials, 11% (1/9) randomized pretest-posttest design studies, and 11% (1/9) clinical trials were analyzed. These studies can be found in Table 2. Each app had different features, including digital diaries [7,44-49], coach feedback [7,44-47,49-51], personalized physical exercise programs and nutrition plans [49,50], general nutrition and physical exercise guidelines [45,51], psychological support courses [45,51], a secure message portal to their health care teams [7,48], health knowledge education [46,48,51], and daily tasks [48,51].

Table 2. Included studies on cancer (n=9).

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Stubbins et al [7]	United States	Prospective, single-arm, open-label clinical trial	To evaluate the feasibility and usability of MOCHA[a] to improve daily accounting of PA[b] and food intake in survivors of breast cancer and improve engagement with health practitioners and peers	MOCHA—Apple- and Android-based app; features: food and PA tracking, coach feedback, and clinician portal	None	The average number of daily uses was approximately 3.5 times per day; participants lost an average of 2 lbs. The average usability score was 77.4, which was greater than the acceptable level. More than 90% of patients found MOCHA easy to navigate, and 84% were motivated to use MOCHA daily.
Lozano-Lozano et al [49]	Spain	Prospective test-retest quasi-experimental study	To investigate the feasibility of BENECA[c] mHealth[d] in an ecological clinical setting with survivors of breast cancer by studying (1) its feasibility and (2) pretest-posttest differences regarding the lifestyles, QOL[e], and PA motivation of survivors of breast cancer	BENECA mHealth—all-smartphone app; features: food and PA tracking, coach feedback, and personalized programming	Theories of learning, Goal-Setting Theory, and Social Cognitive Theory	BENECA was considered feasible by survivors of breast cancer in terms of use (76%, 58/76), adoption (69%, 80/116), and satisfaction (positive NPS[f]). BENECA mHealth improved the QOL, EAF[g] score, and daily moderate to vigorous PA of the participants and reduced their body weight.
Lozano-Lozano et al [50]	Spain	Prospective quasi-experimental pre-post study	(1) Check whether it is feasible to find changes in inflammation biomarkers through an mHealth strategy app as a delivery mechanism of an intervention to monitor energy balance and (2) discover potential predictors of change in these markers in survivors of breast cancer	BENECA mHealth—all-smartphone app; features: food and PA tracking, coach feedback, and personalized programming	Theory of learning, Goal-Setting Theory, and Social Cognitive Theory	Differences between pre- and postassessment CRP[h] and IL-6[i] showed a significant decrease in both markers. Stepwise regression analyses revealed that changes in global QOL, as well as uMARS[j] score and hormonal therapy, were possible predictors of change in CRP concentration after using the mHealth app. Participants showed moderate satisfaction with the mHealth app; high app use (mean 47.9; maximum 56 days); and moderate to low scores in general QOL, fatigue, and pain.
Cheng et al [45]	China	Prospective, single-arm, nonrandomized pilot study	To evaluate the feasibility, safety, and efficacy of a comprehensive intervention model using an mHealth system (CIMmH[k]) in patients with esophageal cancer after esophagectomy	CIMmH—web-based program; features: food and PA tracking, knowledge, and psychological support courses	Comprehensive intervention model	At the 3-month follow-up, except for pain, eating difficulty, dry mouth, and trouble with talking, all other QOL dimensions returned to the preoperative level. There were significant reductions in weight and BMI throughout the study, and no significant changes were observed for physical fitness measured by change in the 6-minute Walk Test distance between baseline and the 1-month follow-up or between baseline and the 3-month follow-up. Depressive symptoms significantly increased 1 month after surgery, whereas other psychological measures did not show relevant changes. Although there were declines in many measures 1 month after surgery, these were much improved at the 3-month follow-up, and the recovery was more profound and faster than with traditional rehabilitation programs. Participants viewed, on average, 84% (3.38/4) of the web-based video intervention content and completed, on average, 14% (3.20/23) and 34% (9.44/28) of the web-based audio and article content, respectively. Participants completed, on average, 63% (5.01/8), 100% (1/1), and 24% (10.89/46) of the web-based nutrition, physical exercise, and psychological intervention content, respectively.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Soh et al [48]	Korea	Prospective study	To develop and validate a multidisciplinary mobile care system that can provide health education and self-management features to improve multiple clinical measures for patients with advanced gastrointestinal cancer	Life Manager—all-smartphone app; features: food and PA tracking, clinician portal, knowledge, and daily tasks	None	For the gastric cancer group, the "general gastric cancer education" was the most frequently viewed (322 times), and for the colon cancer group, the "warming-up exercise" was the most viewed (340 times). Of 13 measurements taken from participants, 9 were taken offline (response rate: 52%-90.1%), and 3 were taken on the web (response rate: 17.6%-57.4%). The overall satisfaction rate among participants was favorable and ranged from 3.93 to 4.01 on a 5-point Likert scale.
Cairo et al [44]	United States	Nonrandomized 2-group controlled study design with pre-post repeated measures	To evaluate if a readily available mHealth intervention (ie, Vida) can lead to healthier lifestyle habits for survivors of breast cancer	Vida—all-smartphone app; features: food and PA tracking and coach feedback	Behavior change theory	At 6 months, more patients in the app group experienced weight loss and had a significantly greater reduction in overall BMI. The app group also demonstrated statistically significant improvements in "strenuous" PA and had significant improvements in their dietary patterns as compared with the self-guided group. The app group had greater reduction in fatigue and improvement in depression, but these changes were not statistically significant. At 12 months, none of the app users were still using the app, but many were still following their wellness plan and had maintained their weight loss.
Yang et al [47]	Korea	Prospective pilot study	To evaluate the usefulness of a health care mobile app in preventing malnutrition and excessive muscle loss in patients with esophageal cancer receiving NACRT[l]	Noom—Android- and Apple-based app; features: food and PA tracking, coach feedback, and knowledge	None	The use (or activation) of the app was noted in approximately 70% (25/36) of the patients until the end of the trial. Compared with the 1:2-matched usual care group by propensity scores balanced with their age, primary tumor location, tumor stage, pre-RT[m] BMI, and pre-RT SMI[n] level, 30 operable patients showed less aggravation of the prognostic nutritional index. However, there was no significant difference in the SMI change or the number of patients with excessive muscle loss. In patients with excessive muscle loss, the walk steps significantly decreased in the last 4 weeks compared with those in the first 4 weeks. Age affected the absolute number of walk steps, whereas pre-RT sarcopenia was related to the recovery of the reduced walk steps.
Keum et al [46]	Korea	Randomized controlled study	To evaluate the efficacy of a mobile app–based program, Noom, in patients receiving chemotherapy for PDAC[o]	Noom—Android- and Apple-based app; features: food and PA tracking, coach feedback, and knowledge	None	All the study participants showed a significant improvement in nutritional status according to the PG-SGA[p] score regardless of Noom app use. Noom users showed statistically significant improvements on the global health status and QOL scales compared with non–Noom users based on the EORTC QLQ[q]. The SMI decreased in both groups during chemotherapy. The decrement was higher in the non–Noom user group than in the Noom user group, but it was not statistically significant.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Çınar et al [51]	Turkey	Single-blinded, single-centered, and randomized design	Mobile app–based training given to cope with the side effects of EHT[r] and how managing the symptoms and the disease process will affect the QOL of women with breast cancer	Name not provided—all-smartphone app; features: coach feedback, knowledge, psychological support courses, and daily tasks	None	QOL of the treatment group after the intervention increased, and distress level was lower compared with the control group; these results were statistically significant. Most of the patients reported that the mobile app was "informative and useful."

[a]MOCHA: Methodist Hospital Cancer Health Application.

[b]PA: physical activity.

[c]BENECA: the Energy Balance on Cancer mobile health system.

[d]mHealth: mobile health.

[e]QOL: quality of life.

[f]NPS: Net Promoter Score.

[g]EAF: Spanish self-efficacy scale for physical activity.

[h]CRP: C-reactive protein.

[i]IL-6: interleukin-6.

[j]uMARS: User Version of the Mobile App Rating Scale.

[k]CIMmH: Comprehensive Intervention Model Using a Mobile Health System.

[l]NACRT: neoadjuvant chemoradiotherapy.

[m]RT: radiotherapy.

[n]SMI: skeletal muscle index.

[o]PDAC: pancreatic ductal adenocarcinoma.

[p]PG-SGA: Patient-Generated Subjective Global Assessment.

[q]EORTC QLQ: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire.

[r]EHT: adjuvant endocrine hormonal therapy.

Target Behavior and Outcome Measures

The most targeted behaviors were physical activity, observed in 56% (5/9) of the included studies, and dietary behavior, observed in 78% (7/9) of the studies. The questionnaires measured both behaviors. Furthermore, QOL was measured in 67% (6/9) of the studies, followed by app use and satisfaction in 56% (5/9) of the studies. Less frequently measured outcomes included app adherence in 44% (4/9) of the studies, depression in 33% (3/9) of the studies, and fatigue in 22% (2/9) of the studies.

BCTs Used

Self-monitoring and feedback were the most frequently used in mHealth apps for cancer (8/9, 89% of the studies), followed by goal setting and motivation in 44% (4/9) of the studies each. However, only 22% (2/9) of the studies used knowledge.

Behavioral Theory or Model

Most of the analyzed studies (6/9, 67%) seemed not to be based on common theories. Only 33% (3/9) of the studies reported at least one theory. The only reported theories were Social Cognitive Theory, Goal-Setting Theory, and the comprehensive intervention model. This result is surprising given that most apps used feedback, goal setting, and motivation to conduct behavior change. Therefore, further research could use behavioral theory or models to enhance their apps for this chronic disease. However, in physical interventions (no digital technology involved), some evidence showed that the transtheoretical model combined with other models was successful in breast cancer screening behavior as it is based on the stages of behavior change and interventions can be tailored to each individual, which increases their empowerment to make change [52].

Behavior Change Effectiveness

Cancer-specific mHealth apps helped significantly improve QOL in 44% (4/9) of the studies, followed by changes in anthropometrics in 33% (3/9) of the studies. A total of 11% (1/9) of the studies reported an increase in QOL and a decrease in distress level. In contrast, physical activity and nutritional status were only significantly improved in 22% (2/9) of the studies. Similar to the CVD section, linking BCTs and theoretical frameworks to behavior change and cancer health measures is challenging. Therefore, this area of study should be investigated further.

DM Results

Demographic, Participant, and Study Design Details

In total, 13 studies comprising 1559 patients were included. Of the 13 studies, we analyzed 10 (77%) randomized controlled trials, 1 (8%) single-arm feasibility study, 1 (8%) case report, and 1 (8%) uncontrolled study. The studies included in this review involving nutrition apps for diabetes can be found in Table 3. Overall, there was a diverse population of patients diagnosed with prediabetes (glucose: 5.55-6.94 mmol/L or 100-125 mg/dL; HbA_{1c}: 39-46 mmol/mol or 5.7%-6.4%), type

1 DM, T2DM, and gestational DM (2-hour oral glucose tolerance test level of ≥9 mmol/L). The features of each app included food tracking [53-62], education, knowledge, or recommendations [54-56,59-65], physical activity tracking [54-56,59-62], weight monitoring [56,59-62], glucose monitoring [55,57,60,61,64], blood pressure monitoring [60], community support [54,59,60], feedback [54-56,59,61,62,64], health coaches [55], clinician portals [60-62], connectivity to digital health devices [56,57], scheduled reminders [56,58,59,62], gamification [59,63], and goal setting [59]. In addition, several targeted behavioral domains and outcomes were measured as a result of the mHealth app nutrition intervention. This factor revealed some interesting insights for future investigations into digital health interventions among people with DM.

Table 3. Included studies on diabetes (n=13).

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Hong et al [53]	Korea	Case study	Examine the effect of a mobile app program ("Diabetes & Nutrition") developed between 2011 and 2012 for self-management in patients with T2DM[a] and recommend important considerations when the mobile app program is developed	Diabetes & Nutrition—all-smartphone app; features: food tracking	None	At 3 months, body weight had decreased by 4.4 kg, waist circumference had decreased by 5 cm, and HbA_{1c} level had decreased from 7.9% to 6.1%. The medication was reduced from the dose of 850 mg to 500 mg of metformin twice a day. Since then, the patient did not continue to use the "Diabetes & Nutrition" app as their level of blood glucose had stabilized and the patient felt that it was inconvenient and annoying to use the program. At 6 months, no significant change in body weight and body composition was observed in comparison with those at 3 months.
Xu et al [54]	China	2-arm RCT[b] with TTM[c]-based social media intervention	Examine the effectiveness of a 6-month mobile-based intervention (DHealthBar, a WeChat applet) combined with behavioral theory compared with a printed intervention in improving dietary behaviors, physical activity, and intention to change these behaviors among populations at high risk of T2DM	DHealthBar—all-smartphone app; features: food and physical activity tracking, knowledge, community support, and coach feedback	TTM	Participants in both groups reported a statistically significant decrease in energy intake at the 2 follow-up assessments compared with baseline. At 6 months, a significantly larger decrease was observed in the intervention group in energy, fat, and carbohydrate intake accompanied by a significantly larger increase in moderate-intensity physical activity compared with the control group. After 6 months of the intervention, participants were more likely to be at higher stages of dietary behaviors and physical activity than the control group.
Koot et al [55]	Singapore	6-month (24-week), single-arm, preintervention (baseline) and follow-up evaluation	Using the RE-AIM[d] evaluation framework, this study assessed the potential effectiveness and feasibility of GlycoLeap, a mobile lifestyle management program for people with T2DM, as an add-on to standard care.	GlycoLeap—all-smartphone app; features: food and physical activity tracking, knowledge, glucose monitoring, and coach feedback	Several theoretical frameworks, including the TTM and Health Belief Model	Program engagement (implementation) started out high but decreased with time for all evaluated components. Self-reported survey data suggest that participants monitored their blood glucose on more days in the previous week at follow-up compared with baseline and reported positive changes to their diet because of app engagement. Statistically significant improvements were observed for HbA_{1c}, with greater improvements for those who logged their weight more often. Participants had a 2.3% reduction in baseline weight. User satisfaction was high, with 74% (59/80) and 79% (63/80) of participants rating the app as good or very good and claiming that they would probably or definitely recommend it to others.
Griauzde et al [56]	United States	12-week, parallel, 3-arm, mixed methods pilot RCT	Examine the feasibility and acceptability of an mHealth[d] intervention designed to increase autonomous motivation and healthy behaviors among adults with prediabetes who had previously declined participation in a diabetes prevention program; in addition, the study aimed to examine changes in autonomous motivation among adults who were offered 2 versions of the mHealth program compared with an information-only control group	mHealth—all-smartphone app; features: food and physical activity tracking, knowledge, weight monitoring, coach feedback, connectivity to digital health devices, and reminders	Self-determination theory	No significant differences were observed in adherence rates between app-only and app-plus participants. Among all participants, mean autonomous motivation measures were relatively high at baseline (6.0 out of a 7.0 scale), with no statistically significant within- or between-group differences in follow-up scores.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Torbjørnsen et al [57]	Norway	3-armed RCT with 2 intervention groups and 1 control group	This study aimed to explore associations between the level of acceptability of a mobile diabetes app and the initial ability to self-manage in patients with T2DM.	Diabetes diary app (no name)—all-smartphone app; features: food tracking, glucose monitoring, and connectivity to digital health devices	None	The study found statistically significant associations between 5 of the 8 self-management domains and "perceived benefit," being one of the acceptability factors. However, when adjusting for age, gender, and frequency of use, only 1 domain, "skill and technique acquisition," remained independently associated with "perceived benefit." Frequency of use of the app was the factor that revealed the strongest association with the acceptability domain "perceived benefit." Moreover, an association was revealed between gender and frequency of use where 69% (25/36) of the high-frequency users were men.
Alfonsi et al [58]	Canada	Iterative usability testing (3 cycles)	Test the app's usability and potential impact on carbohydrate counting accuracy	iSpy—all-smartphone app; features: food tracking and reminders	None	Use of iSpy was associated with improved carbohydrate counting accuracy (total grams per meal), reduced frequency of individual counting errors of >10 g, and lower HbA_{1c} levels. Qualitative interviews and acceptability scale scores were positive. Moreover, 43% (9/21) of iSpy participants were still engaged, with use at least once every 2 weeks at the end of the study.
Block et al [59]	United States	Clinical trial	The aim was to evaluate the effectiveness of a fully automated, algorithm-driven behavioral intervention for diabetes prevention, Alive-PD, delivered via the internet, mobile phone, and automated phone calls.	Alive-PD—web-based application; features: food and physical activity tracking, knowledge, weight monitoring, community support, coach feedback, reminders, gamification, and goal setting	Learning theory, Social Cognitive Theory, and theory of planned behavior	Alive-PD participants achieved significantly greater reductions than controls in fasting glucose, HbA_{1c}, and body weight. Reductions in BMI, waist circumference, and TG[e] HDL[f] ratio were also significantly greater in Alive-PD participants than in the control group. At 6 months, the Alive-PD group reduced their Framingham 8-year diabetes risk from 16% to 11%, significantly more than the control group. Participation and retention were good; intervention participants interacted with the program a median of 17 out of 24 weeks, and 71.1% (116/163) were still interacting with the program in month 6.
Kodmaneh et al [63]	Iran	Interventional study	The aim of this study was to evaluate the effect of mobile game-based learning apps on improving dietary information in patients with T2DM.	Amoo—mobile phone game for all smartphones; features: knowledge and gamification	None	The results indicated a statistically significant difference between the pre- and posttest scores in the intervention group. However, there was no significant difference in fasting blood sugar.
Yu et al [60]	China	24-week, 4-arm, parallel group, non-blinded randomized trial	The aim of this study was to evaluate the effects of an MPA[g] combined with or without SMBG[h] on glycemic control in patients with diabetes.	Diabetes-Carer—all-smartphone app; features: food and physical activity tracking; knowledge; weight, blood pressure, and glucose monitoring; community support; and clinician portal	None	The HbA_{1c} levels in patients of all groups decreased significantly from baseline. There were significant differences in the proportions of patients that achieved HbA_{1c} <7% between groups, especially in group C and group D, compared with group A at week 24. 1,5-anhydroglucitol changes were obvious in group A and group C at week 24 from baseline. Factorial ANOVA showed that the MPA intervention was the main effective factor for HbA_{1c} change, and there was no effect on HbA_{1c} change for the SMBG intervention.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Garnweidner-Holme et al [64]	Norway	2-arm, multicenter, nonblinded RCT	The study analyzed secondary data from a 2-arm, multicenter, and nonblinded RCT to determine whether a smartphone app with targeted dietary information and blood glucose monitoring had an effect on the dietary behavior of women with GDM[i].	Pregnant+—all-smartphone app; features: knowledge, glucose monitoring, and coach feedback	None	All the participants showed improvements in their HDS-P+[j] from baseline. However, the Pregnant+ app did not have a significant effect on their HDS-P+. The control group reported a higher weekly frequency of choosing fish meals. No other significant differences were found between the intervention and control groups.
Agarwal et al [61]	Canada	Multicenter pragmatic RCT	The primary objective of this study was to conduct a pragmatic RCT of the Bluestar mobile app to determine if app use led to improved HbA_{1c} levels among diverse participants in real-life clinical contexts. The authors hypothesized that this mobile app would improve self-management and HbA_{1c} levels compared with controls.	BlueStar—all-smartphone app; features: food and physical activity tracking, knowledge, weight and glucose monitoring, coach feedback, and clinician portal	TTM	The results of an analysis of covariance controlling for baseline HbA_{1c} levels did not show evidence of intervention impact on HbA_{1c} levels at 3 months. Similarly, there was no intervention effect on secondary outcomes measuring diabetes self-efficacy, quality of life, and health care use behaviors. An exploratory analysis of 57 ITG[k] participants investigating the impact of app use on HbA_{1c} levels showed that each additional day of app use corresponded with a 0.016-point decrease in participants' 3-month HbA_{1c} levels. App use varied significantly by site as participants from one site logged in to the app a median of 36 days over 14 weeks; those at another site used the app significantly less.
Lim et al [62]	Singapore	Randomized clinical trial conducted at multiple primary care centers	Compare the effects of a culturally contextualized smartphone-based intervention with usual care on weight and metabolic outcomes	Nutritionist Buddy Diabetes—all-smartphone app; features: food and physical activity tracking, knowledge, weight monitoring, coach feedback, clinician portal, and reminders	Several theoretical models combined that included accountability, communication, and motivation to help adherence	Compared with the control group, intervention participants achieved significantly greater reductions in weight and HbA_{1c} levels, with a greater proportion experiencing a reduction in diabetes medications at 6 months. The intervention led to a greater HbA_{1c} reduction among participants with HbA_{1c} levels of ≥8%. Intergroup differences favoring the intervention were also noted for fasting blood glucose, diastolic blood pressure, and dietary changes.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Juan et al [65]	Spain	Uncontrolled study	Users would have a statistically significant increase in knowledge about the carbohydrate choices of real packaged foods after using the app.	Augmented Reality—all-smartphone app; features: knowledge	None	The results reported that their initial knowledge about carbohydrate choices was very low. This indicates that education about nutritional information in packaged foods is needed. An analysis of the pre- and postknowledge questionnaires showed that users had a statistically significant increase in knowledge of carbohydrate choices after using the app. Gender and age did not influence the knowledge acquired. The participants were highly satisfied with the app.

[a]T2DM: type 2 diabetes mellitus.

[b]RCT: randomized controlled trial.

[c]TTM: transtheoretical model.

[d]RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance.

[e]TG: triglyceride.

[f]HDL: high-density lipoprotein.

[g]MPA: mobile phone app.

[h]SMBG: self-monitoring of blood glucose.

[i]GDM: gestational diabetes mellitus.

[j]HDS-P+: healthy dietary score for Pregnant+.

[k]ITG: immediate treatment group.

Targeted Behavior and Outcome Measures

The most frequently targeted behaviors were glycemic control measured using HbA_{1c} and DM self-efficacy or self-management in 46% (6/13) of the reviewed studies. In addition, dietary behavior, levels of engagement, user acceptability, or motivation, and weight or BMI were measured in 46% (6/13) of the studies, followed by waist circumference in 23% (3/13) of the studies. Less frequently measured outcomes included physical activity, stages of change, carbohydrate counting accuracy, QOL, health care use behavior, and lipids in 23% (3/13) of the studies.

BCTs Used

Numerous BCTs can be used to induce behavior change. All the DM studies reviewed (13/13, 100%) used knowledge and education, followed by self-monitoring in 77% (10/13) of the studies. Both social support or encouragement and autonomous personalized feedback were used in 54% (7/13) of the studies. Prompts or cues were used in 31% (4/13) of the studies, followed by graded tasks and gamification in 15% (2/13) of the studies. Real-time feedback and goal setting were used in 15% (2/13) of the studies reviewed. Unfortunately, no standard definition of techniques is included in the BCTs, making it challenging to identify the techniques used in the interventions [66].

Behavioral Theory or Model

Some of the analyzed studies (7/13, 54%) seemed not to be based on behavioral theories or models. In total, 46% (6/13) of the studies reported at least one theory. The transtheoretical model was most frequently used in 23% (3/13) of the studies, followed by Social Cognitive Theory, self-determination theory, models centering on cues and triggers, theory of planned behavior, behavioral economics, positive psychology, and motivational interviewing as other reported theories. Some research has shown that Social Cognitive Theory has been used in feasibility studies among populations with diabetes as it can increase confidence and promote greater sustained effort to change, making it a guide for digital technology interventions. This theory also includes skill training, which can benefit diabetes management and education programs [67,68].

Behavior Change Effectiveness

DM-specific mHealth apps improved glycemic control by significantly reducing HbA_{1c} values in 46% (6/13) of the studies. In addition, 15% (2/13) of the interventions [53,62] resulted in a decrease in the medications used for glycemic control. Some studies (6/13, 46%) showed significant engagement; however, 17% (1/6) of these studies showed a decline in engagement over time, and 17% (1/6) did not have follow-up data for engagement. Sustainability needs to be considered for the effectiveness of these types of interventions. There were also significant changes in anthropometrics in 31% (4/13) of the reviewed studies. DM self-efficacy and self-management, decrease in energy intake, and increase in physical activity were observed in 8% (1/13) of the studies. Not all studies analyzed the same outcomes for each intervention, making it difficult to link BCTs and theoretical frameworks to behavior change and health measures.

Obesity Results

Demographic, Participant, and Study Design Details

A total of 18 studies comprising 253,775 patients were included. Among these 18 studies, we analyzed 5 (28%) randomized controlled trials, 5 (28%) experimental studies, 3 (17%) feasibility studies, 2 (11%) observations, and 1 (6%) prospective cohort study. The studies involving nutrition apps among people with obesity can be found in Table 4. The studies focused on obesity prevention and treatment in many diverse populations that ranged in age, socioeconomic background, and physical status (ie, pregnancy and post partum). App features included physical activity tracking [69-78], food tracking [70,72-83], knowledge, education, or recommendations [70,72-76,78,79,82-86], push notifications or scheduled reminders [74-76,78,79,81-84,86], weight monitoring [75,76,78,81,85,86], behavior demonstration [69,75,83], motivational challenges [69,75-78,83], goal setting [69,73-75,78,81,84,85], feedback [69,70,73,75,76,81-83,86], community support [70,75,78,85], connectivity to digital health devices [70,72,81,86], clinician portals [72,73,82], and access to a health coach [76-78,81]. Furthermore, numerous behaviors were targeted, and the outcomes were analyzed to determine the effectiveness of a mobile nutrition intervention in promoting healthy weight. This information will help develop approaches and techniques for digital health behavior change interventions to prevent and treat obesity.

Table 4. Included studies on obesity (n=18).

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Lubans et al [69]	Australia	Cluster RCT[a]	Focused on the promotion of lifetime (eg, resistance training) and lifestyle (eg, active transport) physical activities and was aligned with current physical activity guidelines, which include a recommendation to engage in muscle- and bone-strengthening physical activities at least 3 days per week	ATLAS[b]—all-smartphone app; features: physical activity tracking, behavior demonstration, challenges, goal setting, and coach feedback	Self-determination theory and Social Cognitive Theory	Focus group participants reported enjoyment of the program and felt that it had provided them with new skills, techniques, and routines for the future. However, their engagement with the smartphone app was limited. Barriers to the implementation and evaluation of the app included limited access to smartphone devices, technical problems with the push notifications, lack of access to use data, and the challenges of maintaining participants' interest in using the app.
Griffin et al [84]	United States	1-group, pre- to posttest study design	To evaluate changes in dietary and physical activity behaviors and weight after implementation of a 12-week SMS text messaging initiative (My Quest)	SMS text messaging initiative (My Quest)—all smartphones; features: knowledge, reminders, and goal setting	Social Cognitive Theory	Participants significantly improved dietary and physical activity behaviors and food environment, increased their dietary and physical activity goal setting, and reduced their body weight. A total of 56 posttest assessments were completed (84% response rate).
Van der Pligt et al [71]	Australia	Pilot intervention study nested within a cluster RCT	Effectiveness of the Mums OnLiNE[c] intervention with respect to reducing PPWR[d] and improving diet, physical activity, and sedentary behavior	Mums OnLiNE Combined and In-FANT Extend[e]—web-based application or smartphone app with telephone-based support; features: physical activity tracking	Social Cognitive Theory	Mean PPWR decreased in the intervention group and the comparison group 2 although the changes were not significant. Mean waist circumference for all groups exceeded recommendations at baseline but decreased to below recommendations for women in the intervention group and significantly for the intervention group compared with comparison groups 1 and 2. Changes in diet, physical activity, or sedentary behaviors were not significant.
Hull et al [79]	United States	Observational design based on data collected after the testing period	This paper describes the development and beta testing of the CHEW[f] smartphone app. The objective of beta testing was to test the CHEW app prototype with target users focusing on use, usability, and perceived barriers and benefits of the app.	CHEW smartphone app—Android-based app; features: food tracking and knowledge	Socioecological model	Study participants used the app on average once a week for approximately 4.5 minutes per session. Use of specific features averaged at 1-2 times per month for shopping-related activities and 2-4 times per month for the snack gallery. Mothers classified as users rated the app's WIC[g] Shopping Tools relatively high on usability and benefits. The Yummy Snack Gallery and Healthy Snacking Tips scored higher on usability than on benefits, suggesting that the nutrition education components may have been appealing.
Bughin et al [72]	France	Randomized controlled study	The aim of this study was to compare the changes in body composition, anthropometric parameters, exercise capacity, and QOL[h] within 12 weeks of patients in the TR[i] program with those of usual care patients with obesity.	TR Program—Android-based or web-based; features: food and physical activity tracking, knowledge, connectivity to digital health devices, and clinician portal	None	No significant group × time interaction was found for fat mass. Compared with the UCG[j], TRG[k] patients tended to significantly improve their waist-to-hip ratios and improved their QOL physical impact. Significant time effects were observed for body composition, 6-minute Walk Test distance, exercise metabolism, sedentary time, and QOL. Adherence (95%) and satisfaction in the TRG were good.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Toro-Ramos et al [78]	Korea	Intervention	This study investigated the efficacy of a smartphone intervention using a designated app with a lifestyle intervention–focused approach, including a human coaching element, toward weight loss in Korean adults who were overweight or obese.	Noom—Android- and Apple-based app; features: food and physical activity tracking, knowledge, reminders, weight monitoring, challenges, goal setting, community support, and coach feedback	None	Participants showed a clinically significant weight loss effect of −7.5% at the end of the 15-week program, and at a 52-week follow-up, a weight loss effect of −5.2% was maintained. At 15 weeks, percentage of body fat and visceral fat decreased by −6.0% to –5.4% and −3.4 kg to –2.7 kg, respectively. Fasting blood glucose level also decreased significantly by −5.7 to –14.6 mg/dL at 15 weeks. Lipid parameters showed significant improvements except for high-density lipoprotein cholesterol. The frequency of logging meals and exercise was associated with body fat loss.
Pellegrini et al [70]	United States	6-month technology-supported weight loss trial	Examine within-person variation in dietary self-monitoring during a 6-month technology-supported weight loss trial as a function of time-varying factors, including time in the study, day of the week, and month of the year	ENGAGED[l]—smartphone provided with the app; features: food and physical activity tracking, knowledge, feedback, community support, and connectivity to digital health devices	None	Participants recorded less as time in the study progressed. Fewer foods were reported on the weekends compared with on weekdays. More foods were self-monitored in January compared with in October; however, a seasonal effect was not observed.
Dodd et al [74]	Australia	Multicenter, nested randomized trial	The objective was to evaluate the impact of a smartphone app as an adjunct to face-to-face consultations in facilitating dietary and physical activity change among pregnant women.	Name not provided—smartphone was provided with the app; features: food and physical activity tracking, knowledge, reminders, and goal setting	None	Mean difference in HEI[m] score was 0.01 at 28 weeks of pregnancy and −1.16 at 36 weeks of pregnancy. There was no significant additional benefit from the provision of the smartphone app in improving HEI score. Although all women improved dietary quality throughout their pregnancy, use of the smartphone app was poor.
Stasinaki et al [77]	Switzerland	Unblinded RCT	The objective of the study was to assess novel obesity management that moved the focus from on-site consultations in a specialized childhood obesity center to an appealing, youth-friendly, low-threshold mobile intervention (PathMate2) under the supervision of pediatric obesity experts.	PathMate2—smartphone provided with the app; features: food and physical activity tracking, challenges, and coach feedback	None	At intervention start, median BMI SDS[n] of all patients was 2.61. BMI-SDS decreased significantly in the control group at time 1 but not at time 2 and did not decrease in the intervention group during the study. Muscle mass, strength, and agility improved significantly in both groups at time 2; only the intervention group significantly reduced their body fat at time 1 and time 2. Average daily PathMate2 app use rate was 71.5%. Cortisol serum levels decreased significantly after biofeedback but with no association between stress parameters and BMI-SDS.
Ali et al [73]	UAE[o]	Nonrandomized, 2-arm feasibility study	Develop and test a nutrition education intervention delivered via a website and mobile apps to university students in the UAE who were overweight and obese	Rashakaty-Basic and Rashakaty-Enhanced—web-based application and all-smartphone app; features: food and physical activity tracking, knowledge, goal setting, feedback, and clinician portal	Social Cognitive Theory	There was no significant difference in weight loss between the 2 arms. However, waist circumference decreased more in the Rashakaty-Enhanced group. Changes in knowledge related to sources of nutrients and diet-disease relationships were significantly higher among the Rashakaty-Enhanced group. Rashakaty-Enhanced participants reported increased number of days spent on moderate physical activity and minutes walked. They also reported higher scores in social support from friends to reduce fat intake and from family and friends to increase physical activity.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Senecal et al [86]	China	Retrospective observational analysis	To evaluate whether individuals following a weight loss program based on a mobile app, wireless scale, and nutritional program but no face-to-face care could achieve clinically significant weight loss in a large cohort	MetaWell—Android- and Apple-based app; features: knowledge, reminders, weight monitoring, feedback, and connectivity to digital health	None	251,718 individuals (79% female) were included with a mean weight loss of 4.3 kg and a mean follow-up of 120 days. Mean weight loss at 42, 60, 90, and 120 days was 4.1 kg, 4.9 kg, 5.6 kg, and 5.4 kg, respectively. At 120 days, 62.7% of participants had lost at least 5% of their initial weight. Both genders and all use frequencies showed statistically significant weight loss from baseline at each interval, and this loss was greater in men than in women. The frequency of recording was associated with greater weight loss when comparing high-, medium-, and low-use groups at all time intervals investigated.
Delisle Nyström et al [82]	Sweden	2-arm parallel RCT	Investigate the 12-month after-baseline measurements of the MINISTOP[p] intervention	MINISTOP—both web-based application and all-smartphone app; features: food tracking, knowledge, reminders, feedback, and clinician portal	Social Cognitive Theory	At the 12-month follow-up, no statistically significant difference was observed between the intervention and control groups for FMI[q], and no maintained effect for the change in composite score was observed.
Stein and Brooks [81]	United States	Longitudinal observational study	Evaluate weight loss, changes in meal quality, and app acceptability among users of the HCAI[r] with the overarching goal of increasing access to compassionate health care via mHealth[s]	Lark Weight Loss HCAI—Android- and Apple-based app; features: food tracking, reminders, weight monitoring, goal setting, coach feedback, and connectivity to digital health devices	None	Weight loss was 2.38% of baseline weight. The average duration of app use was 15 weeks, and users averaged 103 sessions each. The percentage of healthy meals increased by 31%. The in-app user trust survey had a 100% response rate and positive results, with a satisfaction score of 87 out of 100 and Net Promoter Score of 47.
Prasad et al [80]	United States	Open-label, nonrandomized, prospective 90-day TRE[t] intervention	The primary aim was to test the feasibility of a TRE intervention administered via a smartphone app aimed at reducing the eating window by 4 hours in individuals with a habitually prolonged eating window and determine the efficacy of a 90-day TRE intervention in reducing body weight and blood pressure in adults who were overweight and obese. A secondary aim was to monitor the adherence to the intervention over time.	MyCircadianClock—all-smartphone app; features: food tracking	None	The mean duration of the baseline eating window was 14 hours and 32 minutes (SD 2 hours and 36 minutes), with 56% of participants with a duration of ≥14 hours. TRE participants successfully decreased their eating window from 16 hours and 4 minutes (SD 1 hour and 24 minutes) to 11 hours and 54 minutes (SD 2 hours and 6 minutes) and reduced the number of daily eating occasions by half. Adherence to logging and to the reduced eating window was 64% (SD 22%) and 47% (SD 19%). TRE resulted in decreases in body weight, waist circumference, and systolic blood pressure.
Simpson et al [85]	Scotland	Feasibility RCT	To develop and assess the feasibility and acceptability of an app-, web-, and social support–based intervention in supporting adults with obesity to achieve weight loss goals	HelpMeDoIt!—all-smartphone app and web-based application; features: knowledge, reminders, weight monitoring, goal setting, and community support	Social Cognitive Theory and control theory	Of the 54 (74%) participants who downloaded the app, 48 (89%) used it. Objective physical activity measures perhaps showed the most potential (daily step count [1187 steps] and sedentary time [−60.8 min]). However, these outcomes were poorly completed.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Lin et al [75]	United States	RCT	To compare an mHealth intervention delivered via a CP[u] app with usual care controls and compare with an in-person and phone-supplemented personal coaching intervention enhanced by CP self-monitoring with usual care	CITY[v]—all-smartphone app; features: food and physical activity tracking, knowledge, reminders, weight monitoring, behavior demonstration, challenges, goal setting, feedback, and community support	Rooted in theoretical models and behavioral framework	Use of the app was highest during month 1 for both arms; thereafter, use dropped substantially and continuously until the study end. During the first 6 months, the mean percentage of days that any app component was used was higher for the CP arm (74.2%) than for the personal coaching arm (48.9%). The CP arm used the apps an average of 5.3 times per day, whereas the personal coaching participants used them 1.7 times per day. The former self-weighed more than the latter (57.1% of days vs 32.9% of days). Furthermore, the percentage of days that any app component was used, number of app uses per day, and percentage of days self-weighed all showed significant differences across the 4 weight categories for both arms. Pearson correlation showed a negative association between weight change and the percentage of days that any app component was used, number of app uses per day, and percentage of days self-weighed.
Chew et al [76]	Singapore	Prospective single-cohort study	Assess the effectiveness of adolescent engagement with a mobile app–based lifestyle intervention program as an early intervention before enrollment in a clinic-based multidisciplinary weight management program	Kurbo—all-smartphone app; features: food and physical activity tracking, knowledge, reminders, weight monitoring, challenges, and coach feedback	None	Kurbo engagement was high, with 83% (33/40) of participants completing at least 7 coaching sessions. In total, 78% (18/23) of participants rated the app as good to excellent, and 70% (16/23) stated that they would recommend it to others. There were no statistically significant changes in BMI *z* scores at 3 or 6 months. Participants showed statistically significant improvements in measured body fat percentage, self-reported QOL, and self-reported caloric intake from the 3-day food diaries at 3 and 6 months.

Study	Country	Study design	Purpose	Name of mobile app and features	Behavior change theory or model	Results
Kay et al [83]	United States	Randomized controlled feasibility trial	Comparing app-based diet tracking (active comparator) with app-based diet tracking plus feedback on DASH[w] adherence via SMS text message (intervention)	DASH Cloud—all-smartphone app; features: food tracking, knowledge, reminders, behavior demonstration, challenges, and feedback	None	DASH Cloud did not enhance DASH adherence over diet tracking alone but resulted in greater reductions in blood pressure.

[a]RCT: randomized controlled trial.

[b]ATLAS: Active Teen Leaders Avoiding Screen-Time.

[c]OnLiNE: Online, Lifestyle, Nutrition, and Exercise.

[d]PPWR: postpartum weight retention.

[e]InFANT Extend: Extended Melbourne Infant Feeding Activity and Nutrition Trial.

[f]CHEW: Children Eating Well.

[g]WIC: Women, Infants, and Children.

[h]QOL: quality of life.

[i]TR: telerehabilitation.

[j]UCG: usual care group.

[k]TRG: TR group.

[l]ENGAGED: e-Networks Guiding Adherence to Goals in Exercise and Diet.

[m]HEI: healthy eating index.

[n]SDS: SD score.

[o]UAE: United Arab Emirates.

[p]MINISTOP: Mobile-Based Intervention Intended to Stop Obesity in Preschoolers.

[q]FMI: fat mass index.

[r]HCAI: Health Coach AI.

[s]mHealth: mobile health.

[t]TRE: time-restricted eating.

[u]CP: cell phone.

[v]CITY: Cell Phone Intervention For You.

[w]DASH: Dietary Approaches to Stop Hypertension.

Targeted Behavior and Outcome Measures

Anthropometric measurements (ie, weight, BMI, and waist-to-hip ratio) were the most targeted health outcomes, with behavior change in 78% (14/18) of the reviewed studies, followed by diet quality and physical activity in 61% (11/18) of the studies. In addition, engagement, user acceptability, or motivation levels and body composition were measured in 39% (7/18) and 28% (5/18) of the studies, respectively. QOL, behaviors (ie, goal setting and self-efficacy), and CVD measures (ie, blood pressure, heart rate, glucose, and lipids) were included as outcomes in 22% (4/18) of the studies. Less frequent measures included stress parameters (ie, chronic stress and cortisol levels), diet knowledge, reduced screen time, and behavior and health maintenance, observed in 6% (1/18) of the studies.

BCTs Used

BCTs were incorporated into each intervention. Self-monitoring was the most commonly used BCT (14/18, 78% of the reviewed studies), followed by knowledge or education (12/18, 67% of the studies). Goal setting was used in 56% (10/18) of the studies, feedback and encouragement were used in 56% (10/18) of the studies, and prompts or cues and intention formation were used in 50% (9/18) of the studies. Feedback through private messages was used in 39% (7/18) of the studies, followed by community or social support and demonstration of behavior in 33% (6/18) of the studies each. Graded tasks and gamification or incentives were used in 17% (3/18) of the studies. Finally, information about health benefits and consequences was used in 6% (1/18) of the studies in this section. Clinician portals were used in 17% (3/18) of the interventions to improve patient care. Although it is not a BCT, it is worth mentioning as a vital intervention component.

Behavioral Theory or Model

Some interventions (10/18, 56%) did not mention the basis of behavioral theories or models. A total of 6% (1/18) of the studies mentioned that they were rooted in theoretical models but did not specify which ones. Social Cognitive Theory was the most frequently used behavior model in 33% (6/18) of the studies. Self-determination theory was used in 11% (2/18) of the studies, followed by the socioecological model, control theory, and social support theory in 6% (1/18) of the studies. Furthermore, Social Cognitive Theory has been used in feasibility studies on mHealth technology lifestyle interventions for obesity, for example, by using the principle of verbal persuasion through methods such as personalized encouragement that help individuals realize that they have the capability to make the necessary healthy lifestyle changes to lose weight [87].

Behavior Change Effectiveness

Obesity-specific mHealth apps improved weight through significant reductions in 39% (7/18) of the studies. Some studies (4/18, 22%) showed a substantial decrease in body fat percentage and waist circumference. There were also significant changes in cardiovascular measurements (ie, blood pressure and lipids) and QOL in 11% (2/18) of the reviewed studies. Changes in self-efficacy, decreases in energy intake and screen time, and increases in knowledge of nutrients and physical capabilities were observed in 6% (1/18) of the studies. In total, 11% (2/18) of the studies analyzed and reported significance in the maintenance of weight loss, whereas 11% (2/18) of the studies reported a decline in changes over time. The sustainability of these interventions is an area for future research to determine the effectiveness of these types of interventions.

Discussion

Principal Findings

Overall, mHealth apps used for various chronic disease populations did improve health. These studies showed significant improvements in QOL, cardiac rehabilitation completion, glycemic control (ie, HbA_{1c}), weight reduction, and reduction in physiological measures (ie, blood pressure and lipids). In addition, some studies (3/46, 7%) showed improved self-efficacy and self-management of chronic diseases. Although many studies (35/46, 76%) did not measure long-term effectiveness, some (3/46, 7%) showed significance in maintaining weight loss, whereas others (3/46, 7%) showed a decline in changes or engagement in app use over time. A total of 24% (11/46) of the studies measured maintenance of health behavior change. Of these 11 studies, 7 (64%) sustained behavior change for approximately 6 to 12 months, and 4 (36%) showed a decline in behavior change or discontinued app use.

The main difference between the sustainability of health behavior change was the inclusion of a clinician portal or access to a health coach. Griauzde et al [56] measured postintervention qualitative data that provided reasons for app satisfaction, dissatisfaction, and ways to improve. Reasons for satisfaction included "encouraged self-reflection," reasons for app dissatisfaction included "did not consider personal circumstances," and strategies to improve the intervention included "increased interpersonal contact." Of the 46 studies, 9 (20%) included a clinician portal as a feature to enhance the app intervention, allowing for additional communication between clinicians and patients. Of the 9 studies, 7 (78%) had an effective health behavior change. Silk et al [88] reported acceptability and easy-to-use features regarding the integration of the clinician portal into their mHealth intervention. This component should be considered for health behavior change interventions and investigated further.

Of the 46 nutrition apps in this review, 37 (80%) included some type of self-monitoring feature. Of those 37 apps, 35 (95%) had at least one significant improvement or good usability rating to promote engagement with mobile apps for chronic disease self-management. This is in alignment with other research that shows that the most common self-management application for mHealth is a tracking feature [89]. In addition, when surveying clinicians working in diabetes and weight management patient care settings, the most adoptable apps included self-monitoring features [90]. However, the apps that did not include self-monitoring or tracking features showed significant improvements as well. In addition, there was variability in other features combined with tracking features (ie, knowledge, goal setting, coach feedback, and clinician portals), making it difficult to attribute behavior change to self-monitoring features alone. More research is needed to correlate specific features with health behavior change.

Except for cancer populations, a key finding in this review was that 41% (19/46) of the nutrition app interventions targeted weight management, and 58% (11/19) of those studies were effective in health behavior change. This finding is similar to that of Fakih El Khoury et al [91], who reported that dietary mobile apps positively affected measured nutritional outcomes in chronic diseases, especially weight loss. The key finding for cancer populations was that mHealth nutrition apps can significantly improve QOL. A similar result was found in a review analyzing the use of mHealth involving nutrition in a chronic kidney disease population [92]. A total of 28% (13/46) of the studies examined changes in glycemic control, HbA_{1c}, and other biometric measurements (ie, blood pressure and lipids). In total, 69% (9/13) of those studies showed a significant reduction in their measurements when using a nutrition app in their intervention. In addition, 15% (2/13) of those studies showed a decrease in medication treatment for glycemic control [53,62]. Eberle et al [93] concluded that mHealth apps with a nutrition intervention effectively enhanced diabetes management, which is comparable with our review results.

There was diversity in the chronic disease state; study design; number of participants; and variety of in-app features, BCTs, and behavior models used in the studies. However, linking these factors to a particular behavior change and health outcome posed to be difficult, which is not different from Taj et al [94] in their review of digital health behavior change technology. This review provides insight into the theories and theoretical frameworks or models used in nutrition-focused mHealth interventions to increase understanding and translate them into practice, which is essential for developing behavior change for the sustainability of health improvement. For example, this scoping review found that most apps used BCTs, such as self-monitoring in 80% (37/46) of the interventions, knowledge in 72% (33/46) of the interventions, feedback in 54% (25/46) of the interventions, and goal setting in 20% (9/46) of the interventions, for effective health behavior change, as another review [31] demonstrated.

In addition, our review found that less than half (19/46, 41%) of the studies based their nutrition apps on a behavioral theory or its constructs. Some studies (14/46, 30%) that based their interventions on theories had improved engagement, satisfaction, or app use among participants and suggested that incorporating a behavioral theory in mHealth interventions is an effective strategy, which is a comparable finding with that of another review [91]. However, other factors play a role in usability, acceptability, and overall user satisfaction. The app features, BCTs, and theory-based interventions will all affect the effectiveness of mHealth nutrition apps in health behavior change toward chronic diseases.

Another significant development in the sector of health technology is that smartphones can also be embedded with sensors or coupled with wearable sensors for health monitoring, which could enhance a nutrition app's effectiveness in health behavior change. Examples of these devices that could be coupled with nutrition apps include motion sensors such as accelerometers, gyroscopes, and magnetometers that measure motion and physical activity [95]; wearable devices such as glasses (analyzing the intake pattern with smart glasses) and rings (ring-type tactile sensors to detect food mass) for food intake measurement [96]; portable and handheld single-lead electrocardiogram devices in addition to fitness trackers to measure heart rate and heart rate variability [95]; wristwatches to measure glucose levels extracted from skin interstitial fluid through reverse ionophoresis or through saliva-, sweat-, or tear-based wearable biosensors [97]; a wrist-wearable watch with the function of a pulsimeter without a cuff to measure blood pressure [97]; and ingestible capsules that can be used for medical health monitoring [98]. These enhance personal health care and performance monitoring with the potential to complement nutrition apps and have a broad impact on our society.

Limitations

Although the results of this study showed health improvements achieved when using mHealth nutrition apps for behavior change in chronic diseases, some limitations need to be addressed. An important limitation is the lack of research on mobile apps' long-term effects (>1 year) on disease state populations. Therefore, there is no conclusive result on their long-term behavior change to determine whether the behavior continues to improve or reverses compared with the baseline. Consequently, there need to be more long-term studies conducted. Second, many studies (24/46, 52%) did not include a control group for comparison, and their sample size was small, limiting the interpretation of the results of the studies. Third, many apps are at risk of becoming rapidly obsolete owing to the fast pace at which technologies progress and, therefore, new technological innovations must be considered [8]. For example, the latest mobile technologies can connect and interact with each other, update and track personal health data in real time, and send alerts to users [8]. Similarly, most health apps have encountered serious usability problems or have not undergone usability assessment [99]. Usability affects the efficiency and efficacy of the app (ie, time to complete tasks and errors) [8]. Usability must be considered to increase the chance of the app being successfully adopted by patients [8]. Barriers to using mHealth apps include the patient's lack of integration of technology into everyday life [12] and difficulties using mobile apps [11]. In the older adult population, health problems such as cognitive changes related to aging, disability, and lack of confidence are reasons for not using digital technology [18,19]. Further research is needed to evaluate patients' experiences with apps and the benefits gained as a result. There could be a slight bias from the user perspective as there were 2% (1/46) of apps that were Android-only or web-based applications. The others (45/46, 98%) were web-based applications or all-smartphone apps. There were a few apps (4/46, 9%) that were provided through the smartphone that participants were given as part of the study, and the smartphone was not specified. A few studies (4/46, 9%) did not report the data privacy rules or the effects of users and funders on the app interventions. In addition, there are important confidentiality and funding issues that must be considered when designing interventions [100]. Finally, proficient health care providers should be involved in the app development stage to address safety during self-management and health education [101]. There is a need for comprehensive, efficient, and flexible mobile apps for the self-management of disease states with more features to increase the number of long-term users and induce better self-management and patient empowerment [101,102]. We did not conduct a systematic review or meta-analysis and, thus, did not weigh the quality of evidence or study design against the reported results. Some studies (6/46, 13%) included few participants, and the diversity of study objectives, designs, and outcomes made it difficult to compare them. We reviewed the current evidence to expand the knowledge base regarding the impact of nutrition apps on chronic disease management and assess the effectiveness of health behavior change.

Conclusions

In this scoping review, the use of mHealth nutrition apps and their effects on health behavior change were analyzed for 4 diseases (ie, cancer, CVD, DM, and obesity). The results suggest that mHealth apps involving nutrition can significantly improve health outcomes for people with chronic diseases. The study design, demographics, targeted behavior, health outcomes, BCTs, behavioral theories, and behavior change effectiveness were profoundly diverse among these studies, indicating that a *one-size-fits-all* approach for designing and implementing nutrition apps as part of chronic disease treatment is not possible. Tailoring nutrition apps to specific populations is recommended for effective behavior change and improvement of health outcomes. In addition, some studies (7/46, 15%) showed sustained health behavior change, and some (4/46, 9%) showed a decline in the use of the nutrition apps. These results indicate a need for further investigation on the sustainability of the health behavior change effectiveness of disease-specific nutrition apps.

Conflicts of Interest

None declared.

References

1. Ghosal A, Subramanian VS, Said HM. Role of the putative N-glycosylation and PKC-phosphorylation sites of the human sodium-dependent multivitamin transporter (hSMVT) in function and regulation. Biochimica et Biophysica Acta (BBA) - Biomembranes 2011 Aug;1808(8):2073-2080. [doi: 10.1016/j.bbamem.2011.04.014]
2. Hurst JR, Dickhaus J, Maulik PK, Miranda JJ, Pastakia SD, Soriano JB, et al. Global Alliance for Chronic Disease researchers' statement on multimorbidity. Lancet Global Health 2018 Dec;6(12):e1270-e1271. [doi: 10.1016/s2214-109x(18)30391-7]
3. About Chronic Diseases. Center for Disease Control and Prevention. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2022-05-02]
4. Zou Z, Cini K, Dong B, Ma Y, Ma J, Burgner DP, et al. Time trends in cardiovascular disease mortality across the BRICS: an age-period-cohort analysis of key nations with emerging economies using the global burden of disease study 2017. Circulation 2020 Mar 10;141(10):790-799. [doi: 10.1161/CIRCULATIONAHA.119.042864] [Medline: 31941371]
5. Agarwal MA, Fonarow GC, Ziaeian B. National trends in heart failure hospitalizations and readmissions from 2010 to 2017. JAMA Cardiol 2021 Aug 01;6(8):952-956 [FREE Full text] [doi: 10.1001/jamacardio.2020.7472] [Medline: 33566058]
6. World health statistics 2022: monitoring health for the SDGs, sustainable development goals. World Health Organization. URL: https://www.who.int/publications/i/item/9789240051157 [accessed 2022-05-02]
7. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00054] [Medline: 30652617]
8. Agher D, Sedki K, Despres S, Albinet J, Jaulent M, Tsopra R. Encouraging behavior changes and preventing cardiovascular diseases using the prevent connect mobile health app: conception and evaluation of app quality. J Med Internet Res 2022 Jan 20;24(1):e25384. [doi: 10.2196/25384]
9. Mobile phone ownership over time. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-05-02]
10. Alenazi HA, Jamal A, Batais MA. Identification of type 2 diabetes management mobile app features and engagement strategies: modified delphi approach. JMIR Mhealth Uhealth 2020 Sep 11;8(9):e17083. [doi: 10.2196/17083]
11. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013 Jan;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
12. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
13. Stein K. Remote nutrition counseling: considerations in a new channel for client communication. J Academy Nutrition Dietetics 2015 Oct;115(10):1561-1576. [doi: 10.1016/j.jand.2015.06.375]
14. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484. [doi: 10.1111/obr.12903] [Medline: 31353783]
15. Chow CK, Ariyarathna N, Islam SM, Thiagalingam A, Redfern J. mHealth in cardiovascular health care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
16. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your "App"etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017 Jul 10;5(7):e95 [FREE Full text] [doi: 10.2196/mhealth.7410] [Medline: 28694241]
17. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e16695. [doi: 10.2196/16695]
18. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
19. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
20. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
21. Groen WG, van Harten WH, Vallance JK. Systematic review and meta-analysis of distance-based physical activity interventions for cancer survivors (2013–2018): we still haven't found what we're looking for. Cancer Treatment Rev 2018 Sep;69:188-203. [doi: 10.1016/j.ctrv.2018.07.012]
22. Robertson MC, Tsai E, Lyons EJ, Srinivasan S, Swartz MC, Baum ML, et al. Mobile health physical activity intervention preferences in cancer survivors: a qualitative study. JMIR Mhealth Uhealth 2017 Jan 24;5(1):e3 [FREE Full text] [doi: 10.2196/mhealth.6970] [Medline: 28119278]

23. Pope Z, Lee JE, Zeng N, Lee HY, Gao Z. Feasibility of smartphone application and social media intervention on breast cancer survivors' health outcomes. Transl Behav Med 2019 Jan 01;9(1):11-22. [doi: 10.1093/tbm/iby002] [Medline: 29471477]
24. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
25. Heitkemper EM, Mamykina L, Travers J, Smaldone A. Do health information technology self-management interventions improve glycemic control in medically underserved adults with diabetes? A systematic review and meta-analysis. J Am Med Inform Assoc 2017 Sep 01;24(5):1024-1035. [doi: 10.1093/jamia/ocx025] [Medline: 28379397]
26. Lian J, McGhee S, Chau J, Wong CK, Lam CL, Wong WC. Systematic review on the cost-effectiveness of self-management education programme for type 2 diabetes mellitus. Diabetes Res Clin Pract 2017 May;127:21-34. [doi: 10.1016/j.diabres.2017.02.021]
27. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
28. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of digital interventions for improving glycemic control in persons with poorly controlled type 2 diabetes: a systematic review, meta-analysis, and meta-regression analysis. Diabetes Technol Therapeutics 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
29. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
30. Alvarez-Moya EM, Mirallas J, Fontanals C, Quintana M, Cusidó J, Rimbau J, et al. Development of a mobile application for people with obesity. J Nutrition Educ Behav 2021 Jan;53(1):79-83. [doi: 10.1016/j.jneb.2020.07.004]
31. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
32. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: a systematic literature review. J Telemed Telecare 2015 Feb 12;21(5):243-253. [doi: 10.1177/1357633x15572203]
33. Porter J, Huggins CE, Truby H, Collins J. The effect of using mobile technology-based methods that record food or nutrient intake on diabetes control and nutrition outcomes: a systematic review. Nutrients 2016 Dec 17;8(12):1-13 [FREE Full text] [doi: 10.3390/nu8120815] [Medline: 27999302]
34. Campbell J, Porter J. Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: a systematic review. Nephrology (Carlton) 2015 Oct;20(10):744-751. [doi: 10.1111/nep.12500] [Medline: 25959301]
35. Rejeski WJ, Fanning J. Models and theories of health behavior and clinical interventions in aging: a contemporary, integrative approach. Clin Interventions Aging 2019 May;Volume 14:1007-1019. [doi: 10.2147/cia.s206974]
36. Naslund JA, Aschbrenner KA, Kim SJ, McHugo GJ, Unützer J, Bartels SJ, et al. Health behavior models for informing digital technology interventions for individuals with mental illness. Psychiatr Rehabil J 2017 Sep;40(3):325-335. [doi: 10.1037/prj0000246] [Medline: 28182469]
37. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:1-9 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
38. Moher D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269. [doi: 10.7326/0003-4819-151-4-200908180-00135]
39. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep;24(13):1435-1444. [doi: 10.1177/2047487317715713] [Medline: 28631933]
40. Steinberg DM, Kay MC, Svetkey LP, Askew S, Christy J, Burroughs J, et al. Feasibility of a digital health intervention to improve diet quality among women with high blood pressure: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2020 Dec 07;8(12):1-15 [FREE Full text] [doi: 10.2196/17536] [Medline: 33284116]
41. Choi BG, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Image-based mobile system for dietary management in an American cardiology population: pilot randomized controlled trial to assess the efficacy of dietary coaching delivered via a smartphone app versus traditional counseling. JMIR Mhealth Uhealth 2019 Apr 23;7(4):1-13. [doi: 10.2196/10755]
42. Indraratna P, Biswas U, McVeigh J, Mamo A, Magdy J, Vickers D, et al. A smartphone-based model of care to support patients with cardiac disease transitioning from hospital to the community (TeleClinical care): pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 28;10(2):e32554. [doi: 10.2196/32554]
43. Schmaderer M, Miller JN, Mollard E. Experiences of using a self-management mobile app among individuals with heart failure: qualitative study. JMIR Nursing 2021 Aug 9;4(3):e28139. [doi: 10.2196/28139]
44. Cairo J, Williams L, Bray L, Goetzke K, Perez AC. Evaluation of a mobile health intervention to improve wellness outcomes for breast cancer survivors. J Patient Centered Res Rev 2020 Oct 23;7(4):313-322. [doi: 10.17294/2330-0698.1733]

45. Cheng C, Ho RT, Guo Y, Zhu M, Yang W, Li Y, et al. Development and feasibility of a mobile health-supported comprehensive intervention model (CIMmH) for improving the quality of life of patients with esophageal cancer after esophagectomy: prospective, single-arm, nonrandomized pilot study. J Med Internet Res 2020 Aug 18;22(8):e18946 [FREE Full text] [doi: 10.2196/18946] [Medline: 32808933]
46. Keum J, Chung MJ, Kim Y, Ko H, Sung MJ, Jo JH, et al. Usefulness of smartphone apps for improving nutritional status of pancreatic cancer patients: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e21088 [FREE Full text] [doi: 10.2196/21088] [Medline: 34463630]
47. Yang K, Oh D, Noh JM, Yoon HG, Sun J, Kim HK, et al. Feasibility of an interactive health coaching mobile app to prevent malnutrition and muscle loss in esophageal cancer patients receiving neoadjuvant concurrent chemoradiotherapy: prospective pilot study. J Med Internet Res 2021 Aug 27;23(8):e28695. [doi: 10.2196/28695]
48. Soh JY, Cha WC, Chang DK, Hwang JH, Kim K, Rha M, et al. Development and validation of a multidisciplinary mobile care system for patients with advanced gastrointestinal cancer: interventional observation study. JMIR Mhealth Uhealth 2018 May 07;6(5):e115 [FREE Full text] [doi: 10.2196/mhealth.9363] [Medline: 29735478]
49. Lozano-Lozano M, Cantarero-Villanueva I, Martin-Martin L, Galiano-Castillo N, Sanchez M, Fernández-Lao C, et al. A mobile system to improve quality of life via energy balance in breast cancer survivors (BENECA mHealth): prospective test-retest quasiexperimental feasibility study. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14136 [FREE Full text] [doi: 10.2196/14136] [Medline: 31237570]
50. Lozano-Lozano M, Melguizo-Rodríguez L, Fernández-Lao C, Galiano-Castillo N, Cantarero-Villanueva I, Martín-Martín L, et al. Association between the use of a mobile health strategy app and biological changes in breast cancer survivors: prospective pre-post study. J Med Internet Res 2019 Aug 14;21(8):e15062. [doi: 10.2196/15062]
51. Çınar D, Karadakovan A, Erdoğan AP. Effect of mobile phone app-based training on the quality of life for women with breast cancer. Eur J Oncol Nurs 2021 Jun;52:101960 [FREE Full text] [doi: 10.1016/j.ejon.2021.101960] [Medline: 33882446]
52. Saei Ghare Naz M, Simbar M, Rashidi Fakari F, Ghasemi V. Effects of model-based interventions on breast cancer screening behavior of women: a systematic review. Asian Pac J Cancer Prev 2018 Aug 24;19(8):2031-2041 [FREE Full text] [doi: 10.22034/APJCP.2018.19.8.2031] [Medline: 30139040]
53. Hong MK, Cho YY, Rha MY, Kim JH, Lee M. Six-month outcomes of mobile phone application-based self-management in a patient with type 2 diabetes. Clin Nutr Res 2015;4(3):201. [doi: 10.7762/cnr.2015.4.3.201]
54. Xu Z, Geng J, Zhang S, Zhang K, Yang L, Li J, et al. A mobile-based intervention for dietary behavior and physical activity change in individuals at high risk for type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 03;8(11):e19869 [FREE Full text] [doi: 10.2196/19869] [Medline: 33141092]
55. Koot D, Goh PS, Lim RS, Tian Y, Yau TY, Tan NC, et al. A mobile lifestyle management program (GlycoLeap) for people with type 2 diabetes: single-arm feasibility study. JMIR Mhealth Uhealth 2019 May 24;7(5):e12965. [doi: 10.2196/12965]
56. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 09;7(1):e11267. [doi: 10.2196/11267]
57. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth app intervention for persons with type 2 diabetes and its associations with initial self-management: randomized controlled trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125. [doi: 10.2196/mhealth.8824]
58. Alfonsi JE, Choi EE, Arshad T, Sammott SS, Pais V, Nguyen C, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e22074. [doi: 10.2196/22074]
59. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
60. Yu Y, Yan Q, Li H, Li H, Wang L, Wang H, et al. Effects of mobile phone application combined with or without self-monitoring of blood glucose on glycemic control in patients with diabetes: a randomized controlled trial. J Diabetes Investig 2019 Sep;10(5):1365-1371 [FREE Full text] [doi: 10.1111/jdi.13031] [Medline: 30815973]
61. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321. [doi: 10.2196/10321]
62. Lim SL, Ong KW, Johal J, Han CY, Yap QV, Chan YH, et al. Effect of a smartphone app on weight change and metabolic outcomes in Asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021 Jun 01;4(6):e2112417 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]
63. Koohmareh Z, Karandish M, Hadianfard AM. Effect of implementing a mobile game on improving dietary information in diabetic patients. Med J Islam Repub Iran 2021;35:68 [FREE Full text] [doi: 10.47176/mjiri.35.68] [Medline: 34277505]
64. Garnweidner-Holme L, Henriksen L, Torheim LE, Lukasse M. Effect of the pregnant+ smartphone app on the dietary behavior of women with gestational diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 4;8(11):e18614. [doi: 10.2196/18614]

65. Juan M, Charco J, García-García I, Mollá R. An augmented reality app to learn to interpret the nutritional information on labels of real packaged foods. Front. Comput. Sci 2019 Jun 20;1:1-16 [FREE Full text] [doi: 10.3389/fcomp.2019.00001]
66. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
67. McKibbin CL, Patterson TL, Norman G, Patrick K, Jin H, Roesch S, et al. A lifestyle intervention for older schizophrenia patients with diabetes mellitus: a randomized controlled trial. Schizophrenia Res 2006 Sep;86(1-3):36-44. [doi: 10.1016/j.schres.2006.05.010]
68. Patterson TL, Mausbach BT, McKibbin C, Goldman S, Bucardo J, Jeste DV. Functional Adaptation Skills Training (FAST): a randomized trial of a psychosocial intervention for middle-aged and older patients with chronic psychotic disorders. Schizophrenia Res 2006 Sep;86(1-3):291-299. [doi: 10.1016/j.schres.2006.05.017]
69. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014 May;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]
70. Pellegrini CA, Conroy DE, Phillips SM, Pfammatter AF, McFadden HG, Spring B. Daily and seasonal influences on dietary self-monitoring using a smartphone application. J Nutr Educ Behav 2018 Jan;50(1):56-61.e1 [FREE Full text] [doi: 10.1016/j.jneb.2016.12.004] [Medline: 29325663]
71. van der Pligt P, Ball K, Hesketh KD, Teychenne M, Crawford D, Morgan PJ, et al. J Hum Nutr Diet 2018 Jun;31(3):314-328. [doi: 10.1111/jhn.12521] [Medline: 29034545]
72. Bughin F, Bui G, Ayoub B, Blervaque L, Saey D, Avignon A, et al. Impact of a mobile telerehabilitation solution on metabolic health outcomes and rehabilitation adherence in patients with obesity: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 6;9(12):e28242. [doi: 10.2196/28242]
73. Ali HI, Attlee A, Alhebshi S, Elmi F, Al Dhaheri AS, Stojanovska L, et al. Feasibility study of a newly developed technology-mediated lifestyle intervention for overweight and obese young adults. Nutrients 2021 Jul 26;13(8):2547 [FREE Full text] [doi: 10.3390/nu13082547]
74. Dodd JM, Louise J, Cramp C, Grivell RM, Moran LJ, Deussen AR. Evaluation of a smartphone nutrition and physical activity application to provide lifestyle advice to pregnant women: the SNAPP randomised trial. Matern Child Nutr 2018 Jan;14(1):e12502 [FREE Full text] [doi: 10.1111/mcn.12502] [Medline: 28836373]
75. Lin P, Grambow S, Intille S, Gallis JA, Lazenka T, Bosworth H, et al. The association between engagement and weight loss through personal coaching and cell phone interventions in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 18;6(10):e10471 [FREE Full text] [doi: 10.2196/10471] [Medline: 30341051]
76. Chew CS, Davis C, Lim JK, Lim CM, Tan YZ, Oh JY, et al. Use of a mobile lifestyle intervention app as an early intervention for adolescents with obesity: single-cohort study. J Med Internet Res 2021 Sep 28;23(9):e20520. [doi: 10.2196/20520]
77. Stasinaki A, Büchter D, Shih CI, Heldt K, Güsewell S, Brogle B, et al. Effects of a novel mobile health intervention compared to a multi-component behaviour changing program on body mass index, physical capacities and stress parameters in adolescents with obesity: a randomized controlled trial. BMC Pediatr 2021 Jul 09;21(1):308 [FREE Full text] [doi: 10.1186/s12887-021-02781-2] [Medline: 34243738]
78. Toro-Ramos T, Lee D, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a smartphone application for the management of metabolic syndrome components focusing on weight loss: a preliminary study. Metab Syndr Relat Disord 2017 Dec;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
79. Hull P, Emerson JS, Quirk ME, Canedo JR, Jones JL, Vylegzhanina V, et al. A smartphone app for families with preschool-aged children in a public nutrition program: prototype development and beta-testing. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e102 [FREE Full text] [doi: 10.2196/mhealth.7477] [Medline: 28768611]
80. Prasad M, Fine K, Gee A, Nair N, Popp CJ, Cheng B, et al. A smartphone intervention to promote time restricted eating reduces body weight and blood pressure in adults with overweight and obesity: a pilot study. Nutrients 2021 Jun 23;13(7):2148 [FREE Full text] [doi: 10.3390/nu13072148]
81. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28. [doi: 10.2196/diabetes.8590]
82. Delisle Nyström C, Sandin S, Henriksson P, Henriksson H, Maddison R, Löf M. A 12-month follow-up of a mobile-based (mHealth) obesity prevention intervention in pre-school children: the MINISTOP randomized controlled trial. BMC Public Health 2018 May 24;18(1):658 [FREE Full text] [doi: 10.1186/s12889-018-5569-4] [Medline: 29793467]
83. Kay MC, Miller HN, Askew S, Spaulding EM, Chisholm M, Christy J, et al. Patterns of engagement with an application-based dietary self-monitoring tool within a randomized controlled feasibility trial. AJPM Focus 2022 Dec;1(2):100037 [FREE Full text] [doi: 10.1016/j.focus.2022.100037]
84. Griffin JB, Struempler B, Funderburk K, Parmer SM, Tran C, Wadsworth DD. My quest, an intervention using text messaging to improve dietary and physical activity behaviors and promote weight loss in low-income women. J Nutr Educ Behav 2018 Jan;50(1):11-8.e1 [FREE Full text] [doi: 10.1016/j.jneb.2017.09.007] [Medline: 29325657]
85. Matthews L, Pugmire J, Moore L, Kelson M, McConnachie A, McIntosh E, et al. Study protocol for the 'HelpMeDoIt!' randomised controlled feasibility trial: an app, web and social support-based weight loss intervention for adults with obesity. BMJ Open 2017 Oct 25;7(10):e017159. [doi: 10.1136/bmjopen-2017-017159]

86. Senecal C, Widmer RJ, Larrabee BR, de Andrade M, Lerman LO, Lerman A, et al. A digital health weight loss program in 250,000 individuals. J Obesity 2020 Mar 26;2020:1-8. [doi: 10.1155/2020/9497164]
87. Aschbrenner KA, Naslund JA, Shevenell M, Mueser KT, Bartels SJ. Feasibility of behavioral weight loss treatment enhanced with peer support and mobile health technology for individuals with serious mental illness. Psychiatr Q 2015 Oct 13;87(3):401-415. [doi: 10.1007/s11126-015-9395-x]
88. Silk JS, Pramana G, Sequeira SL, Lindhiem O, Kendall PC, Rosen D, et al. Using a smartphone app and clinician portal to enhance brief cognitive behavioral therapy for childhood anxiety disorders. Behav Ther 2020 Jan;51(1):69-84 [FREE Full text] [doi: 10.1016/j.beth.2019.05.002] [Medline: 32005341]
89. Drake C, Cannady M, Howley K, Shea C, Snyderman R. An evaluation of mHealth adoption and health self-management in emerging adulthood. AMIA Annu Symp Proc 2019;2019:1021-1030 [FREE Full text] [Medline: 32308899]
90. Karduck J, Chapman-Novakofski K. Results of the clinician apps survey, how clinicians working with patients with diabetes and obesity use mobile health apps. J Nutr Educ Behav 2018 Jan;50(1):62-9.e1 [FREE Full text] [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
91. Fakih El Khoury C, Karavetian M, Halfens RJ, Crutzen R, Khoja L, Schols JM. The effects of dietary mobile apps on nutritional outcomes in adults with chronic diseases: a systematic review and meta-analysis. J Acad Nutr Diet 2019 Apr;119(4):626-651 [FREE Full text] [doi: 10.1016/j.jand.2018.11.010] [Medline: 30686742]
92. Sarker MH, Moriyama M, Rashid HU, Rahman MM, Chisti MJ, Das SK, et al. Health education through a campaign and mHealth to enhance knowledge and quality of life among patients with chronic kidney disease in Bangladesh: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Nov 19;10(11):e30191 [FREE Full text] [doi: 10.2196/30191] [Medline: 34806998]
93. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mHealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e23477. [doi: 10.2196/23477]
94. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311. [doi: 10.2196/13311]
95. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors 2019 May 09;19(9):2164. [doi: 10.3390/s19092164]
96. Hong W, Lee WG. Wearable sensors for continuous oral cavity and dietary monitoring toward personalized healthcare and digital medicine. Analyst 2020;145(24):7796-7808. [doi: 10.1039/d0an01484b]
97. Sharma A, Badea M, Tiwari S, Marty JL. Wearable biosensors: an alternative and practical approach in healthcare and disease monitoring. Molecules 2021 Feb 01;26(3):748. [doi: 10.3390/molecules26030748]
98. Miley D, Machado LB, Condo C, Jergens AE, Yoon K, Pandey S. Video capsule endoscopy and ingestible electronics: emerging trends in sensors, circuits, materials, telemetry, optics, and rapid reading software. Advanced Devices Instrumentation 2021 Nov 10;2021:1-30. [doi: 10.34133/2021/9854040]
99. Islam MN, Karim MM, Inan TT, Islam AK. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020 Feb 03;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]
100. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013 Dec 05;2013(12):CD007458 [FREE Full text] [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
101. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
102. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]

Abbreviations

BCT: behavior change technique
CVD: cardiovascular disease
DM: diabetes mellitus
HbA_{1c}: glycated hemoglobin
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 19.07.22; peer-reviewed by S Pandey, KS Sahu, H Mehdizadeh; comments to author 30.09.22; revised version received 15.11.22; accepted 30.11.22; published 13.01.23.

Please cite as:
Salas-Groves E, Galyean S, Alcorn M, Childress A
Behavior Change Effectiveness Using Nutrition Apps in People With Chronic Diseases: Scoping Review
JMIR Mhealth Uhealth 2023;11:e41235
URL: https://mhealth.jmir.org/2023/1/e41235
doi:10.2196/41235
PMID:36637888

Review

Effectiveness of mHealth on Adherence to Antiretroviral Therapy in Patients Living With HIV: Meta-analysis of Randomized Controlled Trials

Liang Sun[1], DPH; Mengbing Qu[1], MPH; Bing Chen[2], MPH; Chuancang Li[1], MPH; Haohao Fan[1], MPH; Yang Zhao[1], DPH

[1]College of Public Health, Zhengzhou University, Zhengzhou, China
[2]Sanmenxia Center for Disease Control and Prevention, Sanmenxia, China

Corresponding Author:
Yang Zhao, DPH
College of Public Health
Zhengzhou University
New campus of Zhengzhou University
100 Science Avenue
Zhengzhou, 450001
China
Phone: 86 13803845359
Email: zhaomiemie@126.com

Abstract

Background: The World Health Organization recommends that all adults with HIV adhere to antiretroviral therapy (ART). Good adherence to ART is beneficial to patients and the public. Furthermore, mHealth has shown promise in improving HIV medication adherence globally.

Objective: The aim of this meta-analysis is to analyze the effectiveness of mHealth on adherence to antiretroviral therapy in patients living with HIV.

Methods: Randomized controlled trials (RCTs) of the association between mHealth and adherence to ART published until December 2021 were searched in electronic databases. Odds ratios (ORs), weighted mean differences, and 95% CIs were calculated. This meta-analysis was performed using the Mantel-Haenszel method or the inverse variance test. We evaluated heterogeneity with the I^2 statistic. If I^2 was ≤50%, heterogeneity was absent, and a fixed effect model was used. If I^2 was >50%, heterogeneity was present, and a random effects model was used.

Results: A total of 2163 participants in 8 studies were included in this meta-analysis. All included studies were RCTs. The random effects model was used for a meta-analysis of the effects of various intervention measures compared to routine nursing; the outcome was not statistically significant (OR 1.54, 95% CI 0.99-2.38; *P*=.05). In the subgroups, only short messaging service (SMS)-based interventions significantly increased adherence to ART (OR 1.76, 95% CI 1.07-2.89; *P*=.03). Further analysis showed that only interactive or bidirectional SMS could significantly increase ART adherence (OR 1.69, 95% CI 1.22-2.34; *P*=.001). After combining the difference in CD4 cell count before and after the interventions, we concluded that there was no statistical heterogeneity among the studies (I^2=0%; tau^2=0.37; *P*=.95).

Conclusions: Interactive or bidirectional SMS can enhance intervention effects. However, whether mHealth can improve adherence to ART in patients with HIV needs further study. Owing to a lack of the required significant staff time, training, and ongoing supervision, there is still much more to do to apply mHealth to the clinical use of ART for patients living with HIV.

Trial Registration: PROSPERO CRD42022358774; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=358774

(JMIR Mhealth Uhealth 2023;11:e42799) doi:10.2196/42799

KEYWORDS

HIV; mHealth; antiretroviral therapy; meta-analysis

Introduction

HIV is a public health issue that every country needs to address. There were an estimated 37.7 million people living with HIV at the end of 2020 [1]. The World Health Organization (WHO) recommends that all adults with HIV should adhere to antiretroviral therapy (ART) [2]. ART does not cure HIV infection but strongly suppresses viral replication within a person's body and modifies HIV from a terminal illness to a manageable chronic disease [3]. One of the most significant factors in the effectiveness of ART is adherence [4]. Good adherence to ART is beneficial to patients and public health [5,6]. In contrast, lack of adherence increases the risk of progression to AIDS and the creation of drug-resistant strains of HIV [7]. Therefore, it is important to promote ART adherence through special techniques [8,9]. However, traditional ART adherence interventions are limited in their ability to maintain behavior modification [10]. Mobile health (mHealth) technology, which refers to the use of mobile and wireless technologies to improve health, has shown promise in improving HIV medication adherence, both globally and domestically [6,11,12]. Therefore, we performed a meta-analysis of the effectiveness of mHealth for improving adherence to ART in patients living with HIV.

Methods

Ethical Considerations

This paper contains no primary data obtained directly from research participants. Data obtained from previously published resources have been acknowledged with references. Ethical approval was not required.

Protocol Registration

The review protocol was prospectively registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42022358774).

Search Strategy

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines [13]. Searching was conducted using the electronic databases PubMed, EMbase, CINAHL, ScienceDirect, the Cochrane Library, Web of Science, and ClinicalTrials.gov to identify original articles meeting the evaluation criteria for inclusion and exclusion published until December 2021. Searching was conducted and evaluated by 2 independent reviewers. The search strategy for identifying studies included the key terms *mobile health*, *human immunodeficiency virus*, *medication adherence*, *randomized controlled trial*, and other related terms. These keywords were also combined using the OR and AND operators (Multimedia Appendix 1).

Inclusion and Exclusion Criteria

The screening process was divided into 2 phases: a preliminary selection by title and abstract and a second phase that screened the full text of the remaining articles. Articles were included based on the following criteria: (1) they reported the results of a randomized controlled trial (RCT); (2) they included HIV-positive persons receiving antiretroviral treatment regardless of age, sex, or nationality; (3) the intervention measures included, but were not limited to, short message service (SMS) texts and voice calls; (4) an mHealth intervention was used in the experimental group with no limits on intervention frequency, time, content, or period and the control group received routine nursing at the same time to help patients improve their treatment compliance; and (5) the primary outcome was adherence to ART. This was measured directly (eg, by pill count) or indirectly (eg, by self-reporting). If the article reported the use of a variety of measurement methods, priority was given to measurement results obtained with the self-report method. The secondary outcome was CD4 cell count.

The exclusion criteria included the following: (1) the study was a duplicate, (2) the study was a systematic review or meta-analysis, (3) the study was missing outcome measures, (4) the experimental group used a variety of interventions, and (5) the control group did not receive routine nursing.

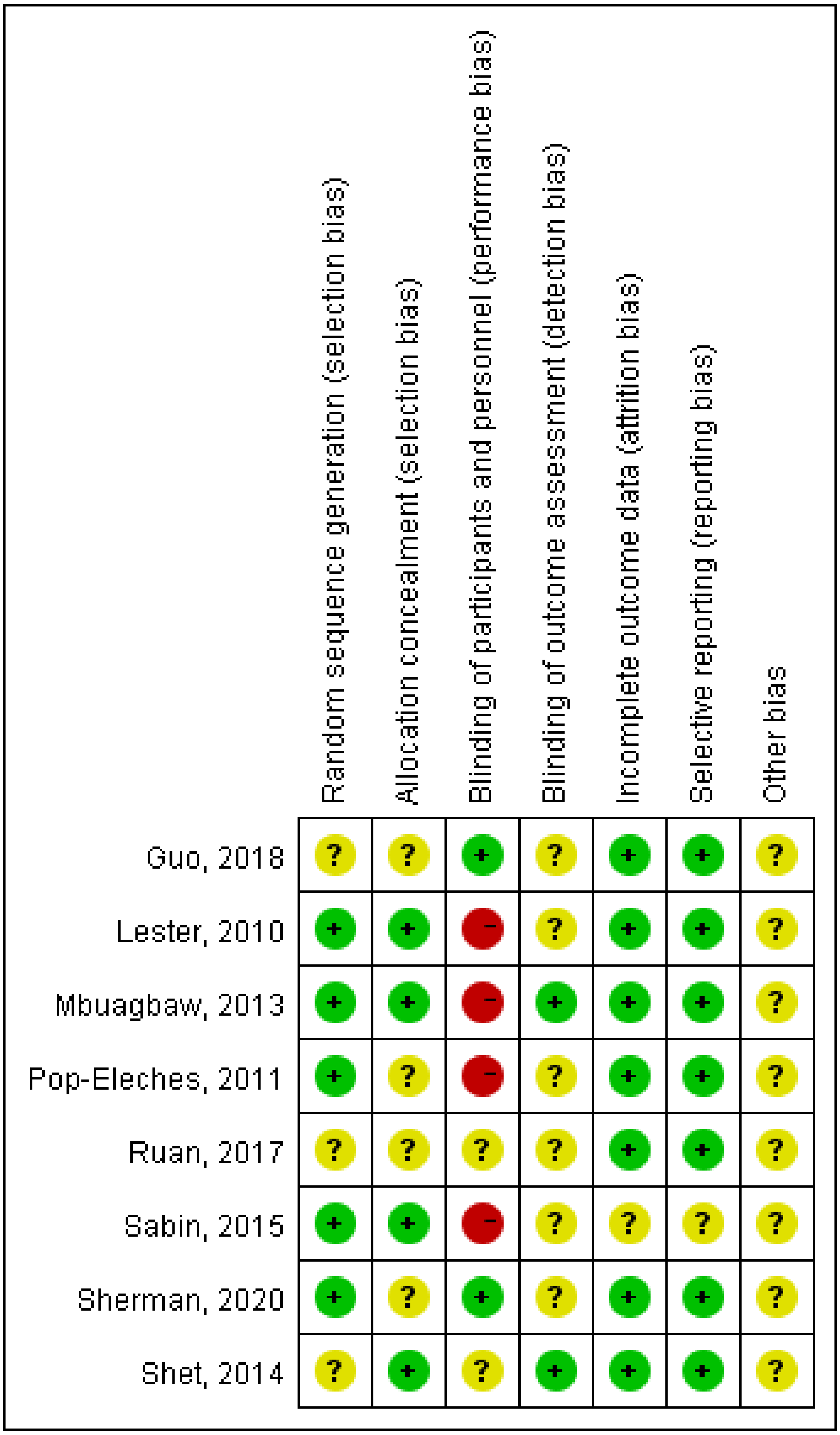
Data Extraction

A predesigned Excel sheet was used to extract and organize the data into categories by 2 independent researchers. These data included (1) authors, (2) location, (3) publication date, (4) intervention details (ie, intervention mode and duration), (5) outcome measures, including ART adherence and CD4 cell counts, and (6) risk of bias.

Risk of Bias and Quality Assessment

Two of the authors assessed the risk of bias using RevMan (version 5.4; Cochrane Collaboration); the results are summarized in Figure 1. The Cochrane Collaboration's Risk of Bias tool was also used to assess quality and risk of bias. This tool assesses bias in 7 domains: random sequence generation (for selection bias), concealment of the allocation sequence (for selection bias), blinding of participants and personnel (for performance bias), blinding of outcome assessment (for detection bias), incomplete outcome data (for attrition bias), selective outcome reporting (for reporting bias), and other biases. Studies are assigned a low risk of bias, an unclear risk of bias, or a high risk of bias.

Figure 1. Risk of bias of the studies included in the meta-analysis. Green indicates a low risk of bias, yellow indicates an unclear risk of bias, and red indicates a high risk of bias.



Data Analysis

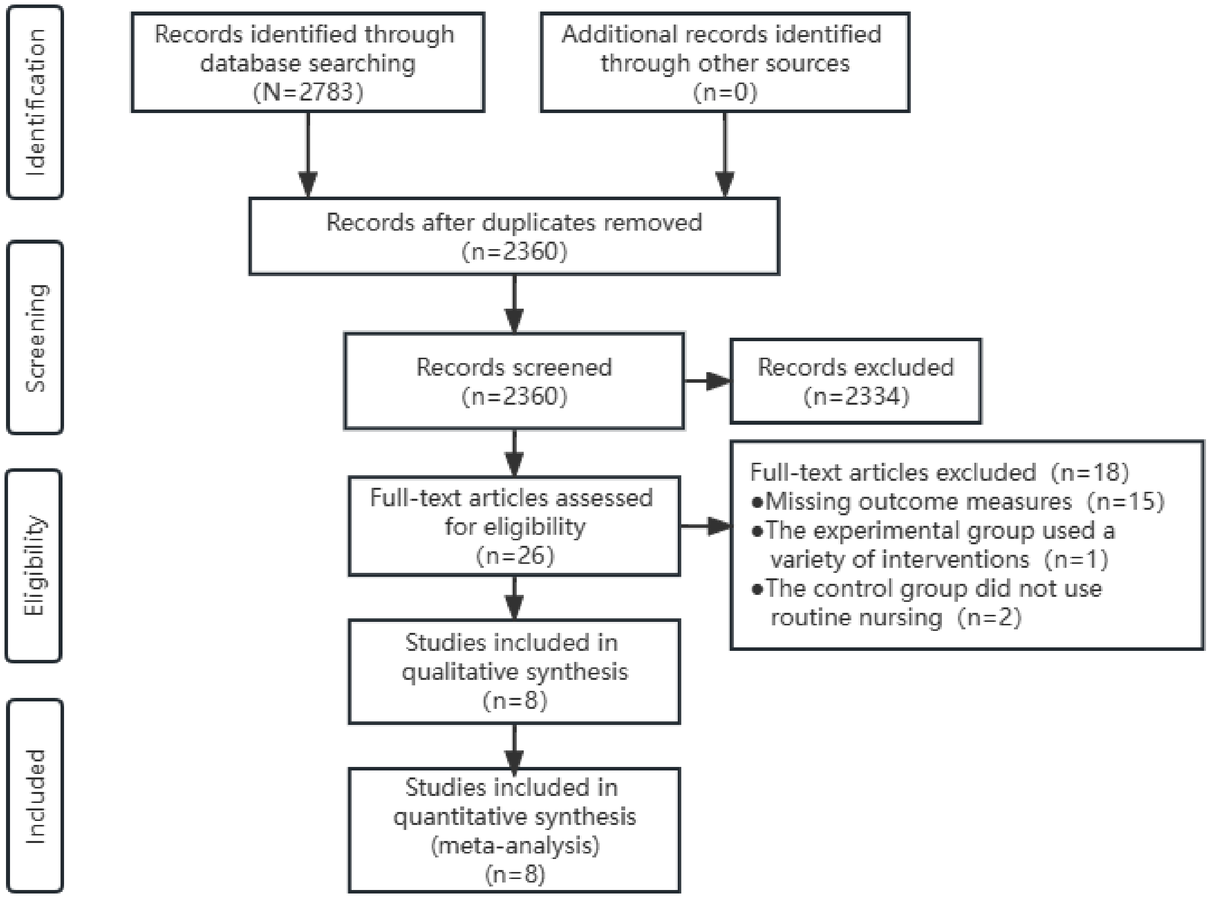
The meta-analysis was conducted using RevMan. Measures of effect are presented as odds ratios (ORs) with the 95% CI. For continuous data, we calculated the sample-size weighted mean difference (WMD). This meta-analysis was performed using the Mantel-Haenszel method or the inverse variance test. We evaluated heterogeneity with the I^2 statistic. If I^2 was ≤50%, heterogeneity was absent, and a fixed effect model was used. If I^2 was >50%, heterogeneity was present, and a random effects model was used.

Results

Study Selection Process

The search strategy identified 2783 articles from the electronic databases. In total, 423 articles were excluded because of duplication. We screened the titles and abstracts of the remaining articles and included 26 for full-text review based on the inclusion and exclusion criteria. Of these 26 studies, 8 met the inclusion criteria, and 18 studies were excluded: 15 because they were missing outcome measures, 1 because a variety of interventions were used in the experimental group, and 2 because they did not use routine nursing in the control group. Therefore, 8 studies were selected for the current meta-analysis [4,9,14-19] (Figure 2).

Figure 2. Flow chart of study selection for the meta-analysis.



Study Characteristics

The characteristics of the studies are summarized in Table 1. All included studies were RCTs. A total of 2163 participants in 8 studies were included in this meta-analysis. Except for the study by Mbuagbaw et al [19], all participants in the studies were aged 18 years or older. Study duration ranged between 1 and 12 months. SMS was used as the basis for the intervention in 6 studies. One of these studies used an mHealth intervention program that included text messages and WeChat. The remaining studies used voice calls.

Table 1. Characteristics of the included studies. All studies used routine nursing in the control group.

First author, year	Location	Participants, n (age, years)	Recruitment period (duration)	Intervention	Outcome measures
Ruan, 2017 [4]	Hengyang, China	100 (≥18)	Mar 2013-Mar 2014 (6 months)	SMS[a]	ART[b] adherence measured by VAS[c], Community Programs for Clinical Research on AIDS Antiretroviral Medication Self-Report, and CD4 cell count
Guo, 2018 [14]	South China	53 (≥18)	Oct 2016-Mar 2017 (3 months)	SMS and WeChat	CD4 cell count
Shet, 2014 [15]	South India	631 (18-60)	Jul 2010-Aug 2011 (6 months)	Customized motivational voice call	ART adherence measured by pill count
Sherman, 2020 [16]	South Florida, US	94 (>18)	Sept 2011-Apr 2014 (3-6 months)	Automated 1-way medication reminders delivered via SMS	ART adherence measured by VAS and CD4 cell count
Lester, 2010 [17]	Kenya	538 (>18)	May 2007-Oct 2008 (12 months)	SMS	ART adherence measured by self-report
Sabin, 2015 [9]	Guangxi, China	119 (≥18)	Dec 2012-Apr 2013 (6 months)	SMS	ART adherence measured by a medication device
Pop-Eleches, 2011 [18]	Kenya	428 (>18)	Jun 2007-Aug 2008 (3-12 months)	SMS	ART adherence measured by a medication event monitoring system and CD4 cell count
Mbuagbaw, 2013 [19]	Cameroon	200 (≥21)	Nov 2010-Dec 2010 (3-6 months)	Weekly standardized motivational text message	VAS and self-reported adherence

[a]SMS: short message service.

[b]ART: antiretroviral therapy.

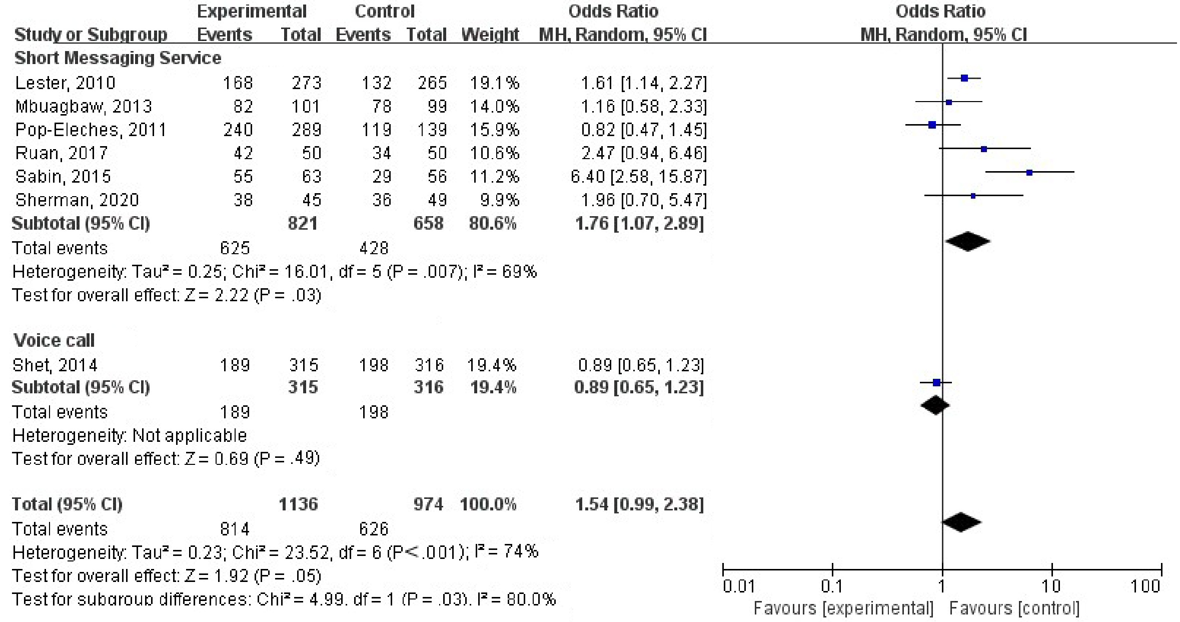
[c]VAS: visual analog scale.

Meta-analysis

Medication Adherence

Adherence to ART was measured as a primary outcome in 7 studies. The method and frequency of measuring adherence varied across the studies. The details are listed in Table 1. A random effects model was used for a meta-analysis of the effects of various intervention measures and routine nursing; the outcome was not statistically significant (OR 1.54, 95% CI 0.99-2.38; *P*=.05). There was also evidence of heterogeneity among the studies (I^2=74%; tau^2=0.23; *P*<.001; Figure 3). In the subgroups, only SMS interventions significantly increased adherence to ART (OR 1.76, 95% CI 1.07-2.89; *P*=.03).

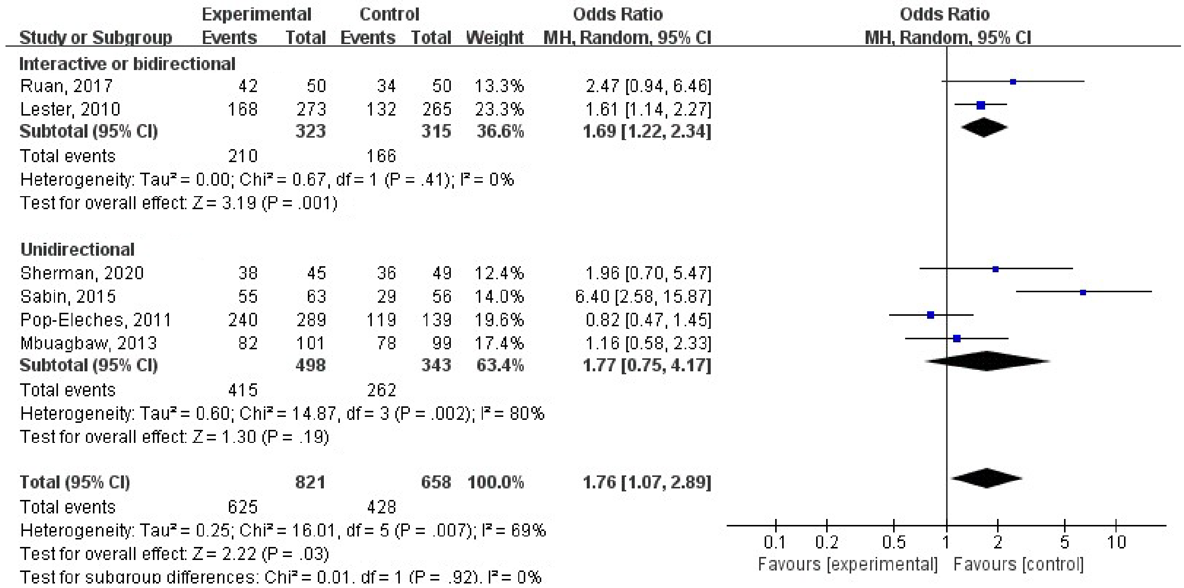
Figure 3. Forest plot of odds ratios with 95% CIs for the effect of various intervention measures and routine nursing on adherence to antiretroviral therapy. MH: Mantel-Haenszel method.



Medication Adherence With SMS Intervention

Six studies were included in a meta-analysis of the effect of SMS on adherence to ART; this showed that SMS could improve adherence (OR 1.76, 95% CI 1.07-2.89; *P*=.03) but also revealed considerable heterogeneity among the included studies (I^2=69%; tau^2=0.25; *P*=.007; Figure 4). In the subgroups, only interactive or bidirectional SMS interventions could significantly increase ART adherence (OR 1.69, 95% CI 1.22-2.34; *P*=.001). However, these studies did not show statistical heterogeneity (I^2=0%; tau^2=0.00; *P*=.41). A subgroup analysis of studies with unidirectional SMS interventions showed no statistical heterogeneity, but the analysis also showed a lack of effect in improving adherence (Figure 4).

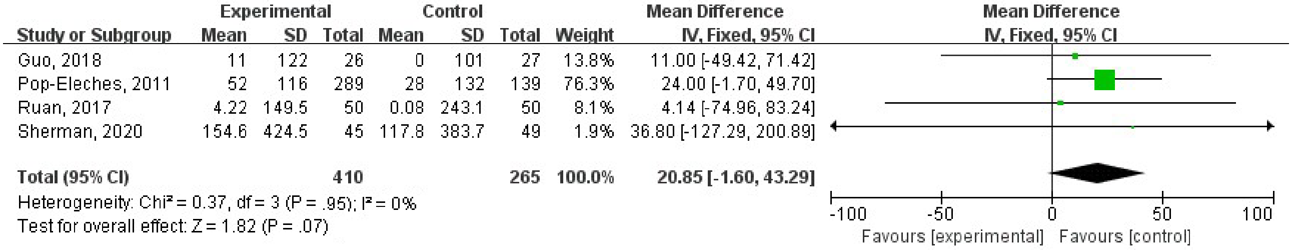
Figure 4. Forest plot of odds ratios with 95% CIs for the effect of short message service interventions on adherence to antiretroviral therapy. MH: Mantel-Haenszel.



CD4 Cell Count

Four studies reported CD4 cell count as a secondary outcome of medication adherence. Combining the differences in CD4 cell count before and after the interventions revealed no statistical heterogeneity among the studies (I^2=0%; tau^2=0.37; *P*=.95; Figure 5). Meta-analysis showed that CD4 cell count measures after mHealth interventions revealed no significant difference in medication adherence among HIV patients compared with routine nursing (WMD=20.85, 95% CI 1.60-43.29; *P*=.07; Figure 5). Only 2 studies reported viral load as an outcome to evaluate the intervention effect, so we did not perform a meta-analysis.

Figure 5. Forest plot of pooled odds ratios with 95% CIs for the effect of mHealth interventions on CD4 cell count. IV: inverse variance.



Discussion

Principal Findings

A total of 2163 participants in 8 studies were included in this meta-analysis. The main result of the meta-analysis was that the pooled OR was 1.54. However, the outcome was not statistically significant, and there was considerable heterogeneity among the studies (I^2=74%). Within-study heterogeneity reduces study robustness and relevance [20]. Our current results are different from those of a 2015 study [21]. That study found that mHealth interventions did seem to have been beneficial. We speculate that one of the possible reasons for this difference is that the number of articles included in our study was limited. Also, the different studies used a variety of methods to measure results. However, compared with the 2015 study, this meta-analysis combined the outcomes (separately for primary and secondary outcomes) to increase the reliability of the results.

A subgroup analysis showed that SMS interventions improved adherence to ART. Further analysis suggested that interactive or bidirectional SMS interventions could enhance intervention effects. This result matches that of a 2014 study [22]. Interactive or bidirectional SMS could increase medication adherence by enhancing engagement and the patient-provider relationship. Follow-up research could further study the content, time, and frequency of text messages. In this study, due to limitations arising from incomplete reporting of the relevant content, these aspects of the intervention were not analyzed.

CD4 cell count and viral load are good indicators of treatment success [23]. Several previous reviews did not obtain the same results as this study for CD4 cell count. One study [22] showed that compared to a control group, a group receiving text message–based support was more likely to maintain an adherence threshold at follow-up and meet the clinical goal of a higher CD4 cell count. However, another study [24], like ours, did not obtain this result. Our meta-analysis of the 4 studies that reported CD4 cell count showed considerable heterogeneity and no significant pooled mean difference. This may be attributable to the heterogeneity of the studies. We did not perform a meta-analysis of studies that measured viral load, as there was an insufficient number of these studies.

The WHO recommends mHealth strategies for improving ART adherence [25]. At the same time, against the background of the coronavirus epidemic, telemedicine has gradually continued to develop [26]. Currently, 95% of people use mobile phones and 77% of people use smartphones in different parts of the world [27]; mHealth has shown promise in increasing the accessibility of self-management interventions [28] and improving HIV health outcomes [6]. At the same time, research has indicated that patients living with HIV are interested in mobile apps to support HIV self-management [29]. However, most current mHealth interventions lack functionality, offering only medication reminders [28] or voice calls. Therefore, more comprehensive mHealth interventions that address multiple self-management needs of patients living with HIV are needed [30]. At the same time, the implementation of ART interventions in real-world clinical settings has been severely limited by a lack of the resources required to initiate and maintain the interventions [31], such as staff time, training, and ongoing supervision. Future research should focus on how to apply personalized mHealth interventions to the management of patients living with HIV.

Limitations

There are several limitations of the current review. The interventions used in the included studies differed in form and frequency. At the same time, these studies used diverse methods for measuring their primary outcomes. This may have produced bias. The robustness and relevance of results increase with the number of distinct outcome measures that show the same result [20]. However, few of the included studies reported CD4 cell count or viral load, and our analysis of these outcomes is thus insufficient.

Conclusion

Interactive or bidirectional SMS interventions can enhance intervention effects. However, whether mHealth can improve adherence to ART in patients with HIV is a question that needs further study. Owing to a lack of staff time, training, and ongoing supervision, there is still much work to be done to use mHealth in the clinic for ART adherence among patients living with HIV.

Acknowledgments

This work has been partly supported by the Program for Youth Backbone Teacher Training in University of Henan Province (2019GGJS021), and Science and Technology Project in Henan province in 2018 (182102310201).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords.
[DOCX File , 46 KB - mhealth_v11i1e42799_app1.docx]

References

1. HIV/AIDS: key facts. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids [accessed 2022-07-20]
2. Nachega J, Skinner D, Jennings L, Magidson J, Altice F, Burke J, et al. Acceptability and feasibility of mHealth and community-based directly observed antiretroviral therapy to prevent mother-to-child HIV transmission in South African pregnant women under Option B+: an exploratory study. Patient Prefer Adherence 2016;10:683-690 [FREE Full text] [doi: 10.2147/PPA.S100002] [Medline: 27175068]
3. Chory A, Callen G, Nyandiko W, Njoroge T, Ashimosi C, Aluoch J, et al. A pilot study of a mobile intervention to support mental health and adherence among adolescents living with HIV in western Kenya. AIDS Behav 2022 Jan;26(1):232-242 [FREE Full text] [doi: 10.1007/s10461-021-03376-9] [Medline: 34292429]
4. Ruan Y, Xiao X, Chen J, Li X, Williams AB, Wang H. Acceptability and efficacy of interactive short message service intervention in improving HIV medication adherence in Chinese antiretroviral treatment-naïve individuals. Patient Prefer Adherence 2017;11:221-228 [FREE Full text] [doi: 10.2147/PPA.S120003] [Medline: 28228652]
5. De Costa A, Shet A, Kumarasamy N, Ashorn P, Eriksson B, Bogg L, HIVIND study team. Design of a randomized trial to evaluate the influence of mobile phone reminders on adherence to first line antiretroviral treatment in South India--the HIVIND study protocol. BMC Med Res Methodol 2010 Mar 26;10:25 [FREE Full text] [doi: 10.1186/1471-2288-10-25] [Medline: 20346136]
6. Duthely LM, Sanchez-Covarrubias AP, Mohamed AB, Potter JE. A multilingual, culturally competent mobile health intervention to improve treatment adherence among women living with hiv: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Jun 19;9(6):e17656 [FREE Full text] [doi: 10.2196/17656] [Medline: 32438338]
7. DeSilva M, Vu CN, Bonawitz R, Hai LT, Van Lam N, Yen LT, et al. The Supporting Adolescent Adherence in Vietnam (SAAV) study: study protocol for a randomized controlled trial assessing an mHealth approach to improving adherence for adolescents living with HIV in Vietnam. Trials 2019 Feb 28;20(1):150 [FREE Full text] [doi: 10.1186/s13063-019-3239-1] [Medline: 30819228]
8. Baranoski AS, Meuser E, Hardy H, Closson EF, Mimiaga MJ, Safren SA, et al. Patient and provider perspectives on cellular phone-based technology to improve HIV treatment adherence. AIDS Care 2014 Jan;26(1):26-32 [FREE Full text] [doi: 10.1080/09540121.2013.802282] [Medline: 23742640]
9. Sabin L, Bachman DeSilva M, Gill C, Zhong L, Vian T, Xie W, et al. Improving adherence to antiretroviral therapy with triggered real-time text message reminders: the China Adherence Through Technology Study. J Acquir Immune Defic Syndr 2015 Aug 15;69(5):551-559 [FREE Full text] [doi: 10.1097/QAI.0000000000000651] [Medline: 25886927]
10. Ramsey SE, Ames EG, Uber J, Habib S, Clark S, Waldrop D. A preliminary test of an mHealth facilitated health coaching intervention to improve medication adherence among persons living with HIV. AIDS Behav 2021 Nov;25(11):3782-3797 [FREE Full text] [doi: 10.1007/s10461-021-03342-5] [Medline: 34117965]
11. Fuchs JD, Stojanovski K, Vittinghoff E, McMahan VM, Hosek SG, Amico KR, et al. A mobile health strategy to support adherence to antiretroviral preexposure prophylaxis. AIDS Patient Care STDS 2018 Mar;32(3):104-111 [FREE Full text] [doi: 10.1089/apc.2017.0255] [Medline: 29565183]
12. Martin CA, Upvall MJ. A mobile phone HIV medication adherence intervention: Acceptability and feasibility study. J Assoc Nurses AIDS Care 2016;27(6):804-816. [doi: 10.1016/j.jana.2016.07.002] [Medline: 27497514]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (text message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]
15. Shet A, De Costa A, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, HIVIND Study Team. Effect of mobile telephone reminders on treatment outcome in HIV: evidence from a randomised controlled trial in India. BMJ 2014 Oct 24;349:g5978 [FREE Full text] [doi: 10.1136/bmj.g5978] [Medline: 25742320]
16. Sherman EM, Niu J, Elrod S, Clauson KA, Alkhateeb F, Eckardt P. Effect of mobile text messages on antiretroviral medication adherence and patient retention in early HIV care: an open-label, randomized, single center study in south Florida. AIDS Res Ther 2020 May 13;17(1):16 [FREE Full text] [doi: 10.1186/s12981-020-00275-2] [Medline: 32404130]

17. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
18. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
19. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]
20. Bärnighausen T, Chaiyachati K, Chimbindi N, Peoples A, Haberer J, Newell M. Interventions to increase antiretroviral adherence in sub-Saharan Africa: a systematic review of evaluation studies. Lancet Infect Dis 2011 Dec;11(12):942-951 [FREE Full text] [doi: 10.1016/S1473-3099(11)70181-5] [Medline: 22030332]
21. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
22. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
23. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: a new health care mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
24. Mehra N, Tunje A, Hallström IK, Jerene D. Effectiveness of mobile phone text message reminder interventions to improve adherence to antiretroviral therapy among adolescents living with HIV: A systematic review and meta-analysis. PLoS One 2021;16(7):e0254890 [FREE Full text] [doi: 10.1371/journal.pone.0254890] [Medline: 34293033]
25. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach, 2nd ed. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241549684 [accessed 2022-07-20]
26. Ameli V, Haberer J, Sabin L, Meinck F, Barlow J, Taj L, et al. Tailored mHealth intervention for improving treatment adherence for people living with HIV in Iran (HamRaah): protocol for a feasibility study and randomised pilot trial with a nested realist evaluation. BMJ Open 2021 Jun 23;11(6):e042296 [FREE Full text] [doi: 10.1136/bmjopen-2020-042296] [Medline: 34162631]
27. Lee SB, Valerius J. mHealth interventions to promote anti-retroviral adherence in HIV: narrative review. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e14739 [FREE Full text] [doi: 10.2196/14739] [Medline: 32568720]
28. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
29. Lucero RJ, Frimpong JA, Fehlberg EA, Bjarnadottir RI, Weaver MT, Cook C, et al. The relationship between individual characteristics and interest in using a mobile phone app for HIV self-management: observational cohort study of people living with HIV. JMIR Mhealth Uhealth 2017 Jul 27;5(7):e100 [FREE Full text] [doi: 10.2196/mhealth.7853] [Medline: 28751298]
30. Ramsey S, Ames E, Uber J, Habib S, Clark S. A mobile health app to improve HIV medication adherence: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Nov 13;8(11):e15356 [FREE Full text] [doi: 10.2196/15356] [Medline: 31719030]
31. Claborn KR, Fernandez A, Wray T, Ramsey S. Computer-based HIV adherence promotion interventions: a systematic review. Transl Behav Med 2015 Sep;5(3):294-306 [FREE Full text] [doi: 10.1007/s13142-015-0317-0] [Medline: 26327935]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMS: short message service
VAS: visual analog scale
WMD: weighted mean difference

Edited by L Buis; submitted 19.09.22; peer-reviewed by E Sanders, A Bwanika Naggirinya; comments to author 18.11.22; revised version received 08.12.22; accepted 21.12.22; published 23.01.23.

Please cite as:
Sun L, Qu M, Chen B, Li C, Fan H, Zhao Y
Effectiveness of mHealth on Adherence to Antiretroviral Therapy in Patients Living With HIV: Meta-analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2023;11:e42799
URL: https://mhealth.jmir.org/2023/1/e42799
doi:10.2196/42799
PMID:36689267

Review

mHealth Intervention for Improving Pain, Quality of Life, and Functional Disability in Patients With Chronic Pain: Systematic Review

Marta Moreno-Ligero[1,2], MSc; Jose A Moral-Munoz[2,3,4], PhD; Alejandro Salazar[2,4,5], PhD; Inmaculada Failde[1,2,4], PhD

[1]Preventive Medicine and Public Health Area, Department of Biomedicine, Biotechnology and Public Health, University of Cádiz, Cádiz, Spain
[2]Observatory of Pain, University of Cádiz, Cádiz, Spain
[3]Department of Nursing and Physiotherapy, University of Cádiz, Cádiz, Spain
[4]Institute of Research and Innovation in Biomedical Sciences of the Province of Cadiz (INiBICA), Cádiz, Spain
[5]Department of Statistics and Operational Research, University of Cádiz, Cádiz, Spain

Corresponding Author:
Jose A Moral-Munoz, PhD
Department of Nursing and Physiotherapy
University of Cádiz
Avda. Ana de Viya, 52
Cádiz, 11009
Spain
Phone: 34 956015699
Email: joseantonio.moral@uca.es

Abstract

Background: Chronic pain (CP) is 1 of the leading causes of disability worldwide and represents a significant burden on individual, social, and economic aspects. Potential tools, such as mobile health (mHealth) systems, are emerging for the self-management of patients with CP.

Objective: A systematic review was conducted to analyze the effects of mHealth interventions on CP management, based on pain intensity, quality of life (QoL), and functional disability assessment, compared to conventional treatment or nonintervention.

Methods: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines were followed to conduct a systematic review of randomized controlled trials (RCTs) published in PubMed, Web of Science, Scopus, and Physiotherapy Evidence Database (PEDro) databases from February to March 2022. No filters were used. The eligibility criteria were RCTs of adults (≥18 years old) with CP, intervened with mHealth systems based on mobile apps for monitoring pain and health-related outcomes, for pain and behavioral self-management, and for performing therapeutic approaches, compared to conventional treatments (physical, occupational, and psychological therapies; usual medical care; and education) or nonintervention, reporting pain intensity, QoL, and functional disability. The methodological quality and risk of bias (RoB) were assessed using the Checklist for Measuring Quality, the Oxford Centre for Evidence-Based Medicine Levels of Evidence, and the Cochrane RoB 2.0 tool.

Results: In total, 22 RCTs, involving 2641 patients with different CP conditions listed in the *International Classification of Diseases 11th Revision* (ICD-11), including chronic low back pain (CLBP), chronic musculoskeletal pain (CMSP), chronic neck pain (CNP), unspecified CP, chronic pelvic pain (CPP), fibromyalgia (FM), interstitial cystitis/bladder pain syndrome (IC/BPS), irritable bowel syndrome (IBS), and osteoarthritis (OA). A total of 23 mHealth systems were used to conduct a variety of CP self-management strategies, among which monitoring pain and symptoms and home-based exercise programs were the most used. Beneficial effects of the use of mHealth systems in reducing pain intensity (CNP, FM, IC/BPS, and OA), QoL (CLBP, CNP, IBS, and OA), and functional disability (CLBP, CMSP, CNP, and OA) were found. Most of the included studies (18/22, 82%) reported medium methodological quality and were considered as highly recommendable; in addition, 7/22 (32%) studies had a low RoB, 10/22 (45%) had some concerns, and 5/22 (23%) had a high RoB.

Conclusions: The use of mHealth systems indicated positive effects for pain intensity in CNP, FM, IC/BPS, and OA; for QoL in CLBP, CNP, IBS, and OA; and for functional disability in CLBP, CMSP, CNP, and OA. Thus, mHealth seems to be an alternative to improving pain-related outcomes and QoL and could be part of multimodal strategies for CP self-management. High-quality studies are needed to merge the evidence and recommendations of the use of mHealth systems for CP management.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42022315808; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=315808

(JMIR Mhealth Uhealth 2023;11:e40844) doi:10.2196/40844

KEYWORDS

chronic pain; mHealth; mobile health; mobile app; health app; digital intervention; monitoring; pain intensity; quality of life; functionality; disability; disabilities; systematic review; review methodology; search strategy; library science; RCT; randomized controlled trial; pain; health outcome; self-management

Introduction

Chronic pain (CP) is a leading cause of disability worldwide [1], affecting approximately 20% of the global population [2]. Moreover, in developed countries, up to 1 of 5 adults suffers from CP of any type [3]. This condition implies a substantial burden for people, and it also has a social and economic impact on health care systems and employment activity [2]. In fact, although the direct health care costs of managing CP conditions are important, the indirect costs, such as disability compensation and work absenteeism, are higher [4].

CP is defined as pain that persists or recurs for longer than 3 months, including a broad range of pain conditions collected in the *International Classification of Diseases 11th Revision* (ICD-11) [5]. It is a new and pragmatic classification system to apply in primary care and clinical settings for specialized pain management [6]. Current pain management interventions are based on multimodal and biopsychosocial models, which include pain education programs, exercise programs, cognitive and behavioral strategies, relaxation techniques, goal setting strategies, self-monitoring symptoms, and self-tailoring strategies [7-9]. Moreover, emotional distress, functional disability, and sleep disturbances are closely linked to the perception of pain and the pain-related outcomes in patients with CP [10,11]. Therefore, strategies for CP management should address all biopsychosocial aspects of this health condition.

Recently, innovative and potential alternatives to support the self-management of patients with CP have emerged, such as mobile device–based health care, or mobile health (mHealth) [12]. mHealth involves the practice of medicine and public health based on mobile devices to improve and promote health status [13]. According to the target of mHealth systems in CP, they can be grouped into 3 categories [12,14]: (1) education, including general information about pain, symptom identification, and treatment planning; (2) monitoring, tracking daily pain episodes and severity, symptoms, mood, activity, and medication use; and (3) treatment, involving several management strategies. These systems empower patients to become more engaged and encourage self-management [15], improving some pain-related outcomes. In line with this, several pain-related apps have been identified from scientific databases and app stores for the management of a wide range of pain (chronic and acute) conditions [16,17]. Nevertheless, there is a lack of scientific and health professional support in many of the mHealth systems, highlighting the need for developing appropriate apps based on the patient's requirements, also in the management of CP [18].

The available evidence points out promising effects of internet-delivered interventions on different biopsychosocial aspects of CP. Gandy et al [19] studied the use of these interventions using any type of device and technology for CP, showing small effects on pain intensity and disability outcomes in patients with mixed CP conditions, chronic low back pain (CLBP), fibromyalgia (FM), arthritic conditions, peripheral neuropathy, spinal cord injury, migraine, and chronic pancreatitis. In a similar vein, Moman et al [14] discussed the effects of both electronic health (eHealth), based on web apps, and mHealth technologies in patients with CP (general CP, CLBP, FM, and osteoarthritis [OA]), showing significant improvements in pain intensity outcomes at short-term follow-up. Nevertheless, the study was mainly based on eHealth systems, and few findings were obtained from mobile apps. Du et al [20] analyzed the use of web-health-based interventions and mHealth interventions in patients with CLBP, showing better effects on both pain and disability outcomes in favor of mHealth systems. According to the effects of mHealth, a recent review [21] evaluated the effectiveness of app-based interventions on several CP conditions (general CP, CLBP, chronic neck pain (CNP), rheumatoid arthritis, OA, menstrual pain, frozen shoulder pain, and migraine), stating that these apps are significantly more effective, with a small effect size in reducing pain in comparison to control groups. Thurnheer et al [22] analyzed the efficacy of app usage in the management of patients with cancer and noncancer pain (chronic cancer pain, general CP, CLBP, CNP, menstrual pain, and acute pain), reporting beneficial effects on pain, particularly in an out-clinic setting. The evidence of the use of mHealth systems is still emerging and focusing mainly on its effects on pain intensity. Moreover, commonly studied pain conditions (cancer and noncancer pain) and different types of pain (acute and chronic) are mixed, leading to heterogeneity in their findings.

In view of this background and to the best of our knowledge, none of the published reviews has examined the effects of the use of mHealth systems on pain intensity along with the effects on the functional disability and quality of life (QoL) of patients with CP. Therefore, the main purpose of this systematic review is to determine the effects of the use of mHealth systems on different CP conditions listed in the ICD-11, based on the improvement of pain intensity, QoL, and functional disability, according to the findings reported with randomized controlled trials (RCTs). Furthermore, we provide an overview of the available mHealth systems for CP management, their purposes, and their features.

Methods

Study Design

The protocol of this systematic review was registered on the International Prospective Register of Systematic Reviews (PROSPERO) database (CRD42022315808) [23]. It was conducted following the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews of RCTs [24].

Search Strategy

The search strategy was based on CP diseases according to the ICD-11 [25]. The search was conducted from February to March 2022 in the following databases: PubMed, Web of Science, Scopus, and Physiotherapy Evidence Database (PEDro). The search strategy was first developed for the PubMed database using Medical Subject Headings, and it was adapted for other databases. The search was not filtered either by language or by date of publication. The search strategy for each database is provided in Multimedia Appendix 1.

Eligibility Criteria

The eligibility criteria were defined according to the PICOS (Population, Intervention, Comparison, Outcomes, Study type) framework [26]. The population included adults (≥18 years old) with any CP condition listed in the ICD-11 [25]. Interventions were mHealth systems based on mobile apps (smartphone or tablet) used for monitoring pain and health-related outcomes, for pain and behavioral self-management, and for performing therapeutic approaches. The rationale for including monitoring apps as an intervention was their effects on modifying the user's behavior, expectation, and performance for disease management or health promotion [27]. Some of the apps' features for promoting behavior changes are reminders and notifications, tracking activity, goal planning, and tailored information [28]. For comparison, the control group included conventional treatments (physical, occupational, and psychological therapies; care medical; and education) or nonintervention. Primary outcomes were based on pain intensity, QoL, and functional disability, and only RCTs were included as study designs.

Studies with a sample of children or adolescents; including a pain condition with a duration less than 3 months; based on the management of cancer-related pain or pre- and postsurgery trauma interventions (eg, knee arthroplasty, carpal tunnel syndrome); including websites, text messages, or other devices (eg, smartwatches, laptops); and those in which all studied groups used the mHealth system for the intervention were excluded.

Study Selection Process

After retrieving the documents from different databases, duplicated documents were removed using Rayyan QCRI (Qatar Computing Research Institute) [29] and manual screening. Studies were first screened by title and abstract by 2 researchers (authors MML and JAMM) according to the eligibility criteria. Next, the full text of potentially relevant papers was reviewed by MML and JAMM to decide whether they should be included in the analysis. Disagreements were discussed and resolved by consensus with a third researcher (author IF).

Data Extraction

The following data were extracted from the included studies: author, year of publication, and country; CP conditions; total number of participants; demographic information, including age and gender, for each study group; intervention details (type, follow-up assessments, and total study duration); and primary and secondary outcomes, as well as outcome measurements or tools. Furthermore, data of the main findings related to pain intensity, QoL, and functional disability were collected. Finally, specific information about the purpose and main features of the mHealth systems used as interventions was identified.

Risk of Bias, Methodological Quality, and Level of Evidence Assessment

First, the risk of bias (RoB) was assessed using the Cochrane RoB 2.0 tool [30], including 5 domains and an overall judgment. The 5 domains are (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcome, and (5) bias in selection of the reported result. Each domain was categorized as "low risk," "high risk," or "unclear risk" based on the answers to signaling questions. An overall RoB assessment of the RCTs was performed following the recommendations in the guidance document.

Second, the Checklist for Measuring Quality [31] was used. It includes 26 items categorized by 5 subscales: reporting (9 items), external validity (3 items), bias (7 items), confounding (6 items), and power (1 item). Each item is scored 0 or 1, except for 1 item in the reporting subscale whose score ranges from 0 to 2 and the single item in the power subscale whose score ranges from 0 to 5, with a maximum overall score of 31. A score less than 50% indicates low methodological quality, 50%-65% indicates medium methodological quality, and >65% indicates high methodological quality.

Finally, the levels of evidence were reported according to the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM), concerning the subject area or clinical setting and the study design involving the clinical question [32]. The level of evidence ranged from 1 (strong evidence) to 5 (weak evidence).

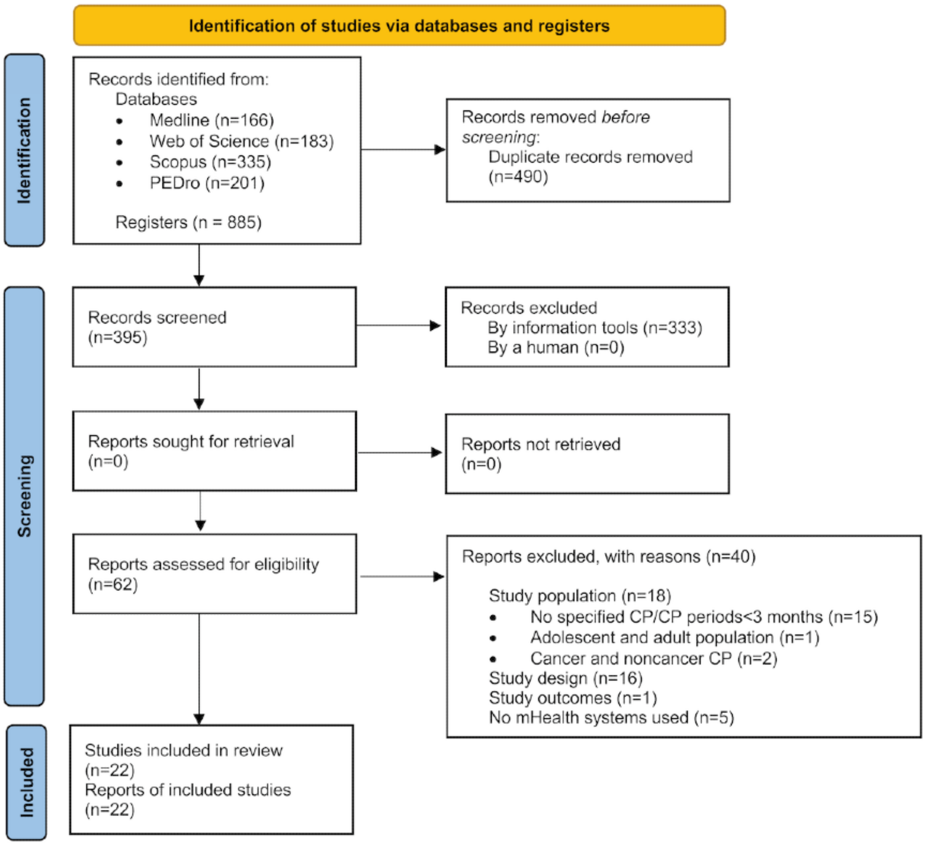
These assessments were performed by 2 authors (MML and JAMM), and the discrepancies were solved by agreement with a third researcher (author AS). These discrepancies appeared mainly in the RoB assessment, specifically in some questions related to deviations from intended interventions and measurement of outcomes. We also discussed some items of the Checklist for Measuring Quality corresponding to external (source population) and internal (blinding and concealment) validity and the power effect.

Results

Study Selection

A total of 885 studies were retrieved from the systematic literature review, of which 490 (55.4%) were duplicates and so deleted automatically. After the first screening by title and abstract, 62/395 (15.7%) studies were selected for full-text reviewing. According to the pre-established selection criteria, a total of 22 (35.5%) studies were finally included in the qualitative analysis. The full screening process and the main reasons for exclusion are shown in Figure 1.

Figure 1. Information flow diagram of the selection process of the systematic review. CP: chronic pain; mHealth: mobile health; PEDro: Physiotherapy Evidence Database.



Risk of Bias, Methodological Quality, and Level of Evidence

Regarding the results of RoB assessment by domain, 18/22 (82%) studies had a low RoB for the random allocation domain and 16/22 (73%) studies had a RoB for the missing outcome data domain. For the second (bias due to deviations from intended interventions) and fourth (measurement of outcomes) domains, 11 (50%) and 14 (64%) studies had some concerns, respectively. Last, in the selection of the reported results domain, 16 (73%) studies had a low RoB but 3 (14%) studies had a high RoB. For overall judgment, 7/22 (32%) studies had a low RoB for their outcomes, 10/22 (45%) studies had some concerns, and 5/22 (23%) studies had a high RoB.

Regarding the Checklist for Measuring Quality, 18 (82%) studies [33-50] reported medium methodological quality (between 50% and 65%), and the rest [51-54] scored high on methodological quality (>65%). Based on the clinical settings of the included studies, which concern therapy or treatment, the OCEBM level of evidence was based on systematic reviews of RCTs or, failing that, individual RCTs with narrow 95% CIs. Thus, all included papers yielded an OCEBM level of 2 for a clinical question of treatment benefits, considering them as highly recommendable.

Detailed results of the RoB assessment are shown in Figures 2 and 3. The methodological quality and the level of evidence and degrees of recommendation of the included studies are detailed in Multimedia Appendix 2 [33-54] .

Figure 2. RoB assessment: traffic light plot. RoB: risk of bias.

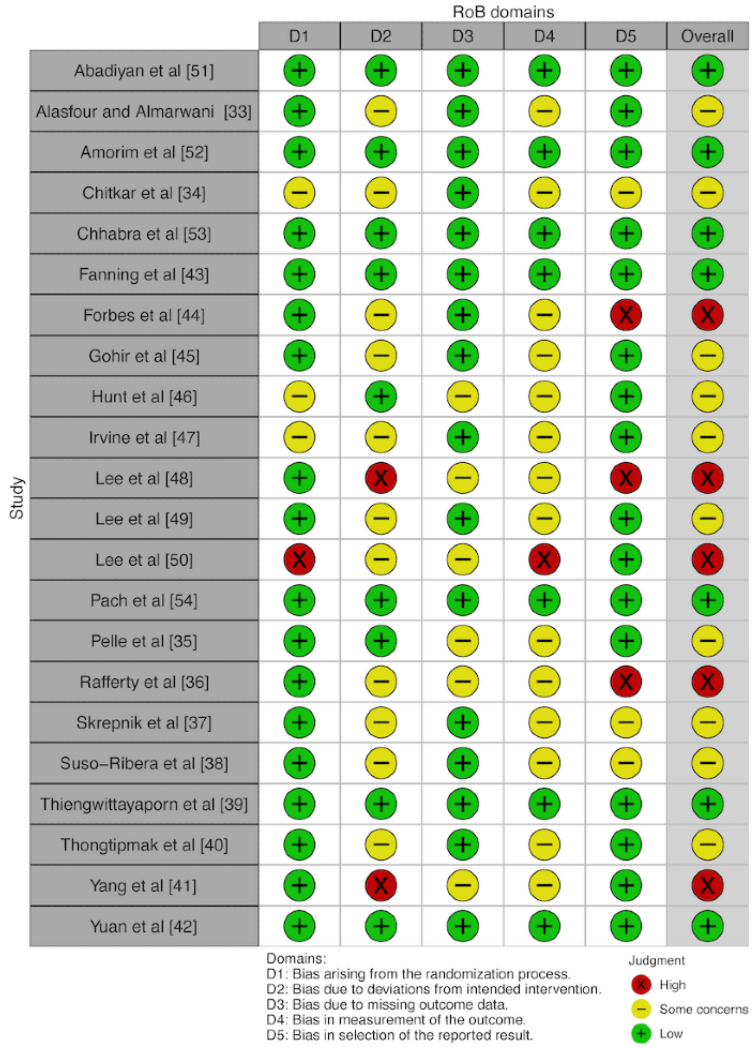
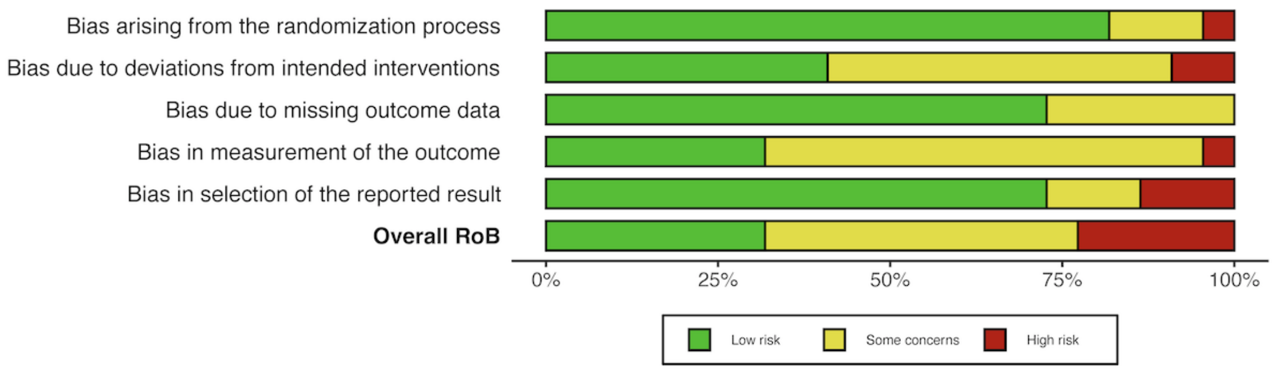


Figure 3. RoB assessment: summary plot. RoB: risk of bias.



Study Characteristics

The main characteristics of the studies included are shown in Table 1. Publication dates ranged from 2015 to 2022. A total of 2641 patients with CP were involved in this present systematic review, 70.6% (1793/2539) being female. The average age was 38.93 (SD 59.29) years, excluding 1 (5%) study [47] in which this information was not available.

According to CP conditions listed in the ICD-11, OA is the condition most studied in the literature, followed by CLBP [41,47,52,53] and CNP [40,49,51,54]. The lowest studies were chronic pelvic pain (CPP) [44] and interstitial cystitis/bladder pain syndrome (IC/BPS) [50].

Table 1. Study characteristics.

Characteristics	Value
Year of publication (N=22), n (%)	
2015-2018	6 (27.3)
2019-2022	16 (72.7)
Region where the study took place (N=22), n (%)	
Asia	11 (50.0)
Europe	4 (18.2)
North America	5 (22.7)
South America	1 (4.5)
Oceania	1 (4.5)
Age (years)[a], mean (SD)	38.93 (59.29)
Gender (N=2539)[b], n (%)	
Female	1793 (70.6)
Male	746 (29.4)
CP[c] conditions (N=22), n (%)	
CLBP[d] [41,47,52,53]	4 (18.2)
CMSP[e] [38]	1 (4.5)
CNP[f] [40,49,51,54]	4 (18.2)
CP (unspecified) [43]	1 (4.5)
CPP[g] [44]	1 (4.5)
FM[h] [42,48]	2 (9.1)
IC/BPS[i]]50]	1 (4.5)
IBS[j] [36,46]	2 (9.1)
OA[k] [33-35,37,39,45]	6 (27.3)
Interventions based on mHealth[l] systems (N=22), n (%)	
Home-based PA[m] program	9 (40.9)
Education	8 (36.4)
CBT[n]	4 (18.2)
Monitoring pain-related outcomes and symptoms	10 (45.5)
Monitoring PA parameters	11 (50.0)
Mind relaxation techniques	5 (22.7)
Intervention period (N=22), n (%)	
<3 months	15 (68.2)
3-6 months	7 (31.8)
Outcomes assessed (N=22), n (%)	
Pain intensity	17 (77.3)
QoL[o]	15 (68.2)
Functional disability	17 (77.3)

[a]Average age of available data except for 1 study.

[b]Gender proportion of available data except for 1 study.

[c]CP: chronic pain.
[d]CLBP: chronic low back pain.
[e]CMSP: chronic musculoskeletal pain.
[f]CNP: chronic neck pain.
[g]CPP: chronic pelvic pain.
[h]FM: fibromyalgia.
[i]IC/BPS: interstitial cystitis/bladder pain syndrome.
[j]IBS: irritable bowel syndrome.
[k]OA: osteoarthritis.
[l]mHealth: mobile health.
[m]PA: physical activity.
[n]CBT: cognitive behavioral therapy.
[o]QoL: quality of life.

Types of mHealth and Comparison Interventions

Several approaches for the self-management of patients with CP involved mHealth systems. On the one hand, we found the monitorization of pain-related variables and symptoms as part of the interventions, either isolated [37,38,41,48] or in combination with other management strategies [39,40,42,47,49,50]. Similarly, the tracking of physical activity (PA) parameters (daily PA and mobility, PA-related goals achieved, and adherence) was also used in 11 (50%) studies [33,35,37,40,41,43,45,49,51-53] aiming to record PA-related goals and to enhance PA performance and behaviors. On the other hand, self-management of CP focused on home-based PA programs as the most common intervention [33,35,39,40,45,49,51,53,55], including a wide variety of exercises, both general and specific for this population. Other common self-management approaches were educational sessions and materials [34-36,45-47,50,55]. Less frequent strategies were cognitive behavioral therapy (CBT) [43,46,47,54] and relaxation and mind-body techniques [43,44,46,54,55]. A total of 23 mHealth systems were used for monitoring [37,38,41,48,52], treatment strategies [34,36,44,46,54], and a combination of both [33,35,39,40,42,43,45,47,49-51,53]. Detailed information about the mHealth systems, their purpose of use, and the principal features are summarized in Multimedia Appendix 3.

In the control groups, interventions were based on usual health care (medical and physical therapies), being the most common comparison intervention [35,37,38,41,44,45,47-54]. Other papers performed the same intervention in both groups, one using mHealth and the other using traditional methods [33,34,36,39,42]. Finally, only 3 (14%) studies [40,43,46] did not involve any intervention.

Study Outcomes and Measurement Tools Used

Pain intensity was assessed in a total of 17 (77%) studies (N=1780). The numeric rating scale (NRS) [33,37,38,45,52-54] and the visual analogue scale (VAS) [40-42,48-51] were the most used. Regarding the OA condition, the Knee injury and Osteoarthritis Outcome Score (KOOS), the Hip injury and Osteoarthritis Outcome Score (HOOS), and the Western Ontario and McMaster (WOMAC) questionnaires were specific tools also used to assess pain intensity [34,35,39].

There was a wide range of tools used in 15 (68%) studies (N=1744) for assessing the QoL. The most repeated instruments were the 36-item Short Form Health Survey (SF-36) [34,41,43,49-51], followed by the EuroQoL-5D [35,48].

In the case of functional disability [33-35,38,39,41-45,47-53], 17 (77%) studies (N=1928) assessed it. Although there are different tools for assessing this outcome, they usually focus on a specific condition (Multimedia Appendix 3). For example, for patients with CNP [49,51], the Neck Disability Index (NDI) was used; for patients with FM, the Fibromyalgia Impact Questionnaire (FIQ) [42,48] was used; and for patients with OA, WOMAC [33,34,45], KOOK [35,39] and HOOS [35] were used.

Effects of mHealth Interventions vs Control Groups

To provide an overview of the differences found between mHealth interventions and control groups in the included studies, a visual representation is shown in Tables 2-5. The "*" sign indicates significance in favor of the mHealth intervention group, and the "=" sign indicates no significant differences between groups. No significant differences in favor of the control groups were reported.

Results of home-based PA programs delivered by mHealth systems led to a significant improvement in pain intensity in patients with CNP [49,51] and OA [35,45] when compared to usual care. Likewise, this type of intervention had significant effects on functional disability [35,45,49,51,53], but only Abadiyan et al [51] showed significant differences in the QoL between groups. In addition, when home-based PA programs delivered by mHealth systems were compared with similar traditional methods, a significant improvement in favor of mHealth for pain intensity [33], QoL, and functional disability outcomes [39] was observed in patients with OA and CNP. Nevertheless, no significant differences were obtained for any of the outcomes measured in patients with FM [42].

In relation to educational interventions based on mHealth, improvements in the QoL in OA [34], IBS [36], and IC/BPS [50] conditions were observed when compared either to usual care or to similar intervention by traditional methods. This intervention also showed improvements in functionality and pain intensity in patients with OA [34] but not for pain intensity in patients with IC/BPS [50].

CBT based on mHealth systems showed some significant improvements in QoL [46,47] and functional disability [47] in favor of the mHealth group when compared to usual care or no intervention. Nevertheless, this intervention neither reduced pain intensity in CNP [54] nor improved the QoL and functional disability in CP significantly [43].

Finally, the results of mHealth interventions focused on monitoring pain and symptoms, compared to usual care, were inconclusive. Thus, significant improvements in reducing pain were reported for patients with OA [37] and FM [48] but not for those with CMSP [38] and CLBP [41]. In patients with CMSP and CLBP, functional disability outcomes significantly improved in favor of mHealth groups [38,41], while those diagnosed with FM did not achieve significant improvements in this outcome [48]. No significant changes in the overall QoL were observed between groups with this type of intervention [41,48].

Other interventions, such as isolated monitoring of PA parameters [52] and mindfulness meditation alone [44], did not show significant differences between the mHealth and control groups for any of the studied outcomes.

With regard to the reporting of adverse events or treatment reactions of the studied interventions, only 6 (27%) of the 22 studies [37,38,45,50,51,54] provided this information, of which only 1 (17%) [54] recorded serious adverse events (cancer, sudden hearing loss, nerve injury and spinal tap, tonsillectomy, and accident causing a fracture), but none of them was considered related to the trial intervention.

Table 2. Characteristics of participants and study interventions (studies 1-11).

Study	CP[a] condition	Participants, N, intervention group (IG), n (%), control group (CG), n (%); age (years), mean (SD); gender (% female)	Intervention: mHealth[b]	Intervention: Control	Total study duration (weeks); follow-up period
Alasfour and Almarwani [33]	Knee OA[c]	N=40; 54.40 (4.33); 100% IG: n=20; 53.65 (3.96); 100% CG: n=20; 55.15 (4.64); 100%	Home-based PA[d] program (lower-limb-strengthening exercises) with the My Dear Knee app; also, exercise adherence and completed sessions recorded by the app	Home-based PA program through paper handouts	6; 3rd and 6th weeks
Arfaei Chitkar et al [34]	Knee OA	N=60; 58.17 (7.55); 100% IG: n=31; 57.84 (8.63); 100% CG: n=29; 58.52 (6.33); 100%	Educational content through the mobile app; usual medical care	Educational content without the app; usual medical care	8; 2nd month
Pelle [35]	Knee or hip OA	N=427 IG: n=214; 62.1 (7.7); 68.7% CG: n=213; 62.1 (7.0); 74.7%	Home-based PA program and education content provided by the Dr. Bart app, with also PA-related goals, self-monitoring, and motivational reminders	Usual care with no active treatment	24; 3rd and 6th months
Rafferty et al [36]	IBS[e]	N=25 IG: n=14; 27.2 (9.5); 86% CG: n=11; 25.7 (11.9); 91%	Nutrition information and recommendations based on patient-specific and individualized diet plans through the Heali app; standard dietary education materials (online)	Standard dietary education materials (online)	4; 1st month
Skrepnik et al [37]	Knee OA	N=211; 62.6 (9.4); 50.2% IG: n=107; 61.6 (9.5); 55.1% CG: n=104; 63.6 (9.3); 45.2%	Monitoring pain, PA parameters, and mood data with feedback and motivational messages from the OA GO app; standard-of-care instructions and education; unblinded wearable device	Standard-of-care instructions and education; blinded wearable device	12; 1 week, 1st and 3rd months
Suso-Ribera et al [38]	CMSP[f]	N=165; 52.1 (11.2); 73.8% IG-1: 53 IG-2: 56 CG: 56	IG-1: monitoring pain-related outcomes using the Pain Monitor app with alarms and usual care; IG-2: monitoring pain-related outcomes with the Pain Monitor app without alarms and usual care	Usual care	4; 1st month
Thiengwittayaporn et al [39]	Knee OA	N=82 IG: n=42; 62.2 (6.8); 85.7% CG: n=40; 63.0 (9.7); 92.5%	Home-based PA program and education, and disease monitoring (symptoms and stages) with the Rak Kao app	Standard education and exercise instructions through handouts	4; 1st month
Thongtipmak et al [40]	CNP[g]	N=100 IG: n=50; 22.86 (1.99); 82% CG: n=50; 22.68 (2.23); 76%	Home-based PA program and monitoring pain level before and after exercises with the NeckProtector app	Rest	Same day
Yang et al [41]	CLBP[h]	N=8 IG: n=5; 35 (10.93); 20% CG: n=3; 50.33 (9.29); 100%	Monitoring pain intensity and activity levels using the Pain Care app; self-management program based on individualized exercises and physiotherapy treatment	Only physiotherapy treatment	4; 2nd and 4th weeks
Yuan et al [42]	FM[i]	N=40 IG: n=20; 43.3 (8.4); 95% CG: n=20; 42.1 (11.8); 100%	Self-care management based on education, home-based PA, and sleep hygiene and relaxation techniques using the ProFibro app, with also self-monitoring disease impact according to FIQ domains; usual medical care	Traditional paper book of similar content; usual medical care	6; 6th week
Fanning et al [43]	CP	N=28; 70.21 (5.22); 78.6% IG: n=15; 70.12 (5.43); 86.7% CG: n=13; 70.32 (5.20); 69.2%	Monitoring PA-related goals, CBT[j], and mindfulness-based relapse prevention using the Mobile Health Intervention to Reduce Pain and Improve Health [MORPH] Companion and Fitbit apps	Waitlist	12; 3rd month

[a]CP: chronic pain.
[b]mHealth: mobile health.
[c]OA: osteoarthritis.
[d]PA: physical activity.
[e]IBS: irritable bowel syndrome.
[f]CMSP: chronic musculoskeletal pain.
[g]CNP: chronic neck pain.
[h]CLBP: chronic low back pain.
[i]FM: fibromyalgia.
[j]CBT: cognitive behavioral therapy.

Table 3. Characteristics of participants and study interventions (studies 12-22).

Study	CP[a] condition	Participants, N, intervention group (IG), n (%), control group (CG), n (%); age (years), mean (SD); gender (% female)	Intervention: mHealth[b]	Intervention: Control	Total study duration (weeks); follow-up period
Forbes et al [44]	CPP[c]	N=90 IG-1: n=31; 34.8 (9.9); 100% IG-2: n=30; 35.7 (5.7); 100% CG: n=29; 35.0 (8.6); 100%	IG-1: mindfulness meditation course delivered by the Headspace app and usual care; IG-2: muscle relaxation techniques in the app and usual care	Usual care	8; 2nd, 3rd, and 6th months
Gohir et al [45]	Knee OA[d]	N=105 IG: n=48; 65.2 (9.7); 70.8% CG: n=57; 68.0 (8.6); 64.9%	Home-based PA[e] program, including strengthening, core stability and balance exercises, and educational sessions, provided by the Hereafter app	Usual care	6; 6th week
Hunt et al [46]	IBS[f]	N=121; 32 (10.2); 75.2% IG: n=62 CG: n=59	Psychoeducation, CBT[g], relaxation techniques, and information about diet, provided by the Zemedy app	Waitlist	8; 2nd month
Irvine et al [47]	CLBP[h]	N=597 IG-1: n=199; 58.3% IG-2: n=199; 58.8% CG: n=199; 62.8%	IG-1: CBT and education through the FitBack app, with also recording of pain-related outcomes; IG-2: alternative care by emails with internet resources (both groups received weekly reminder prompts and emails for assessments)	Usual care; only contacted for assessments	8; 2nd and 4th months
Lee et al [48]	FM[i]	N=25 IG: n=14; 42.8 (7.2); 100% CG: n=11; 41.7 (11.2); 100%	Monitoring pain-related outcomes (intensity, frequency, and environmental factors) with the Pain Assessment and Analysis System [PAAS] Clinic app	Usual care	12; 1st and 3rd months
Lee et al [49]	CNP[j]	N=20 IG: n=11; 27.09 (4.83); 55% CG: n=9; 27.56 (4.67); 45%	McKenzie neck exercise program with a smartphone app in the workplace environment, with also a self-feedback function and monitoring pain	Written instructions about postural hygiene	8; 2nd month
Lee et al [50]	IC/BPS[k]	N=56 IG: n=29; 42.9 (10.4); 100% CG:n=27; 46.3 (14.2); 100%	Health education and symptom self-management with the Taiwan Interstitial Cystitis Association [TICA] app; patients could continue using usual care	Usual care	8; 2nd month
Abadiyan et al [51]	CNP	N=60; 38.5 (9.1) IG-1: n=20; 41.3 (8.1); 50% IG-2: n=20; 40.3 (7.9); 50% CG: n=20; 37.4 (9.8); 35%	IG-1: home-based PA program, global posture re-education (GPR), and self-managed work time with the Seeb app, with also recording of PA parameters; IG-2: GPR alone	Traditional neck education and exercise therapy	8; 8th week
Amorim et al [52]	CLBP	N=68 IG: n=34; 59.5 (11.9); 44% CG: n=34; 57.1 (14.9); 56%	Monitoring PA-related goals with the IMPACT app, with motivational messages; telephone-based coaching sessions; PA and sedentary behavior information booklet	PA information booklet and advice to stay active	24; weekly and 6th month
Chhabra et al [53]	CLBP	N=93 IG: n=45; 41.4 (14.2) CG: n=48; 41.0 (14.2)	Home-based PA program, including specific back exercises and aerobic PA; monitoring daily PA parameters with the Snapcare app; written prescription and usual medical care	Written prescription, including PA advice; usual medical care	12; 3rd month
Pach et al [54]	CNP	N=220 IG: n=110; 37.9 (11); 67.3% CG: n=110; 39.8 (11.6); 71.8%	Relaxation exercises (autogenic training, mindfulness meditation, and guided imagery) and CBT strategies with the RelaxNeck app; follow-up data collected using app-based questionnaires	Usual care; app for data entry only	24; 3rd and 12th months

[a]CP: chronic pain.

[b]mHealth: mobile health.

[c]CPP: chronic pelvic pain.

[d]OA: osteoarthritis.
[e]PA: physical activity.
[f]IBS: irritable bowel syndrome.
[g]CBT: cognitive behavioral therapy.
[h]CLBP: chronic low back pain.
[i]FM: fibromyalgia.
[j]CNP: chronic neck pain.
[k]IC/BPS: interstitial cystitis/bladder pain syndrome.

Table 4. Overall RoB[a] assessment, study outcomes, and main results (studies 1-11).

Study	CP[b] condition	Study outcomes (measurement tools)		RoB	Outcome results[c]		
		Primary	Secondary		Pain intensity	QoL[d]	Functional disability
Alasfour and Almarwani [33]	Knee OA[e]	• Self-reported exercise adherence (percentage of completed exercises)	• Pain intensity (Arabic version of the numeric pain rating scales [ANPRS]) • Physical function (ArWOMAC[f]) • Lower-limb muscle strength (Five-Times Sit-to-Stand Test [FTSST])	–	*	N/A[g]	=
Arfaei Chitkar et al [34]	Knee OA	• Physical functioning (WOMAC[h]) • QoL (SF-36[i])	N/A	–	*	*	*
Pelle [35]	Knee/hip OA	• Number of self-reported consultations in health care	• Pain intensity, symptoms, and functional limitations (KOOS[j]/HOOS[k]) • QoL (EuroQoL-5D-3L) • PA[l] level (Short Questionnaire to Assess Health-enhancing physical activity [SQUASH]) • Patient’s cognitive and emotional perceptions (brief Illness Perception Questionnaire [IPQ]) • Knowledge, skills, and confidence (PAM-13)	–	*	=	*
Rafferty et al [36]	IBS[m]	• IBS symptoms (5-item IBS Symptom Severity Scale [IBSS-SSS]; Rome IV) • QoL (World Health Organization Quality of Life [WHOQOL-BREF]) • LFD knowledge (low FODMAP dietary consumption questionnaire [LFDA Quest.]) • LFD adherence (low FODMAP dietary knowledge questionnaire [LFDK Quest.])	N/A	X	N/A	*	N/A
Skrepnik et al [37]	Knee OA	Mobility (6-minute walking test [6MWT]; steps/day)	• Pain intensity (NRS[n]) • Patient and physical satisfaction (PAM-13) • Quality of sleep (wearable activity monitor) • Mood states (visual analogue mood scale [VAMS]) • Treatment-emergent adverse events (treatment-emergent adverse events [TEAEs])	–	*	N/A	N/A
Suso-Ribera et al [38]	CMSP[o]	• Pain intensity (NRS) • Medication side effects (NRS)	• Pain-related interference (NRS) • Fatigue (NRS) • Depression, anxiety, and anger (NRS)	–	=	N/A	*

Study	CP[b] condition	Study outcomes (measurement tools)		RoB	Outcome results[c]		
		Primary	Secondary		Pain intensity	QoL[d]	Functional disability
Thiengwittayaporn et al [39]	Knee OA	• Patient's ability to correctly perform the exercises (80% completed exercise repetitions)	• Range of motion (goniometer) • Pain intensity, symptoms, daily life activities, PA and sports performed, and QoL (KOOS) • Satisfaction/expectation with functional ability (Knee Society Score [KSS])	+	=	*	*
Thongtipmak et al [40]	CNP[p]	• Pain intensity (VAS[q]) • Muscle tension (VAS) • Pressure pain threshold (pressure algometry) • Cervical range of motion (CROM; device) • Acceptability assessment (System Usability Scale [SUS])	N/A	–	*	N/A	N/A
Yang et al [41]	CLBP[r]	• Pain intensity (VAS) • Disability (Roland-Morris Disability Questionnaire [RMDQ]) • QoL (SF-36) • Self-efficacy (Pain Self-Efficacy Questionnaire [PSEQ])	N/A	X	=	=	*
Yuan et al [42]	FM[s]	• QoL (FIQ[t])	• Pain intensity (VAS) • Function (FIQ-Function) • Painful body regions (Widespread Pain Index [WPI]) • Symptom Severity (SS) scale • Self-care (Appraisal of Self-Care Agency Scaled-Revised [ASAS-R])	+	=	=	=
Fanning et al [43]	CP	• QoL (SF-36) • Physical functioning (SF-36: physical functioning subscale) • Self-efficacy for walking (8-item scale) • Satisfaction with physical functioning (7-item scale)	N/A	+	N/A	=	=

[a]RoB: risk of bias; interpretation of RoB: +, low RoB; –, some concerns; X, high RoB.

[b]CP: chronic pain.

[c]Interpretation of outcome results: *, significant differences (*P*<.05) in favor of the mHealth group; =, nonsignificant differences between groups.

[d]QoL: quality of life.

[e]OA: osteoarthritis.

[f]ArWOMAC: Arabic version of Western Ontario and McMaster.

[g]N/A: not applicable.

[h]WOMAC: Western Ontario and McMaster.

[i]SF-36: 36-item Short-Form Health Survey.

[j]KOOS: Knee injury and Osteoarthritis Outcome Score.

[k]HOOS: Hip injury and Osteoarthritis Outcome Score.

[l]PA: physical activity.

[m]IBS: irritable bowel syndrome.

[n]NRS: numeric rating scale.
[o]CMSP: chronic musculoskeletal pain.
[p]CNP: chronic neck pain.
[q]VAS: visual analogue scale.
[r]CLBP: chronic low back pain.
[s]FM: fibromyalgia.
[t]FIQ: Fibromyalgia Impact Questionnaire.

Table 5. Overall RoB[a] assessment, study outcomes, and main results (studies 12-22).

Study	CP[b] condition	Study outcomes (measurement tools)		RoB	Outcome results[c]		
		Primary	Secondary		Pain intensity	QoL[d]	Functional disability
Forbes et al [44]	CPP[e]	• Pain-related disability (Chronic Pain Grade-Disability subscale) • QoL (RAND 36) • Pain acceptance (chronic pain acceptance questionnaire [CPAQ]) • Depression and anxiety (HAD) • Self-efficacy (Pain Self-efficacy Quest.) • Sexual health (sexual health outcomes in women questionnaire [SHOW-Q]) • Mindfulness (Cognitive and mindfulness-revised scale) • Individualized outcome (Measure yourself medical outcome profile [MYMOP])	• Study feasibility (CPAQ) • App usability (System Usability Scale [SUS]) • Adherence to the app (frequency of app use)	X	N/A[f]	=	=
Gohir et al [45]	Knee OA[g]	• Pain intensity (NRS[h])	• Physical functioning (WOMAC[i], Timed Up & Go [TUG], and 30-second sit-to-stand test) • QoL (Musculoskeletal Health Questionnaire [MSK-HQ]) • Symptoms sensory (pressure pain threshold [PPT])	–	*	=	*
Hunt et al [46]	IBS[j]	• QoL (Irritable Bowel Syndrome Quality of Life [IBS-QOL]) • Symptom severity (Gastrointestinal Symptom Rating Scale-IBS [GSRS-IBS])	• Diagnostic criteria for IBS (Rome IV) • Fear of food (Fear of Food Questionnaire [FFQ]) • Gastrointestinal (GI) symptom–specific anxiety (Visceral Sensitivity Index [VSI]) • Cognitions-related impact (Gastrointestinal Cognition Questionnaire [GI-COG]) • Depression and anxiety (Depression Anxiety Stress Scale [DASS]) • Diagnosis and depressive symptom severity (Patient Health Questionnaire [PHQ]) • Dose (number of app modules completed)	–	N/A	*	N/A

Study	CP[b] condition	Study outcomes (measurement tools)		RoB	Outcome results[c]		
		Primary	Secondary		Pain intensity	QoL[d]	Functional disability
Irvine et al [47]	CLBP[k]	Physical outcomes: • Pain intensity, episodes, and duration (back pain scales) • Daily pain management activities • Functionality (10-item scale based on Multidimensional Pain Inventory Interference Scale [MPI] and Brief Pain Inventory [BPI]) • QoL (Dartmouth Primary Care Cooperative Information Project [Dartmouth CO-OP] scale)	Prevention-helping behaviors Worksite outcomes: • Worker productivity (4-item Work Limitations Questionnaire [WLQ]) • Presenteeism (Stanford Presenteeism Scale) Other outcomes: • Responsibility of own health (Patient Activation Measure [PAM]) • Behavior constructs (knowledge, behavioral intentions, and self-efficacy) • Attitudes toward pain (10-item Survey of Pain Attitudes [SOPA]) • Catastrophizing of pain (Tampa scale)	–	N/A	*	*
Lee et al [48]	FM[l]	• Pain intensity (VAS[m])	• QoL (EuroQoL-5D) • Disease impact (FIQ[n]) • Depression index (Beck's Depression Index [BDI]) • Patient global assessment (patient global assessment [PtGA])	X	*	=	=
Lee et al [49]	CNP[o]	• Pain intensity (VAS) • Functional disability (NDI[p])	• QoL (SF-36[q]) • Maximal voluntary strength (digital handheld dynamometer) • Fear avoidance belief (Fear-Avoidance Belief Questionnaire [FABQ]) • Exercise adherence (app)	–	*	=	*
Lee et al [50]	IC/BPS[r]	• QoL (SF-36)	• Pain intensity (VAS) • Symptoms (O´Leary-Sant symptom) • Physical function, role physical, bodily pain, vitality, social function, role emotional and mental health (SF-36 subscales)	X	*	=	=
Abadiyan et al [51]	CNP	• Pain intensity (VAS)	• Disability (NDI) • QoL (SF-36) • Endurance (progressive isoinertial lifting evaluation [PILE] test) • Forward head posture (craniovertebral angle)	+	*	*	*
Amorim et al [52]	CLBP	• Pain intensity (NRS) • Disability (Roland-Morris Disability Questionnaire [RMDQ]) • Care seeking (health care consultations)	• Self-reported PA[s] level (International Physical Activity Questionnaire [IPAQ]) • PA data (accelerometer)	+	=	N/A	=
Chhabra et al [53]	CLBP		Only for GI: • Daily PA (activity tracker built within the app) • Progress in symptoms (Current Symptom Score [CSS])	+	=	N/A	*

Study	CP[b] condition	Study outcomes (measurement tools)		RoB	Outcome results[c]		
		Primary	Secondary		Pain intensity	QoL[d]	Functional disability
		• Pain intensity (numeric pain rating scales [NPRS]) • Disability (Modified Oswestry Disability Index [MODI])					
Pach et al [54]	CNP	• Pain intensity during first 3 months (NRS)	• Pain intensity, weekly and during the 6 months (NRS) • Pain acceptance (CPAQ) • Neck pain–related stress • Sick leave days • Pain medication intake • Adherence	+	=	N/A	N/A

[a]RoB: risk of bias; interpretation of RoB: +, low RoB; –, some concerns; X, high RoB.

[b]CP: chronic pain.

[c]Interpretation of outcome results: *, significant differences (*P*<.05) in favor of the mHealth group; =, nonsignificant differences between groups.

[d]QoL: quality of life.

[e]CPP: chronic pelvic pain.

[f]N/A: not applicable.

[g]OA: osteoarthritis.

[h]NRS: numeric rating scale.

[i]WOMAC: Western Ontario and McMaster.

[j]IBS: irritable bowel syndrome.

[k]CLBP: chronic low back pain.

[l]FM: fibromyalgia.

[m]VAS: visual analogue scale.

[n]FIQ: Fibromyalgia Impact Questionnaire.

[o]CNP: chronic neck pain.

[p]NDI: Neck Disability Index.

[q]SF-36: 36-item Short-Form Health Survey.

[r]IC/BPS: interstitial cystitis/bladder pain syndrome.

[s]PA: physical activity.

Discussion

Principal Findings

This study provided an overview of the use of mHealth systems for the self-management of patients with different CP conditions. To the best of our knowledge, this is the first systematic review that identifies the available mHealth interventions and their effects on pain intensity, QoL, and functional disability in patients with CP. Results showed that some interventions based on mHealth systems have beneficial effects on reducing pain and functional disability and improving the QoL. Thus, the scientific evidence suggests that these systems could be a promising alternative in CP self-management through multimodal approaches.

Regarding the analyzed outcomes, 9 of the 17 studies assessing pain intensity [33-35,37,40,45,48,49,51] showed significant effects in reducing pain in favor of mHealth groups. There are several systematic reviews and meta-analyses that support these findings. Pfeifer et al [21] showed that mHealth apps are more effective in reducing pain when compared to control interventions in patients with different CP conditions, such as general CP, CLBP, CNP, arthritis (rheumatoid and OA), menstrual pain, frozen shoulder pain, and migraine. Nevertheless, the authors stated that most of the included studies used cointerventions (eg, physiotherapy, self-management booklets, pharmacological approach, and wearable activity monitors), in addition to using mHealth systems. Likewise, Moman et al [14] observed significant short- and intermediate-term improvements in pain-related outcomes in patients with CP, CLBP, FM and OA, and Thurnheer et al [22] reported a decrease in pain severity in patients with several CP diagnoses (chronic cancer pain, general CP, CLBP, CNP, menstrual pain, and also acute pain) using mobile apps for their management. Furthermore, focusing on the CP condition, Du et al [20] indicated that mHealth-based self-management programs for reducing pain show clinically important effects. Similarly, Chen et al [56] showed that the use of mobile apps for delivering PA programs is associated with significant pain relief in patients with knee OA or chronic knee pain.

Regarding the QoL, improvements were observed in 7 of 15 studies [34,36,39,46,47,50,51] involving several CP conditions (OA, CNP, CLBP, IBS, and IC/BPS). This result agrees with a previous systematic review [22] reporting that patients using a mHealth app for their self-management have a higher QoL compared to patients not using that system. Nevertheless, in the meta-analysis carried out by Chen et al [56], when analyzing the type of technology used for delivering PA programs, they observed that the use of the web is associated with significant improvements in the QoL in patients with knee OA or chronic knee pain, but the use of mobile apps or smartphones is not. This may be because only few of the studies included in this meta-analysis used mobile apps to deliver the interventions, making it difficult to examine the effects of this type of technology.

In the case of functional disability, we found some significant differences between mHealth and control groups in patients with musculoskeletal pain (CLBP [41,47,53], CNP [49,51], and CMSP [38]) and OA [34,35,45]. Nevertheless, these findings are not in line with the available literature. Chen et al [56] did not find evidence for a significant improvement in physical function with technology-supported PA programs. Likewise, results of meta-analyses of telehealth-based interventions, including mHealth and eHealth systems, have suggested that these interventions have no significant effects on physical functionality [14] and disability [57] at short- and intermediate-term follow-up. However, these results are provided by different technology-based interventions and not specifically mHealth systems, which are more recent technologies not sufficiently researched yet.

The types of intervention of the studies included in this systematic review were home-based PA programs, education, CBT, mind-body therapies, and monitoring. This is in line with a large review of the recommendations from clinical practice guidelines (CPG) for musculoskeletal pain, where 3 pillar interventions were identified as key self-management approaches: education, PA, and psychosocial therapies [58]. Similarly, Geraghty et al [59] analyzed the available self-management interventions for chronic widespread pain, with PA programs and medical information being the 2 most common components, followed by psychological approaches. Our findings reported that depending on the type of interventions carried out by mHealth, there are differences in their effects on study outcomes. In this regard, home-based PA programs and education, combined or isolated, showed significant effects on all outcomes compared to other interventions, especially in the case of functional disability. We also found that PA programs and education are commonly considered as cointerventions.

The use of PA as a clinical intervention is suggested as being adequate for several of the conditions included in this systematic review. In patients with OA, it showed a moderate effect on physical functioning, with high patient acceptability and limited side effects, being strongly recommended as conservative management [60]. Similarly, van Doormaal et al [61] reported that PA reduces pain and improves physical function and QoL, with strong-to-moderate evidence. Finally, the CPG for OA include specific exercise programs as core treatment of the nonsurgical management of this condition [58,62]. Moreover, for CLBP and CNP self-management, PA showed significant improvements in pain intensity and functional disability outcomes and slightly more effects on the QoL. In line with this, Bertozzi et al [63] and Price et al [64] reported significant improvement effects of PA programs on CNP in the short and intermediate terms. Nevertheless, both studies have mentioned that the effects of PA are not maintained in the long term, although no high-quality trials are available [63,64]. In the case of FM, although only Yuan et al [42] performed a home-based PA program, this type of intervention is strongly recommended in clinical guidelines for the management of this pathology [65,66]. In fact, previous evidence supports the effectiveness of different modalities of exercise (aerobic, strength, and functional training) in common symptoms of FM and QoL [67].

Education is also considered an essential component of conservative management. In fact, the included studies on several CP conditions applied this approach in isolation or in combination with other interventions, showing improvements in pain-related outcomes, functional disability, and QoL. Education usually includes information about the condition, its prognosis, possible consequences, associated factors, the importance of maintaining a healthy lifestyle, and self-care management options [58,60]. Education promotes feelings of hope and optimism and a positive expectation of the treatment benefits in patients with CP [62].

As previously mentioned, another key purpose of the CPG was to address the psychosocial factors related to CP, for which the internet-delivered interventions may be 1 means of increasing remote access to psychological care. In fact, the previous literature shows beneficial effects of internet-delivered cognitive and behavioral interventions for CP on pain intensity, disability, mood states, and QoL, supporting the use of technological devices for pain management outcomes [19,68]. In that line, CBT is the most studied and used and is especially important in some CP conditions, such as FM [65]. Evidence showed that patients with FM who received CBT showed reduced pain and improved health-related QoL and functional disability more than patients receiving usual care, no treatment, or other nonpharmacological interventions [69]. Similarly, Mascarenhas et al [70] found high-quality evidence in favor of CBT for pain in the short term but with a small effect size that did not reach the minimum clinically important change. Although CBT is a common treatment strategy in FM, the studies included in our systematic review did not apply this type of intervention for FM. However, CBT was applied to patients with both IBS and IC/BPS, showing improvements in QoL and functional disability outcomes. Guidelines recommend that the management of these CP conditions should include multimodal behavioral, physical, and psychological techniques [71].

Other self-management interventions delivered by mHealth systems found in the studies included in this review were the monitoring of pain, other symptoms (mood states, disease stages and impact, and adverse events), and PA parameters, isolated or as cointerventions of other therapies. In addition, mind-body components encompassing meditation, mindfulness, and relaxation techniques were found. Nevertheless, the results of these strategies were heterogeneous, showing only some slight differences when compared to usual care or similar intervention by traditional methods. Thus, it suggests that these interventions have insufficient evidence in CP to provide conclusive findings.

Regarding the overall methodological quality of the studies included, almost all of them reported medium methodological quality according to the Checklist for Measuring Quality. Nevertheless, some items related to internal and external validity were frequently scored as “null” or “unable to determine,” which could limit the interpretation and generalization of the results. Likewise, the results of the Cochrane RoB 2.0 assessment tool showed some concerns and a high RoB in the domain related to deviations from intended interventions due to the nature of the study design itself. Lack of blinding of participants is a common issue reported in research where the implementation of interventions depends on the participants, making it difficult to blind them. Similarly, lack of blinding of outcome assessors poses some concerns and a high RoB in the measurement outcome domain, which could also influence the interpretation of findings. Therefore, a future RCT should address these issues to strengthen the evidence on mHealth-based interventions for the self-management of patients with CP.

Study Limitations and Recommendations for Future Research

Although this systematic review provides a wide perspective on the use of mHealth for self-management of CP, some limitations should be remarked. First, due to the inclusion criteria of the study population, the heterogeneity among CP conditions and patient characteristics makes generalization of the findings not suitable for a specific CP condition. In addition, the high heterogeneity in terms of study interventions and outcome measures makes a meta-analysis not congruent enough to extract a quantitative synthesis of the findings. Third, due to the nature of the RCT, patients in most studies were aware of the interventions, so the effect of a placebo cannot be rejected and could suppose a risk of performance bias. Similarly, the lack of blinding outcome assessors poses a risk of detection bias, which could influence the interpretation of results. Therefore, future research with higher quality in these methodological aspects is needed. Fourth, in some studies, the sample size was small, in addition to losses to follow-up during ongoing research, which could limit the interpretation of the results and limit the drawing of conclusive evidence. Last, because we focused our study on the adult population with CP conditions, the review did not provide information about the effects that the mHealth systems might have on children and adolescents. This could be of interest for future research, as this type of intervention may be attractive and motivating for those populations who are currently familiar with the use of mobile technologies.

Conclusion

This systematic review analyzed the effects of mHealth systems on self-management interventions in patients with different CP conditions, showing beneficial effects on pain intensity, QoL, and functional disability. Concretely, mHealth systems showed positive effects on pain intensity in CNP, FM, IC/BPS, and OA; on the QoL in CLBP, CNP, IBS, and OA; and on functional disability in CLBP, CMSP, CNP, and OA. No statistically significant changes for any of the study outcomes were observed in patients with unspecific CP and CPP. Despite the methodological limitations, mHealth systems seem to be a promising alternative for the management of patients with CP through a biopsychosocial framework. Indeed, there is a wide variety of mHealth systems for the management of CP, ranging from the monitoring of pain and symptoms to therapeutic approaches, mainly based on exercise, education, and psychosocial components. However, further clinical studies of high methodological quality are needed to consolidate the scientific evidence and recommendations for the use of mHealth systems in patients with CP.

Acknowledgments

This work was supported by the Spanish Ministry of Science, Innovation and Universities (grant FPU19/04377 held by MM-L); the Consejería de Salud y Familia, Junta de Andalucía, Spain (no. PI-0112-2018); the Biomedical Research, and Innovation Institute of Cádiz (INiBICA; no. LI19/04IN-CO13); the University of Cádiz (no. PR2017-038); and the Grünenthal Foundation. The Grünenthal Foundation is a private, nonprofit organization that promotes the dissemination of scientific knowledge and supports research.

Authors' Contributions

MM-L, JAM-M, AS, and IF were responsible for conceptualization, methodology, and writing—review and editing, and MM-L and JAM-M for writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Complete search strategy.

[PDF File (Adobe PDF File), 49 KB - mhealth_v11i1e40844_app1.pdf]

Multimedia Appendix 2

Methodological quality (Checklist for Measuring Quality) and grade of recommendation (OCEBM). OCEBM: Oxford Centre for Evidence-Based Medicine.
[DOCX File , 27 KB - mhealth_v11i1e40844_app2.docx]

Multimedia Appendix 3
mHealth systems used and their features.
[PDF File (Adobe PDF File), 186 KB - mhealth_v11i1e40844_app3.pdf]

References

1. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. . [doi: 10.1016/S0140-6736(16)31678-6]
2. Goldberg DS, McGee SJ. Pain as a global public health priority. BMC Public Health 2011 Oct 06;11(1):770 [FREE Full text] [doi: 10.1186/1471-2458-11-770] [Medline: 21978149]
3. Reid KJ, Harker J, Bala MM, Truyers C, Kellen E, Bekkering GE, et al. Epidemiology of chronic non-cancer pain in Europe: narrative review of prevalence, pain treatments and pain impact. Curr Med Res Opin 2011 Feb;27(2):449-462. [doi: 10.1185/03007995.2010.545813] [Medline: 21194394]
4. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
5. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the International Classification of Diseases (ICD-11). Pain 2019 Jan;160(1):19-27. [doi: 10.1097/j.pain.0000000000001384] [Medline: 30586067]
6. Treede R, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain 2015 Jun;156(6):1003-1007 [FREE Full text] [doi: 10.1097/j.pain.0000000000000160] [Medline: 25844555]
7. Miaskowski C, Blyth F, Nicosia F, Haan M, Keefe F, Smith A, et al. A biopsychosocial model of chronic pain for older adults. Pain Med 2020 Sep 01;21(9):1793-1805. [doi: 10.1093/pm/pnz329] [Medline: 31846035]
8. Cheatle MD. Biopsychosocial approach to assessing and managing patients with chronic pain. Med Clin North Am 2016 Jan;100(1):43-53. [doi: 10.1016/j.mcna.2015.08.007] [Medline: 26614718]
9. Ampiah PK, Hendrick P, Moffatt F, Ahenkorah J. Operationalisation of a biopsychosocial approach for the non-pharmacological management of patients with chronic musculoskeletal pain in low- and middle-income countries: a systematic review. Musculoskelet Care 2020 Sep 14;18(3):227-244. [doi: 10.1002/msc.1462] [Medline: 32056363]
10. Nicholas M, Vlaeyen JWS, Rief W, Barke A, Aziz Q, Benoliel R, IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019 Jan;160(1):28-37. [doi: 10.1097/j.pain.0000000000001390] [Medline: 30586068]
11. Sun Y, Laksono I, Selvanathan J, Saripella A, Nagappa M, Pham C, et al. Prevalence of sleep disturbances in patients with chronic non-cancer pain: a systematic review and meta-analysis. Sleep Med Rev 2021 Jun;57:101467. [doi: 10.1016/j.smrv.2021.101467] [Medline: 33827029]
12. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
13. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
14. Moman RN, Dvorkin J, Pollard EM, Wanderman R, Murad MH, Warner DO, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain-is it time to start prescribing electronic health applications? Pain Med 2019 Nov 01;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]
15. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730 [FREE Full text] [doi: 10.2196/14730] [Medline: 31769758]
16. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. A systematic review of pain-related apps. PLoS One 2014 Jul;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
17. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011 Aug;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
18. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
19. Gandy M, Pang ST, Scott AJ, Heriseanu AI, Bisby MA, Dudeney J, et al. Internet-delivered cognitive and behavioural based interventions for adults with chronic pain: a systematic review and meta-analysis of randomized controlled trials. Pain 2022 Oct 01;163(10):e1041-e1053. [doi: 10.1097/j.pain.0000000000002606] [Medline: 35121696]
20. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]

21. Pfeifer A, Uddin R, Schröder-Pfeifer P, Holl F, Swoboda W, Schiltenwolf M. Mobile application-based interventions for chronic pain patients: a systematic review and meta-analysis of effectiveness. J Clin Med 2020 Nov 05;9(11):3557 [FREE Full text] [doi: 10.3390/jcm9113557] [Medline: 33167300]
22. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
23. National Institute for Health Research. PROSPERO: International prospective register of systematic reviews. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2023-01-23]
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
25. World Health Organization. ICD-11: International Classification of Diseases (11th revision). URL: https://icd.who.int/ [accessed 2023-01-23]
26. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
27. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
28. Haase J, Farris KB, Dorsch MP. Mobile applications to improve medication adherence. Telemed J E Health 2017 Feb;23(2):75-79. [doi: 10.1089/tmj.2015.0227] [Medline: 27248315]
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun 01;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
32. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/explanation-of-the-2011-ocebm-levels-of-evidence/ [accessed 2023-01-23]
33. Alasfour M, Almarwani M. The effect of innovative smartphone application on adherence to a home-based exercise programs for female older adults with knee osteoarthritis in Saudi Arabia: a randomized controlled trial. Disabil Rehabil 2022 Jun 25;44(11):2420-2427. [doi: 10.1080/09638288.2020.1836268] [Medline: 33103499]
34. Arfaei Chitkar SS, Mohaddes Hakkak HR, Saadati H, Hosseini SH, Jafari Y, Ganji R. The effect of mobile-app-based instruction on the physical function of female patients with knee osteoarthritis: a parallel randomized controlled trial. BMC Womens Health 2021 Sep 14;21(1):333 [FREE Full text] [doi: 10.1186/s12905-021-01451-w] [Medline: 34521400]
35. Pelle T, Bevers K, van der Palen J, van den Hoogen F, van den Ende C. Effect of the Dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthr Cartil 2020 Apr;28(4):418-427 [FREE Full text] [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
36. Rafferty AJ, Hall R, Johnston CS. A novel mobile app (Heali) for disease treatment in participants with irritable bowel syndrome: randomized controlled pilot trial. J Med Internet Res 2021 Mar 02;23(3):e24134 [FREE Full text] [doi: 10.2196/24134] [Medline: 33650977]
37. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
38. Suso-Ribera C, Castilla D, Zaragozá I, Mesas Á, Server A, Medel J, et al. Telemonitoring in chronic pain management using smartphone apps: A randomized controlled trial comparing usual assessment against app-based monitoring with and without clinical alarms. Int J Environ Res Public Health 2020 Sep 09;17(18):6568 [FREE Full text] [doi: 10.3390/ijerph17186568] [Medline: 32916983]
39. Thiengwittayaporn S, Wattanapreechanon P, Sakon P, Peethong A, Ratisoontorn N, Charoenphandhu N, et al. Development of a mobile application to improve exercise accuracy and quality of life in knee osteoarthritis patients: a randomized controlled trial. Arch Orthop Trauma Surg 2021 Aug 28:1-10 [FREE Full text] [doi: 10.1007/s00402-021-04149-8] [Medline: 34453570]
40. Thongtipmak S, Buranruk O, Eungpinichpong W, Konharn K. Immediate effects and acceptability of an application-based stretching exercise incorporating deep slow breathing for neck pain self-management. Healthc Inform Res 2020 Jan;26(1):50-60 [FREE Full text] [doi: 10.4258/hir.2020.26.1.50] [Medline: 32082700]
41. Yang J, Wei Q, Ge Y, Meng L, Zhao M. Smartphone-based remote self-management of chronic low back pain: a preliminary study. J Healthc Eng 2019;2019:4632946 [FREE Full text] [doi: 10.1155/2019/4632946] [Medline: 30881606]

42. Yuan SLK, Couto LA, Marques AP. Effects of a six-week mobile app versus paper book intervention on quality of life, symptoms, and self-care in patients with fibromyalgia: a randomized parallel trial. Braz J Phys Ther 2021 Jul;25(4):428-436 [FREE Full text] [doi: 10.1016/j.bjpt.2020.10.003] [Medline: 33248904]
43. Fanning J, Brooks AK, Hsieh KL, Kershner K, Furlipa J, Nicklas BJ, et al. The effects of a pain management-focused mobile health behavior intervention on older adults' self-efficacy, satisfaction with functioning, and quality of life: a randomized pilot trial. Int J Behav Med 2022 Apr 20;29(2):240-246 [FREE Full text] [doi: 10.1007/s12529-021-10003-3] [Medline: 34018138]
44. Forbes G, Newton S, Cantalapiedra Calvete C, Birch J, Dodds J, Steed L, et al. MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. BMJ Open 2020 Mar 12;10(3):e030164 [FREE Full text] [doi: 10.1136/bmjopen-2019-030164] [Medline: 32165549]
45. Gohir SA, Eek F, Kelly A, Abhishek A, Valdes AM. Effectiveness of internet-based exercises aimed at treating knee osteoarthritis: the iBEAT-OA randomized clinical trial. JAMA Netw Open 2021 Feb 01;4(2):e210012 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0012] [Medline: 33620447]
46. Hunt M, Miguez S, Dukas B, Onwude O, White S. Efficacy of Zemedy, a mobile digital therapeutic for the self-management of irritable bowel syndrome: crossover randomized controlled trial. JMIR Mhealth Uhealth 2021 May 20;9(5):e26152 [FREE Full text] [doi: 10.2196/26152] [Medline: 33872182]
47. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
48. Lee J, Park S, Ju JH, Cho JH. Application of a real‐time pain monitoring system in Korean fibromyalgia patients: a pilot study. Int J Rheum Dis 2019 Feb 21;22(5):934-939. [doi: 10.1111/1756-185x.13472]
49. Lee J, Lee M, Lim T, Kim T, Kim S, Suh D, et al. Effectiveness of an application-based neck exercise as a pain management tool for office workers with chronic neck pain and functional disability: a pilot randomized trial. Eur J Integr Med 2017 Jun;12:87-92. [doi: 10.1016/j.eujim.2017.04.012]
50. Lee M, Wu H, Tseng C, Ko T, Weng T, Chen Y. Health education and symptom flare management using a video-based m-Health system for caring women with IC/BPS. Urology 2018 Sep;119:62-69. [doi: 10.1016/j.urology.2018.05.027] [Medline: 29894774]
51. Abadiyan F, Hadadnezhad M, Khosrokiani Z, Letafatkar A, Akhshik H. Adding a smartphone app to global postural re-education to improve neck pain, posture, quality of life, and endurance in people with nonspecific neck pain: a randomized controlled trial. Trials 2021 Apr 12;22(1):274 [FREE Full text] [doi: 10.1186/s13063-021-05214-8] [Medline: 33845880]
52. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
53. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov 15;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
54. Pach D, Blödt S, Wang J, Keller T, Bergmann B, Rogge AA, et al. App-based relaxation exercises for patients with chronic neck pain: pragmatic randomized trial. JMIR Mhealth Uhealth 2022 Jan 07;10(1):e31482 [FREE Full text] [doi: 10.2196/31482] [Medline: 34994708]
55. Yuan SLK, Marques AP. Development of ProFibro - a mobile application to promote self-care in patients with fibromyalgia. Physiotherapy 2018 Sep;104(3):311-317. [doi: 10.1016/j.physio.2018.04.005] [Medline: 30025715]
56. Chen T, Or C, Chen J. Effects of technology-supported exercise programs on the knee pain, physical function, and quality of life of individuals with knee osteoarthritis and/or chronic knee pain: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2021 Feb 15;28(2):414-423 [FREE Full text] [doi: 10.1093/jamia/ocaa282] [Medline: 33236109]
57. Dario AB, Moreti Cabral A, Almeida L, Ferreira ML, Refshauge K, Simic M, et al. Effectiveness of telehealth-based interventions in the management of non-specific low back pain: a systematic review with meta-analysis. Spine J 2017 Sep;17(9):1342-1351. [doi: 10.1016/j.spinee.2017.04.008] [Medline: 28412562]
58. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020 Jan 02;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
59. Geraghty AWA, Maund E, Newell D, Santer M, Everitt H, Price C, et al. Self-management for chronic widespread pain including fibromyalgia: a systematic review and meta-analysis. PLoS One 2021 Jul 16;16(7):e0254642 [FREE Full text] [doi: 10.1371/journal.pone.0254642] [Medline: 34270606]
60. van Doormaal MCM, Meerhoff GA, Vliet Vlieland TP, Peter WF. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. Musculoskelet Care 2020 Dec 09;18(4):575-595. [doi: 10.1002/msc.1492] [Medline: 32643252]
61. Kraus VB, Sprow K, Powell KE, Buchner D, Bloodgood B, Piercy K, et al. Effects of physical activity in knee and hip osteoarthritis. Med Sci Sports Exerc 2019;51(6):1324-1339. [doi: 10.1249/mss.0000000000001944]

62. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
63. Bertozzi L, Gardenghi I, Turoni F, Villafañe JH, Capra F, Guccione AA, et al. Effect of therapeutic exercise on pain and disability in the management of chronic nonspecific neck pain: systematic review and meta-analysis of randomized trials. Phys Ther 2013 Aug;93(8):1026-1036. [doi: 10.2522/ptj.20120412] [Medline: 23559524]
64. Price J, Rushton A, Tyros I, Tyros V, Heneghan NR. Effectiveness and optimal dosage of exercise training for chronic non-specific neck pain: a systematic review with a narrative synthesis. PLoS One 2020 Jun 10;15(6):e0234511 [FREE Full text] [doi: 10.1371/journal.pone.0234511] [Medline: 32520970]
65. Sarzi-Puttini P, Giorgi V, Marotto D, Atzeni F. Fibromyalgia: an update on clinical characteristics, aetiopathogenesis and treatment. Nat Rev Rheumatol 2020 Nov 06;16(11):645-660. [doi: 10.1038/s41584-020-00506-w] [Medline: 33024295]
66. Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Häuser W, Fluß E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017 Feb 04;76(2):318-328 [FREE Full text] [doi: 10.1136/annrheumdis-2016-209724] [Medline: 27377815]
67. Winkelmann A, Häuser W, Friedel E, Moog-Egan M, Seeger D, Settan M, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. [Physiotherapy and physical therapies for fibromyalgia syndrome. Systematic review, meta-analysis and guideline]. Schmerz 2012 Jun 5;26(3):276-286. [doi: 10.1007/s00482-012-1171-3] [Medline: 22760460]
68. Eccleston C, Hearn L, Williams A. Psychological therapies for the management of chronic neuropathic pain in adults. Cochrane Database Syst Rev 2015 Oct 29;2015(10):CD011259 [FREE Full text] [doi: 10.1002/14651858.CD011259.pub2] [Medline: 26513427]
69. Bernardy K, Klose P, Welsch P, Häuser W. Efficacy, acceptability and safety of cognitive behavioural therapies in fibromyalgia syndrome - a systematic review and meta-analysis of randomized controlled trials. Eur J Pain 2018 Feb 06;22(2):242-260 [FREE Full text] [doi: 10.1002/ejp.1121] [Medline: 28984402]
70. Mascarenhas RO, Souza MB, Oliveira MX, Lacerda AC, Mendonça VA, Henschke N, et al. Association of therapies with reduced pain and improved quality of life in patients with fibromyalgia: a systematic review and meta-analysis. JAMA Intern Med 2021 Jan 01;181(1):104-112 [FREE Full text] [doi: 10.1001/jamainternmed.2020.5651] [Medline: 33104162]
71. Malde S, Palmisani S, Al-Kaisy A, Sahai A. Guideline of guidelines: bladder pain syndrome. BJU Int 2018 Nov 13;122(5):729-743. [doi: 10.1111/bju.14399] [Medline: 29777618]

Abbreviations

CBT: cognitive behavioral therapy
CLBP: chronic low back pain
CMSP: chronic musculoskeletal pain
CNP: chronic neck pain
CP: chronic pain
CPG: clinical practice guidelines
CPP: chronic pelvic pain
FIQ: Fibromyalgia Impact Questionnaire
FM: fibromyalgia
HOOS: Hip injury and Osteoarthritis Outcome Score
IBS: irritable bowel syndrome
ICD-11: *International Classification of Diseases 11th Revision*
IC/BPS: interstitial cystitis/bladder pain syndrome
KOOS: Knee injury and Osteoarthritis Outcome Score
mHealth: mobile health
NDI: Neck Disability Index
NRS: numeric rating scale
OA: osteoarthritis
OCEBM: Oxford Centre for Evidence-Based Medicine
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
QoL: quality of life
RCT: randomized controlled trial
RoB: risk of bias
SF-36: 36-item Short Form Health Survey
VAS: visual analogue scale
WOMAC: Western Ontario and McMaster

Edited by L Buis; submitted 07.07.22; peer-reviewed by D Lucena, E Sadeghi-Demneh, R de la Vega; comments to author 06.10.22; revised version received 24.11.22; accepted 14.12.22; published 02.02.23.

Please cite as:
Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I
mHealth Intervention for Improving Pain, Quality of Life, and Functional Disability in Patients With Chronic Pain: Systematic Review
JMIR Mhealth Uhealth 2023;11:e40844
URL: https://mhealth.jmir.org/2023/1/e40844
doi:10.2196/40844
PMID:36729570

Review

The Feasibility of Using Smartphone Sensors to Track Insomnia, Depression, and Anxiety in Adults and Young Adults: Narrative Review

Doaa Alamoudi[1], PhD; Emma Breeze[2], BSc; Esther Crawley[3], PhD; Ian Nabney[1], PhD

[1]Department of Computer Science, University of Bristol, Bristol, United Kingdom

[2]Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]Centre for Academic Child Health, University of Bristol Medical School, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Doaa Alamoudi, PhD
Department of Computer Science
University of Bristol
Merchant Venturers' Building
Woodland Road
Bristol, BS8 1UB
United Kingdom
Phone: 44 117 928 3000
Email: d.alamoudi@bristol.ac.uk

Abstract

Background: Since the era of smartphones started in early 2007, they have steadily turned into an accepted part of our lives. Poor sleep is a health problem that needs to be closely monitored before it causes severe mental health problems, such as anxiety or depression. Sleep disorders (eg, acute insomnia) can also develop to chronic insomnia if not treated early. More specifically, mental health problems have been recognized to have casual links to anxiety, depression, heart disease, obesity, dementia, diabetes, and cancer. Several researchers have used mobile sensors to monitor sleep and to study changes in individual mood that may cause depression and anxiety.

Objective: Extreme sleepiness and insomnia not only influence physical health, they also have a significant impact on mental health, such as by causing depression, which has a prevalence of 18% to 21% among young adults aged 16 to 24 in the United Kingdom. The main body of this narrative review explores how passive data collection through smartphone sensors can be used in predicting anxiety and depression.

Methods: A narrative review of the English language literature was performed. We investigated the use of smartphone sensors as a method of collecting data from individuals, regardless of whether the data source was active or passive. Articles were found from a search of Google Scholar records (from 2013 to 2020) with keywords including "mobile phone," "mobile applications," "health apps," "insomnia," "mental health," "sleep monitoring," "depression," "anxiety," "sleep disorder," "lack of sleep," "digital phenotyping," "mobile sensing," "smartphone sensors," and "sleep detector."

Results: The 12 articles presented in this paper explain the current practices of using smartphone sensors for tracking sleep patterns and detecting changes in mental health, especially depression and anxiety over a period of time. Several researchers have been exploring technological methods to detect sleep using smartphone sensors. Researchers have also investigated changes in smartphone sensors and linked them with mental health and well-being.

Conclusions: The conducted review provides an overview of the possibilities of using smartphone sensors unobtrusively to collect data related to sleeping pattern, depression, and anxiety. This provides a unique research opportunity to use smartphone sensors to detect insomnia and provide early detection or intervention for mental health problems such as depression and anxiety if insomnia is detected.

(JMIR Mhealth Uhealth 2023;11:e44123) doi:10.2196/44123

KEYWORDS

mHealth; digital; health; mental health; insomnia; technology; sleep; risk; cardiovascular disease; diabetes; men; mortality; sleep disorder; anxiety; depression; heart disease; obesity; dementia; sensor; intervention; young adult

Introduction

Background

Insomnia is defined as inadequate sleep, with the most common causes being poor sleep conditions and stress [1]. It has also been defined as the presence of long sleep latency, also called sleep onset latency, the elapsed time from being fully awake to sleeping [2]. Sleep latency differs from person to person. Sleep latency and how quickly we reach rapid eye movement (REM) sleep can be indicators of the quantity and quality of sleep. Good sleep quality is measured by the time falling asleep (the ideal is 15 to 20 minutes), the ability to stay asleep all night without waking up, and the ability to spend at least 85% of the time asleep rather than awake [3,4].

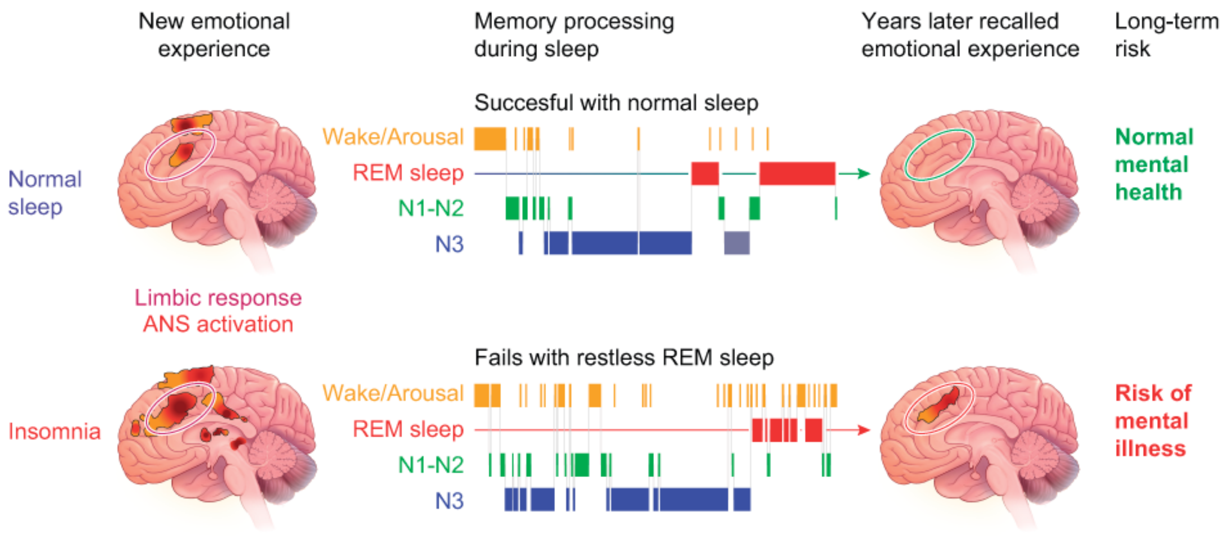
About 40% of people who are diagnosed with insomnia symptoms also report mental health problems [5]. Mental health problems and insomnia are linked in significant ways, where insomnia is a common diagnostic symptom for depression and anxiety [5]. Compared to the longstanding perspective that regarded sleep issues as related to symptoms of mental problems, there is growing research evidence that the relationship between mental disorders and insomnia is problematic and includes bidirectional causation.

The Risk of Insomnia

Mendelson et al [6] were the first to publish findings on this topic; they found that over 90% of depressed patients complained about impaired sleep quality. Several early epidemiological studies found similar strong associations: Ford and Kamerow [7], in 1989, found that people with chronic insomnia were 40 times more likely to have major depression and 6 times more likely to have an anxiety disorder. Mellinger et al [8], in 1985, found that there was a significant association between insomnia and depression. This led to the generally accepted concept that insomnia is one of the core symptoms of psychopathology.

Insomnia and depression share multiple underlying mechanisms. Both conditions have been shown to be triggered by psychosocial stressors, which can then cause overactivity of arousal-inducing neurons in the central nervous system (CNS) compared to the sleep-promoting areas, leading to hyperarousal (Figure 1) [9,10]. Another hypothesis is that insomnia could disrupt synaptic plasticity and neural network function, both of which could precipitate depression [11].

Figure 1. Brain mechanisms of insomnia (adapted from Someren [9], with permission from the American Physiological Society). ANS: autonomic nervous system; REM: rapid eye movement.



More recent studies started to find that insomnia can be an independent indicator for depression. This is highlighted by a 2011 meta-analysis by Baglioni et al [12] of 21 longitudinal epidemiological studies, which found “an overall odds ratio for insomnia to predict depression” of 2.60 (95% CI 1.98-3.42). Results from a 2016 meta-analysis were consistent with these findings (risk ratio 2.27, 95% CI 1.89-2.71) [13]. The recent evidence triggered a change in international guidelines, with insomnia being included in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, a commonly used diagnostic tool, as an independent disorder. This therefore means that insomnia, although still very closely linked to depression, is no longer classified as “primary” or “secondary” and is now considered a disorder in its own right [14,15].

Nyer et al [16] conducted a study to explore the association of sleep disturbance and depressive symptoms in 287 college students with depressive symptoms. The study assumed that students with depressive symptoms and sleep disturbance would have a significant burden of psychiatric symptoms compared to those who had depressive symptoms without sleep disturbance. The study addressed its aims by using different self-report scales, such as the Beck Depression and Anxiety inventory [17,18], the Beck Hopelessness Scale [19], the Anxiety Symptoms Questionnaire [20], the Massachusetts General Hospital Cognitive and Physical Functioning questionnaire [21], and the Quality-Of-Life Enjoyment and Satisfaction questionnaire [22]. For all measures, descriptive statistics showed that the total number of students who had depression with sleep disturbance was 220, and the remaining 67 students had depression without sleep disturbance problems. The results further showed that students who had depressive symptoms with sleep disturbance experienced a significant burden of anxiety symptoms compared to those with depressive symptoms without sleep disturbance [16].

A similar study conducted by Samaranayake et al [23] among university students found that a large number of students were affected by depression and anxiety due to sleep disorders. A total of 1933 students completed a self-report survey. The study found that 39.4% of the students had a sleep disorder lasting over a month. Moreover, depression and anxiety were present in 17.3% and 19.7% of students, respectively, and 7.3% of the students had thought of self-harm.

Studies have shown the close relationship between sleep disorders and mental health problems. Poor sleep quality, defined as waking up frequently during the night, may cause mental health issues such as depression and anxiety. Also, mental health problems could affect an individual's sleep pattern. Anxiety due to worries or repetitive thoughts might keep their brain awake. These symptoms have to occur at least 3 times a week over a period of at least 3 months in order to be diagnosed as insomnia disorder [11]. Furthermore, studies have shown that insomnia is associated with an increased risk of depression.

Methods

The main research criterion we looked at in the publications was the use of smartphone sensors to track sleep. We excluded studies that monitored sleep using other tools, such as wearable devices, electroencephalography headsets, or dream headbands. The reason for excluding these factors was the possibility of finding a method that can track sleep unobtrusively. According to Deloitte's seventh annual mobile consumer survey, around 79% of young adults check their phones before going to sleep [24].

We looked at different studies that monitored sleep using smartphone sensors. The total sample population was up to 205 university students and adults. The designed studies used the Android operating system to collect the data. We categorized the studies based on whether they monitored sleep with a single or multiple sensors.

Anxiety and depression are common, with a prevalence of 18% to 21% among young adults aged 16 to 24 in the United Kingdom [25]. As we wanted to study the relationship between poor sleep and anxiety and depression, we looked to a range of different studies that used mobile sensors to detect individuals' behavior. They predicted changes in mood using behavioral signals such as location, mobility, speech pattern, phone use, and activities. We excluded papers that discussed other mental health issues, such as bipolar disorder or schizophrenia, and articles that examined changes in mood using other methods, such as physiological and social signals.

Results

In the field of measuring sleep via smartphone sensors, we found a total of 7 published articles that used mobile sensors to passively collect data [8,14,26-30]. A total of 5 articles discussed the use of mobile sensors to track depression and anxiety [31-35].

Smartphone Sensors for Sleep Detection

With the widening reach of technology, several groups have carried out sleep studies relying on smartphone sensors [36-38]. The accelerometer embedded in every smartphone has been used to measure phone (hence body) movement to understand sleep stage [27]. Room environment conditions can be measured and monitored by other smartphone sensors to estimate sleep quality. Room environmental variables that can be measured by smartphone sensors to assess the quality of the environment include noise [27,39,40], luminosity or darkness [40], and temperature [41]. Screen on/off timing has been used in several studies as an indicator of sleeping time and duration [27,28,30]. Microphone sensors can also measure an individual's snoring, which might impact their sleep quality [26]. Snoring can lead to fragmented and unrefreshing sleep, which often causes poor daytime functioning (tiredness and sleepiness) [41]; moreover, snoring is common among people aged 60 years or older [42,43].

Single-Sensor Method

The iSleep app, which was developed by Hao et al [26], aims to replace wearable devices in sleep tracking by relying on the built-in microphone. The algorithm was able to classify several types of extraneous noise, such as that from fans and air conditioners. However, in order to collect sleep events, the user is required to turn on the app and place the phone in a certain location, and this requires user intervention on a daily basis. That means that in order to calculate sleep duration, the user needs to manually start and close the app. Two matrices (snoring and coughing detection) are used to evaluate the accuracy of the app. During the experiment, the smartphone or tablet is used to collect acoustic data and should be placed 1.5 meters away. A second omnidirectional microphone attached on the headboard and connected to a laptop collects high-quality audio. The microphone should be placed 5 meters away from the bed to avoid noises coming from the laptop fan that might impact the quality of the collected data. Additionally, an iPod is used to record any movement on the bed. The iSleep app uses microphone sensors to collect data related to sleep pattern without considering privacy-sensitive data sets from the user's smartphone. However, the complicated procedure used to detect sleeping patterns will be difficult to implement in daily life, as it requires a detailed room setup and user intervention, which may be difficult for young adults. Also, although the app has been tested among young adults aged between 20 and 30 years, the prevalence of snoring among young adults is only around 10%, although it increases with age [42,43].

A “tappigraphy” sensor has also been used to measure sleeping patterns [29]. Tappigraphy involves measuring touchscreen events and comparing them with both actigraphy and a daily sleep diary. The study design is based on calculating the number of times the user touches their phone and follows a 24-hour sleep-wake cycle. The longest period of not using the phone is considered to be sleeping time. Tappigraphy overestimates sleeping time compared to actigraphy when a low number of touchscreen touches are measured per day [29]. The sample size of this study was 79 users aged between 16 and 45 years. The large variation in the subjects' age may have made the data collected an inaccurate reflection of reality, as students and workers have different phone-use styles. Usually, older people spend less time using their phone compared to younger people [44]. The 24-hour sleep-wake cycle might also not have been accurate, especially if the user was a worker or a student, as sometimes, they may need to avoid using their phone during working or class time, affecting the accuracy of predictions of sleeping time, whether this was during the day or the night.

Mobile apps that use screen on/off events to measure sleeping pattern have been discussed in several research papers [28,30,38]. A study based on computing circadian rhythm focused on detecting the chronotype, which means the activity and sleep preferences of an individual within a 24-hour period, and the impact of social jet lag on sleep duration using an unobtrusive and low-cost method based on on/off screen sensors among 9 subjects aged between 19 and 25 years [28]. In order to obtain accurate results, the sleep duration was calculated using a ground truth determined by sleep diaries that were collected daily from participants. When comparing the app performance with the sleep-diary data, there was less than 45 minutes of error. The study showed that participants with an early chronotype experienced the most social jet lag, due to social pressure applied by people with later chronotypes on weekends. However, the study was only interested in determining sleep onset that happened between late night and early morning, without taking into consideration individual preferences. For example, international students who have different sleep times in their home country may adjust their sleep time to be able to contact their family and friends when they are available.

The Know Addiction app [38] was developed to monitor the link between circadian rhythm and individual sleeping patterns using smartphone screen on/off events during a predetermined sleeping window; this window was the period between 10 PM and 10 AM among 61 subjects who were non–shift workers and aged between 20 and 56 years. The app collected total sleep duration according to different parameters. The 3 parameters that were used to measure sleeping pattern were as follows: reactive use episode, proactive use episode, and nonuse episode. A reactive use episode was defined as any notification within 1 minute prior to a screen-on event. In contrast, a proactive use episode was defined as no notifications within 1 minute prior to a screen-on event. The app excluded all reactive use episodes in the sleep indicators calculation. The third parameter was nonuse episodes, defined as a screen on/off event. Sleeping time was determined by measuring the maximal nonuse episode within the predetermined sleeping window. Overall, the study results showed that total daily duration of smartphone use was statistically significantly correlated with delayed sleep onset (correlation 0.0808, 95% CI 0.0434-0.1182; $P<.05$) The study was limited to tracking sleeping patterns and did not attempt to measure other health or mental health issues. However, the predetermined sleeping window of 10 PM to 10 AM could not accurately estimate all individuals' sleep duration, as having mental health problems such as depression or anxiety means that sleep may not occur during the defined sleeping time window. An individual with depression symptoms would prefer to sleep for a longer time than the estimated sleeping time window.

The iSenseSleep app [30] lists all screen on/off events and then analyzes them to estimate sleep duration; it was validated in different groups (4 working mothers and 10 university students). The algorithm provides a list of all time points related to sleep episodes; the iSenseSleep app then considers the longest period to be the sleeping time, ignoring any disturbance that is less than 5 minutes. The app was designed to estimate sleeping time occurring during the night from 10 PM onwards. The app was designed to estimate sleep duration during the normal sleeping time, ignoring those who have changes in their sleeping pattern, for example, weekdays versus weekend days, where the sleeping time may differ. The iSenseSleep app estimates the wake-up time by checking the last screen-on event in the morning that was at least 4 hours since the last screen-off event. The app was able to predict sleep duration with an average error of 24 (SD 17) minutes (7%, SD 4% of the total duration). The estimation of sleep duration was more accurate among university students than working mothers, with an average error of 68 minutes (17% of the total) and 83 minutes (20% of the total), respectively. However, the iSenseSleep app was only used for 2 days to predict the sleeping pattern of the user for the rest of the nights, and that may not reflect the reality of changes in their sleeping pattern over time, especially for people who have problems with depression and anxiety.

Although the above studies conducted their research to understand individual sleeping patterns based on screen on/off events and tappigraphy, sleep quality was not determined in these studies. Sleep quality measures disturbance time, which is when the user wakes up in the middle of the night before going back to sleep. The user can also be asked to enter their sleeping time for working days and weekend days, so their sleeping pattern and sleep quality can be accurately predicted.

Multiple Sensor Modalities

Chen et al [27] developed Best Effort Sleep (BES), an Android mobile app to estimate sleep duration without user intervention by collecting data from multiple mobile sensors, including the light sensor and the microphone, as well from phone use and whether the device was in stationary mode (ie, not moving). The authors also developed another mobile app, known as Sleep With the Phone (SWP). SWP was developed to collect sleeping data using the accelerometer, with a strict protocol the user needs to follow in order to ensure accuracy while collecting the data, which includes placing the phone on the pillow when the user intends to sleep. Over 1 week with 8 participants, the accuracy of the BES app was tested with SWP and other commercial wearable systems, such as a Jawbone wristband and Zeo headband. When evaluating the accuracy of sleep monitoring using BES, SWP, Jawbone, and Zeo, the authors found that BES achieved a sleep duration error of plus or minus 42 minutes. The use of BES can be considered as an ideal approach to sleep monitoring in terms of its low cost and reduced need for user intervention to record the data, avoiding putting a burden on the user. However, room environment observations might not be considered a good predictor to rely on when predicting sleeping time, as some people may sleep while the room light is turned on. In addition, in the case that the user forgets to recharge the phone, the app will consider that the user is sleeping, as it is in the phone-off mode. The study showed that light and phone-off features contributed to lowering error.

Toss ’N’ Turn (TNT) is an Android mobile app developed to investigate how smartphones can detect bedtime, wake time, and sleep duration without requiring changes in people’s behavior and thus estimate the regularity of sleep quality [36]. A total of 27 participants were studied who were aged between 20 and 59 years. The algorithm observes the sensor logs for the accelerometer, screen on/off events, light, microphone, and battery in a 10-minute window to classify a sleep or not-sleep state. It then eliminates possible sleep detection errors, such as noise or disturbance states, between quiet and stationary situations. The app produces errors of plus or minus 35 minutes, 31 minutes, and 49 minutes in detecting bedtime, wake time, and sleep duration, respectively. However, when the mobile battery is low, the app will stop collecting data, and using multiple sensors can reduce the battery life. The study showed that changes in the screen status (ie, on/off events) and accelerometer were related to wake time, while changes in the light sensor were not always related.

Both studies (ie, BES and TNT) used multiple sensors to understand sleeping patterns, but apps that are based on multiple sensors can reduce the battery life. In addition, both studies showed that the presence of an accelerometer and screen on/off events provided good results in predicting sleeping time, while light sensors could vary in predicting the sleeping pattern. Moreover, the studies used microphone sensors to collect data without leveraging the privacy and sensitivity of the collected data from the users’ smartphones.

Table 1 shows a summary of methods to determine sleep duration; column 3 shows the sample size for validation and column 4 shows the duration of the validation study.

Earlier studies have used unobtrusive methods to predict sleeping patterns without user intervention and have used various sensors to predict sleeping time and various techniques to determine sleep duration and quality. Moreover, earlier papers [28,30,38] have shown that sleep duration, bedtime, and wake time could be identified over a significant period with screen on/off events instead of complex sensors that require additional software, use protocols, or collect sensitive data, such as from the microphone.

Table 1. Studies of mobile sensors to monitor sleep.

Study characteristics				Sensors used						
Authors	Year	Sample size, n	Study length, days	Accelerometer	Screen on/off event	Light sensor	Stationary mode	Microphone	Battery	Touch screen event
Hao et al [26]	2013	7	51					✓[a]		
Borger et al [29]	2019	79	1400							✓
Abdullah et al [28]	2014	9	97		✓					
Lin et al [38]	2019	61	14		✓					
Ciman and Wac [30]	2019	14	180		✓					
Chen et al [27]	2016	8	7		✓	✓	✓	✓		
Min et al [36]	2014	27	30	✓	✓	✓		✓	✓	

[a]✓: indicates the type of sensor used in the study.

Smartphone Sensors for Detecting Depression and Anxiety

The interest in studying the effectiveness of using smartphones for tracking individuals’ sleeping patterns and activities and the relationship with mental well-being has increased at a brisk pace. Over this period of technological advancement, a considerable amount of literature has been published using mobile sensors to categorize mental health and well-being. These studies have used mobile sensors to collect behavioral signals and later relate them to mental health problems.

GPS has been used to study mental health problems such as depression [32,33,35]. These studies discussed a correlation between physical activities and mental health problems. DeMasi et al [32] aimed to track depression symptoms using GPS. Sleep duration was measured by estimating the longest period an individual was not physically active after 9 PM and individual physical activities during the day. A total of 47 undergraduate students installed the app over an 8-week period and completed a self-report survey related to depression and bipolar symptoms. The study showed a positive correlation between physical activities and estimated mental health status. There were limitations to activity recognition, especially that the smartphone was not in a fixed position, participants performed nonstandard activities, and the phones were set down, such as when they were left in a gym locker. Sleep duration and sleep disturbance could not be identified with a single sensor such as GPS.

Saeb et al [33] performed a study of 40 adult subjects over a 2-week period to detect daily life behavior and depression symptoms using GPS, including circadian movement (over 24 hours), location variance, and use of phone features. The data were collected and compared with self-report surveys. The accuracy achieved from this study in predicting depression symptoms through GPS and phone use was 86.5%. A similar study was conducted to study changes in mental health by tracking individuals' activities and sleeping patterns using mobile sensors [31]. This study aimed to understand changes in depression and stress level using data collected from smartphone sensors. The data were collected from 47 young adults over a 10-week period. GPS was used to track individual daily activities while sleep duration was tracked based on light, microphone data, mobile use, and accelerometer data. The study used the algorithm developed for BES [27] to calculate sleep duration. Self-report surveys of stress and depression were collected on a daily basis. The result of the analysis showed a correlation between individual activities and sleep duration with daily stress and depression level.

StudentLife is an app that is run on smartphones and wearable devices [31,34]. The study was conducted for a 10-week period among a single class of 48 students whose age was between 19 and 30 years. The study aimed to track the engagement and performance of students on an individual level. StudentLife used different types of smartphone sensors, such as the accelerometer, microphone, light sensor, and GPS [31,34]. The microphone was coded to capture sounds every 2 minutes. In contrast, GPS was activated every 10 minutes to calculate the total daily distance moved by the individual. The accelerometer embedded in the smartphone was used to detect the ratio of movement versus being stationary. The study used data retrieved from device lock duration, the accelerometer, the microphone, and the light sensor to calculate sleep duration. A self-report was used to measure mental health status and depression (with the Patient Health Questionnaire-9) [45], perceived stress [46], flourishing [47], and loneliness scales [48]. However, the study did not consider that not carrying the phone during the day would prevent the app from accurately predicting data. For example, if the user left the phone at home, all the data from the microphone and accelerometer would not be collected on that day, and the system would assume that the person was stationary and quiet, which was interpreted as sleeping. In consequence, the predicted mental health state would not be accurate. Other difficulties arose from the app measuring sleep duration using the light level. If the user was from a high-latitude location, which is dark for most of the day, then the app would consider the dark time to be sleeping time. Also, if the user preferred to sleep while keeping the room light on, it would not predict bedtime and sleeping duration accurately.

From the previous studies, we can see that the proliferation of digital technologies and mobile sensors can provide a feasible and unobtrusive method to continuously collect behavioral data from individuals, which can help in better understanding the mental health condition of individuals. Using accelerometers and other phone [31,33] features has been shown to be an efficient way of understanding individuals' behavior and mental well-being.

Discussion

Principal Findings

Mental health problems and sleep are linked in significant ways. Compared to the longstanding perspective that sleep issues are symptoms of mental problems, there is growing research evidence that the relationship between mental disorders and sleep is problematic and includes bidirectional causation.

Many people may not be aware of how their sleeping pattern can impact their health and well-being, seeking treatment only when physical and mental symptoms have started to manifest. Furthermore, both children and adults are reluctant to seek help, with only 1 in 3 receiving treatments for common mental health problems. These reasons, along with the increasing difficulty of accessing primary care services, leave room for an alternative method of insomnia identification. Smartphone sensor technologies in users' phones may be a suitable method to track sleeping patterns and early sleep disorders. Using these technologies may prevent more serious symptoms arising.

Research has demonstrated the effectiveness of using mobile phone sensors to record personal data and predict mental health and well-being. Several apps have been developed to track behavioral signals and link them with individual mental health and well-being. These apps are based on wearable and nonwearable devices that collect data using accelerometers, microphones, light sensors, and screen on/off events.

Conclusion

This review describes the effectiveness of several sleep apps that have been used to track insomnia, which can cause depression and anxiety. Furthermore, the studies in this review found that using smartphone sensors to detect mental health problems can be useful for monitoring behavioral patterns that can cause depressive symptoms. Further study is needed in this area to understand the feasibility of using mobile sensors to track sleep disorders and provide early intervention and treatment when insomnia is detected, so as to reduce mental health problems.

Conflicts of Interest

None declared.

References

1. Insomnia | definition, types and symptoms. Britannica. URL: https://www.britannica.com/science/insomnia [accessed 2022-08-31]
2. Hirshkowitz M, Sarwar A, Sharafkhaneh A. Evaluating sleepiness. In: Principles and Practice of Sleep Medicine, Fifth Edition. Amsterdam, Netherlands: Elsevier; 2010:1624-1631.
3. Sleep Quality. The Sleep Charity. URL: https://sleepcouncil.org.uk/advice-support/sleep-hub/sleep-matters/sleep-quality/ [accessed 2020-10-14]
4. How To Determine Poor Sleep Quality. Sleep Foundation. URL: https://www.sleepfoundation.org/articles/5-signs-your-sleep-quality-poor-and-how-fix-it [accessed 2020-10-14]
5. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 2007 Aug 15;3 [FREE Full text] [doi: 10.5664/jcsm.26929] [Medline: 17824495]
6. Mendelson WB, Gillin JC, Wyatt RJ. Sleep: Human Sleep and Its Disorders. New York, NY: Springer; 1978.
7. Ford DE, Kamerow DB. Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention? JAMA 1989 Sep 15;262(11):1479-1484. [doi: 10.1001/jama.262.11.1479] [Medline: 2769898]
8. Mellinger GD, Balter MB, Uhlenhuth EH. Insomnia and its treatment. Prevalence and correlates. Arch Gen Psychiatry 1985 Mar;42(3):225-232. [doi: 10.1001/archpsyc.1985.01790260019002] [Medline: 2858188]
9. Van Someren EJW. Brain mechanisms of insomnia: new perspectives on causes and consequences. Physiol Rev 2021 Jul 01;101(3):995-1046 [FREE Full text] [doi: 10.1152/physrev.00046.2019] [Medline: 32790576]
10. Riemann D, Spiegelhalder K, Feige B, Voderholzer U, Berger M, Perlis M, et al. The hyperarousal model of insomnia: a review of the concept and its evidence. Sleep Med Rev 2010 Feb;14(1):19-31. [doi: 10.1016/j.smrv.2009.04.002] [Medline: 19481481]
11. Riemann D, Krone LB, Wulff K, Nissen C. Sleep, insomnia, and depression. Neuropsychopharmacology 2020 Jan;45(1):74-89 [FREE Full text] [doi: 10.1038/s41386-019-0411-y] [Medline: 31071719]
12. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
13. Li L, Wu C, Gan Y, Qu X, Lu Z. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. BMC Psychiatry 2016 Nov 05;16(1):375 [FREE Full text] [doi: 10.1186/s12888-016-1075-3] [Medline: 27816065]
14. Insomnia. National Institute for Health and Care Excellence. URL: https://cks.nice.org.uk/topics/insomnia/background-information/definition/ [accessed 2022-06-16]
15. Seow LSE, Verma SK, Mok YM, Kumar S, Chang S, Satghare P, et al. Evaluating DSM-5 insomnia disorder and the treatment of sleep problems in a psychiatric population. J Clin Sleep Med 2018 Feb 15;14(2):237-244 [FREE Full text] [doi: 10.5664/jcsm.6942] [Medline: 29394962]
16. Nyer M, Farabaugh A, Fehling K, Soskin D, Holt D, Papakostas GI, et al. Relationship between sleep disturbance and depression, anxiety, and functioning in college students. Depress Anxiety 2013 Sep;30(9):873-880 [FREE Full text] [doi: 10.1002/da.22064] [Medline: 23681944]
17. Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. Beck Depression Inventory. APA PsycNet. URL: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/beck-depression [accessed 2022-09-08]
18. Beck A, Epstein N, Brown G, Steer R. Beck Anxiety Inventory. American Psychological Association. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft02025-000 [accessed 2022-09-08]
19. Beck AT, Weissman A, Lester D, Trexler L. The measurement of pessimism: the hopelessness scale. J Consult Clin Psychol 1974 Dec;42(6):861-865. [doi: 10.1037/h0037562] [Medline: 4436473]
20. Baker A, Simon N, Keshaviah A, Farabaugh A, Deckersbach T, Worthington JJ, et al. Anxiety Symptoms Questionnaire (ASQ): development and validation. Gen Psychiatr 2019;32(6):e100144 [FREE Full text] [doi: 10.1136/gpsych-2019-100144] [Medline: 31922090]
21. Fava M, Iosifescu DV, Pedrelli P, Baer L. Reliability and validity of the Massachusetts general hospital cognitive and physical functioning questionnaire. Psychother Psychosom 2009;78(2):91-97. [doi: 10.1159/000201934] [Medline: 19218827]
22. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321-326. [Medline: 8290681]
23. Samaranayake C, Arroll B, Fernando A. Sleep disorders, depression, anxiety and satisfaction with life among young adults: a survey of university students in Auckland, New Zealand. N Z Med J 2014 Aug 01;127(1399):13-22. [Medline: 25145302]
24. Dewa LH, Lavelle M, Pickles K, Kalorkoti C, Jaques J, Pappa S, et al. Young adults' perceptions of using wearables, social media and other technologies to detect worsening mental health: A qualitative study. PLoS One 2019;14(9):e0222655 [FREE Full text] [doi: 10.1371/journal.pone.0222655] [Medline: 31532786]
25. Young people's wellbeing. Office for National Statistics. 2017. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/youngpeopleswellbeingandpersonalfinance/2017 [accessed 2020-04-28]
26. Hao T, Xing G, Zhou G. iSleep: Unobtrusive sleep quality monitoring using smartphones. In: SenSys '13: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Presented at: SenSys '13: The 11th ACM Conference on Embedded Network Sensor Systems; November 11-15, 2013; Rome, Italy p. 1-14. [doi: 10.1145/2517351.2517359]

27. Chen Z, Lin M, Chen F. Unobtrusive sleep monitoring using smartphones. 2013 May 5 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; May 5-8, 2013; Venice, Italy p. 145. [doi: 10.4108/icst.pervasivehealth.2013.252148]
28. Abdullah S, Matthews M, Murnane E. Towards circadian computing: "early to bed and early to rise" makes some of us unhealthy and sleep deprived. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; Sept 13-17, 2014; Seattle, WA p. 673-684. [doi: 10.1145/2632048.2632100]
29. Borger JN, Huber R, Ghosh A. Capturing sleep-wake cycles by using day-to-day smartphone touchscreen interactions. NPJ Digit Med 2019;2:73 [FREE Full text] [doi: 10.1038/s41746-019-0147-4] [Medline: 31372507]
30. Ciman M, Wac K. Smartphones as sleep duration sensors: validation of the iSenseSleep algorithm. JMIR Mhealth Uhealth 2019 May 21;7(5):e11930 [FREE Full text] [doi: 10.2196/11930] [Medline: 31115341]
31. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
32. DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B. Well-being tracking via smartphone-measured activity and sleep: cohort study. JMIR Mhealth Uhealth 2017 Oct 05;5(10):e137 [FREE Full text] [doi: 10.2196/mhealth.7820] [Medline: 28982643]
33. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
34. Wang R, Chen F, Chen Z. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; September 13-17, 2014; Seattle, WA p. 14. [doi: 10.1145/2632048.2632054]
35. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Predicting depressive symptoms using smartphone data. Smart Health 2020 Mar;15:100093. [doi: 10.1016/j.smhl.2019.100093]
36. Min J, Doryab A, Wiese J. Toss 'n' turn: smartphone as sleep and sleep quality detector. In: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, ON p. 477. [doi: 10.1145/2556288.2557220]
37. Wilde-Frenz J, Schulz H. Rate and distribution of body movements during sleep in humans. Percept Mot Skills 1983 Feb;56(1):275-283. [doi: 10.2466/pms.1983.56.1.275] [Medline: 6844073]
38. Lin Y, Wong B, Lin S, Chiu Y, Pan Y, Lee Y. Development of a mobile application (app) to delineate "digital chronotype" and the effects of delayed chronotype by bedtime smartphone use. J Psychiatr Res 2019 Mar;110:9-15. [doi: 10.1016/j.jpsychires.2018.12.012] [Medline: 30611008]
39. Fietze I, Barthe C, Hölzl M, Glos M, Zimmermann S, Bauer-Diefenbach R, et al. The effect of room acoustics on the sleep quality of healthy sleepers. Noise Health 2016;18(84):240-246 [FREE Full text] [doi: 10.4103/1463-1741.192480] [Medline: 27762252]
40. External Factors that Influence Sleep. Harvard Medical School. URL: http://healthysleep.med.harvard.edu/healthy/science/how/external-factors [accessed 2020-04-16]
41. Snoring. Sleep Disorders Guide. URL: https://www.sleepdisordersguide.com/article/sleep-disorders/snoring-stop-snoring-causes-of-snoring-snoring-remedies [accessed 2020-05-27]
42. Adebusoye LA, Ogunbode AM, Olowookere OO. Factors associated with reported snoring among elderly patients attending the geriatric centre in Nigeria. Pan Afr Med J 2014;19:309 [FREE Full text] [doi: 10.11604/pamj.2014.19.309.5244] [Medline: 25883736]
43. Çelik A, Işik A, Ural A, Arslan S, Bahadir O, İmamoğlu M. Prevalence and risk factors of snoring, obstructive sleep apnea symptoms, and excessive daytime somnolence in trabzon. KBB-Forum 2010;9(4):78-87.
44. Howarth J. Time Spent Using Smartphones (2022 Statistics). Exploding Topics. URL: https://explodingtopics.com/blog/smartphone-usage-stats#smartphone-usage-by-demographic [accessed 2022-09-08]
45. Depression Test Questionnaire. Egton Medical Information Systems. URL: https://patient.info/doctor/patient-health-questionnaire-phq-9 [accessed 2020-04-17]
46. Perceived Stress Scale. APA PsycNet. 1986. URL: https://psycnet.apa.org/doiLanding?doi=10.1037/t02889-000 [accessed 2020-05-27]
47. Diener E, Biswas-Diener R. Flourishing Scale. Measurement Instrument Database for the Social Sciences. 2009. URL: https://www.midss.org/sites/default/files/fs.pdf [accessed 2020-05-27]
48. Weeks DG, Michela JL, Peplau LA, Bragg ME. Relation between loneliness and depression: a structural equation analysis. J Pers Soc Psychol 1980 Dec;39(6):1238-1244. [doi: 10.1037/h0077709] [Medline: 7205551]

Abbreviations

BES: Best Effort Sleep
CNS: central nervous system
REM: rapid eye movement
SWP: Sleep With the Phone
TNT: Toss 'N' Turn

Edited by L Buis; submitted 07.11.22; peer-reviewed by T Martino, K Gupta, OR Patil; comments to author 03.01.23; revised version received 25.01.23; accepted 03.02.23; published 17.02.23.

Please cite as:
Alamoudi D, Breeze E, Crawley E, Nabney I
The Feasibility of Using Smartphone Sensors to Track Insomnia, Depression, and Anxiety in Adults and Young Adults: Narrative Review
JMIR Mhealth Uhealth 2023;11:e44123
URL: https://mhealth.jmir.org/2023/1/e44123
doi:10.2196/44123
PMID:36800211

Review

Effectiveness of Remote Fetal Monitoring on Maternal-Fetal Outcomes: Systematic Review and Meta-Analysis

Suya Li[1*], MSN; Qing Yang[1*], MSN; Shuya Niu[2], BS; Yu Liu[1], BS

[1]Nursing Department, Tongji Hospital, Tongji Medical College, HuaZhong University of Science and Technology, Wuhan, China

[2]Zhongnan Hospital of Wuhan University, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Yu Liu, BS
Nursing Department
Tongji Hospital, Tongji Medical College
HuaZhong University of Science and Technology
No. 1095 Jiefang Rd
Wuhan, 430030
China
Phone: 86 13995579713
Email: hust512@sohu.com

Abstract

Background: To solve the disadvantages of traditional fetal monitoring such as time-consuming, cumbersome steps and low coverage, it is paramount to develop remote fetal monitoring. Remote fetal monitoring expands time and space, which is expected to popularize fetal monitoring in remote areas with the low availability of health services. Pregnant women can transmit fetal monitoring data from remote monitoring terminals to the central monitoring station so that doctors can interpret it remotely and detect fetal hypoxia in time. Fetal monitoring involving remote technology has also been carried out, but with some conflicting results.

Objective: The review aimed to (1) examine the efficacy of remote fetal monitoring in improving maternal-fetal outcomes and (2) identify research gaps in the field to make recommendations for future research.

Methods: We did a systematic literature search with PubMed, Cochrane Library, Web of Science, Embase, MEDLINE, CINAHL, ProQuest Dissertations and Theses Global, ClinicalTrials.gov, and Open Grey in March 2022. Randomized controlled trials or quasi-experimental trials of remote fetal monitoring were identified. Two reviewers independently searched articles, extracted data, and assessed each study. Primary outcomes (maternal-fetal outcomes) and secondary outcomes (health care usage) were presented as relative risks or mean difference. The review was registered on PROSPERO as CRD42020165038.

Results: Of the 9337 retrieved literature, 9 studies were included in the systematic review and meta-analysis (n=1128). Compared with a control group, remote fetal monitoring reduced the risk of neonatal asphyxia (risk ratio 0.66, 95% CI 0.45-0.97; P=.04), with a low heterogeneity of 24%. Other maternal-fetal outcomes did not differ significantly between remote fetal monitoring and routine fetal monitoring, such as cesarean section (P=.21; I^2=0%), induced labor (P=.50; I^2=0%), instrumental vaginal birth (P=.45; I^2=0%), spontaneous delivery (P=.85; I^2=0%), gestational weeks at delivery (P=.35; I^2=0%), premature delivery (P=.47; I^2=0%), and low birth weight (P=.71; I^2=0%). Only 2 studies performed a cost analysis, stating that remote fetal monitoring can contribute to reductions in health care costs when compared with conventional care. In addition, remote fetal monitoring might affect the number of visits and duration in the hospital, but it was not possible to draw definite conclusions about the effects due to the limited number of studies.

Conclusions: Remote fetal monitoring seems to reduce the incidence of neonatal asphyxia and health care costs compared with routine fetal monitoring. To strengthen the claims on the efficacy of remote fetal monitoring, further well-designed studies are necessary, especially in high-risk pregnant women, such as pregnant women with diabetes, pregnant women with hypertension, and so forth.

(JMIR Mhealth Uhealth 2023;11:e41508) doi:10.2196/41508

KEYWORDS

remote fetal monitoring; maternal outcomes; fetal outcomes; review

Introduction

Fetal safety has always been a top priority for perinatal care. According to the World Health Organization, as of 2019, there were an estimated 2 million stillbirths, most of which can be prevented by safe and quality care, timely emergency care, and accurate recording [1]. Fetal monitoring is the primary means of monitoring to assess fetal safety and contributes to reducing the risk of stillbirth by detecting fetal hypoxia as early as possible [2,3]. Previous studies have repeatedly demonstrated the clinical value of fetal monitoring in reducing adverse perinatal outcomes (eg, neonatal cerebral palsy, hypoxic-ischemic encephalopathy, or stillbirth) [4,5].

Traditional antenatal care is resource intensive and not friendly to underserved settings. Beyond that, routine prenatal monitoring is only suitable for hospital settings, which means that pregnant women require regular outpatient follow-up [6]. Recurrent outpatient visits also pose additional travel risks (eg, falls, collisions, and bumps), especially for high-risk pregnant women. Telemedicine refers to the long-distance transmission of medical information between medical workers and patients through telecommunication technology [7], which has many potential advantages such as reducing outpatient time, alleviating the shortage of medical resources, reducing transportation costs and medical costs, and so forth [8-10]. Remote monitoring using telephones, websites, portable devices, and so forth during pregnancy is becoming more and more popular [11,12].

Systematic reviews have demonstrated the feasibility and superiority of telemedicine in obstetrics [13], focusing on blood pressure (BP) management [14,15], blood glucose management [16], and weight management [17] during pregnancy. However, we are not yet clear about the benefits or dangers of remote fetal monitoring. The primary objective of this systematic review was to assess the effectiveness of remote fetal monitoring for improving maternal-fetal outcomes. In addition, we also sought to analyze the cost-effectiveness of remote fetal monitoring compared to conventional prenatal monitoring.

Methods

Reporting Standards

This systematic review and meta-analysis was carried out according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines of 2009 [18] and was registered on PROSPERO as CRD42020165038.

Literature Retrieval

In total, 9 web-based databases were searched in March 2022, including PubMed (January 1966-March 2022), Cochrane Library (January 1947-March 2022), Web of Science (January 1990-Mar 2022), Embase (January 1974-March 2022), MEDLINE (January 1950-March 2022), CINAHL (January 1982-March 2022), ProQuest Dissertations and Theses Global (January 1899-March 2022), ClinicalTrials.gov (January 1997-March 2022), and Open Grey (January 1980-March 2022). Search terms generated from the inspection of relevant papers were wielded to search for eligible studies, such as fetal, remote, telemetry, monitor, and so forth. The full search strategy was available in Multimedia Appendix 1 and was rerun before the final analysis.

Inclusion Criteria

Studies were considered eligible if they simultaneously met the following criteria: (1) pregnant women; (2) randomized controlled trials (RCTs) or quasi-experimental trials; (3) fetal monitoring data were transmitted to the central monitoring station by remote monitoring terminals; and (4) outcomes included at least 1 maternal-fetal outcome or health resource usage. There were no restrictions on language, nationality, or publication status.

Exclusion Criteria

Studies meeting any of the following criteria were excluded: (1) no control group in the study; (2) comparative studies of 2 or more remote monitoring technologies; and (3) the full text was still unavailable after contacting the original authors. Studies were not excluded due to monitoring settings (hospital, home, community setting, or mixed).

Outcome Measures

The primary outcomes were maternal-fetal outcomes (cesarean section, induced labor or miscarriage, instrumental vaginal birth, spontaneous delivery, gestational weeks at delivery, premature delivery, birth weight, and so on). The secondary outcomes were health care usage, which was assessed by on-site appointments, home visits, duration in the hospital, prenatal costs, and so on.

Study Selection

A 3-step screening identified articles that met the inclusion and exclusion criteria were literature retrieval, preliminary screening (title and abstract), and full-text screening. Literature retrieval was conducted by 2 investigators. All searched articles were uploaded into the reference management tool of EndNote. Articles with the same author, year, title, and so on were identified and removed by EndNote. Subsequently, 2 independent investigators (SYL and QY) selected all articles by evaluating the title and abstract after the removal of duplicates. Finally, the same 2 investigators (SYL and QY) identified the ultimately eligible articles by screening independently the full text according to the inclusion and exclusion criteria. In addition, the first author (SYL) hand-searched the references of the ultimately included literature to identify further publications. Any discrepancies and disagreements were finally resolved by consultation with a third reviewer (YL). We also contacted the original authors for verification if there were any uncertain technical types.

Data Extraction

Data from included studies were extracted by SYL and then cross-checked by another author (QY). A standardized data extraction form was designed by the research team and included the following data: (1) basic information of included studies (first author, year of publication, country, and study design); (2) characteristics of participants (maternal age, gestational weeks, sample size, and attrition rate); (3) characteristics of interventions (trial settings, duration of the intervention, monitoring personnel, monitoring content, feedback types, and technical support); and (4) outcomes measurement (maternal-fetal outcomes and health care usage). For insufficient data, we contacted the original authors via email. The standardized data extraction form was available in Multimedia Appendix 2.

Quality Assessment

Independently, the quality of eligible studies was assessed by 2 reviewers (SYL and QY) according to the Cochrane Risk of Bias Tool [19], which consisted of 7 items (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias) with the responses of “low risk,” “high risk,” and “unclear risk.” The research was considered high quality with a low risk score on at least 4 domains, which must include 3 key domains (random sequence generation, allocation concealment, and incomplete outcome data) [20]. Consensuses between 2 investigators (SYL and QY) were reached by discussion with a third reviewer (YL).

Data Synthesis and Statistical Analysis

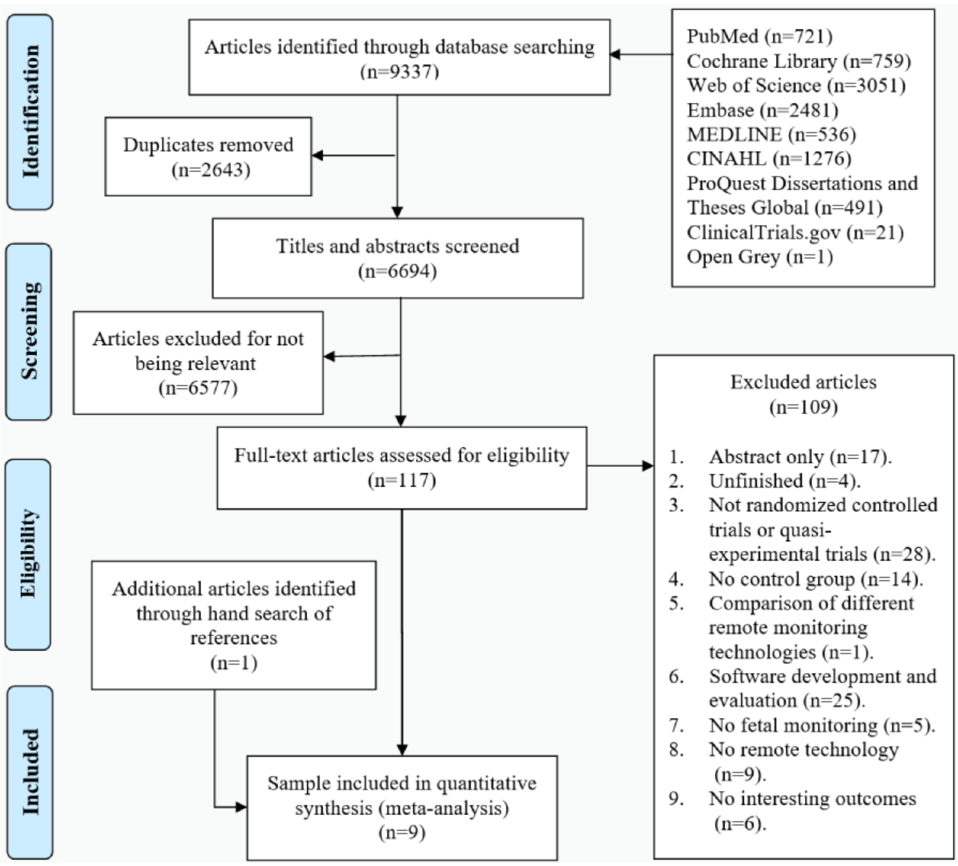
Quantitative analysis of included studies was carried out in Review Manager (RevMan) software (version 5.4). Continuous variables were presented as mean difference (MD), and dichotomous variables were described as risk ratio (RR) with a 95% CI. The statistical heterogeneity of selected studies was assessed by the chi-square test combined with I^2. Heterogeneity was divided into nonsignificant heterogeneity (I^2 ranging from 0% to 40%), moderate heterogeneity (I^2 ranging from 30% to 60%), substantial heterogeneity (I^2 ranging from 50% to 90%) and considerable heterogeneity (I^2 ranging from 75% to 100%) [19]. When I^2<40%, the fixed-effects model was adopted; otherwise, a random effect model was considered. In addition, sensitivity analysis and subgroup analysis were used to explore the sources of heterogeneity if needed.

Results

Study Selection

A total of 9337 studies were initially retrieved by searching 9 databases. After the 3-step screening, 8 studies met the inclusion and exclusion criteria. From a manual search of related references, 1 additional study was included. Finally, 9 RCTs were included in the systematic review and meta-analysis. The results of 1 study were published in 2 articles [21,22]. The detailed flow diagram of study selection is shown in Figure 1.

Figure 1. Flow diagram of study selection.



Study Characteristics

The characteristics of 9 RCTs are outlined in Table 1, involving 1128 participants from 6 countries. Seven studies were from developed countries (1 from the United States [23], 3 from the United Kingdom [24-26], 2 from the Netherlands [21,22], and 1 from Finland [27]). Only 2 studies were performed in developing countries (1 from China [28] and 1 from Mexico [29]). Eight of the screened studies were monocentric, and 1 was multicenter [26]. On the duration of interventions, 7 studies were carried out in the prenatal period [21-23,25,26,28,29], and 2 studies were conducted during labor [24,27]. In terms of participants, most of the included studies recruited high-risk pregnant women [21,22,25,26], and the remaining studies recruited low-risk pregnant women [23], late pregnant women [28], and pregnant women facing labor [24,25], respectively. The pooled mean age of pregnant women was 29.28 (SD 5.03) years in 6 RCTs [21-23,25-27].

Table 1. Characteristics of included studies.

Author, year, country	Study design	Participants	Duration	Sample, N	Attrition rate (%)	Major characterization	Major results
Butler Tobah et al, 2019, United States [23]	2-arm RCT,[a] monocentric	Low-risk pregnancies (IG：29.5±3.3 years; CG：29.7±3.6 years)	<13 weeks of gestation to deliver	IG[b]: N=134; CG[c]: N=133	11	IG: OB Nest care (8 on-site appointments, 6 remote visits via phone or web-based communication) CG: usual care (12 prescheduled prenatal clinic appointments)	Pregnancy outcomes (cesarean, delivery, miscarriage, and preterm delivery); neonatal outcomes (low birth weight, and neonatal asphyxia); health care usage (on-site appointments, remote visits, and inpatient days)
Wang et al, 2019, China [28]	2-arm RCT, monocentric	Late pregnancies (IG：22-40 years; CG：22-38 years)	36-41 weeks of gestation to deliver	IG: N=80; CG: N=80	0	IG: remote FHR[d] monitoring (3-4 times daily); CG: own fetal movement count (3 times daily) and routine outpatient FHR monitoring	Neonatal outcomes (neonatal asphyxia and nonstress test）
Tapia-Conyer et al, 2015, Mexico [29]	2-arm RCT, monocentric	High-risk pregnancies (<19 or >35 years)	27-29 weeks of gestation to deliver	IG: N=74; CG: N=79	12	IG: wireless maternal-fetal monitoring (1- to 2-week intervals); CG: conventional care (standard midwifery visits)	Pregnancy outcomes (preterm, preeclampsia, and eclampsia); neonatal outcomes (low birth weight); adherence
Dawson et al, 1999, United Kingdom [26]	2-arm RCT, multicenter	High-risk pregnancies (IG：25.7 ± 5.0 years; CG：27.2 ± 6.3 years)	12 weeks of gestation to deliver	IG: N=43; CG: N=38	0	IG: domiciliary monitoring daily via DFM[e] system; CG: conventional care (standard midwifery visits)	Pregnancy outcomes (weeks of gestation at delivery, spontaneous delivery, cesarean delivery, operative vaginal delivery, and induced labor); neonatal outcomes (neonatal asphyxia); health care usage (on-site appointments, home visits, inpatient days, and cost-effectiveness)
Birnie et al, 1997, the Netherlands [21]	2-arm RCT, monocentric	High-risk pregnancies (IG：29.6±5.8 years; CG：30.9±5.8 years)	32-43 weeks of gestation to deliver	IG: N=76; CG: N=74	0	IG: domiciliary monitoring daily via portable cardiotocography; CG: in-hospital monitoring daily	Pregnancy outcomes (weeks of gestation at delivery, cesarean delivery, and induced labor); neonatal outcomes (birth weight); health care usage (inpatient days and cost-effectiveness)
Monincx et al, 1997, the Netherlands [22]	2-arm RCT, monocentric	High-risk pregnancies (IG：29.6±5.8 years; CG：30.9±5.8 years)	32-43 weeks of gestation to deliver	IG: N=76; CG: N=74	0	IG: domiciliary monitoring daily via portable cardiotocography; CG: in-hospital monitoring daily	Pregnancy outcomes (spontaneous delivery, operative vaginal delivery, and perinatal mortality); neonatal outcomes (neonatal asphyxia and neurological optimality scores)
Dawson et al, 1989, United Kingdom [25]	2-arm RCT, monocentric	High-risk pregnancies (IG：28.78±5.85 years; CG：26.06±3.51 years)	26-41 weeks of gestation to deliver	IG: N=40; CG: N=17	5	IG: domiciliary monitoring daily via DFM system; CG: conventional hospital care	Pregnancy outcomes (weeks of gestation at delivery, cesarean delivery, and induced labor)

Author, year, country	Study design	Participants	Duration	Sample, N	Attrition rate (%)	Major characterization	Major results
Calvert et al, 1982, United Kingdom [24]	3-arm RCT, monocentric	Patients facing labor (≤37 weeks of gestation)	During labor	IG: N=100; CG: N=100	0	IG: remote monitor cardiotocography (patients could get out of bed to walk or sit); CG: conventional bedside cardiotocography	Pregnancy outcomes (spontaneous delivery, cesarean delivery, and operative vaginal delivery); Neonatal outcomes (neonatal asphyxia)
Haukkamaa et al, 1982, Finland [27]	2-arm RCT, monocentric	Patients facing labor (IG: 28.35±3.75 years; CG: 28.1±3.7 years)	During labor	IG: N=31; CG: N=29	0	IG: FHR monitored by telemetry (patients were encouraged to sit or walk); CG: FHR monitored by conventional cardiotocography	Pregnancy outcomes (cesarean delivery, operative vaginal delivery, and induced labor)

[a]RCT: randomized controlled trial.

[b]IG: intervention group.

[c]CG: control group.

[d]FHR: fetal heart rate.

[e]DFM: domiciliary fetal monitoring.

Characteristics of Interventions

The characteristics of interventions are described in Table 2. Most of the included studies were undertaken at home [21-23,25,26,28], with 3 exceptions occurring in rural clinics [29] and hospitals [24,27]. Pregnant women in the control groups received "conventional care," including routine outpatient monitoring, in-hospital monitoring, or conventional bedside cardiotocography. Pregnant women in the intervention groups received remote fetal monitoring with web, Bluetooth, or telephone. Of the included studies, 5 RCTs only supervised fetal heart rate [24-28], and the remaining 4 RCTs monitored extra BP [21-23,29], blood glucose [29], height [29], weight [29], or temperature [21,22].

The frequency of fetal monitoring and guidance varied among the included studies as did the form of feedback. Due to the different stages of pregnancy, the frequency of fetal monitoring ranged from 3 to 4 times daily to biweekly. There were many ways to achieve one-to-one, personalized, and exclusive guidance, including phone visits, on-site appointments, or family visits. In addition, 2 other studies, which occurred during labor, used the obstetrical telemetry system to remotely monitor the fetus in real time [24,27]. During the birth process, the pregnant women in the conventional group were nursed in bed, whereas those with telemetry equipment were encouraged to get out of bed to walk or sit on a chair.

Table 2. Characteristics of interventions.

Author, year, country	Monitoring personnel	Monitoring locus	Monitoring content	Feedback	Technical support
Butler Tobah et al, 2019, United States [23]	Patient, nurse, and obstetrician	Domiciliary	FHR,[a] BP[b]	• Transmission of data via a phone or the institution's electronic medical record system • Personalized guidance by telephone visits or on-site appointments	Home digital sphygmomanometer, handheld fetal Doppler, and patient web portal
Wang et al 2019, China [28]	Patient and obstetrician	Domiciliary	FHR	• Transmission of data via phone • Personalized guidance via telephone if necessary	Portable intelligent medical terminal system
Tapia-Conyer et al, 2015, Mexico [29]	Nurse and obstetrician	Rural clinics	FHR, BP, blood glucose, height, and weight	• Transmission of data through a Bluetooth interface and web access • Personalized consultations via fetal monitoring visits	MiBebe fetal remote monitor prototype, Bluetooth, and patient web portal
Dawson et al 1999, United Kingdom [26]	Patient, community midwife	Domiciliary	FHR	• Transmission of data via telephone using modems • Personalized surveillance and care for each pregnant woman	DFM[c] system
Birnie et al 1997, the Netherlands [21]	Investigator, midwife, and physician	Domiciliary	FHR, BP, and temperature	• Transmission of data via telephone • Personalized consultations via telephone if necessary	Portable cardiotocography and public telephone network
Monincx et al 1997, the Netherlands [22]	Investigator, midwife, and physician	Domiciliary	FHR, BP, and temperature	• Transmission of data via telephone • Personalized consultations via telephone if necessary	Portable cardiotocography and public telephone network
Dawson et al 1989, United Kingdom [25]	Patient, midwife	Domiciliary	FHR	• Transmission of data via telephone fetal monitoring systems • Personalized guidance via regular family visits	DFM system
Calvert et al 1982, United Kingdom [24]	Midwife	Hospital	FHR	• Transmission of data via an obstetrical telemetry system	Obstetrical telemetry system
Haukkamaa et al 1982, Finland [27]	Midwife	Hospital	FHR	• Transmission of data via an obstetrical telemetry system	Obstetrical telemetry system

[a]FHR: fetal heart rate.

[b]BP: blood pressure.

[c]DFM: domiciliary fetal monitoring.

Risk of Bias

Overall, the quality of included studies was moderate, 4 of which (44%) were high-quality research [21-23,26]. The studies showed the main bias in the blinding of participants and personnel, which might be caused by the nature of interventions. In addition, 1 study (11%) showed a high risk of bias for random sequence generation because of grouping according to the hospital number [24]. Fortunately, all outcomes were obtained from medical records, so the outcome assessment would not be influenced by the lack of blinding. Based on the above reasons, the blinding of outcome assessment of included studies was assessed as "low risk of bias." Three RCTs (22%) reported clear data loss, with attrition of 11% [23], 12% [29], and 5% [25], respectively. One of the studies had a relatively large difference in attrition between the groups (20% and 4%, respectively), and it was unclear whether the loss to follow-up varied [29]. Three studies (22%) used intention-to-analysis [21-23] (Figures 2 and 3).

Figure 2. Risk of bias in each study.

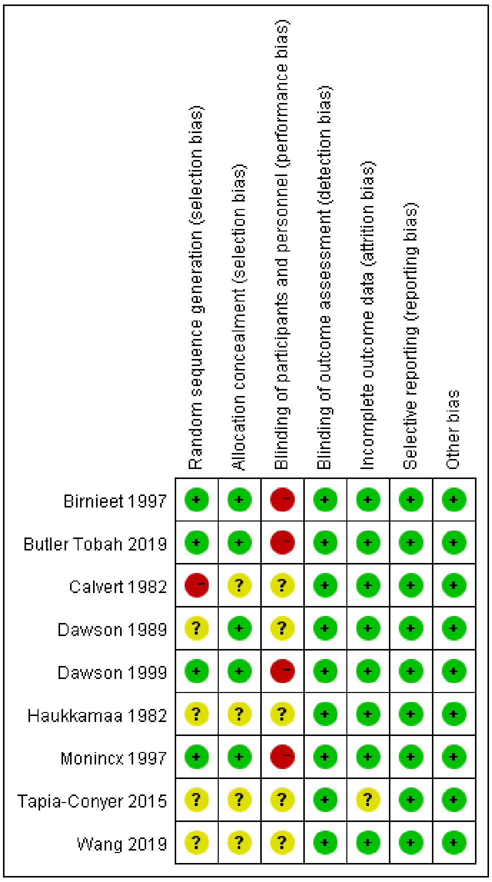
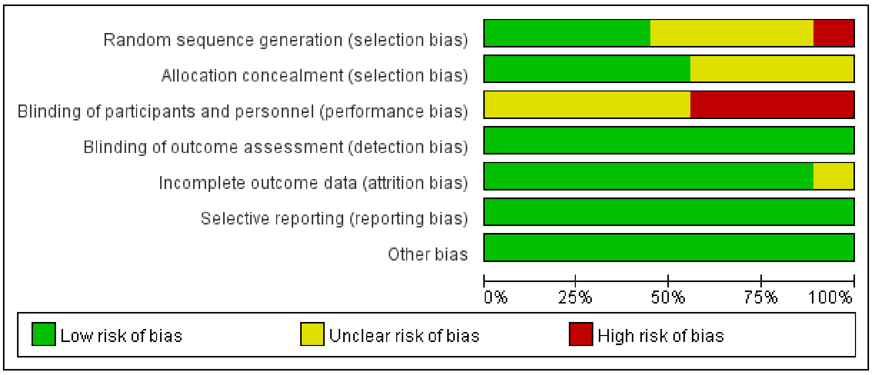


Figure 3. Overall risk of each type of bias.



Synthesis of Results

The review extracted 8 maternal-fetal outcomes and the pooled analyses are presented in Table 3.

Table 3. Effect estimates of 8 outcomes.

Outcomes	Studies, n	Participants, n	Statistical methods	Effect estimates
Cesarean section	6	815	Risk ratio (M-H,[a] fixed, 95% CI)	0.81 (0.59 to 1.12)
Neonatal asphyxia	5	859	Risk ratio (M-H, fixed, 95% CI)	0.66 (0.45 to 0.97)[b]
Instrumental vaginal birth	4	492	Risk ratio (M-H, fixed, 95% CI)	1.21 (0.74 to 1.98)
Induced labor	4	348	Risk ratio (M-H, fixed, 95% CI)	0.90 (0.66 to 1.22)
Spontaneous delivery	3	432	Risk ratio (M-H, fixed, 95% CI)	0.99 (0.89 to 1.10)
Gestational weeks at delivery	3	288	Mean difference (IV,[c] fixed, 95% CI)	−0.28 (−0.86 to 0.30)
Premature delivery	2	420	Risk ratio (M-H, fixed, 95% CI)	0.80 (0.44 to 1.46)
Low birth weight	2	420	Risk ratio (M-H, fixed, 95% CI)	1.20 (0.45 to 3.20)

[a]M-H: Mantel-Haenszel.

[b]Statistically significant at *P*=.04 level.

[c]IV: inverse variance.

Maternal Outcomes

Cesarean section was the most assessed in the included studies, involving 815 pregnant women from 6 RCTs [21,23-27]. Under the fixed effect model, the pooled results showed a nonsignificant difference between the intervention group and the control group (RR 0.81, 95% CI 0.59-1.12; *P*=.21), without any heterogeneity (I^2=0%; *P*=.93; Figure 4).

Instrumental vaginal birth was mentioned in 4 studies involving 492 pregnant women [22,24,26,27]. There was no evidence of heterogeneity when pooling the 4 studies (I^2=0%; *P*=.88). With a fixed effect model, the prevalence of instrumental vaginal birth did not significantly differ between the remote monitoring group and the routine monitoring group (RR 1.21, 95% CI 0.74-1.98; *P*=.45; Figure 5).

Four RCTs (n=348) reported induced labor with an overall rate of 32% [21,25-27]. Moreover, no significant difference (RR 0.90, 95% CI 0.66-1.22; *P*=.50) between groups and the heterogeneity (I^2=0%; *P*=.42) in pooling 4 studies was demonstrated (Figure 6).

Similarly, no significant difference was found in the risk of spontaneous delivery (RR 0.99, 95% CI 0.89‐1.10; *P*=.85) [22,24,26] or premature delivery (RR 0.80, 95% CI 0.44‐1.46; *P*=.47) [23,29], both with no heterogeneity (I^2=0%; *P*=.68 and *P*=.45, respectively; Figures 7 and 8). For gestational weeks at delivery, the overall effect of 3 studies [21,25,26] was also insignificant (MD −0.28, 95% CI −0.86 to 0.30; *P*=.35) in the absence of heterogeneity (I^2=0%; *P*=.68; Figure 9).

Figure 4. Forest plot of cesarean section.

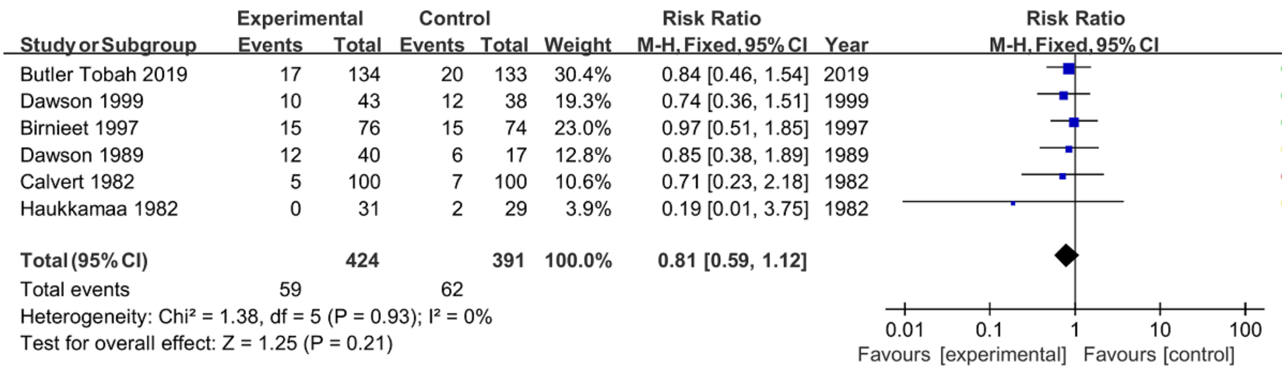


Figure 5. Forest plot of instrumental vaginal birth.

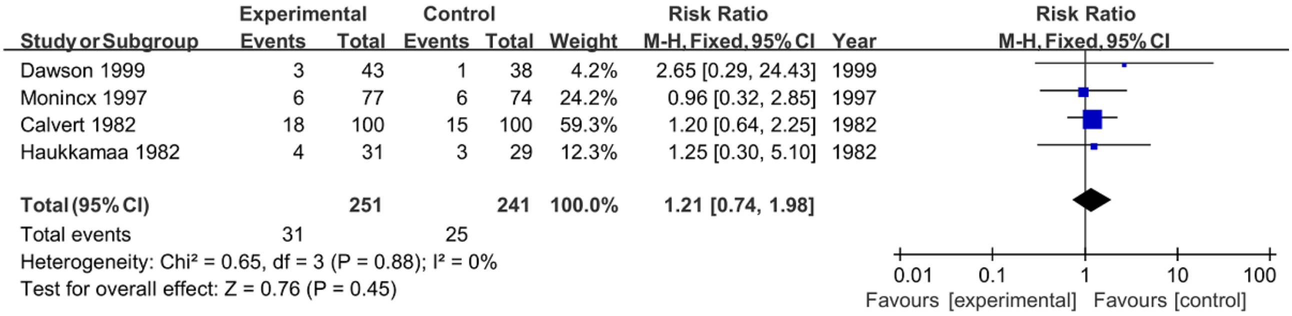


Figure 6. Forest plot of induced labor.

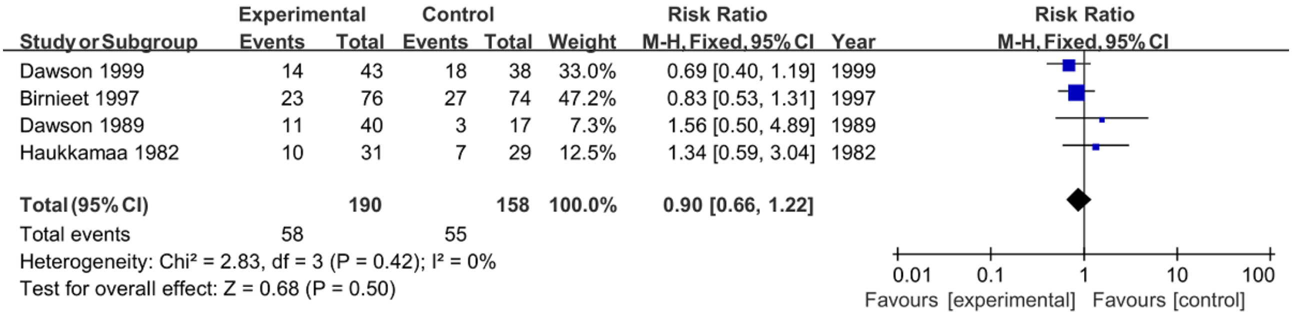


Figure 7. Forest plot of spontaneous delivery.

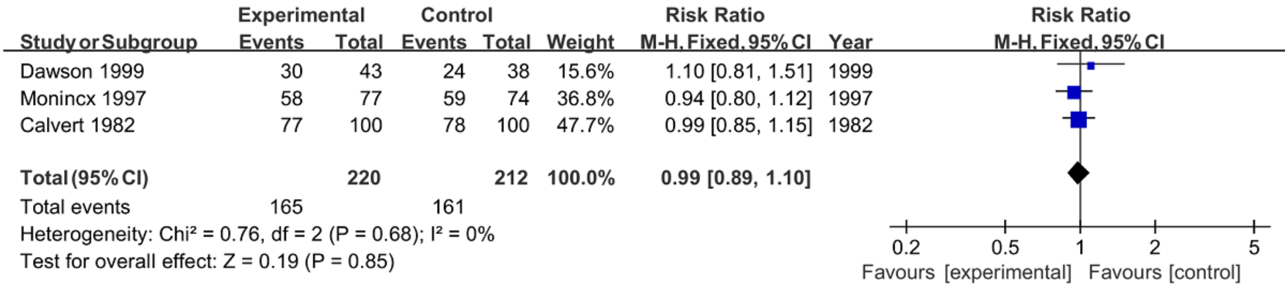
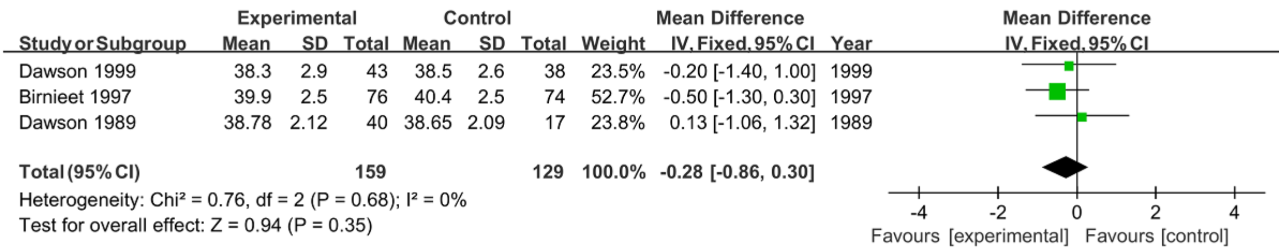


Figure 8. Forest plot of premature delivery.

Figure 9. Forest plot of gestational weeks at delivery.



Outcomes of Infants

Five studies (n=859) compared the incidence of neonatal asphyxia between the intervention group and the control groups, with an overall prevalence of 11% [22-24,26,28]. Furthermore, the overall effect of neonatal asphyxia was significant, and the combined risk ratio was 0.66 (95% CI 0.45-0.97; *P*=.04) with an acceptable heterogeneity across studies (I^2=24%; *P*=.26; Figure 10).

For low birth weight, the pooled results of 2 studies involving 420 newborns showed that no significant difference was discovered between the intervention group and the control group (RR 1.20, 95% CI 0.45-3.20; *P*=.71), without any heterogeneity (I^2=0%; *P*=.41; Figure 11) [23,29].

Figure 10. Forest plot of neonatal asphyxia.

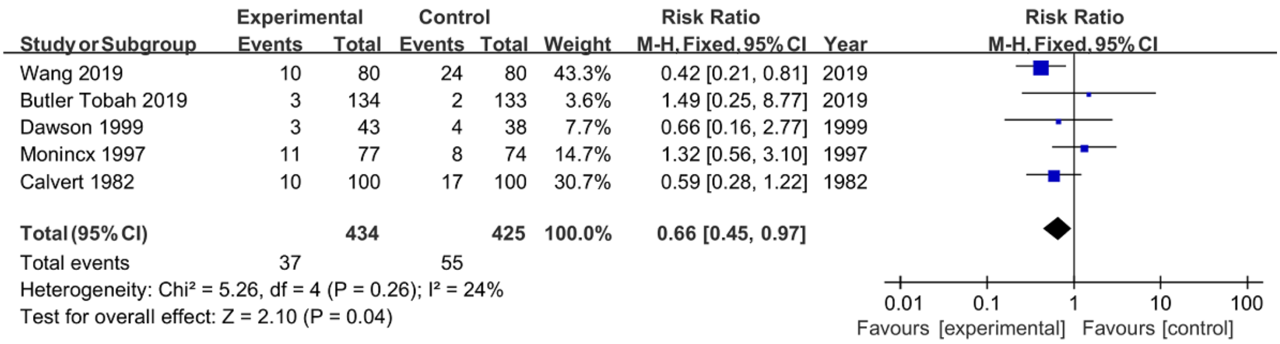
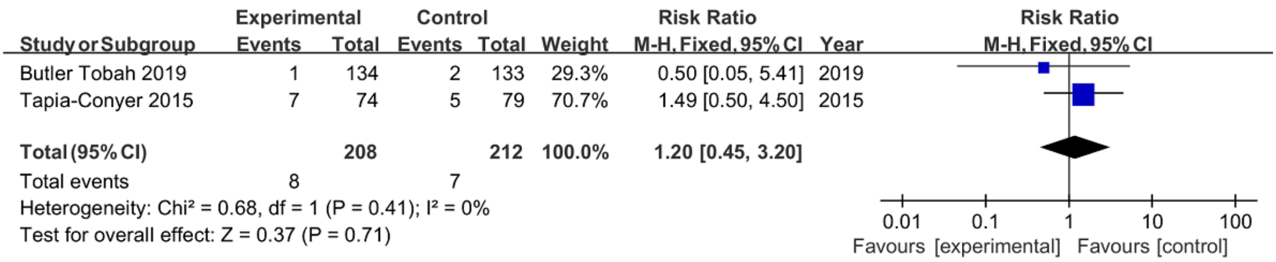


Figure 11. Forest plot of low birth weight.



Health Care Usage

The outcomes of health care usage were investigated in 3 studies [21,23,26], involving the number of on-site appointments or home visits, duration in hospital, medical cost, and so on. However, none of them was suitable for meta-analysis due to heterogeneity of evaluation methods and assessment timing or to a lack of sufficient data.

Butler Tobah et al [23] reported that compared with conventional nursing, the number of on-site appointments with clinicians and nurses decreased significantly in the intervention group (11.25 vs 14.69 visits; *P*<.01), while the duration of time spent on coordinating care and connected care appointments by phone or on the internet was higher in the intervention group (401.20 vs 167.10 minutes per woman; *P*<.01). Similarly, Dawson et al [26] also reported that the remote group received more home visits (3.7 vs 1.4 visits; *P*=.002) and longer home visits (33.5 vs 12.8 minutes per visit; *P*<.001). There was no significant difference in the number of antenatal clinic visits between the 2 groups (2.4 vs 3.2 visits; *P*=.11) [26]. For antenatal inpatient days, Dawson et al [26] found there were no significant differences between the 2 groups (3.58 vs 5.13 days; *P*=.12), whereas Birnie et al [21] reported longer hospital stays in the control group (1 vs 7 days; *P*<.001). Furthermore, no significant differences in hospital length of stay after delivery [21,23] were observed across groups.

Two studies reported cost-effectiveness [21,26]. Birnie et al [21] indicated that domiciliary monitoring had lower prenatal costs than in-hospital monitoring (US $1521 vs US $3558 per woman; *P*<.001), mainly focusing on nursing care, domiciliary monitoring, and informal family care. Dawson et al [26] also supported that the total cost of domiciliary care was €223.83 (US $239.89 in 2023) per woman less than that of conventional care, consisting of community midwife (time and travel), home monitoring equipment (capital cost and maintenance), lost productive output (women and partners), and antenatal clinic attendances (visits, ultrasound scans, and antenatal inpatient care) [26].

Discussion

Principal Findings

As far as we know, this is the first article to quantitatively analyze the effects of remote fetal monitoring. The systematic review and meta-analysis highlighted that remote fetal monitoring significantly reduced the risk of neonatal asphyxia by 34%. Beyond that, remote fetal monitoring was also beneficial for reducing prenatal costs, which showed some potential for greater cost-effectiveness.

Comparison With Prior Studies

In previous reviews, the superiority of obstetric remote monitoring has also been repeatedly emphasized because of real-time, periodic, and remote monitoring [3,30,31]. By integrating 14 studies involving blood glucose, fetal heart rate, and uterine activity, Lanssens et al found that remote monitoring reduced low neonatal birth weight and neonatal intensive care unit admissions, as well as prolonged gestational age [31]. Likewise, a recent systematic review, focusing on obstetric remote monitoring of BP, uterine contractions, weight, heart rate, and so forth also supported that telemonitoring during pregnancy had great potential for promoting better pregnancy outcomes [3]. However, due to limited research on prenatal remote monitoring, no further quantitative analysis was carried out in the above reviews.

This systematic review and meta-analysis, the first to focus remote monitoring on the fetus, revealed that remote fetal monitoring reduced the risk of neonatal asphyxia by 34%. Remote fetal monitoring can identify signs of fetal hypoxia in time by monitoring wherever and whenever, which is essential to reduce neonatal asphyxia, especially in high-risk pregnant women [32]. In terms of cost-effectiveness, only 2 RCTs out of 9 studies reported cost-effectiveness [21,26]. Both demonstrated that remote monitoring significantly reduced prenatal costs, which was consistent with previous studies [31,33,34]. In Lanssens' [31] review, 2 retrospective studies found that remote monitoring significantly reduced health care costs. In the studies reviewed, cost analysis focused on health care costs, patient costs, caregiver costs, and productivity costs. Remote fetal monitoring had additional equipment costs and maintenance costs, but in the long run, it saved much more than that, such as time costs, travel costs, or outpatient costs.

In addition, the disadvantages of remote fetal monitoring remained controversial, such as whether additional cesarean sections would be added. In this regard, this meta-analysis covering 9 studies found no consistent evidence of adverse effects on maternal and infant outcomes, with a small heterogeneity ranging from 0% to 24%. This might be related to accurate guidance from midwives or obstetricians on the remote monitoring team. Nonetheless, a recent review in 2019 evaluated information involving decreased fetal movement in 24 mobile applications, revealing that the information varied widely and lacked evidence-based clinical advice [35]. Accurate information about fetal movement is essential for improving maternal and infant outcomes. Therefore, it is recommended that health care personnel cooperate with software developers to jointly develop high-quality prenatal education tools, which will help to promote more pregnant women to obtain timely and accurate guidance.

Notably, in the current systematic review and meta-analysis, 7 of the 9 studies were carried out in developed countries, which were inseparable from the rich medical resources and advanced medical technologies of developed countries. The latest global figures showed that in 2020, there were 26 and 17 deaths per 1000 live births in low- and middle-income countries (LMICs), respectively. However, in high-income countries, the rate only stood at 3 per 1000 [36]. Given the higher perinatal mortality rate, the need for remote fetal monitoring in developing countries may be more urgent. Furthermore, a recent review focused on LMICs concluded that mobile technology can overcome economic and geographic barriers by transmitting clinical information collected using low-cost devices, thereby increasing the perinatal care coverage of LMICs [5]. It can be argued that remote fetal monitoring supported by mobile technology appears to have greater potential in LMICs, where antenatal care services need to be improved. Therefore, we encourage remote fetal monitoring in LMICs to alleviate the shortage of medical resources and further complement the benefits of remote fetal monitoring.

Suggestions for Clinical Practice

This systematic review has demonstrated that remote fetal monitoring has a significant effect on improving maternal and infant outcomes, but this does not mean that remote fetal monitoring can replace face-to-face communication between doctors and patients, which is necessary for shared decision-making. Remote monitoring breaks through the barriers of time and distance, so it is reasonable as an effective complement to traditional outpatient monitoring [37]. Especially during the COVID-19 pandemic, pregnant women, as a high-risk group, should not gather in outpatient clinics for a long time. At this time, remote fetal monitoring not only realizes noncontact medical services but also ensures the safety of mothers and babies. Unfortunately, remote fetal monitoring is rarely implemented in developing countries, especially in areas with limited medical resources [3]. Therefore, the development and implementation of remote monitoring technology urgently need to be put on the agenda. Aside from the technical issues, another concern of remote fetal monitoring is that authentication rules, reimbursement policies, data security, legal responsibilities, and so forth are not yet clear [38]. Although remote fetal monitoring has not yet shown adverse consequences, it is still necessary to conduct relevant research cautiously in combination with the local medical level.

Limitations

There were some limitations worth noting. The diversity of pregnant women in the current systematic review was the major limitation, involving low-risk pregnancies, high-risk pregnancies, late pregnancies, and patients facing labor. Future research can continue to explore which types of pregnant women are more suitable for remote fetal monitoring. In addition, several RCTs included in this meta-analysis were relatively old, which might limit the direct applicability of the evidence to current clinical practice. Finally, due to the limited literature, it was difficult to quantitatively analyze the efficacy of remote fetal monitoring in health resource usage. Future studies are expected to assess the cost-effectiveness of remote fetal monitoring, including implementation costs (technology costs, medical costs, etc), intervention costs (patient resource costs, commuting costs, etc), and downstream costs (productivity costs, future costs, etc) [39]. Likewise, the number of consultations, length of hospital stay, and patient compliance or satisfaction cannot be ignored and need to be explored further.

Conclusions

The present systematic review and meta-analysis of 9 studies highlighted that remote fetal monitoring had a favorable effect on reducing neonatal asphyxia. Remote fetal monitoring has not yet found hidden dangers, but more large-scale, multicenter, and high-quality studies are still expected to explore its safety and efficacy. At the same time, more research is also recommended to further carry out the cost analysis of remote fetal monitoring, which will help alleviate the huge medical expenses.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[DOCX File , 15 KB - mhealth_v11i1e41508_app1.docx]

Multimedia Appendix 2

Data extraction form.

[DOCX File , 25 KB - mhealth_v11i1e41508_app2.docx]

References

1. Ending preventable maternal mortality (EPMM): a renewed focus for improving maternal and newborn health and well-being. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240040519 [accessed 2022-05-22]
2. Crawford A, Hayes D, Johnstone ED, Heazell AEP. Women's experiences of continuous fetal monitoring: a mixed-methods systematic review. Acta Obstet Gynecol Scand 2017;96(12):1404-1413 [FREE Full text] [doi: 10.1111/aogs.13231] [Medline: 28902389]
3. Alves DS, Times VC, da Silva É, Melo PSA, Novaes MA. Advances in obstetric telemonitoring: a systematic review. Int J Med Inform 2020;134:104004. [doi: 10.1016/j.ijmedinf.2019.104004] [Medline: 31816495]
4. Martin JK, Price-Haywood EG, Gastanaduy MM, Fort DG, Ford MK, Peterson SP, et al. Unexpected term neonatal intensive care unit admissions and a potential role for centralized remote fetal monitoring. Am J Perinatol 2023;40(3):297-304. [doi: 10.1055/s-0041-1727214] [Medline: 33882588]
5. Valderrama CE, Ketabi N, Marzbanrad F, Rohloff P, Clifford GD. A review of fetal cardiac monitoring, with a focus on low- and middle-income countries. Physiol Meas 2020;41(11):11TR01 [FREE Full text] [doi: 10.1088/1361-6579/abc4c7] [Medline: 33105122]
6. van den Heuvel JFM, Ayubi S, Franx A, Bekker MN. Home-based monitoring and telemonitoring of complicated pregnancies: nationwide cross-sectional survey of current practice in the Netherlands. JMIR Mhealth Uhealth 2020;8(10):e18966 [FREE Full text] [doi: 10.2196/18966] [Medline: 33112250]
7. Kern-Goldberger AR, Srinivas SK. Telemedicine in obstetrics. Clin Perinatol 2020;47(4):743-757. [doi: 10.1016/j.clp.2020.08.007]
8. Pflugeisen BM, McCarren C, Poore S, Carlile M, Schroeder R. Virtual visits: managing prenatal care with modern technology. MCN Am J Matern Child Nurs 2016;41(1):24-30 [FREE Full text] [doi: 10.1073/pnas.73.5.1398] [Medline: 5721]
9. Magann EF, McKelvey SS, Hitt WC, Smith MV, Azam GA, Lowery CL. The use of telemedicine in obstetrics: a review of the literature. Obstet Gynecol Surv 2011;66(3):170-178. [doi: 10.1097/OGX.0b013e3182219902] [Medline: 21689487]
10. Xydopoulos G, Perry H, Sheehan E, Thilaganathan B, Fordham R, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population: cost-minimization study. Ultrasound Obstet Gynecol 2019;53(4):496-502 [FREE Full text] [doi: 10.1002/uog.19041]
11. Runkle J, Sugg M, Boase D, Galvin SL, Coulson CC. Use of wearable sensors for pregnancy health and environmental monitoring: descriptive findings from the perspective of patients and providers. Digit Health 2019;5:2055207619828220 [FREE Full text] [doi: 10.1177/2055207619828220] [Medline: 30792878]
12. Schramm K, Grassl N, Nees J, Hoffmann J, Stepan H, Bruckner T, et al. Women's attitudes toward self-monitoring of their pregnancy using noninvasive electronic devices: cross-sectional multicenter study. JMIR Mhealth Uhealth 2019;7(1):e11458. [doi: 10.2196/11458]
13. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
14. Tran K, Padwal R, Khan N, Wright MD, Chan WS. Home blood pressure monitoring in the diagnosis and treatment of hypertension in pregnancy: a systematic review and meta-analysis. CMAJ Open 2021;9(2):E642-E650 [FREE Full text] [doi: 10.9778/cmajo.20200099] [Medline: 34131027]
15. Kalafat E, Benlioglu C, Thilaganathan B, Khalil A. Home blood pressure monitoring in the antenatal and postpartum period: a systematic review meta-analysis. Pregnancy Hypertens 2020;19:44-51. [doi: 10.1016/j.preghy.2019.12.001] [Medline: 31901652]

16. Li SY, Ouyang YQ, Qiao J, Shen Q. Technology-supported lifestyle interventions to improve maternal-fetal outcomes in women with gestational diabetes mellitus: a meta-analysis. Midwifery 2020;85:102689. [doi: 10.1016/j.midw.2020.102689]
17. Lau Y, Klainin-Yobas P, Htun TP, Wong SN, Tan KL, Ho-Lim ST, et al. Electronic-based lifestyle interventions in overweight or obese perinatal women: a systematic review and meta-analysis. Obes Rev 2017;18(9):1071-1087. [doi: 10.1111/obr.12557] [Medline: 28544551]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535 [FREE Full text] [Medline: 19622551]
19. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions (version 6.3). 2022. URL: https://training.cochrane.org/handbook/current [accessed 2022-05-22]
20. Grzeskowiak LE, Smithers LG, Amir LH, Grivell RM. Domperidone for increasing breast milk volume in mothers expressing breast milk for their preterm infants: a systematic review and meta-analysis. BJOG 2018;125(11):1371-1378 [FREE Full text] [doi: 10.1111/1471-0528.15177]
21. Birnie E, Monincx WM, Zondervan HA, Bossuyt PM, Bonsel GJ. Cost-minimization analysis of domiciliary antenatal fetal monitoring in high-risk pregnancies. Obstet Gynecol 1997;89(6):925-929. [doi: 10.1016/s0029-7844(97)00150-6] [Medline: 9170466]
22. Monincx WM, Zondervan HA, Birnie E, Ris M, Bossuyt PM. High risk pregnancy monitored antenatally at home. Eur J Obstet Gynecol Reprod Biol 1997;75(2):147-153. [doi: 10.1016/s0301-2115(97)00122-x] [Medline: 9447367]
23. Butler Tobah YS, LeBlanc A, Branda ME, Inselman JW, Morris MA, Ridgeway JL, et al. Randomized comparison of a reduced-visit prenatal care model enhanced with remote monitoring. Am J Obstet Gynecol 2019;221(6):638.e1-638.e8. [doi: 10.1016/j.ajog.2019.06.034] [Medline: 31228414]
24. Calvert JP, Newcombe RG, Hibbard BM. An assessment of radiotelemetry in the monitoring of labour. Br J Obstet Gynaecol 1982;89(4):285-291. [doi: 10.1111/j.1471-0528.1982.tb04697.x] [Medline: 7073996]
25. Dawson AJ, Middlemiss C, Coles EC, Gough NA, Jones ME. A randomized study of a domiciliary antenatal care scheme: the effect on hospital admissions. Br J Obstet Gynaecol 1989;96(11):1319-1322. [doi: 10.1111/j.1471-0528.1989.tb03230.x] [Medline: 2611171]
26. Dawson A, Cohen D, Candelier C, Jones G, Sanders J, Thompson A, et al. Domiciliary midwifery support in high-risk pregnancy incorporating telephonic fetal heart rate monitoring: a health technology randomized assessment. J Telemed Telecare 1999;5(4):220-230. [doi: 10.1258/1357633991933756] [Medline: 10829372]
27. Haukkamaa M, Purhonen M, Teramo K. The monitoring of labor by telemetry. J Perinat Med 1982;10(1):17-22. [doi: 10.1515/jpme.1982.10.1.17] [Medline: 7062228]
28. Wang Q, Yang W, Li L, Yan G, Wang H, Li J. Late pregnancy analysis with Yunban's remote fetal monitoring system. Int J Distrib Sens Netw 2019;15(3):1550147719832835. [doi: 10.1177/1550147719832835]
29. Tapia-Conyer R, Lyford S, Saucedo R, Casale M, Gallardo H, Becerra K, et al. Improving perinatal care in the rural regions worldwide by wireless enabled antepartum fetal monitoring: a demonstration project. Int J Telemed Appl 2015;2015:794180 [FREE Full text] [doi: 10.1155/2015/794180] [Medline: 25691900]
30. Snoswell CL, Chelberg G, De Guzman KR, Haydon HH, Thomas EE, Caffery LJ, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2021:1357633X211022907. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
31. Lanssens D, Vandenberk T, Thijs IM, Grieten L, Gyselaers W. Effectiveness of telemonitoring in obstetrics: scoping review. J Med Internet Res 2017;19(9):e327 [FREE Full text] [doi: 10.2196/jmir.7266] [Medline: 28954715]
32. Signorini MG, Lanzola G, Torti E, Fanelli A, Magenes G. Antepartum fetal monitoring through a wearable system and a mobile application. Technologies 2018;6(2):44. [doi: 10.3390/technologies6020044]
33. Buysse H, De Moor G, Van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. Int J Med Inform 2008;77(7):470-476. [doi: 10.1016/j.ijmedinf.2007.08.009] [Medline: 17923433]
34. Lemelin A, Paré G, Bernard S, Godbout A. Demonstrated cost-effectiveness of a telehomecare program for gestational diabetes mellitus management. Diabetes Technol Ther 2020;22(3):195-202. [doi: 10.1089/dia.2019.0259] [Medline: 31603351]
35. Daly LM, Boyle FM, Gibbons K, Le H, Roberts J, Flenady V. Mobile applications providing guidance about decreased fetal movement: review and content analysis. Women Birth 2019;32(3):e289-e296. [doi: 10.1016/j.wombi.2018.07.020] [Medline: 30139669]
36. UNICEF, WHO. Mortality rate, neonatal (per 1,000 live births). The World Bank. URL: https://data.worldbank.org/indicator/SH.DYN.NMRT [accessed 2022-05-22]
37. Zizzo AR, Hvidman L, Salvig JD, Holst L, Kyng M, Petersen OB. Home management by remote self-monitoring in intermediate- and high-risk pregnancies: a retrospective study of 400 consecutive women. Acta Obstet Gynecol Scand 2022;101(1):135-144 [FREE Full text] [doi: 10.1111/aogs.14294]
38. Greiner AL. Telemedicine applications in obstetrics and gynecology. Clin Obstet Gynecol 2017 Dec;60(4):853-866. [doi: 10.1097/GRF.0000000000000328] [Medline: 28990981]
39. Gold HT, McDermott C, Hoomans T, Wagner TH. Cost data in implementation science: categories and approaches to costing. Implement Sci 2022;17(1):11 [FREE Full text] [doi: 10.1186/s13012-021-01172-6] [Medline: 35090508]

Abbreviations

BP: blood pressure
LMIC: low- and middle-income country
MD: mean difference
RCT: randomized controlled trial
RR: risk ratio

Edited by L Buis; submitted 31.07.22; peer-reviewed by M Kapsetaki, M Bhatta; comments to author 10.11.22; revised version received 29.12.22; accepted 23.01.23; published 22.02.23.

Please cite as:
Li S, Yang Q, Niu S, Liu Y
Effectiveness of Remote Fetal Monitoring on Maternal-Fetal Outcomes: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2023;11:e41508
URL: https://mhealth.jmir.org/2023/1/e41508
doi:10.2196/41508
PMID:36811944

Review

Community Health Worker Use of Smart Devices for Health Promotion: Scoping Review

Merlin Greuel[1], BA, MGH, MD; Frithjof Sy[1], BSc, MD; Till Bärnighausen[1,2,3], MSc, MD, SCD; Maya Adam[1,4], BA, MD; Alain Vandormael[1], MSc, PhD; Jennifer Gates[5], BA; Guy Harling[2,6,7,8,9], BSc, MA, MPH, SCD

[1]Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Germany

[2]Africa Health Research Institute, KwaZulu-Natal, South Africa

[3]Department of Global Health and Population, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

[4]Department of Pediatrics, School of Medicine, Stanford University, Stanford, CA, United States

[5]Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Institute for Global Health, University College London, London, United Kingdom

[7]Department of Epidemiology and Harvard Center for Population and Development Studies, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

[8]Medical Research Council/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[9]School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, KwaZulu-Natal, South Africa

Corresponding Author:
Merlin Greuel, BA, MGH, MD
Heidelberg Institute of Global Health
Heidelberg University
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 17625498934
Email: merlin.greuel@gmail.com

Abstract

Background: Community health workers (CHWs) have become essential to the promotion of healthy behaviors, yet their work is complicated by challenges both within and beyond their control. These challenges include resistance to the change of existing behaviors, disbelief of health messages, limited community health literacy, insufficient CHW communication skills and knowledge, lack of community interest and respect for CHWs, and CHWs' lack of adequate supplies. The rising penetration of "smart" technology (eg, smartphones and tablets) in low- and middle-income countries facilitates the use of portable electronic devices in the field.

Objective: This scoping review examines to what extent mobile health in the form of smart devices may enhance the delivery of public health messages in CHW-client interactions, thereby addressing the aforementioned challenges and inducing client behavior change.

Methods: We conducted a structured search of the PubMed and LILACS databases using subject heading terms in 4 categories: technology user, technology device, use of technology, and outcome. Eligibility criteria included publication since January 2007, CHWs delivering a health message aided by a smart device, and face-to-face communication between CHWs and clients. Eligible studies were analyzed qualitatively using a modified version of the Partners in Health conceptual framework.

Results: We identified 12 eligible studies, 10 (83%) of which used qualitative or mixed methods approaches. We found that smart devices mitigate challenges encountered by CHWs by improving their knowledge, motivation, and creativity (eg, through self-made videos); their status within the community; and the credibility of their health messages. The technology stimulated interest in both CHWs and clients—and sometimes even in bystanders and neighbors. Media content produced locally or reflecting local customs was strongly embraced. Yet, the effect of smart devices on the quality of CHW-client interactions was inconclusive. Interactions suffered as CHWs were tempted to replace educational conversations with clients by passively watching video content. Furthermore, a series of technical difficulties experienced especially by older and less educated CHWs compromised some of the advantages brought about by mobile devices. Adequate CHW training ameliorated these difficulties. Only 1 study (8%) considered client health behavior change as an end point, thus revealing a major research gap.

Conclusions: Smart mobile devices may augment CHWs' field performance and enhance face-to-face interactions with clients, yet they also generate new challenges. The available evidence is scarce, mostly qualitative, and focused on a limited range of health outcomes. Future research should include larger-scale interventions across a wide range of health outcomes and feature client health behavior change as an end point.

(JMIR Mhealth Uhealth 2023;11:e42023) doi:10.2196/42023

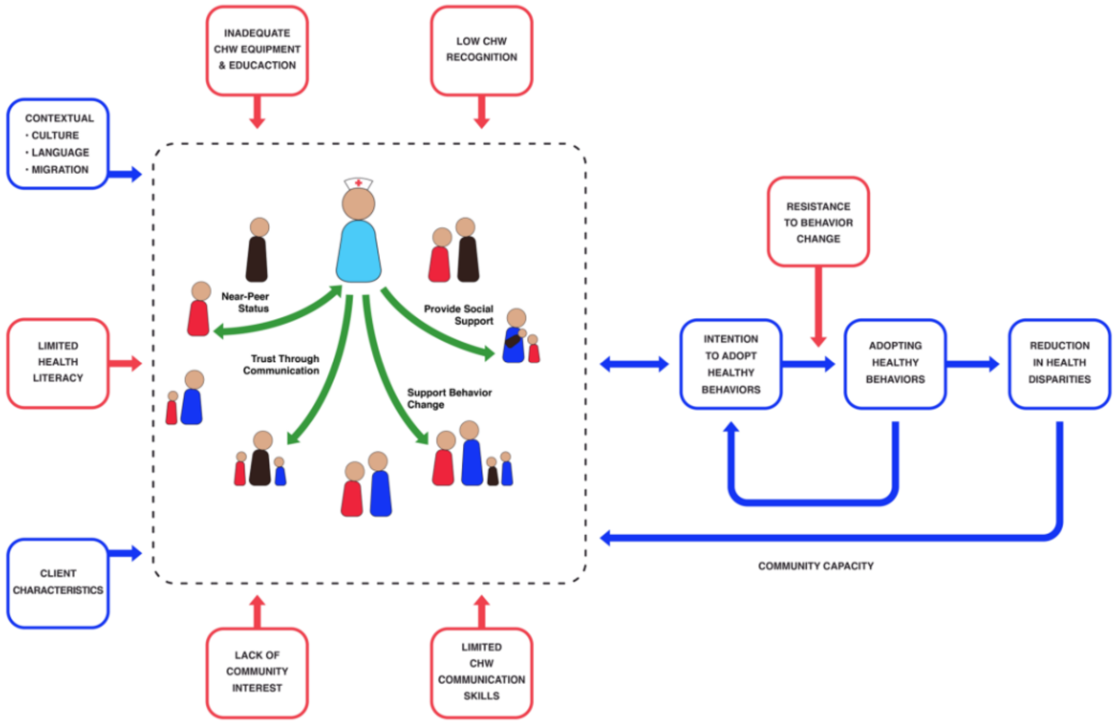
KEYWORDS

mobile health; community health workers; smart phones; tablets; health promotion; public health; health worker; smart devices; health behaviour; smart technology; health message; health outcome

Introduction

Community health workers (CHWs) have become central to health promotion activities, with more than 5 million CHWs working worldwide in 2014 [1]. In part, CHWs' impact is a result of speaking the local language and identifying with the community they serve. Therefore, they have the potential to convey health messages more effectively than other health cadres [2] and may be able to "improve key health-related behaviors" [1]. Katigbak and colleagues [3] have developed the Partners in Health conceptual framework for how CHWs can facilitate the adoption of healthy behaviors. In this framework, client characteristics, the environment, and CHW activities reciprocally influence each other to generate behavior change. Nevertheless, factors both within and beyond CHWs' control can impede their health promotion activities. Based on the literature cited below, we have identified challenges to CHW health promotion activities and have integrated them into the Partners in Health framework (Figure 1).

Figure 1. Conceptual framework of facilitators and barriers to community health workers (CHWs) and patients acting as partners in health. Challenges to the CHW-client interaction are shown in red (adapted from Katigbak et al [3] with permission from the American Journal of Public Health).



One challenge is the way humans manage change, as promoting healthy behavior often entails encouraging changes in existing behavior. Since multiple social, emotional, and cognitive factors interact to mediate [4] and sustain behavior change [5], harmful behaviors are often resistant to change. A second challenge to promoting healthy behaviors is community literacy. In particular, limited health literacy, the ability to comprehend and act on health-related information, is associated with negative health outcomes [6,7] and may complicate health message uptake. In contrast, adequate health literacy can promote healthy behaviors, such as physical activity, by increasing knowledge and self-efficacy related to these behaviors, resulting in positive health outcomes [7]. While low health literacy is certainly a problem in higher-income countries [6], it constitutes a larger problem in low- and middle-income countries (LMIC). For example, basic literacy in many sub-Saharan African countries ranges from 24% to 60% [8].

Other challenges relate to the characteristics of CHWs and their interaction with community members. These include insufficient CHW communication skills [9]. In addition, a lack of community participation and interest, CHWs' own limitations in understanding complex health information due to low levels of education, a lack of respect for CHW knowledge, and disbelief in health promotion messages may complicate the work of CHWs [10]. The lack of community recognition and the low community status of CHWs may pose additional challenges [11], and this problem may be aggravated if CHWs lack adequate supplies and equipment [12]. Facing these challenges, CHWs have demanded educational communication materials that can be carried to the households they visit [9] and suggested using media to reinforce health messages [10].

Mobile electronic media—in particular "smart" devices such as smartphones and tablets—may constitute powerful tools to deliver public health messages. Smart devices can provide learning via videos or mobile apps, providing information through multiple modes (eg, verbally and visually). Learners presented with visual information in addition to verbal information generate a multimedia effect that deepens learning [13]. Dual coding theory suggests that this deeper learning occurs because learners process visual and verbal information separately and then select pieces of information from each before unifying them into a coherent mental representation of knowledge [14]. This theory has been used to optimize multimedia learning materials for e-learning [15] and medical education [16].

The rapidly rising smartphone ownership in LMIC [17] presents an opportunity for increased access to health-related information and resources extending health system reach [18]. However, the comparatively low penetration of smart devices in low-income settings may limit their usefulness as health promotion vehicles; in several sub-Saharan African countries, adult smartphone ownership is less than 20% [17]. Equipping CHWs with mobile smart devices provides an intermediate solution, allowing electronic multimedia education to be accessed even in low device-penetration communities. For example, tablet-displayed videos have transmitted agricultural knowledge and induced abstractive learning in rural Uganda [19].

Given the potential of smart devices as health promotion vehicles, equipping CHWs with such devices may address several of the aforementioned health promotion challenges. Past reviews of mobile health (mHealth) and CHWs have not focused specifically on the use of mHealth as a health promotion tool. Braun and colleagues' [20] review of CHW mHealth use concentrated on how mHealth improved intra-CHW communication and learning. Hall and colleagues' [21] review of mHealth interventions in LMIC included client education and behavior change but focused on the role of mHealth in improving treatment adherence and appointment compliance rather than multimedia applications. Källander and colleagues' [22] review considered SMS text messaging rather than multimedia applications, while the review by White et al [23] focused on how mHealth improved CHW-patient communication broadly but did not focus specifically on educational uses. Thus, there has not been a systematic review of how smart mobile devices can facilitate the delivery of public health messages by CHWs.

Accordingly, we conducted a scoping review of how multimedia features of smart mobile devices have been used to enhance knowledge transfer and behavior change by CHWs. We excluded distance-based media approaches such as SMS text messaging, automated voice messages, and phone calls and focused instead on studies involving direct face-to-face communication between CHWs and community members. Our fundamental question is whether smart mobile devices can enhance CHW face-to-face delivery of public health messages and thereby enhance client health behavior change. The aim of this study was to identify the type of evidence available, point out knowledge gaps, and indicate possible directions for future research. Given these objectives, we deemed a scoping review approach as the most suitable [24].

Methods

Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews) methodology [25], we used a scoping approach to review the literature published between January 1, 2007, and January 5, 2022. The start date was chosen because the first publicly available smartphones using capacitive touch screens were released in 2007 (tablet computers became more common after 2010). The search was performed in English, but no articles were excluded from the full-text assessment if they were published in another language. We consulted the PubMed and LILACS databases, modifying White and colleagues' [23] strategy to capture the intersection of 4 search categories: technology user, technology device, use of technology, and outcome.

We defined our "technology users" as CHWs using smart devices to deliver public health messages and client recipients. As definitions for CHWs vary [26], we employed the common definition used in a World Health Organization study group review, in that CHWs should be "members of the communities where they work, should be selected by the communities, should be answerable to the communities for their activities, should be supported by the health system but not necessarily a part of its organization, and have shorter training than professional workers" [27]. Clients are defined as community and household members of any age or gender who receive a health message outside the context of health facilities.

For "technology device" and "use of technology," we focused on digital content requiring portable computer-like "smart" devices that distinguish themselves from regular cell phones by the ability to run apps, show video content, and connect to the internet. We excluded technological features not requiring "smart" devices that can be used by traditional cell phones, (eg, SMS text messages, automated voice messages, and phone calls). We also required direct, person-to-person communication between CHWs and clients (as opposed to CHWs sending messages from a distance) since we were interested in whether technology enhances the effectiveness of person-to-person communication. The person-to-person communication had to be primary health education (ie, the delivery of a preventative health message such as the promotion of healthy behaviors). We excluded secondary prevention messages such as treatment dissemination or medication reminders. Finally, we included any qualitative or quantitative "outcome" that allowed us to assess the effectiveness, advantages, or disadvantages of the use of smart devices for health promotion from the viewpoint of any stakeholder.

For each category, we identified relevant Medical Subject Headings (MeSH) for PubMed and the corresponding multilingual Descriptores en Ciencias de la Salud in LILACS. The complete search process including all MeSH terms used is shown in Multimedia Appendix 1. For each article fulfilling the inclusion criteria, we screened all references for other potentially relevant articles and used Google Scholar to search for relevant publications citing these included studies, as well as additional publications by the same authors.

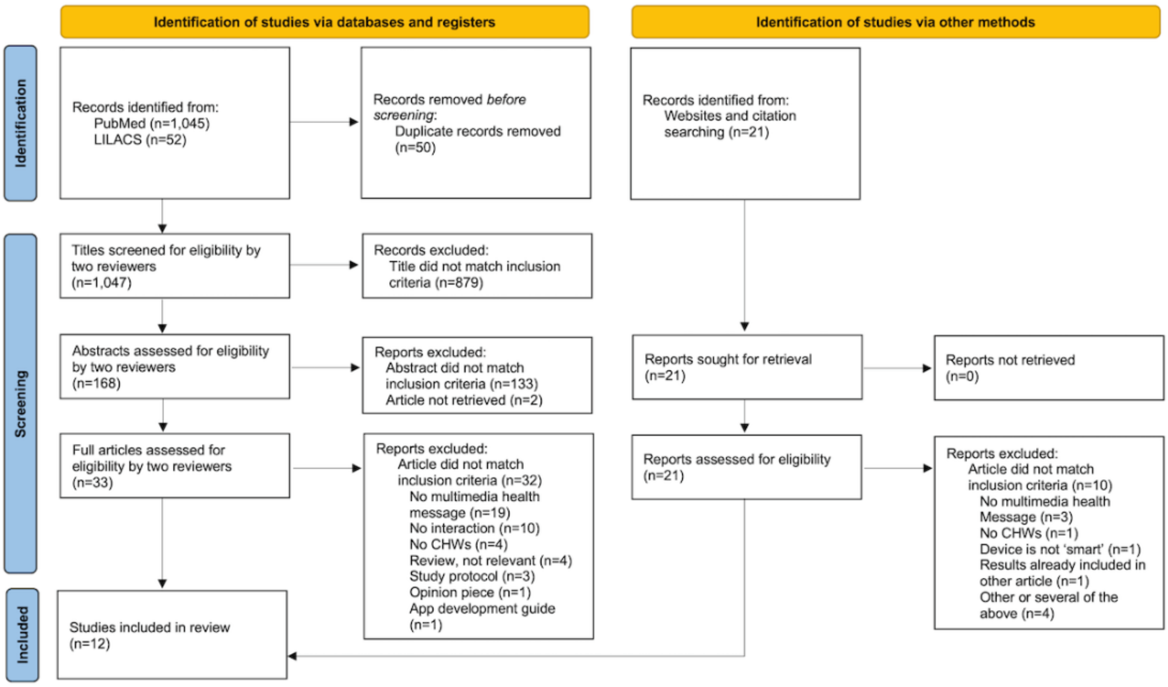
Details of all articles found through these searches were extracted to a spreadsheet and deduplicated. Two authors (MG and FS) independently screened the article titles and abstracts for relevance with respect to the inclusion criteria defined above. If deemed relevant by at least 1 of the 2 authors, the article was included in the following stage of review. The same authors then independently reviewed the full text of all retained articles from the abstract screen. All full-text articles deemed relevant by at least 1 of the 2 authors were then discussed in-depth to reach the final agreement on inclusion. Disagreements were resolved through mutual consent or consultation with a third author (GH). We grouped all included studies by methodology, extracting methods and findings into tables, and used these to qualitatively describe the literature in the context of our original conceptual framework. The PRISMA-ScR checklist [25] that guided our approach can be found in Multimedia Appendix 2.

Results

Identified Articles

The PubMed search yielded 1045 results, and the LILACS search 52, all but 2 (96%) of them duplicates of PubMed articles. Of all articles in either database, 168 articles (16%) were included after the title review, 33 (3%) after the abstract review, and 4 after the full paper review. Of these, 1 was chosen to be included in the results of this publication. From the Google Scholar search and reference screenings, 21 additional abstracts were selected, 11 (52%) of which met our inclusion criteria. All 11 studies were included. Hence, of the 1118 initially identified studies (1068 after deduplication), 12 (1%) were retained for analysis (Figure 2 [28]).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram for review articles. Reasons for exclusion do not sum to total because some categories overlap. CHW: community health worker.



Overview of Studies

The included studies—all published between 2010 and 2021—were conducted in South Africa [29,30], Nigeria [31], Burkina Faso [32], Lesotho [33-35], and India [36-40] (Table 1). Among them, 9 (75%) studies were rural, 2 (17%) urban, and 1 (8%) both. Moreover, 2 (17%) studies were quantitative, 7 (58%) were qualitative, and 3 (25%) were a combination of both. Five (42%) articles were published in peer-reviewed journals, and 7 (58%) were conference papers. Eight (67%) studies addressed maternal and child health (MCH), 1 (8%) addressed polio immunization, and 3 studies (25%) included multiple health themes. Moreover, 11 (92%) studies included between 7 and 81 CHW participants. Of these, CHWs were the sole or primary participants in 6 (50%) studies; in other cases, they constituted 1 group of participants, alongside mothers, field staff, or mobile shop/laptop owners. The remaining 1 (8%) study examined only clients. Note that sample sizes reported both in the text and Table 1 refer to the participants relevant to our research question and in some cases do not reflect the overall sample size of all participants featured in the study.

Table 1. Overview of studies.

Citation	Title	Health issue	Location	Sample	Study type	Publication
Adam et al, 2021 [30]	Evaluation of a community-based mobile video breastfeeding intervention in Khayelitsha, South Africa: The Philani MOVIE[a] cluster-randomized controlled trial	Maternal and child health (exclusive breastfeeding)	South Africa (urban)	1502 pregnant women	Quantitative (RCT[b])	Journal
Birukila et al, 2016 [31]	Reducing resistance to polio immunisation with free health camps and Bluetooth messaging: An update from Kaduna, Northern, Nigeria	Resistance to polio immunization	Nigeria (rural and urban)	12,418 households	Quantitative	Journal
Coetzee et al, 2018 [29]	Community health workers' experiences of using video teaching tools during home visits—A pilot study	Maternal and child health (HIV, alcohol, nutrition, and breastfeeding)	South Africa (urban)	24 CHWs[c]	Qualitative	Journal
Gopalakrishnan et al (2020)	Using mHealth to improve health care delivery in India: A qualitative examination of the perspectives of community health workers and beneficiaries	Maternal and newborn health	India (rural)	32 CHWs, 55 clients	Qualitative (interviews)	Journal
Isler et al, 2019 [32]	Iterative adaptation of a mobile nutrition video-based intervention across countries using human-centered design: Qualitative study	Maternal and child health (nutrition during pregnancy and breastfeeding)	Burkina Faso (rural)	CHWs, mothers (varying N)	Qualitative (focus groups, interviews, observations)	Journal
Kumar et al, 2015 [37]	Projecting health: Community-led video education for maternal health	Maternal and newborn health	India (rural)	CHWs, mothers, field staff (N not specified)	Qualitative (observations, interviews, focus groups)	Conference paper
Molapo et al, 2017 [33]	Video consumption patterns for first-time smartphone users – community health workers in Lesotho	Various	Lesotho (rural)	42 CHWs	Qualitative (observations, interviews, focus groups), quantitative (video views)	Conference paper
Molapo et al, 2016 [34]	Designing with community health workers: enabling productive participation through exploration	Various	Lesotho (rural)	54 CHWs	Qualitative (discussions, focus groups, workshops)	Conference paper
Molapo and Marsden, 2013 [35]	Software support for creating digital health training materials in the field	Various (eg, tuberculosis, sexual health)	Lesotho (rural)	15 CHWs	Qualitative (observations, interviews, focus groups, video logs)	Conference paper
Ramachandran et al, 2010 [38]	Mobile-izing health workers in rural India	Anemia and maternal health	India (rural)	7 CHWs	Qualitative (interviews, observations) and quantitative	Conference paper
Treatman and Lesh, 2012 [39]	Strengthening community health systems with localized multimedia	Maternal health, child nutrition, newborn health	India (rural)	8 CHWs	Qualitative (interviews)	Conference paper
Vashistha and Kumar, 2016 [40]	Mobile video dissemination for community health	Maternal and newborn health (birth preparedness, hand washing, exclusive breastfeeding, thermal care, delayed bathing)	India (rural)	84 mobile phone shop owners, 71 laptop owners, 81 CHWs	Quantitative (number of phone calls) and qualitative (interviews, focus groups, discussions)	Conference paper

[a]MOVIE: Mobile Video Intervention for Exclusive Breastfeeding.

[b]RCT: randomized controlled trial.

[c]CHW: community health worker.

Quantitative Assessments

Birukila et al [31] assessed the acceptance of videos containing messages to promote polio immunization in rural and urban Nigeria. The videos—described as pictorial, digitalized flipcharts—were shown to parents and caregivers in 21,242 households on CHWs' smartphones. Almost all (99.9%) of the 11,612 caregivers who watched the videos claimed that these videos met their health information needs, and 85.4% of the 12,418 mobile phone owners agreed to receive the videos via Bluetooth. Over the study period, CHWs shared the videos around 100 times a day.

The only randomized controlled trial (RCT) we encountered was the intervention by Adam et al [30] in South Africa, wherein 1502 pregnant mothers were randomized into 2 groups. The control group received standard of care (SOC) home-based infant feeding counseling by CHWs. Using tablets, the intervention group was shown videos on infant feeding in addition to the SOC. No differences in behavior (infant feeding practices) were observed between the groups at 1 month and 5 months follow-up, but the videos had replaced around 40% of the CHWs' face-to-face counseling, thereby freeing up time for other health-related tasks. The small increase in maternal knowledge, observed at the 1-month follow-up, was no longer present after 5 months.

Qualitative and Mixed Assessments

Ramachandran et al [38] evaluated portable multimedia content in rural India. In their study, 7 CHWs used smartphones to show educational videos on maternal health and anemia to pregnant women during weekly household visits. Some of the material was produced by CHWs and featured influential community members. CHWs approached the videos with enthusiasm, yet older CHWs struggled with the technical features of the smartphones. The devices were often used in a noninteractive manner due to a lack of training. However, CHW coaching mitigated these issues. A written test administered before and after the intervention revealed improvements in CHW knowledge of pregnancy danger signs and self-efficacy after the intervention.

Treatman et al [39] developed a smartphone app featuring culturally appropriate color illustrations and audio recordings in the local language containing health messages about topics in MCH. In their study, 8 CHWs in rural India tested the app and were then interviewed. The audio messages were considered more significant than the illustrations; an engaging speaker was deemed especially important. CHWs described the devices as fun to use and impressive to clients, who considered the health messages credible and trustworthy. CHWs preferred the phones over other job aides since they were easier to carry. However, CHWs doubted the effectiveness of the multimedia content if presented without facilitation and thus highlighted the importance of interaction with clients. Moreover, smartphones appeared inept for use in noisy environments or with groups of clients.

In urban South Africa, Coetzee et al [29] provided 24 CHWs with tablets to show videos to pregnant women and mothers during home visits. Pre- and postintervention focus groups were conducted. The tablets increased CHW motivation by amplifying the perceived importance of their work. The videos stimulated clients' interest and attention, improved CHW credibility and time efficiency, and triggered interest even among nontargeted household members. However, some CHWs worried about tablet theft and their credibility and social status being compromised by insufficient technologic capability. Sometimes, tablets were regarded as a means to avoid interaction with clients, especially when CHWs were tired. Moreover, some clients were concerned that the tablets might be recording them, compromising confidentiality.

Molapo et al [33-35] carried out a series of qualitative assessments based on interviews, focus groups, and observations in Lesotho, one of which [33] also contained a quantitative component. In the first intervention [35], a computer application allowed rural trainers of CHWs to create educational videos with local content transferable to the smartphones of 15 CHWs via Bluetooth. Repeated video views helped CHWs deepen their knowledge, and CHWs requested video material deemed especially important. Surprisingly, CHWs not only used the videos for their own education, as was intended, but they also shared them with community members and peers who did not possess smartphones. The health workers experienced a sense of pride, respect from others, and empowerment, and the videos helped them talk about topics that made them feel uncomfortable, such as sexual health.

During the second intervention [34], the existing video content was improved through community feedback. Since CHWs had started showing the videos to community members, their trainers created videos catered to this purpose. Different versions of the application were tested in the field by 54 CHWs, each equipped with a smartphone. CHWs usually showed the videos to groups of clients; they disliked pausing the videos for feedback or questions because the interruptions limited their perceived professionality.

In the last of the 3 interventions, Molapo et al [33] analyzed the results of their 17 months of fieldwork both qualitatively and quantitatively using log data of video views. The 42 CHWs preferred to watch the videos to completion and interact with their clients afterward instead of pausing the videos. Older CHWs handled the smartphones as well as their younger peers after appropriate instructions. In general, CHWs found the smartphones easy to use, though a lack of English literacy sometimes caused problems. Explicit graphic images, such as in videos about sexually transmitted infections, were popular and triggered discussion. The average number of views per video per CHW declined 24% to 87% percent in 16 months, which the authors attribute to a waning novelty effect. However, video views tended to increase shortly after CHW educational workshops, with views increasing for around 3 months. The number of views per video depended on the video's perceived importance and individual features, with a preference for videos showing influential community members.

In the study by Kumar et al [37] spanning 24 months, Indian CHWs showed educational videos on MCH to community members in 84 rural villages during monthly group gatherings. These events became so popular that CHWs began to organize them independently, thereby exceeding the researchers' expectations. The videos were played on small, battery-powered projectors; they facilitated CHWs' explanations, generated discussions, and highly elevated their social status. The latter was especially true for videos starring CHWs, giving them "celebrity status" [37]. Locally filmed videos were the most popular, as community members could relate to the content. Other advantages included the videos' repeatability and credibility boost for health messages.

Vashistha et al [40] attempted to identify the most effective means of distributing offline health videos on personal mobile phones. They equipped 81 CHWs, 84 mobile shop owners, and 71 laptop owners in rural India with videos promoting MCH. The videos featured unique phone numbers, and viewers were urged to call if they liked the video. The number of calls was recorded, and callbacks were conducted to gather viewers' opinions. By the end of the 14-week study period, mobile shop owners had distributed the video material to 6 times as many clients compared to laptop owners and CHWs. However, the number of calls received from videos distributed by CHWs far exceeded the calls from those disseminated by mobile shop or laptop owners. The authors provided 3 reasons for this finding: the CHWs had stronger ties with their clients, they were considered experts in their domain, and they seemed to have emphasized the importance of making the phone calls. CHWs regarded the technology as an effective way to enable clients to learn, review, and share health-related knowledge.

Gopalakrishnan et al [36] developed a software for smartphones and tested how its use would affect CHW-client interactions. In their study, 32 CHWs showed videos with health messages on MCH to 55 clients during home visits. Postintervention interviews revealed a divergence in the software's perceived utility between CHWs and beneficiaries; some initial technical difficulties notwithstanding, the former reported increases in CHW authority, the credibility of health messages, client attention, and the involvement of key household decision makers. They also noted a positive impact on behavior change (vaccination rates). However, for most beneficiaries, the software failed to improve interactions with CHWs. On the contrary, CHW-client interactions were harmed by rushed and short visits and the failure of CHWs to mediate interactions appropriately.

Finally, Isler et al [32] adapted a series of videos related to MCH and nutrition from 1 cultural setting (South Africa) to another (Burkina Faso). Animated videos with South African content and design received input from Burkanese CHWs and clients and were then modified to represent the Burkina Faso cultural, linguistic, and physical setting. Clients emphasized that the characters portrayed should reflect community members in appearance, behavior, and financial situation. Moreover, they recommended acknowledging local household structures and hierarchies during video presentations. The ease of tablet usage varied by CHW education and age. Some CHWs expressed technical concerns and preferred reducing the amount of information transmitted in each video viewing session.

Discussion

Principal Findings

Studies to date have been conducted in India and across Africa and have largely evaluated the practicability of smartphones or tablets as health promotion vehicles. Despite the diversity of study designs and cultural contexts, the studies have common findings regarding both the ability of mHealth to alleviate some challenges encountered by CHWs and some key drawbacks, such as equivocal effects on CHW-client interactions and frequent technical difficulties in the field. While locally produced media content proved popular, the potential of smart devices as catalysts for health behavior change remains elusive and merits larger-scale, quantitative interventions in the future.

Overcoming Challenges

mHealth technologies may mitigate some of the challenges experienced by CHWs in the promotion of healthy behaviors. mHealth can increase CHW health knowledge [35,38], thereby addressing CHWs' own educational deficits. CHWs themselves stressed that the ability to carry mobile videos helped them remember crucial health concepts [35], an assertion verified objectively in a written test [38]. Increased CHW health knowledge should benefit their interactions with clients and increase clients' respect for the health workers and their work.

As a novel technology, mHealth use by CHWs may stimulate community interest [29-32,35-37,39]. Clients, bystanders, and neighbors were interested mHealth deliveries [29,39], as were CHWs not possessing electronic devices [35]. While this interest may wane over time as recipients become accustomed to the technology, it appears possible to rekindle interest through the introduction of new material [33]. Constant innovations (in the form of new videos, apps, etc) may therefore be necessary to maintain interest. The rising personal ownership of mobile electronic devices in LMIC may limit such novelty effects in the future, making the case for constant technological or creative progress even stronger.

Part of the creativity required for this progress may well be shown by the CHWs themselves, as the introduction of the technology motivated them [29,38] and was overall enjoyable [39]. CHWs became innovative in the development of new media [38] and took ownership of the technology, for example, by independently organizing events going beyond the aims set out by the researchers [37]. At times, CHW motivation also extended to influential community members who became involved in the projects [38]. The increased levels of CHW motivation may be explained to some extent by their elevated levels of self-efficacy [38] and community recognition [29,30,35-37]. The mHealth technology gave CHWs a sense of pride and empowerment [35] and elevated their social status [30,37]. This, in turn, affected the extent to which clients accepted health messages, as they considered them trustworthy and credible [29,30,37,39]. Thus, by boosting CHW motivation, self-efficacy, and community status and by raising the credibility of health messages, mHealth may help CHWs promote healthy behaviors more effectively in their interactions with clients. However, the question remains whether these interactions improved with the use of mobile technology.

CHW-Client Interactions

The impact of mHealth on CHW-client interactions was ambivalent. On the one hand, CHWs sometimes viewed smart devices as complementary to their interaction with clients [39]. For instance, CHW-moderated group video sessions generated considerable discussion [37]. Devices were particularly beneficial for sensitive topics such as sexual health [35]. On the other hand, CHWs also used portable media to replace client conversations [29,30,38], for example, because they did not know what else to do [38], they were tired [29], or they wanted to save time [30]. In other cases, CHWs preferred to interact with clients only after the videos had played until completion [33,34]. Beneficiaries frequently considered video-assisted health promotion sessions as rushed or too short [36]. In 1 study [36], there was a considerable divergence in how CHWs and clients assessed the use of smartphones; while CHWs embraced them, beneficiaries criticized the quality of interactions. This suggests that mHealth-facilitated health promotion may simplify CHWs' work by replacing discussions with screen time—at the expense of clients' experience.

Personal interaction with clients is one of the factors that distinguish CHWs from other health-promoting agents. It is thus questionable whether mHealth can benefit CHWs' health-promoting efforts should the quality of face-to-face interactions be undermined. However, appropriate CHW training in how to use videos to stimulate discussions may reduce or eliminate the risk of noninteraction, as 1 study showed [38]. Further research should focus on mHealth's influence on the quality of CHW-client interactions and how a stimulating synergy between technology and interactions can be achieved.

Technical Difficulties

Even the best electronic device is of no avail if a CHW lacks the knowledge on how to use it. Some CHWs initially struggled when using portable media [32,36,38]. Younger CHWs generally adapted more quickly than their older colleagues [32,38], and higher education levels facilitated the adoption of the technology [32,38]. Due to the technological challenges, CHWs reported feeling anxious [29], nervous [32], and even worried about their community status being compromised [29,34]. However, the technical difficulties were reduced after CHWs received appropriate training [33,38], and even older CHWs learned to use the devices as effectively as their younger peers [33]. Having overcome the initial problems, CHWs enjoyed using the devices [33,39], considered touchscreens user friendly [33], and preferred portable devices over other, bulkier job aides [39]. Some CHWs reported that their skillful usage of the technology enhanced their perceived authority in the community [30]. Hence, proper CHW instruction is the key to converting smart devices from stressors and sources of discomfort into pleasant companions at work.

Go Local

Both CHWs and clients highlighted the importance of the media featuring local content [32,35,37-39]. When given the opportunity to create their own educational videos, CHWs chose to include testimonials from influential community members [38]. Videos featuring sequences of CHWs raised their social status in the community and enabled clients to identify with the content [37]. Clients preferred locally filmed videos over those shot in different locations [37]. The inclusion of locally appropriate color illustrations and the local language was also appreciated by clients [32,39], who emphasized that animated video content should resemble the local population in appearance and behaviors [32]. Hence, mHealth promotional strategies should adopt local features to maximize client identification with the material. Locally produced content thus has the potential to affect health behaviors more powerfully than material conceptualized elsewhere.

Client Learning and Behaviors

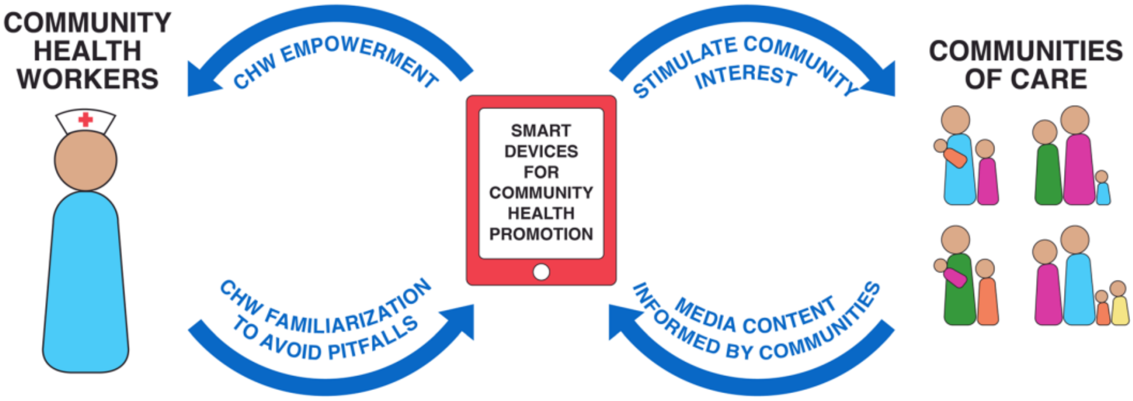
Perhaps the most important question in the context of mHealth and health promotion is whether mHealth helps CHWs improve clients' health behaviors. The reviewed literature contained almost no evidence relating to changes in clients' behaviors or process measures such as behavioral intent or health knowledge. Both clients [31] and CHWs [29,40] regarded the portable devices as enriching educational tools for the recipients of the interventions. Some CHWs provided anecdotal evidence that the use of the devices could have contributed to the adoption of health behaviors such as vaccinations [36]. However, only 1 (8%) of the 12 studies featured behavior change as an end point [30]. In this long-term RCT, the researchers observed no effect of the mHealth video intervention on health behaviors and only a small positive effect on health knowledge [30]. Thus, the potential of mHealth to bring about behavior change when employed by CHWs remains unclear.

Looking Ahead

The studies included in this review contain valuable information on the impact of mobile electronic devices on the delivery of health messages by CHWs. mHealth can empower CHWs and potentially help alleviate many challenges faced in the field. By stimulating community interest, health messages may be conveyed more effectively. Media content informed by the targeted communities themselves has been shown to be especially persuasive. However, the use of mobile media will always require careful training to maximize benefits and minimize potential pitfalls. CHWs should be familiarized with the technology and instructed on employing it as a complement to their interactions with clients, not as a replacement thereof. If used appropriately, smart devices may catalyze health promotion, benefitting CHWs and clients alike (Figure 3).

Nevertheless, our primary research question—whether mHealth improves CHW-led face-to-face delivery of public health messages and behavior change interventions—remains unanswered. Notably, all but 2 (83%) of the studies in this review are qualitative. While these are invaluable for planning, improvement, and evaluation, more large-scale quantitative studies, such as the included RCT [30], are needed with behavior change end points. In addition, 8 (67%) of the 12 studies in this review focus on MCH. This reflects the importance of MCH in low-income settings, but a wider scope of mHealth assessments would be desirable.

Figure 3. Smart devices as catalysts for community health promotion.



Limitations

Our review has some limitations. First, we only used 2 databases which, while wide-ranging in scope, did not capture all the relevant published studies we finally used. Therefore, we may have missed other published studies. Second, capturing all relevant research in this fast-moving field is difficult, as highlighted by most of the included work being only available as conference papers. These findings suggest the importance of future updates to this review.

Conclusions

Novel technological improvements may increase the effectiveness of CHW-led promotion of healthy behaviors. In this review, we show that smart mobile devices have the potential to enhance face-to-face interactions between CHWs and their clients, as these job aides address many of the challenges that CHWs commonly encounter in the field. However, we also find that the available evidence on our research question is scarce, largely qualitative, and focused on a limited scope of health outcomes. In particular, it is unclear whether mHealth helps CHWs change clients' health behaviors. Moreover, the impact of employing mHealth in the field is not all positive, as smart devices may burden CHWs with technological difficulties and lead them to act more passively in their interactions with clients. Further research is required to develop interventions to address this issue, along with large-scale quantitative interventions across a wider range of health outcomes to determine the full potential for interactive mHealth interventions to support CHW behavior change work in low-resource settings.

Acknowledgments

This research was funded in whole, or in part, by the Wellcome Trust through a personal fellowship to GH (Grant 210479/Z/18/Z). For the purpose of open access, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. Additional funding has been received from research grants from the European Union (Horizon 2020 and European Institute of Innovation and Technology Health), German Research Foundation (Deutsche Forschungsgemeinschaft), US National Institutes of Health, German Ministry of Education and Research, Alexander von Humboldt Foundation, Else-Kröner-Fresenius-Foundation, and the World Health Organization.

Authors' Contributions

GH, TB, and MG conceived and designed the study. MG screened the titles, abstracts, and full texts and led the quality assessment, data extraction, and drafting of the manuscript. FS screened titles, abstracts, and full texts and participated in quality assessment and data extraction. GH judged over disagreements. MG, GH, FS, MA, AV, and JG interpreted the results and participated in the manuscript drafts and finalization. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search entries.
[DOCX File , 13 KB - mhealth_v11i1e42023_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews) checklist.
[DOCX File , 84 KB - mhealth_v11i1e42023_app2.docx]

References

1. Perry HB, Zulliger R, Rogers MM. Community health workers in low-, middle-, and high-income countries: an overview of their history, recent evolution, and current effectiveness. Annu Rev Public Health 2014 Mar;35:399-421. [doi: 10.1146/annurev-publhealth-032013-182354] [Medline: 24387091]
2. Tulenko K, Møgedal S, Afzal MM, Frymus D, Oshin A, Pate M, et al. Community health workers for universal health-care coverage: from fragmentation to synergy. Bull World Health Organ 2013 Nov 01;91(11):847-852. [doi: 10.2471/blt.13.118745]
3. Katigbak C, Van Devanter N, Islam N, Trinh-Shevrin C. Partners in health: a conceptual framework for the role of community health workers in facilitating patients' adoption of healthy behaviors. Am J Public Health 2015 May;105(5):872-880. [doi: 10.2105/ajph.2014.302411]
4. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
5. Bouton ME. Why behavior change is difficult to sustain. Prev Med 2014 Nov;68:29-36 [FREE Full text] [doi: 10.1016/j.ypmed.2014.06.010] [Medline: 24937649]
6. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004 Dec;19(12):1228-1239 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.40153.x] [Medline: 15610334]
7. Osborn C, Paasche-Orlow MK, Bailey SC, Wolf MS. The mechanisms linking health literacy to behavior and health status. Am J Health Behav 2011;35(1):118-128 [FREE Full text] [doi: 10.5993/ajhb.35.1.11] [Medline: 20950164]
8. UNESCO e-atlas of literacy. UNESCO Institute for Statistics. 2016. URL: https://tellmaps.com/uis/literacy/#!/tellmap/-601865091 [accessed 2019-03-03]
9. Haq Z, Hafeez A. Knowledge and communication needs assessment of community health workers in a developing country: a qualitative study. Hum Resour Health 2009 Jul 21;7(1). [doi: 10.1186/1478-4491-7-59]
10. Ozano K, Simkhada P, Thann K, Khatri R. Improving local health through community health workers in Cambodia: challenges and solutions. Hum Resour Health 2018 Jan 06;16(1):2 [FREE Full text] [doi: 10.1186/s12960-017-0262-8] [Medline: 29304869]
11. Grossman-Kahn R, Schoen J, Mallett JW, Brentani A, Kaselitz E, Heisler M. Challenges facing community health workers in Brazil's Family Health Strategy: a qualitative study. Int J Health Plann Manage 2018 Apr 21;33(2):309-320 [FREE Full text] [doi: 10.1002/hpm.2456] [Medline: 28940668]
12. Jaskiewicz W, Tulenko K. Increasing community health worker productivity and effectiveness: a review of the influence of the work environment. Hum Resour Health 2012 Sep 27;10(1):38 [FREE Full text] [doi: 10.1186/1478-4491-10-38] [Medline: 23017131]
13. Mayer RE. Multimedia learning: Are we asking the right questions? Educ Psychol 1997 Jan;32(1):1-19. [doi: 10.1207/s15326985ep3201_1]
14. Clark JM, Paivio A. Dual coding theory and education. Educ Psychol Rev 1991 Sep;3(3):149-210. [doi: 10.1007/bf01320076]
15. Clark RC, Mayer RE. e-Learning: promise and pitfalls. In: e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 3rd Edition. Hoboken, NJ: John Wiley & Sons; Aug 16, 2011:7-28.
16. Mayer R. Applying the science of learning to medical education. Med Educ 2010 Jun;44(6):543-549. [doi: 10.1111/j.1365-2923.2010.03624.x] [Medline: 20604850]
17. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center. 2016 Feb 22. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2019-01-10]
18. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015 Nov 24;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
19. Van Campenhout B, Vandevelde S, Walukano W, Van Asten P. Agricultural extension messages using video on portable devices increased knowledge about seed selection, storage and handling among smallholder potato farmers in southwestern Uganda. PLoS One 2017 Jan 25;12(1):e0169557 [FREE Full text] [doi: 10.1371/journal.pone.0169557] [Medline: 28122005]
20. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013 Jun;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
21. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014 Oct 27;7:25606. [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
22. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AHA, et al. Mobile Health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17. [doi: 10.2196/jmir.2130] [Medline: 23353680]
23. White A, Thomas DS, Ezeanochie N, Bull S. Health worker mHealth utilization. Comput Inform Nurs 2016;34(5):206-213. [doi: 10.1097/cin.0000000000000231]

24. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Dec 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Olaniran A, Smith H, Unkels R, Bar-Zeev S, van den Broek N. Who is a community health worker? - a systematic review of definitions. Glob Health Action 2017 Jan 27;10(1):1272223 [FREE Full text] [doi: 10.1080/16549716.2017.1272223] [Medline: 28222653]
27. Strengthening the performance of community health workers in primary health care : report of a WHO Study Group [meeting held in Geneva from 2 to 9 December 1987] . World Health Organization. 1989. URL: https://apps.who.int/iris/handle/10665/39568 [accessed 2019-03-03]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
29. Coetzee B, Kohrman H, Tomlinson M, Mbewu N, Le Roux I, Adam M. Community health workers' experiences of using video teaching tools during home visits-A pilot study. Health Soc Care Community 2018 Mar 05;26(2):167-175 [FREE Full text] [doi: 10.1111/hsc.12488] [Medline: 28872210]
30. Adam M, Johnston J, Job N, Dronavalli M, Le Roux I, Mbewu N, et al. Evaluation of a community-based mobile video breastfeeding intervention in Khayelitsha, South Africa: The Philani MOVIE cluster-randomized controlled trial. PLoS Med 2021 Sep 28;18(9):e1003744 [FREE Full text] [doi: 10.1371/journal.pmed.1003744] [Medline: 34582438]
31. Birukila G, Babale SM, Epstein H, Gugong V, Anger R, Corkum M, et al. Reducing resistance to polio immunisation with free health camps and Bluetooth messaging: An update from Kaduna, Northern, Nigeria. Glob Public Health 2017 Jan 21;12(1):19-30. [doi: 10.1080/17441692.2016.1152283] [Medline: 26998877]
32. Isler J, Sawadogo NH, Harling G, Bärnighausen T, Adam M, Kagoné M, et al. Iterative adaptation of a mobile nutrition video-based intervention across countries using human-centered design: qualitative study. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e13604 [FREE Full text] [doi: 10.2196/13604] [Medline: 31710302]
33. Molapo M, Densmore M, DeRenzi B. Video consumption patterns for first time smartphone users: community health workers in Lesotho. 2017 Presented at: CHI Conference on Human Factors in Computing Systems; May 6-11; Denver, CO p. 6159-6170. [doi: 10.1145/3025453.3025616]
34. Molapo M, Densmore M, Morie L. Designing with community health workers: enabling productive participation through exploration. In: Proceedings of the First African Conference on Human Computer Interaction. 2016 Presented at: 1st African Conference on Human Computer Interaction; November 21-25; Nairobi, Kenya p. 58-68.
35. Molapo M. Software support for creating digital health training materials in the field. 2013 Presented at: 6th International Conference on Information and Communication Technologies and Development; December 7-10; Cape Town, South Africa p. 205-214.
36. Gopalakrishnan L, Buback L, Fernald L, Walker D, Diamond-Smith N, CAS Evaluation Consortium. Using mHealth to improve health care delivery in India: A qualitative examination of the perspectives of community health workers and beneficiaries. PLoS One 2020 Jan 15;15(1):e0227451 [FREE Full text] [doi: 10.1371/journal.pone.0227451] [Medline: 31940326]
37. Kumar N, Perrier T, Desmond M, Israel-Ballard K, Kumar V, Mahapatra S, et al. Projecting health: community-led video education for maternal health. In: Proceedings of the Seventh International Conference on Information and Communication Technologies and Development. 2015 Presented at: 7th International Conference on Information and Communication Technologies and Development (ICTD '15); May 15-18; Singapore p. 1-10. [doi: 10.1145/2737856.2738023]
38. Ramachandran D, Canny J, Dutta Das P, Edward C. Mobile-izing health workers in rural India. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '10); April 10-15; Atlanta, GA p. 1889-1898. [doi: https://doi.org/10.1145/1753326.1753610]
39. Treatman D, Lesh N. Strengthening community health systems with localized multimedia. 2012 Presented at: 3rd International Conference on Mobile Communication for Development (M4D); February 28-29; New Delhi, India p. 7-22.
40. Vashistha A, Kumar N, Mishra A, Anderson R. Mobile video dissemination for community health. In: Proceedings of the 8th International Conference on Information and Communication Technologies and Development. 2016 Presented at: 8th International Conference on Information and Communication Technologies and Development; June 3-6; Ann Arbor, MI. [doi: 10.1145/2909609.2909655]

Abbreviations

CHW: community health worker
LMIC: low- and middle-income countries
MCH: maternal and child health

MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews
RCT: randomized controlled trial
SOC: standard of care

Edited by L Buis; submitted 18.08.22; peer-reviewed by E Syundyukov, C Calvo-Lobo; comments to author 04.11.22; revised version received 06.12.22; accepted 20.12.22; published 22.02.23.

Please cite as:
Greuel M, Sy F, Bärnighausen T, Adam M, Vandormael A, Gates J, Harling G
Community Health Worker Use of Smart Devices for Health Promotion: Scoping Review
JMIR Mhealth Uhealth 2023;11:e42023
URL: https://mhealth.jmir.org/2023/1/e42023
doi:10.2196/42023
PMID:36811947

Review

Design Features Associated With Engagement in Mobile Health Physical Activity Interventions Among Youth: Systematic Review of Qualitative and Quantitative Studies

Ayla Schwarz[1], PhD; Laura H H Winkens[1], PhD; Emely de Vet[1], PhD; Dian Ossendrijver[1], MSc; Kirsten Bouwsema[1], MSc; Monique Simons[1], PhD

Department of Social Sciences, Chair Group Consumption & Healthy Lifestyles, Wageningen University & Research, Wageningen, Netherlands

Corresponding Author:
Ayla Schwarz, PhD
Department of Social Sciences, Chair Group Consumption & Healthy Lifestyles
Wageningen University & Research
Hollandseweg 1
Wageningen, 6706KN
Netherlands
Phone: 31 0639187108
Email: ayla.schwarz@wur.nl

Abstract

Background: Globally, 81% of youth do not meet the physical activity (PA) guidelines. Youth of families with a low socioeconomic position are less likely to meet the recommended PA guidelines. Mobile health (mHealth) interventions are preferred by youth over traditional in-person approaches and are in line with their media preferences. Despite the promise of mHealth interventions in promoting PA, a common challenge is to engage users in the long term or effectively. Earlier reviews highlighted the association of different design features (eg, notifications and rewards) with engagement among adults. However, little is known about which design features are important for increasing engagement among youth.

Objective: To inform the design process of future mHealth tools, it is important to investigate the design features that can yield effective user engagement. This systematic review aimed to identify which design features are associated with engagement in mHealth PA interventions among youth who were aged between 4 and 18 years.

Methods: A systematic search was conducted in EBSCOhost (MEDLINE, APA PsycINFO, and Psychology & Behavioral Sciences Collection) and Scopus. Qualitative and quantitative studies were included if they documented design features associated with engagement. Design features and related behavior change techniques and engagement measures were extracted. Study quality was assessed according to the Mixed Method Assessment Tool, and one-third of all screening and data extraction were double coded by a second reviewer.

Results: Studies (n=21) showed that various features were associated with engagement, such as a clear interface, rewards, multiplayer game mode, social interaction, variety of challenges with personalized difficulty level, self-monitoring, and variety of customization options among others, including self-set goals, personalized feedback, progress, and a narrative. In contrast, various features need to be carefully considered while designing mHealth PA interventions, such as sounds, competition, instructions, notifications, virtual maps, or self-monitoring, facilitated by manual input. In addition, technical functionality can be considered as a prerequisite for engagement. Research addressing youth from low socioeconomic position families is very limited with regard to engagement in mHealth apps.

Conclusions: Mismatches between different design features in terms of target group, study design, and content translation from behavior change techniques to design features are highlighted and set up in a design guideline and future research agenda.

Trial Registration: PROSPERO CRD42021254989; https://tinyurl.com/5n6ppz24

(JMIR Mhealth Uhealth 2023;11:e40898) doi:10.2196/40898

KEYWORDS

systematic review; youth; physical activity; design features; engagement; mHealth; mobile health; mobile phone

Introduction

Background

Physical activity (PA) in youth is associated with a variety of health benefits [1], including physical [2] and mental health benefits [3]. Despite the health benefits, 81% of youth (ie, those in childhood and adolescence) globally do not meet the PA guidelines [4,5] of daily 60 minutes of moderate- to vigorous-intensity physical activity (MVPA) and vigorous activities 3 days per week [1]. In Europe, only 19% of adolescents comply with the MVPA guidelines, and higher family affluence is associated with higher levels of MVPA [6]. Moreover, PA in youth is reported to track into adulthood, underlying the importance of promoting PA in the youth [7,8]. Also, the youth of families with a low socioeconomic position (SEP) are less likely to meet the recommended PA guidelines [9,10]. Therefore, effective and socially acceptable PA interventions are needed [11], especially among youth of families with a low SEP. SEP is often measured in terms of education (attainment), income, and occupation status, which are often interrelated and related to the social and economic resources available [12].

Mobile health (mHealth) tools such as smartphone apps can be cost-effective [13] in changing total PA [14,15] and daily steps [14]. mHealth interventions are preferred by youth over traditional in-person approaches [16-18] and are in line with preferences of youth, with regard to multimedia formats (ie, text, sound, and video) [19]. Mobile devices and apps have gained popularity in the daily life of increasingly younger age groups of youth, starting at the age of 3 or 4 years already [20-24]. Families of young children, especially those with a low-SEP background, indicate concern about inappropriate content [20], which might suggest that appropriate content might rather be supported by guardians. Appropriate content may refer to different forms of app use that might not necessarily increase current screen time use, as the design of the mHealth intervention may have considered a limited use time frame, only stimulating sporadic screen time (eg, integration to the direct physical environment to stimulate children to go outside and be physically active instead of the requirement of using the digital tool when being physically active) or stimulating the co-use of young children and guardians [25]. Earlier systematic reviews indicated that currently, foremost, adults are included in mHealth effectiveness studies, and limited research has been conducted among youth [26]. People with high SEP compared with people with low SEP report foremost positive health effects of the same digital intervention [27]. Apps are primarily designed for people with high SEP, and people with low SEP are often reported as being difficult to reach [28]. In contrast to mHealth apps, youth from low-SEP families engage in more screen time [29] and are more involved on the web [30] and active [31] gaming. This suggests that approaching youth with low SEP via apps (eg, screen and games) can be useful and beneficial, yet other ingredients (eg, tailoring) are needed to reach them effectively in mHealth apps [32].

Despite the promise of digital or mHealth interventions in promoting PA [26], a common challenge is to engage users in the long term [14] or effectively [26]. With regard to including people with a low-SEP background in research, studies identified challenges in reaching the target group [33] and engaging them in research [28] and in the digital health intervention [34]. In addition, eHealth and mHealth studies focus on a general target group and often do not research subgroups (ie, low-SEP groups) [35].

Engagement refers to the involvement and motivation of the user in the intervention [36]. It refers to “(1) the extent (e.g., amount, frequency, duration, depth) of usage and (2) the subjective experience characterized by attention, interest and affect” (ie, enjoyment) [37]. Attention refers to the degree of focus or absorption versus distraction in the intervention. Interest refers to feelings of interest or fascination versus boredom with the intervention. Enjoyment reflects the enjoyable experience of fun or pleasure versus frustration and annoyance while using the intervention [37,38]. Engagement can contribute to the overall effectiveness of digital interventions [26,39], meaning that to be effective, it is important that the user experiences a sufficient degree of engagement. The extent of use or dose is less predictive of engagement than the subjective experience to achieve the intended outcomes of the intervention (ie, effective engagement) [40]. This means that frequently engaging with the intervention is not always required to be effectively engaged with the intervention or the behavior eventually.

Currently, engagement is still not commonly reported in terms of use data (ie, extent) [26] and experience data (ie, subjective engagement) [41]. To improve intervention exposure and related intervention efficacy, it is important to better understand the design features (ie, active ingredients of an app that are in best cases informed by theory and translated into design, often referred to as “persuasive features,” features, and elements) contributing to engagement [26]. A review of commercial apps indicated low engagement scores and suggested investigating features that improve engagement with an app among youth [42], as the features may differ with those studied in adults [37]. Earlier studies reported a common decline in app use over time and called for investigating features that contribute to intervention uptake and engagement [14,26] also among particular subgroups, such as people with a low-SEP background [27,32,43,44]. For example, research among low-SEP groups indicated that frustration with particular design features and navigations (eg, data log) hampered app use [45].

When designing mHealth interventions, several active ingredients, otherwise indicated as behavior change techniques (BCTs), are applied and translated to app features to foster behavior change. In addition to low engagement scores, existing reviews reveal that a limited number of BCTs, such as instructions, encouragement, rewards, and feedback on performance, which are essential components of interventions to promote behavior change [46], are currently applied in mHealth apps [42]. BCTs can be directly translated into gamification or app features (eg, goal-setting translated into challenges), here referred to as design features [47]. Studies suggest that applying either a larger number of BCTs [42]—particular BCTs (eg, self-monitoring, feedback, goal-setting, rewards, reminders, and social support under certain circumstances) or a particular combination of BCTs (eg, problem-solving and rewards) [48]—may contribute to engagement in digital interventions [44,47,48].

Objective

To inform the design process of future mHealth tools, it is important to investigate the design features that can yield effective user engagement. Research has demonstrated that for adults, user guidance, well-designed reminders, self-monitoring, positive feedback, rewards (eg, lottery), goal-setting, personalization, social networking, and health message framing [43,44,49] were associated with higher user engagement. However, most studies included adult samples, targeted mental health instead of PA among adolescents, or did not focus on low-SEP groups [50-52]. To our knowledge, no systematic review that primarily seeks to identify design features in mHealth PA intervention among youth, especially youth from low-SEP families that have been associated with engagement, has yet been conducted. Therefore, this systematic review aimed to identify which features are associated with engagement in mHealth PA interventions among youth who were aged between 4 and 18 years. This could help inform app developers and (digital) behavior change intervention researchers to better integrate engagement during the design process of mobile behavioral change interventions.

Methods

Systematic Review

The research protocol of the review was registered with the International Prospective Register of Systematic Reviews (PROSPERO; #CRD42021254989). Reporting complied with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1 [53]) statement [54].

Search Strategy

A systematic search was conducted on June 2, 2021, updated on June 24, 2022, and included the following databases: EBSCOhost (MEDLINE, APA PsycINFO, and Psychology & Behavioral Sciences Collection) and Scopus. The search terms that included synonyms relating to engagement and the population (youth), the intervention (mHealth), and outcome (PA) were informed by earlier reviews [37,55,56] and were eventually reviewed by an academic librarian. Multimedia Appendix 2 provides the complete search terms and synonyms used.

Eligibility Criteria

Quantitative, qualitative, and mixed methods studies were included in the review, as engagement is an emerging field of research, and the aim of this review was to gain a broad overview of features that are considered engaging. Available full-text research articles or conference papers were eligible in case they (1) described the development (ie, user-testing) or evaluation of an mHealth intervention (or more broadly the use of healthy lifestyles as long as specific information about PA could be retrieved), designed for PA promotion; (2) reported on the extent of use or subjective experience; (3) reported an association (ie, relation of design features with engagement, either assessed qualitatively and referred to "self-perceived association" or quantitatively referred to "association"); (4) included an intervention that was designed for healthy youth who were aged between 4 and 18 years (in line with Dutch PA guidelines [57] and in line with the increasing popularity of mobile devices in the daily life of [increasingly younger] youth) [20-24] or if the mean sample age was within this range; (5) were written in English; and (6) were published between 2010 and 2021 (owing to the smartphone use increase among youth, the use of mHealth interventions has been increasing since 2010).

Papers were excluded in case they (1) included a PA intervention that was not mobile based; (2) did not report on the extent of use or subjective experience; (3) did not report an association between a particular intervention feature and engagement; and (4) included either a study population of children who were aged <4 years, adults, or older adults or a population whose mean sample age was not within the range of 4 to 18 years (in line with Dutch national PA recommendation age range). Moreover, editorials, opinion papers, case studies, research protocols, design papers not including any user-testing, book chapters, systematic reviews, or meta-analyses were excluded (Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Population
  - Youth was defined as individuals who were:
    - healthy
    - aged between 4 and 18 years or the mean age of sample in the same age range
- Intervention
  - Physical activity mobile health interventions that are:
    - developed (ie, offered for user-testing) or evaluated
    - mobile
    - designed for physical activity promotion (or healthy lifestyle, including physical activity)
    - designed for the target group of youth (4-18 years old)
- Comparison
  - Not applicable
- Outcome
  - Drivers of the extent of use or subjective experience with regard to physical activity
- Study type
  - Quantitative, qualitative, and mixed methods studies that:
    - report on behavioral or subjective engagement
    - report a relation or association between the particular feature and engagement
    - are original and available full-text research articles or conference paper

Exclusion criteria

- Population
  - Youth in disease recovery and rehabilitation, experiencing a chronic disease (ie, overweight, obesity, and diabetes)
  - Children aged <4 years or mean age not within the age range of 4 to 18 years and adults or older adults
- Intervention
  - Physical activity interventions that:
    - were still in the early development phase (ie, not offering an app but, for example, only a general discussion about apps, paper mock-ups)
    - used different formats than mobile
    - only included wearables (eg, smartwatches) without a mobile health app
- Comparison
  - Not applicable
- Outcome
  - No drivers of the extent of use or subjective experience with regard to physical activity
- Study type
  - Quantitative, qualitative, and mixed methods studies that:
    - do not report on behavioral or subjective engagement
    - do not report a relation or association between the particular feature and engagement, for example, engagement is only described in general without any relation to the particular feature

- are editorials, opinion papers, case studies, research protocols, or design papers not including any user-testing, book chapters, systematic review, or meta-analysis

Screening and Data Extraction

All the papers identified through the searches were downloaded into Rayyan software (Rayyan Systems Inc), a systematic review software, and duplicates were removed. A multistep strategy was applied. First, titles and abstracts were screened, followed by full-text reading to determine eligibility (DO). One-third of the titles and abstracts and full texts were reviewed by a second reviewer (KB). Interrater reliability for the titles and abstracts and full texts was substantial and moderate (Cohen κ=0.73 and Cohen κ=0.54), respectively. Disagreements regarding the title and abstract and full texts were resolved through discussion between the reviewers (DO and KB). Any disagreement between the 2 reviewers was discussed, and a third reviewer (AS) was consulted if necessary.

Data extraction was completed by both reviewers (DO and KB). Disagreement regarding data extraction was resolved through discussion between the reviewers (DO and KB). Background information (1-3) was extracted by one reviewer (DO), and the rest of the data (4-7) were extracted by two reviewers (DO and KB): (1) publication information (author and year); (2) study information (country and target group [number, percentage of female, percentage of people of low SEP, mean age, and age range]), and 3) app description (name, device, operating system, aim, and theory used; based on the mHealth taxonomy) [58]; (4) engagement measures (engagement measure or definition); (5) presence of BCT and, if present, behavioral change measures (based on the ABACUS scale) [59]; (6) design features (type of feature and 1 feature vs multiple features); and (7) association with engagement (short summary of association of particular features and engagement). The engagement measure of the original individual study was extracted (ie, usability, motivation, enjoyment, and liking) and then coded according to attention (ie, paying attention to mHealth app), interest (ie, feeling interested in mHealth app), and enjoyment (ie, experiencing enjoyment while using it; Textbox 2) [37].

As the researchers of this study were not aware of one framework of design features, 3 (digital) behavior scientist researchers (AS, LW, and MS) developed a coding frame to code all design features based on earlier studies on PA apps; gamification (ie, application of any features [eg, badges, points, avatars etc] in a nongame setting); and game research (ie, full-fledged and rule-based games) [61]. Free codes were created if they did not suit 1 category. Owing to the design process of translating BCTs into design features, obviously, overlap between the features and BCTs occurs (ie, goals vs goal-setting). Design features were coded as interface esthetics, challenges, narrative, levels, feedback, monitoring, customization or personalization, reinforcement, navigation, goals, social, progress, and credibility (Textbox 3).

Textbox 2. Coding scheme of engagement and definitions.

Interest

- Experiencing interest, boredom, fascination, intrigue, or indifference [37]
- Cognitive state that is occurring spontaneously and relates to liking and willful engagement [60]

Attention

- Experiencing focus, inattention, absorption, distraction, or mindfulness [37]
- Cognitive state of focused awareness and relates to focalization and concentration [60]

Enjoyment

- Experiencing frustration, annoyance, enjoyment, fun, or pleasure [37]
- Relates to the sensory experience and relates to pleasure and activation [60]

Textbox 3. Coding framework for design features and definitions.

Interface esthetics

- Graphics, color, sound, music, and language [55,56]

Challenges

- “Time-limited goals or competitions” [55,62-64]

Narrative

- Story [54,61,63]
- Levels [55,62,64]
- Feedback [55,56,62,63]

Monitoring

- Self-monitoring, monitoring, and tracking of user’s behavior (eg, smartphone sensor and wearable) [55,56,63,65]

Customization or personalization

- Assessment of user’s characteristics and provision of personal information and avatars [55,56,62,63]

Reinforcement

- Reminders and rewards (badges, points, unlockable content [“access to enhanced functionality or content for accumulating experience or achieving a specific goal”], leaderboard, certificates, medals, cash or gifts, congratulation messages, and updates) [55,56,62-65]

Navigation

- Navigation support (menu bar, search bar, etc), instructions, and guidance [55,56]

Goals

- “Performance-based measures and targets,” either set by the app or by the user itself [55,63-65]

Social

- Social messages (communicating with others), “performance is publicly displayed,” community, competition (compete with each other), and collaboration (work together) [55,56,62-65]

Progress

- Progress, achievement, and levels [55,62-64]

Credibility

- Absence of advertisement, credible sources or information, privacy policy, and password protection [56,63]

Quality Assessment

Quality was assessed according to the Mixed Method Assessment Tool [66], and it was not used as a basis for study exclusion but to support the interpretation of results (Multimedia Appendix 3 [67-87]). All the studies were coded by 1 reviewer (DO), and one-third of the studies were double coded (KB). Interrater reliability was substantial (κ=0.63). Any disagreement between the 2 reviewers was discussed, and a third reviewer (AS) was consulted if necessary.

Synthesis of Results

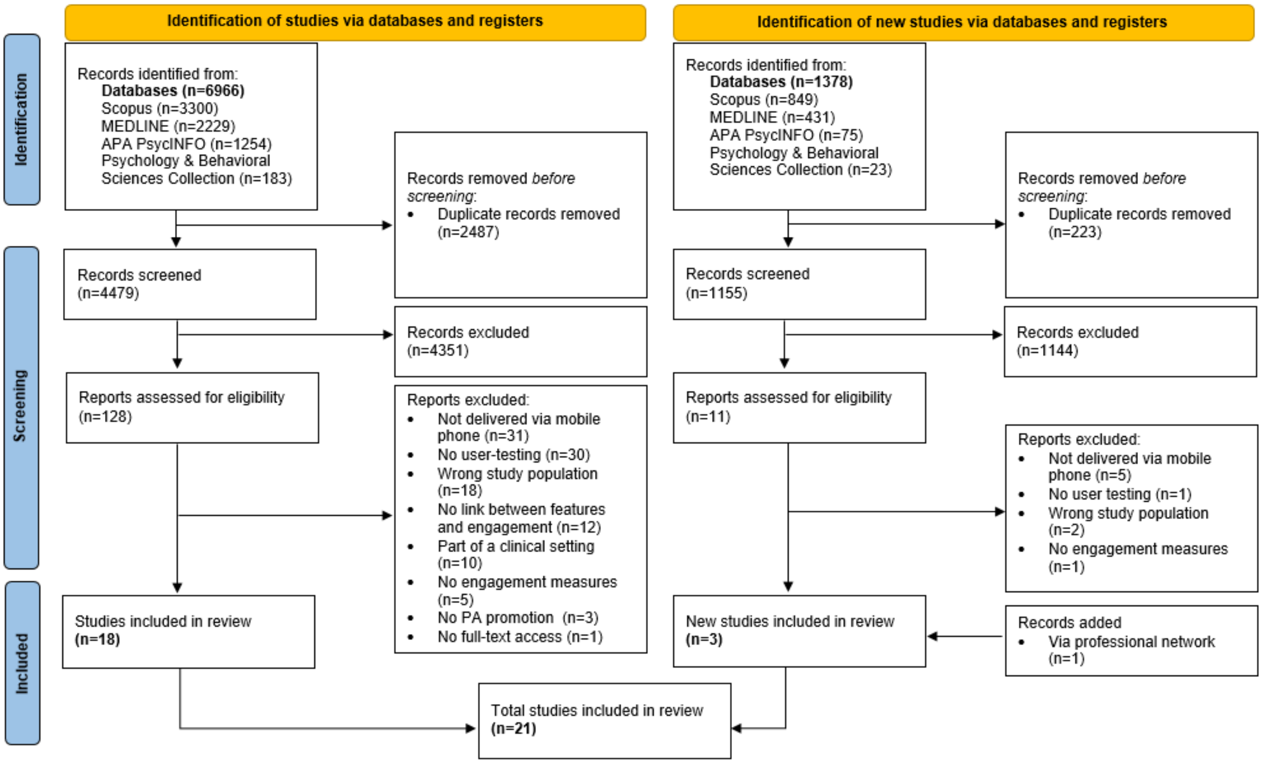
A systematic narrative format [88] structured around the design features that were associated with engagement was applied. All the features that have been actually tested (ie, user-testing) were reported. In case the same study also provided suggestions by participants to further develop the app (ie, hypothetical design features not tested but suggested), this was marked explicitly as a suggestion. In addition, the presence or absence of BCTs was outlined, indicating whether theories had been applied in the design phase of design features or not. A subgroup analysis was conducted for youth from low-SEP families. The categorization of SEP was derived from the classification of the original studies. Research results were not grouped by the origin of the data (ie, qualitative, quantitative, or mixed methods), but it was indicated explicitly for each research outcome.

Results

Description of Included Studies

In total, 21 studies were included in this review (Figure 1). The studies originated from different countries (15) and were conducted in Europe (n=10), North America (n=4), Australia (n=3), Asia (n=3), and South America (n=1). These studies were published between 2012 and 2021. Mobile interventions were used and evaluated by youth who were aged between 3 and 18 years. Sample sizes ranged from 5 to 354 youths, with a total sample size of 1443 youths. All the studies retrieved informed consents from parents and youth participants, except for 5 studies that only collected consent from the youth participants alone [67,68] or from the school principal [70], were conducted in parent-child dyads [69], or did not clearly state this [71]. Studies included different engagement measures varying from acceptability, experience, usability, motivation, feasibility, enjoyment, engagement, satisfaction, or interest. When categorizing the measures in terms of interest, attention, or enjoyment, most studies (15/21, 71%) measured solely interest, especially in terms of features that were in general liked by the users. A combination of interest and attention (2/21, 10%), solely attention (1/21, 5%), solely enjoyment (1/21, 5%), or solely engagement (2/21, 10%) was rarely studied. Engagement was mainly measured as the primary outcome in some studies (14/21, 67%). Most studies (12/21, 57%) used a mixed methods design, ranging from posttest prototype studies to quasi-experimental studies to randomized controlled trials, mixed with designs such as qualitative exit interviews, focus groups, or think-aloud studies. In total, 24% (5/21) of studies included a qualitative study design (interview and focus group study), and 19% (4/21) of studies included a quantitative study design (experimental design with quantitative survey). Studies relevant to this review included qualitative data (10/21, 48%), quantitative data (4/21, 19%), or mixed methods data (7/21, 33%) to determine the association between design features and engagement.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart illustrating the inclusion and exclusion of studies. PA: physical activity.



On the basis of Mixed Method Assessment Tool [66], the quality of the included studies was mixed. In general, the studies that measured engagement as the primary study outcome had a moderate to high quality, especially scoring high on transparency regarding the data collection methods and analysis. Studies that focused on engagement as a secondary outcome scored rather low on the quality of the study design applicable to measure engagement (eg, including nonvalidated methods such as a number of open exit questions in a larger randomized controlled trial), although the studies scored high on the design (eg, randomized controlled trial) to measure the primary outcome (eg, PA).

mHealth Interventions

In general, the mobile interventions tested in the included studies (N=21) were either developed for commercial purposes (n=4, 19%) or for study purposes (n=17, 81%). Most of the interventions (14/21, 67%) were designed to improve PA behavior. In addition, 33% (7/21) of studies focused also on additional health behaviors associated with a healthy lifestyle (eg, nutrition or sleep) in addition to PA behavior. The studies included different types of mHealth interventions, ranging from mobile games (10/21, 48%) to mobile apps (8/21, 38%) to mobile text messaging systems (3/21, 14%). Mobile games mainly focused on PA (9/10, 90% studies); included different PA activities (ie, ranging from simple arm movements such as swinging arm to complex full-body movement such as running); and applied different inputs (ie, ranging from manual input to camera data to Bluetooth, GPS, or PA sensor data). Games ranged from short-term activities (eg, push-ups) to long-term gameplay (eg, location-based treasure hunt games). Mobile apps ranged from narrative-driven PA guidance to GPS-tracking information. Mobile text messaging systems especially focused on generating and adapting goal-setting, facilitating self-monitoring, and providing feedback [72,73] and informative content (ie, quizzes and factoids) [74]. Furthermore, 67% (2/3) of the studies focused also on health behaviors other than PA (Multimedia Appendix 4 [67-87]). The interventions tested in the included studies were perceived as engaging by the youth. Youth liked to play games that aimed to increase their PA level and evaluated apps or text messaging systems as easy to use and supportive in improving PA or lifestyle behaviors. Multiple studies (6/21, 29%) found that poor functionality and technical issues are considered a barrier for engagement with mobile interventions [67,68,75-77] and suggested to include technical improvements [76].

In the following section, all the design features are summarized, starting with the design feature researched most often. A detailed overview of the results is presented in Table 1. All tested design features (ie, features offered to participants via an app and actually tested) are outlined as an association between the particular design feature and engagement (either as self-perceived association in qualitative studies or as tested association in quantitative studies). In case the participants provided further suggestions on design features that might be included in a future version (ie, feature not tested or only hypothetical), it was explicitly referred to as a suggestion. The results section concludes with a subanalysis on BCTs and low-SEP subgroups.

Table 1. Design features and association with user engagement.

Details	Association (ie, tested)	Suggestion (ie, hypothetical)
Interface aesthetics		
Language		
PA[a] and nutrition messages (ranging from information to quizzes)	[74][b,c] [67][b,d] [74][b,c]	[67][d]
Familiar content	[74][c,e]	N/A[f]
Short messages	[78][b,g] [74][b,c]	N/A
Cheerful and personal tone	[78][b,g] [74][b,c]	N/A
Visual		
Clearly presented text	[68][b]	N/A
Large text blocks and complicated language	[68][e,g]	N/A
Clear and realistic visualization (eg, infographics, figures, tables, pictures, symbols, and videos)	[79][b,g] [68][b,g]	[68][g] [67][d] [76][c] [77][d]
Quality not reflecting reality	[70][e,g]	N/A
Sound		
Sounds associated with arm movements or game activities	[75,80][b,c]	N/A
Sounds associated with every click	[81][e,g]	N/A
Monotone and robotic voice	[75][c,e]	N/A
General interface		
Lack of sophistication	[75][c,e] [79][e,g]	N/A
Reinforcement		
Rewards		
Tangible rewards (prize, medal, new Pokémon, and healthy real-world actions)	[75,82][b,c] [78][b,g]	[83][c] [78][g]
Intangible rewards (achievement, encouragement, evolving Pokémon, and comparisons)	[70,81][b,g] [82,84][b,c]	N/A
Messages		
Messages sent at random times, maximum frequency <2 times per day	[74][b,c]	N/A
Frequency and timing of messages	[78][e,g]	N/A
Receiving reminders	[79][g,h]	N/A
Reinforcement		
Leaderboard	[76][h,c] [87][h,d] [85][e,c]	N/A
Navigation		
PA instructions		
Instructions on PA (eg, video)	[67][b,d]	[71][g] [86][c]
Textual and unclear instructions on physical missions	[71][e,g] [83,85,86][c,e]	N/A
In-app instructions		
Clear and intuitive in-app instructions	[84][b,c]	[80][c]
Written manual and unclear instructions	[78][e,g] [76][c,e]	N/A
Layout		
Straightforward and simple layout	[84][b,c]	[67][d] [83][c]
Scrolling to find information	[68][e,g]	N/A

Details	Association (ie, tested)	Suggestion (ie, hypothetical)
Lack of logical flow between different modules	[68][e,g]	N/A
Navigation		
Finding and recognizing items on a map	[70][e,g]	N/A
Difficulty reading the map	[76][c,e]	N/A
Controls reacting slowly	[78][e,g] [85][c,e]	N/A
Social		
Multiplayer		
Socializing and multiplayer capabilities (eg, friends)	[80][b,c] [70,78,81][b,g]	[76,80,82][c] [71][d]
Social messages		
Sharing results and postings	[78][b,g] [83][b,c]	N/A
Social networking	[84][c,e]	N/A
Replying messages or chat	[74][b,c]	[82][c]
Competition		
Competing against classmates	[71][g,h] [82][c,h] [87][d,h]	N/A
Competing with friends	[87][d,h] [78][b,g]	N/A
Cooperation		
Cooperation and togetherness	[82][b,c]	[82][c]
Challenges		
Types of challenges		
Searching for items (eg, QR codes, bombs, Pokémon, and tags)	[70,71,81][b,g] [82][b,c]	N/A
PA missions	[84][b,c] [69,71,78][b,g]	N/A
Variety		
Repetitive game with lack of progression or clear end goal	[78][e,g]	N/A
Larger variety of actions (eg, special events, season-themed challenges, and new missions or minigames)	N/A	[78][g] [76,82][c] [77][d]
Difficulty level		
Difficulty level of the challenges that suit a player's skill level	[81][b,g]	[80][c]
Time limit		
Time given (to diffuse bombs)	[70][g,h]	N/A
Monitoring		
Manual input		
Manual input of multiple activities or behaviors	[68,69][e,g]	N/A
Remembering multiple behaviors	[79][e,g]	N/A
Run logs and PA diary	[84][b,c] [67][b,d]	N/A
Device		
Monitoring via phone (eg, SMS text messaging and smartphone)	[73][b,d] [71][b,g]	N/A
Monitoring via external sensor (eg, pedometer, wearables, and heart rate monitor)	[73][b,d] [71][b,g]	N/A
Monitoring with real-time information	[83][b,c] [77][d,h]	N/A
Self-monitoring		

Details	Association (ie, tested)	Suggestion (ie, hypothetical)
Self-reflection	[72][c,h]	N/A
Ability to track multiple behaviors	[79][g,h]	N/A
Goals		
Self-set goals		
Personal, self-selected goals	[67][b,d] [78,79][b,g] [76][c,h]	N/A
Type of goals		
Choosing between different types of goals	[84][b,c]	N/A
Goals helped to perform PA	[83][b,c] [67][b,d]	N/A
Goals		
Goal reminders	[72,76][b,c]	N/A
Customization		
Types of customization		
Ability to customize user account, own music, choosing the type of PA, and personal goals	[78,79][b,g] [83,84][b,c]	[83][c]
Avatars		
Avatars (and additional sports equipment)	[69][b,g]	[69][g]
Feedback		
Types of feedback		
Feedback message on goals, weekly PA, real-time information, and individualized feedback	[67][b,d] [83][b,c] [79][b,g] [77][d,h]	N/A
Representation		
Feedback presented in graphs or via SMS text messages	[67,73][b,d]	N/A
PA input		
PA input		
Variety of arm movements and physical exercise	[80][b,c] [71][b,g]	N/A
Running, PA, and FMS[i] components	[70,81][g,h]	N/A
Narrative		
Characters		
Variety of virtual components (eg, zombies), weapons, and ways to increase character abilities	[80,84][b,c]	[78][g]
Setting		
Realistic	[80][b,c]	N/A
More world to explore	N/A	[78][g]
Story		
Complex story line	N/A	[70][g] [80,86][c]
Levels		
Increasing levels		
Gradually increasing levels	[84][b,c]	[69][g]
Credibility		
Messages originating from nutrition professionals	[74][b,c]	N/A

Details	Association (ie, tested)	Suggestion (ie, hypothetical)
Other features		
Push notifications		
Additional push notifications	N/A	[80][c]
GPS		
Reduce GPS latency	N/A	[76][c]

[a]PA: physical activity.

[b]Positive association between features and engagement.

[c]Qualitative studies, assessing self-perceived association.

[d]Quantitative studies, assessing association.

[e]Negative association between features and engagement.

[f]N/A: not applicable.

[g]Mixed methods studies.

[h]Positive and negative associations between features and engagement.

[i]FMS: fundamental movement skills.

Interface Esthetics

Clear and short messages with positive and personal tone and relevant content were associated with engagement in youth [67,74,78]. Engagement was hampered by lack of sophistication in the interface (eg, game graphics, sound effects, and camera use) [75,79]. Youth preferred a clear and understandable layout [67,68,76] and text with clear headlines and a large font size supported with infographics, figures, and tables [68]. The design of progress graphs was identified as an important contributor to engagement in terms of readability [79]. Too much text or large text blocks, as well as heavy and complicated language, were considered to hamper usability [68]. Participants suggested to include more pictures, symbols, and videos to make the app more attractive [68]; improved or realistic visuals [67,76,77]; or general interfaces [76]. Youth liked sounds that were associated with their movements or game activities [75,80]. However, robotic and monotonic voices [75] or sound associated with clicks could negatively impact engagement [81].

Reinforcement

Youth enjoyed winning tangible rewards [70,75,78,82] and suggested offering rewards (ie, in apps that have not introduced rewards earlier) [78,83]. Intangible rewards (eg, feeling of achievement [81] and encouragement [84]) were especially interesting for the youth. Messages that were sent at random times were preferred over messages sent at preset times. The preferred frequency of the messages was less than twice per day [74]. In the study by Martin et al [78], participants expressed concerns about the frequency and timing of messages. The timing should be appropriate, for example, not receiving messages at a moment when one is not able to act on the messages [74,78]. Mixed results with regard to leaderboards were noted. Youth enjoyed leaderboards when beating persons placed above them but they did not enjoy being at the top of the leaderboard. In addition, leaderboards could lead to too much competition that could result in teasing [76]. In another study, adolescents particularly indicated that they missed seeing the other player's scores [85].

Navigation

Instructions related on the one hand to instructions of the in-app functions and on the other hand to instructions related to PA activities. In both cases, well-outlined instructions could support engagement [67,84], and suggestions were made to design clear, simple, and intuitive instructions [80] with the addition of visual components [71,86]. Unclear instructions on requested movements [83,85,86] or unclear [76] textual [71,78] instructions could hamper engagement. The wish for clear instructions had been underlined by a straightforward and simple layout [67,68,83,84]. Working with a map was considered difficult [70,76]. Laine and Suk [70] reported that the quality of the map tile was perceived as poor, which made it difficult to identify one's current location. In addition, Robertson et al [76] reported that children experienced difficulties in understanding the map representation in the game interface. Interestingly, both studies integrated the Google Maps (Google Inc) services, which hampered PA and interrupted engagement. To find and recognize targets, the map resolution and targets should be clear. Furthermore, controls facilitating navigation through the app were negatively associated with engagement in case they were reacting slowly [78,85].

Social

Multiplayer gameplay was considered important for engaging, socializing [80,81], and playing with friends [70] and setting challenges for friends [78]. In different studies, multiplayer gameplay was suggested to include in future versions of the app [76,77,80,82]. However, competing with friends revealed mixed results. Martin et al [78] found a positive association with engagement [78], and as outlined previously, leaderboards could affect high competition (see *Reinforcement* section); however, when competition was compared with cooperation, the latter was preferred [71,82,87]. This was also quantitatively assessed by Nuijten et al [87], who found that intergroup competition (ie, collaboration within class and competing against other classes) increased engagement in comparison with intragroup competition (ie, individual competition). Competing with others was suggested by youth for future app versions [80].

Cooperation and togetherness were considered important during gameplay [82,87], and an extension to collaborate with others was suggested for future app versions [82]. Social interaction in terms of messages to the research team [74], postings [83], and sharing results [78] was considered to contribute to engagement. A chat function [82] or a social platform [80] was suggested for a future app version. Direito et al [84] described that children seldom or never used the social networking features and disliked sharing their runs or status updates.

Challenges

Youth suggested increasing the variety of actions or (mini) games available, indicating that they liked a wide variety of challenges to choose from when playing a game [77,78]. In addition, the new content was considered important, for example, season-themed challenges or special events [76,78,82]. Youth liked activities organized around a hunt, including activities of finding and scanning objects [70,71,81,82] or missions, for example, PA-related workout missions [69,71,78,84]. Youth liked games that made a connection between energy expenditure in real life and the energy level in a game [78]. Repetitive challenges with a lack of progression or clear end goals negatively impacted engagement [78]. The difficulty level of the challenges needed to fit the users' level [80,81], as challenges that were not adjusted to a younger age appeared to negatively impact young kids' enjoyment [81]. This finding was supported by the study done by Arteaga et al [80], in which children suggested including multiple challenges that fit the different levels of individual players. The study by Laine and Suk [70] reported mixed results regarding a time limit paired with challenges.

Monitoring

Manual input hampered engagement [68,69,79], although youth liked to make use of logbooks [84] and diaries [67]. Using devices [71,73] to track real-time information was appreciated and perceived as useful [77,83]. Youth endorsed the self-monitoring function but had complaints relating to discomfort while wearing the monitoring device [72]. The accelerometer that was used in this study was not part of the intervention but was used to measure the PA outcome. The perceived discomfort might have influenced the use of the intervention. Including the function to track multiple health behaviors, on the one hand, engaged youth but, on the other hand, challenged them by the need to remember multiple behaviors [79].

Goals

Self-set goals were considered important [67,76,78,79]. However, choosing a goal was considered cognitively challenging [76]. Specific goals on PA or different types of goals were considered to contribute to engagement [67,83,84]. The feeling of achieving a goal contributed to enjoyment [76]. Reminders to set goals were considered useful [72].

Customization

Youth liked to customize their accounts (eg, removing features such as friends, goal-setting, and training plans) [79,83]; choose the type of PA they wanted to perform [83]; sort their personal goals [78]; and listen to their own music [84]. Youth in general suggested including the possibility of customizing one's account [83].

Youth liked to engage with avatars [69], and children (10-12 years old) suggested including dragons (35%), followed by a dinosaur (25%), and suggested to upgrade their avatar with personal equipment [69].

Feedback

Participants considered regular feedback on achievements and goals [67], which is individualized and in real time to contribute to engagement [79,83]. However, feedback based on dynamic tailoring did not reveal to contribute to overall engagement but contributed to narrative sensation (ie, feeling of presence) [77]. Feedback that was shared via SMS text messaging and presented in graphs was highly valued [67,73].

PA Input

Youth enjoyed upper-body movements [80] compared with full-body movement such as running. However, requiring running to proceed in the game achieved mixed results (ie, facilitating and hampering) on engagement [70,71,81].

Narrative

Youth liked the addition of a narrative, including a variety of characters [78,80,84] and settings [78] that are realistic [80], and the addition of a (complex) story, which was suggested by various studies [70,80,86].

Levels

Gradually increasing levels [84] were considered important, and higher upgrade levels, referring to unlocking new levels as the player proceeds, were suggested to be added [69].

Credibility

Youth found it important to receive information from a credible source as nutrition professionals [74].

Other Features

Youth suggested including additional push notifications [80] and reducing the GPS latency [76].

Included BCTs

In total, 19% (4/21) of the studies did not include any reference to BCTs [68,71,82,86]; 19% (4/21) of studies did only include a reference to BCTs in the results [70,73,74,80]; and 33% (7/21) of studies did include a reference to BCTs in the methods and results; however, they did not particularly refer to it as BCT [67,69,75,77,78,84,87]. Furthermore, 29% (6/21) of studies included a particular reference to BCTs and included the outline of the BCT in the methods and results [72,76,79,81,83,85]. In general, 57% (12/21) of studies included theory in the development of the app (4/21, 19% excluded as apps developed for commercial purposes). In addition, 81% (17/21) of studies included several BCTs (ranging from 2-10), which were partly related to the design features outlined in Table 1. The particular BCTs were not adopted in Table 1, as they have been translated into design features. The BCTs that were most often mentioned were feedback (11/21, 52%), social comparison (11/21, 52%), goal-setting (10/21, 48%), rewards (10/21, 48%), monitoring (9/21, 43%), and encouragement (6/21, 29%).

Subanalysis SEP

In total, 29% (6/21) of the studies (all from countries that form the Global North) included a substantial target population of low SEP [72,75-77,86,87]; however, 33% (2/6) of studies eventually included a large percentage of high-SEP children (34% to 75%) [72,76]. In only 33% (2/6) of studies [75,76], the SEP was indicated as a direct study aim, and in 33% (2/6) of studies [77,86], a mix in SEP was desired. Low SEP was defined differently (ie, neighborhood site, family income, family affluence, parents’ education level, and youth education level), which made it difficult to compare studies, and engaging design features were not presented for different SEP categories. Therefore, subanalysis of the different design features could not be conducted.

Discussion

Principal Findings

This systematic review presents the results of 21 studies that assessed the associations between features and engagement with mHealth PA interventions. According to our knowledge, this is the first mHealth intervention review that focused on PA and youth, with an additional focus on low-SEP background.

The results showed that various design features, such as a clear interface; rewards; multiplayer game mode; facilitation of social interaction; variety of challenges with personalized difficulty level; self-monitoring options; and a variety of customization options among others, including self-set goals, personalized feedback, progress, and a narrative, were positively associated with engagement. In contrast, various features, such as sounds, leaderboards, competition, instructions, timing of messages or notifications, maps, or self-monitoring facilitated by manual input, that have been negatively associated with engagement need to be carefully considered while designing mHealth PA interventions. In addition, technical functionality can be considered as a prerequisite for engagement.

When comparing the results with those of research in adults, several features are shared, such as a simple and structured interface, tailored and positive feedback on PA levels [43,44,89,90], progress [90], well-designed reminders (eg, timing and frequency of notifications) [26,43,44,89], rewards [43,44,49,91-93], self-monitoring [26,43,44], goal-setting, [43,44], clear navigation [43,44], accurate tracking function [43], personalization (eg, goals) [94], and the offer of a variety of features [43,95]. Technical difficulties were also negatively influencing engagement in adults [43,89].

Several features that have been in particular noted for adults, yet not in this review on youth, are credential sources, adequate privacy settings [43], and content preventing the occurrence of surprises [26,43]. Sharing accomplishments via social media was considered less engaging in adults [90] than in youth. Although the studies among youth provided mixed results on competition, literature on adults suggests including competition [44,90], leaderboards [91], and hierarchical status (eg, progressive status report) [91]. In gamification research, leaderboards are considered the most common feature implemented (alongside points, goals, and progress) [96,97], and this seems to be considered with care when addressing youth for PA promotion. In studies including adults, no direct links with narratives, avatars, a hunt, difficulties with manual input, sounds, or virtual location maps were found. Research on using these design features for interventions in youth is limited, according to the results of this review.

This review focused on design features that have been tested by youth and included hypothetical or suggested features of the same studies but did not extend the search to studies that only provided ideas or suggestions that have not been tested (ie, cocreation study on app development without user-testing). Earlier studies have indicated that what users may request and what they actually engage with in practice do not have to match [44,98]. In the studies in this review, several features were tested and included as further suggestions, which may underline the importance of the match between what youth would like to have included and what they actually engage with. The following results are in line with existing research on app development (ie, no user-testing) among youth: (1) clear user interface esthetics (ie, youthful visuals) [99,100]; (2) rewards, referring to a fair reward system [100,101], rewards that progressively increase [102], and social rewards (ie, storybooks with interactive questions) [103]; and (3) multiplayer mode, facilitating competition [55,100] and comparison (eg, leaderboards) [55], as well as cooperation [103,104]. Several features that have been suggested and tested in the studies included in this review were not found in existing research, and these include (1) chat functions (although with mixed results), (2) increasing difficulty level, (3) customizing user content, and (4) adding a variety of avatars and characters. In addition, several features have been suggested but have not been tested, and these include (1) meaningful information [99,100], including PA-related tips and plans [101], and (2) variety and content updates [102].

Features that have been solely suggested in this review (ie, not tested) relate to (1) adding video and moving figures that navigate the PA and app instructions, (2) adding a storyline, and (3) including additional push notifications. Other research suggests that a clear navigation that is self-explanatory is important [99], yet that notifications evoke mixed results on engagement [101]. Features that have been suggested in other research, but not in this review, include (1) in-app events [104] and (2) GPS and map editors [105,106]. This review contributes to the body of research and highlights that although maps can be suggested by youth, they can be very challenging when applied in practice, risking hampering engagement. Additional features that especially could hamper engagement include a small number of available minigames [106], no instant feedback [102] or personalized feedback by email [101], lengthy texts, and difficult navigation [99]. Future studies are needed to conduct experimental testing on features that yield inconclusive evidence so far among others competition, leaderboards, and notifications, as it may be possible that the results may differ for different subgroups of youth. We recommend that designers and health researchers consider designing a clear interface and an appropriate and fair reward system, enabling social interaction, and providing a variety of content, which is preferably customizable.

Earlier research indicates how a variety of design features are designed and implemented in mobile apps [65,107]. This is also indicated in other research outlining the challenges of how BCTs are operationalized [65,108-110]. In other words, the content or the active ingredient of the intervention may still differ by how it is delivered, in which context, or in which combinations (ie, design features, including gamification elements and BCTs) it is applied [111]. An ontology, as proposed by West and Michie [112], and tools such as SciModeller that integrate multiple pieces of empirical data [113] can eventually contribute to this knowledge and help researchers and mHealth developers apply and build on this knowledge. This review focused on whether the app design has been informed by or has been based on relating BCTs or behavioral theories and concluded that BCTs are often not outlined in detail, and there is no particular translation of BCTs to app content. This review only included 33% (7/21) of studies that made a clear translation from a particular BCT to a design feature. A scoping review not only summarized the challenges in mHealth design in translating BCTs to an mHealth feature but also raised challenges with regard to integrating ideas from different perspectives (ie, BCTs, user needs, and stakeholder views), which can result in conflicting ideas [110]. Existing research on mHealth, focusing on app features, BCTs, or both, points to the challenge of mapping which features or BCTs work in isolation or in combination with others [47,114]. Furthermore, a wide applicability exists with regard to creating design features that are often designed from scratch. In light of the Open Digital Health initiative [115], we recommend making mHealth apps and data accessible to be able to reuse or continue working on existing design features that have been proven to be effective for engagement. By this, we can start mapping working design features for different user groups in different contexts [115,116].

Different systematic reviews focusing on the effectiveness of mHealth interventions have also focused on BCTs. However, a large number of the reviews map the number of BCTs that have been included in an intervention [42,117] or the BCTs that have been included [42,117]. However, there is dearth of research that identifies the individual or interaction effects of design features or BCTs in mHealth interventions among youth. Several studies have identified that modeling is an effective BCT for children. With regard to adolescents, providing consequences for behavior, providing information on others' approval, facilitating intention formation, self-monitoring, using behavioral contract [118], and providing individualization support [25] were positive predictors of PA effect size. Providing instructions was negatively predictive of study effect size [118]. In future reviews, it would be interesting to map the similarities and differences between design features that contribute to engagement in the mHealth intervention (ie, microengagement) and BCTs that contribute to the behavioral change effectiveness (ie, macroengagement) to identify effective design features [40]. However, limited BCTs are researched in effectiveness studies, especially among youth [119]. Therefore, more individual studies are needed before this comparison can be made. Future studies may, therefore, test the effectiveness of individual BCTs and combinations of BCTs on behavioral changes in factorial designs [120] in youth. Future studies should also consider testing both microengagement and macroengagement in the same mHealth intervention [40] to better inform how to increase engagement in the mHealth intervention (ie, microengagement) and translate this to behavioral change, which can be engaged in the long term in the absence of any intervention (ie, macroengagement).

This review underlines that research addressing youth from low-SEP families is very limited with regard to engagement in mHealth apps. Although it is often argued that apps can be a suitable platform to reach diverse groups of youth, it is striking that only 7 studies have been identified that aim to involve youth from different SEP backgrounds. Unfortunately, it was not possible to conduct a subanalysis because the features were not directly linked to the SEP position. Studies that succeeded in addressing a large percentage of children with low SEP recruited participants via schools [49,77,86]. From this review, it cannot be directly derived which recruitment strategy leads to lower percentages of children with low SEP, as especially, the measures of low SEP differ greatly. We advocate for more research that includes youth from low-SEP families early in the development phase and user-testing and suggest recruiting youth from low-SEP background via schools. Further research in adults suggested that personal contact between study staff and participants is essential and that community sites (comparable with school settings) created a sense of community and support [121]. Thereby, youth who have the potential to book the greatest health gains (ie, often youth with low-SEP or low-PA levels) are addressed appropriately, which may contribute to reducing health inequalities and the digital divide [44]. A scoping review highlighted the need to further investigate user engagement studies in low-SEP groups and called for future in-depth formative studies [95]. Existing research indicates that multimedia, personalization, variation, and gamification [95], such as competition [116], can contribute to engagement in mHealth apps among young adults with low-SEP backgrounds. We advocate to start testing these features in youth with low SEP.

In this review, we identified a heterogeneity of engagement measures, which has also been identified in earlier reviews [55,60,114,122,123]. In addition, several measures only reflect one construct of engagement and do not measure the multidimensional concept of engagement [55]. Furthermore, research studies included in this review did not distinguish between different features that can be considered in different stages of engagement. Research suggests that engagement is a dynamic process rather than a state, although it is often measured as a state (ie, 1 postintervention questionnaire on engagement instead of cyclic measurement) [124]. On the basis of earlier research, it may however be stated that features such as a clear interface are especially important for the initial stage of engagement (ie, attention grabbing) [124]. Features that sustain engagement may relate to social interaction, a variety of challenges with personalized difficulty level, self-monitoring and customization options, and narratives [77]. Disengaging may relate to certain types of sounds, leaderboards, instructions, messages or notifications, competition, and self-monitoring facilitated by manual input. Features that reengaged or nonengaged users have not been identified in this review. However, for example, Janko et al [69] discussed unlockable avatars and upgrading levels as possible features that may contribute to reengagement. Research on young adults indicated that users did not engage (ie, nonengagement) owing to low uptake of the intervention among peers [49]. The reasons for disengagement mentioned in this review relate to factors outside the mHealth intervention (eg, holidays, competing after-school activities, weather, school policies, and unstructured or leisure settings) [76,77,81,83,87], yet nonadherence has not been linked to design features. Future research is needed to identify particular features in different stages of engagement to grasp engagement as a multidimensional and dynamic process [124]. Thereby, designers can improve design features according to the stage of engagement (eg, include particular features to reengage users and prevent them from sustained disengagement). Furthermore, engagement may also be related to the setting in which the mHealth tool is implemented. The activity that competes with the mHealth tool is central. Earlier research indicated that pupils tend to choose to engage with an mHealth tool in order not to participate in class in a school setting [125]. In comparison, mHealth tools are often challenged in leisure time, and the competition with other apps or leisure activities is central [126,127]. Future studies should focus on the implementation of mHealth tool and investigate differences in engagement in either voluntary versus more obligatory settings.

In terms of comparability, the Persuasive Systems Design (PSD) model could have helped map the different design features. PSD is often used in research to outline persuasive design elements. The model maps different persuasive design elements in primary task support, dialogue support, system credibility support, and social support. When comparing the PSD model with the coding scheme applied in this systematic review, a number of similarities can be identified. In terms of primary task support, navigation (PSD model: tunneling), personalization, and self-monitoring are identified. In terms of dialogue support, rewards, messages (PSD model: reminders), and interface aesthetics (PSD model: liking) can be identified. Liking is very broadly defined as "visually attractive" in the PSD model and has been criticized in earlier research [128]. It finds more detail in the coding of our systematic review (ie, interplay of interface elements such as sounds, visuals, and language). In terms of system credibility support, only credibility (PSD model: expertise) can be identified. In terms of social support, reinforcement (PSD model: social comparison), social messages (PSD model: social facilitation), cooperation, and competition can be identified. Design features that are difficult to code in terms of the PSD model are challenges, levels, feedback, avatars, and narratives and are predominately game elements and refer to achievement- and immersion-based features [96]. These have also been identified in earlier research, distinguishing their research from the PSD model [129,130]. On the one hand, this emphasizes the challenge to identify and apply a complete list of design elements. As a proposed solution, Geuens et al [128] created a website that derived from the PSD model and mapped a working list of adaptable features. However, it should be noted that this list is currently not complete in terms of gamification elements. On the other hand, the large number of similarities between the PSD model and the coding frame of this systematic review suggest that various research studies may have found (to a certain degree) a consensus on design features. A revision of the PSD model, adapted by, for example, gamification elements, could help create an updated list of design features that are widely applicable to designers and researchers.

Strengths and Limitations

As identified in an earlier review [55], this review included a larger number of studies reporting the foremost positive associations, which may bias the overview in features that are either unrelated or negatively associated and may not be reported (ie, publication bias). In addition, the strength of associations has not been included in the individual studies, owing to a lack of high-quality experimental studies. In addition, heterogeneity among studies with regard to engagement measures was identified, making direct comparison challenging; this has also been identified in earlier studies [114]. This systematic review maps a large number of features that are associated with engagement individually. The interaction of different features and their effect on engagement is still not researched and needs to be considered in future engagement studies, for example, experimental factorial designs [41]. Individual studies included small sample sizes, and only a small number of studies focused on low SEP. Using different SEP measures and mixing them with high-SEP measures made it challenging to compare studies and prevented us from conducting a subanalysis based on the SEP background. Therefore, the findings might not be transferable to low-SEP groups and this underlines the need for future research. In addition, the variability in SEP measures made it difficult to standardize SEP thresholds and compare between countries, as youth from a low-SEP background may not be equivalent in terms of studies and countries, and disparity occurs even in different parts of one country. Studies that included SEP measures were exclusively from countries of the Global North, emphasizing that in general, a minority of studies focused on the Global South (4/21, 19%), and none focused on SEP. As the search was limited to English, this review may not have included all available and relevant research on this topic. Earlier research outlines the differences in the effectiveness of mHealth intervention on behavioral outcomes and engagement among very young children and older children. Research further suggests that preferences for design features may differ between different age groups of youth. This systematic review did not cluster for different age groups because of the limited number of studies that, for example, focus on the very young children (aged <4 years) [25]. More individual studies are needed among very young children to identify whether age differences exist. Earlier research suggests that gender differences in games may exist [131-137]. None of the included studies in this systematic review distinguished between gender for design features in mHealth interventions. In future research, it can therefore be considered to investigate possible differences in future user-testing. The quality of the studies was mixed, indicating in general that study designs with regard to measuring engagement need to evolve. Especially, studies that focus on engagement as a secondary outcome need to operationalize engagement clearly and report the methods transparently. In this review, no validated design feature taxonomy has been used for coding, although it was based on existing overviews of design features [114]. In this review, all study designs were included, and this provided a comprehensive overview of relating research on engagement and triangulated evidence. This is the first review identifying mHealth design features among youth with the aim of promoting PA. This review was based on the PRISMA guidelines and succeeded in providing a coherent overview of all nuances relating to a good quality assessment and high interrater reliability.

Conclusions

The results indicate that a clear interface; an appropriate and fair reward system; social interaction; and a variety of app content (ie, missions, content, and characters) that are preferably customizable can contribute to engagement in mHealth PA interventions for youth. Design features, such as sounds, competition, instructions, notifications, virtual maps, or self-monitoring facilitated by manual input, that were negatively associated with engagement need to be carefully considered when designing mHealth PA interventions. In addition, technical functionality can be considered a prerequisite for engagement. Research addressing youth from low-SEP families is very limited with regard to engagement in mHealth apps, and design features often lack a sufficient degree of operationalization based on behavioral change theories or techniques. Future studies are needed to further test design features in youth, particularly youth from low-SEP families, and to evolve engagement measures.

Acknowledgments

This review was supported by the mVital@2040 project, which is a part of the Vitality Academy and connected to the EWUU Alliance (Eindhoven University of Technology, Wageningen University & Research, Utrecht University, and University Medical Centre Utrecht). The project and AS were funded by the Wageningen University & Research. MS initiated and acquired the funding for the mVital@2040 project.

Authors' Contributions

AS was responsible for leading all stages of the review. AS established the research protocol, which was reviewed by MS and LHHW. DO conducted the search and data extraction under the supervision of LHHW and drafted the first version of the methods and results. KB served as the second reviewer for search and data extraction. The search was updated by AS. AS was responsible for drafting the manuscript, which was written together with MS, LHHW, and EdV. The manuscript was eventually reviewed and adapted by MS, LHHW, EdV, DO, and KB.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 40 KB - mhealth_v11i1e40898_app1.docx]

Multimedia Appendix 2
Full search strategy.
[DOCX File , 29 KB - mhealth_v11i1e40898_app2.docx]

Multimedia Appendix 3
Quality check Mixed Method Assessment Tool.
[XLSX File (Microsoft Excel File), 58 KB - mhealth_v11i1e40898_app3.xlsx]

Multimedia Appendix 4
Data extraction of the studies included in the systematic review.
[XLSX File (Microsoft Excel File), 53 KB - mhealth_v11i1e40898_app4.xlsx]

References

1. Chaput J, Willumsen J, Bull F, Chou R, Ekelund U, Firth J, et al. 2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5-17 years: summary of the evidence. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):141 [FREE Full text] [doi: 10.1186/s12966-020-01037-z] [Medline: 33239009]
2. Hills AP, King NA, Armstrong TP. The contribution of physical activity and sedentary behaviours to the growth and development of children and adolescents: implications for overweight and obesity. Sports Med 2007;37(6):533-545. [doi: 10.2165/00007256-200737060-00006] [Medline: 17503878]
3. Biddle SJ, Asare M. Physical activity and mental health in children and adolescents: a review of reviews. Br J Sports Med 2011 Sep;45(11):886-895. [doi: 10.1136/bjsports-2011-090185] [Medline: 21807669]

4. Health behaviour in school-aged children (HBSC) study: international report from the 2013/2014 survey. Unicef Bulgaria. URL: https://www.unicef.org/bulgaria/en/reports/health-behaviour-school-aged-children-hbsc-20132014 [accessed 2022-06-30]
5. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet Child Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
6. Spotlight on adolescent health and well-being. Findings from the 2017/2018 Health Behaviour in School-aged Children (HBSC) survey in Europe and Canada. World Health Organization. URL: https://apps.who.int/iris/handle/10665/332104 [accessed 2022-06-30]
7. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: a systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
8. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195 [FREE Full text] [doi: 10.1159/000222244] [Medline: 20054224]
9. Molina-García J, Queralt A, Adams MA, Conway TL, Sallis JF. Neighborhood built environment and socio-economic status in relation to multiple health outcomes in adolescents. Prev Med 2017 Dec;105:88-94. [doi: 10.1016/j.ypmed.2017.08.026] [Medline: 28863871]
10. Stalsberg R, Pedersen AV. Effects of socioeconomic status on the physical activity in adolescents: a systematic review of the evidence. Scand J Med Sci Sports 2010 Jun;20(3):368-383. [doi: 10.1111/j.1600-0838.2009.01047.x] [Medline: 20136763]
11. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Global Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
12. Gibbings J, Blakemore T, Strazdins L. Measuring family socioeconomic position. Aust Soc Pol 2009:121-168.
13. Abraham AA, Chow W, So H, Yip BH, Li AM, Kumta SM, et al. Lifestyle intervention using an internet-based curriculum with cell phone reminders for obese Chinese teens: a randomized controlled study. PLoS One 2015;10(5):e0125673 [FREE Full text] [doi: 10.1371/journal.pone.0125673] [Medline: 25946465]
14. He Z, Wu H, Yu F, Fu J, Sun S, Huang T, et al. Effects of smartphone-based interventions on physical activity in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 01;9(2):e22601 [FREE Full text] [doi: 10.2196/22601] [Medline: 33522980]
15. Shin Y, Kim SK, Lee M. Mobile phone interventions to improve adolescents' physical health: a systematic review and meta-analysis. Public Health Nurs 2019 Nov;36(6):787-799. [doi: 10.1111/phn.12655] [Medline: 31456259]
16. Radovic A, McCarty CA, Katzman K, Richardson LP. Adolescents' perspectives on using technology for health: qualitative study. JMIR Pediatr Parent 2018 Mar 14;1(1):e2. [doi: 10.2196/pediatrics.8677] [Medline: 30740590]
17. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]
18. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]
19. Lau PW, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
20. Okma K, Boeke H. Iene Miene media: monitor 2018. Mediawijzer. URL: https://netwerkmediawijsheid.nl/wp-content/uploads/sites/6/2018/04/Onderzoek-IeneMieneMedia-2018.pdf [accessed 2022-06-30]
21. Parenting children in the age of screens. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2020/07/PI_2020.07.28_kids-and-screens_FINAL.pdf [accessed 2022-06-30]
22. Nikken P. Een onderzoek naar omgang met en gebruik van media(apparaten) van kinderen van 0 t/m 6 jaar en de rol die ouders hierin spelen. Iene Miene Media. 2022 Mar. URL: https://netwerkmediawijsheid.nl/wp-content/uploads/2022/03/Iene-Miene-Media-onderzoek-2022.pdf [accessed 2023-02-24]
23. Iene Miene media: monitor 2020. Netwerk Mediawijsheid. URL: https://netwerkmediawijsheid.nl/wp-content/uploads/sites/6/2020/03/Iene-Miene-Media-onderzoek-2020.pdf [accessed 2022-06-30]
24. Vanwynsberghe H, Joris G, Waeterloos C, Anrijs S, Vanden AM, Ponnet K. Onderzoeksrapport Apestaartjaren : de digitale leefwereld van jongeren. Gent. 2022. URL: https://biblio.ugent.be/publication/8667985 [accessed 2022-06-30]
25. Baumann H, Fiedler J, Wunsch K, Woll A, Wollesen B. mHealth interventions to reduce physical inactivity and sedentary behavior in children and adolescents: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 May 11;10(5):e35920 [FREE Full text] [doi: 10.2196/35920] [Medline: 35544294]

26. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
27. Western MJ, Armstrong ME, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
28. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
29. Männikkö N, Ruotsalainen H, Miettunen J, Marttila-Tornio K, Kääriäinen M. Parental socioeconomic status, adolescents' screen time and sports participation through externalizing and internalizing characteristics. Heliyon 2020 Feb;6(2):e03415 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03415] [Medline: 32099928]
30. Zhan M, Cai Y, Guo D. Discovering the influence of socioeconomic factors on online game behaviors. In: Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure. 2015 Presented at: XSEDE '15: Extreme Science Engineering Discovery Environment 2015 Conference; Jul 26 - 30, 2015; St. Louis Missouri. [doi: 10.1145/2792745.2792752]
31. Simons M, de Vet E, Brug J, Seidell J, Chinapaw MJ. Active and non-active video gaming among Dutch adolescents: who plays and how much? J Sci Med Sport 2014 Nov;17(6):597-601. [doi: 10.1016/j.jsams.2013.10.250] [Medline: 24275124]
32. Coupe N, Cotterill S, Peters S. Tailoring lifestyle interventions to low socio-economic populations: a qualitative study. BMC Public Health 2018 Aug 03;18(1):967 [FREE Full text] [doi: 10.1186/s12889-018-5877-8] [Medline: 30075716]
33. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
34. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
35. Kracht CL, Hutchesson M, Ahmed M, Müller AM, Ashton LM, Brown HM, et al. E-&mHealth interventions targeting nutrition, physical activity, sedentary behavior, and/or obesity among children: A scoping review of systematic reviews and meta-analyses. Obes Rev 2021 Dec 02;22(12):e13331 [FREE Full text] [doi: 10.1111/obr.13331] [Medline: 34476890]
36. Crutzen R, van 't Riet J, Short CE. Enjoyment: a conceptual exploration and overview of experimental evidence in the context of games for health. Games Health J 2016 Feb;5(1):15-20. [doi: 10.1089/g4h.2015.0059] [Medline: 26699455]
37. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
38. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277 [FREE Full text] [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
39. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
40. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
41. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
42. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
43. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
44. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
45. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
46. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change

interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
47. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
48. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. J Med Internet Res 2016 Jun 07;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
49. Nuijten R, Van Gorp P, Hietbrink J, Le Blanc P, Kemperman A, van den Berg P, et al. Pilot evaluation of the impact of lottery-based incentives on engagement levels of male low SES vocational students with an mHealth app. Front Digit Health 2021 Jan 7;3:748588 [FREE Full text] [doi: 10.3389/fdgth.2021.748588] [Medline: 35072150]
50. Bry LJ, Chou T, Miguel E, Comer JS. Consumer smartphone apps marketed for child and adolescent anxiety: a systematic review and content analysis. Behav Ther 2018 Mar;49(2):249-261 [FREE Full text] [doi: 10.1016/j.beth.2017.07.008] [Medline: 29530263]
51. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
52. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
53. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
54. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
55. Schwarz AF, Huertas-Delgado FJ, Cardon G, DeSmet A. Design features associated with user engagement in digital games for healthy lifestyle promotion in youth: a systematic review of qualitative and quantitative studies. Games Health J 2020 Jun;9(3):150-163. [doi: 10.1089/g4h.2019.0058] [Medline: 31923363]
56. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
57. Weggemans RM, Backx FJ, Borghouts L, Chinapaw M, Hopman MT, Koster A, Committee Dutch Physical Activity Guidelines 2017. The 2017 Dutch physical activity guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58 [FREE Full text] [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
58. Olla P, Shimskey C. mHealth taxonomy: a literature survey of mobile health applications. Health Technol 2015 Jan 30;4(4):299-308. [doi: 10.1007/s12553-014-0093-8]
59. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
60. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in ehealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
61. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28 - 30, 2011; Tampere ,Finland. [doi: 10.1145/2181037.2181040]
62. Souza-Júnior M, Queiroz L, Correia-Neto J, Vilar G. Evaluating the use of gamification in m-health lifestyle-related applications. In: New Advances in Information Systems and Technologies. Cham: Springer; 2016.
63. Dallinga J, Janssen M, van der Werf J, Walravens R, Vos S, Deutekom M. Analysis of the features important for the effectiveness of physical activity-related apps for recreational sports: expert panel approach. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e143 [FREE Full text] [doi: 10.2196/mhealth.9459] [Medline: 29914863]
64. Cotton V, Patel MS. Gamification use and design in popular health and fitness mobile applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
65. Sporrel K, Nibbeling N, Wang S, Ettema D, Simons M. Unraveling mobile health exercise interventions for adults: scoping review on the implementations and designs of persuasive strategies. JMIR Mhealth Uhealth 2021 Jan 18;9(1):e16282 [FREE Full text] [doi: 10.2196/16282] [Medline: 33459598]
66. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inform 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]

67. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of gamified mobile service aimed at physical activation in young men: population-based randomized controlled study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
68. Müssener U, Thomas K, Linderoth C, Löf M, Åsberg K, Henriksson P, et al. Development of an intervention targeting multiple health behaviors among high school students: participatory design study using heuristic evaluation and usability testing. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e17999 [FREE Full text] [doi: 10.2196/17999] [Medline: 33118942]
69. Janko V, Cvetković B, Gradišek A, Luštrek M, Štrumbelj B, Kajtna T. e-Gibalec: mobile application to monitor and encourage physical activity in schoolchildren. J Ambient Intell Smart Environ 2017 Aug 11;9(5):595-609. [doi: 10.3233/ais-170453]
70. Laine TH, Suk HJ. Designing mobile augmented reality exergames. Games Culture 2016 May 22;11(5):548-580. [doi: 10.1177/1555412015572006]
71. Lindberg R, Seo J, Laine TH. Enhancing physical education with exergames and wearable technology. IEEE Trans Learning Technol 2016 Oct 1;9(4):328-341. [doi: 10.1109/tlt.2016.2556671]
72. Cushing CC, Bejarano CM, Ortega A, Sayre N, Fedele DA, Smyth JM. Adaptive mHealth intervention for adolescent physical activity promotion. J Pediatr Psychol 2021 Jun 03;46(5):536-546. [doi: 10.1093/jpepsy/jsaa125] [Medline: 33484137]
73. Fassnacht DB, Ali K, Silva C, Gonçalves S, Machado PP. Use of text messaging services to promote health behaviors in children. J Nutr Educ Behav 2015;47(1):75-80. [doi: 10.1016/j.jneb.2014.08.006] [Medline: 25282200]
74. Hingle M, Nichter M, Medeiros M, Grace S. Texting for health: the use of participatory methods to develop healthy lifestyle messages for teens. J Nutr Educ Behav 2013;45(1):12-19. [doi: 10.1016/j.jneb.2012.05.001] [Medline: 23103255]
75. Blackman KC, Zoellner J, Kadir A, Dockery B, Johnson SB, Almeida FA, et al. Examining the feasibility of smartphone game applications for physical activity promotion in middle school students. Games Health J 2015 Oct;4(5):409-419. [doi: 10.1089/g4h.2014.0120] [Medline: 26287931]
76. Robertson J, Jepson R, Macvean A, Gray S. Understanding the importance of context: a qualitative study of a location-based exergame to enhance school childrens physical activity. PLoS One 2016 Aug 22;11(8):e0160927 [FREE Full text] [doi: 10.1371/journal.pone.0160927] [Medline: 27548504]
77. Schwarz A, Cardon G, Chastin S, Stragier J, De Marez L, Consortium SmartLife, et al. Does dynamic tailoring of a narrative-driven exergame result in higher user engagement among adolescents? Results from a cluster-randomized controlled trial. Int J Environ Res Public Health 2021 Jul 12;18(14):7444 [FREE Full text] [doi: 10.3390/ijerph18147444] [Medline: 34299895]
78. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
79. Thornton L, Gardner LA, Osman B, Green O, Champion KE, Bryant Z, Health4Life Team. A multiple health behavior change, self-monitoring mobile app for adolescents: development and usability study of the Health4Life app. JMIR Form Res 2021 Apr 12;5(4):e25513 [FREE Full text] [doi: 10.2196/25513] [Medline: 33843590]
80. Arteaga SM, González VM, Kurniawan S, Benavides RA. Mobile games and design requirements to increase teenagers' physical activity. Pervasive Mobile Comput 2012 Dec;8(6):900-908. [doi: 10.1016/j.pmcj.2012.08.002]
81. McKenzie S, Bangay S, Barnett LM, Ridgers ND, Salmon J. Design elements and feasibility of an organized multiplayer mobile active videogame for primary school-aged children. Games Health J 2014 Dec;3(6):379-387. [doi: 10.1089/g4h.2013.0097] [Medline: 26192645]
82. Lindqvist A, Castelli D, Hallberg J, Rutberg S. The praise and price of Pokémon GO: a qualitative study of children's and parents' experiences. JMIR Serious Games 2018 Jan 03;6(1):e1 [FREE Full text] [doi: 10.2196/games.8979] [Medline: 29298750]
83. Seah ML, Koh KT. The efficacy of using mobile applications in changing adolescent girls' physical activity behaviour during weekends. Eur Physical Educ Rev 2020 Jun 17;27(1):113-131. [doi: 10.1177/1356336x20930741]
84. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
85. Farič N, Smith L, Hon A, Potts HW, Newby K, Steptoe A, et al. A virtual reality exergame to engage adolescents in physical activity: mixed methods study describing the formative intervention development process. J Med Internet Res 2021 Feb 04;23(2):e18161 [FREE Full text] [doi: 10.2196/18161] [Medline: 33538697]
86. McCloskey ML, Thompson DA, Chamberlin B, Clark L, Johnson SL, Bellows LL. Mobile device use among rural, low-income families and the feasibility of an app to encourage preschoolers' physical activity: qualitative study. JMIR Pediatr Parent 2018 Dec 06;1(2):e10858 [FREE Full text] [doi: 10.2196/10858] [Medline: 31518295]
87. Nuijten RC, Van Gorp P, Borghouts T, Le Blanc P, Van den Berg P, Kemperman A, et al. Preadolescent students' engagement with an mHealth intervention fostering social comparison for health behavior change: crossover experimental study. J Med Internet Res 2021 Jul 29;23(7):e21202 [FREE Full text] [doi: 10.2196/21202] [Medline: 34326041]

88. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme. Lancaster University. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2022-06-20]
89. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
90. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 01;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
91. Maher CA, Olds T, Vandelanotte C, Plotnikoff R, Edney SM, Ryan JC, et al. Gamification in a physical activity app: what gamification features are being used, by whom, and does it make a difference? Games Health J 2022 Jun 01;11(3):193-199. [doi: 10.1089/g4h.2021.0207] [Medline: 35501981]
92. Nuijten R, Van Gorp P, Khanshan A, Le Blanc P, Kemperman A, van den Berg P, et al. Health promotion through monetary incentives: evaluating the impact of different reinforcement schedules on engagement levels with a mHealth app. Electronics 2021 Nov 26;10(23):2935. [doi: 10.3390/electronics10232935]
93. Nuijten R, Van GP, Kaymak U, Simons M, Kemperman A, Van DB. Evaluation of the impact of extrinsic rewards on user engagement in a health promotion context. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019 Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 23-27 , 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8856296]
94. Nuijten R, Van Gorp P, Khanshan A, Le Blanc P, van den Berg P, Kemperman A, et al. Evaluating the impact of adaptive personalized goal setting on engagement levels of government staff with a gamified mHealth tool: results from a 2-month randomized controlled trial. JMIR Mhealth Uhealth 2022 Mar 31;10(3):e28801 [FREE Full text] [doi: 10.2196/28801] [Medline: 35357323]
95. Stowell E, Lyson M, Saksono H, Wurth R, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal, QC, Canada. [doi: 10.1145/3173574.3173589]
96. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inform Manage 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
97. Koivisto J, Hamari J. Gamification of physical activity: a systematic literature review of comparison studies. In: Proceedings of the GamiFIN 2019 conference. 2019 Presented at: GamiFIN 2019 conference; Apr, 2019; Levi, Finland.
98. Croon R, Geuens J, Verbert K, Vanden AV. A systematic review of the effect of gamification on adherence across disciplines. In: HCI in Games: Experience Design and Game Mechanics. Cham: Springer; 2021.
99. Nitsch M, Adamcik T, Kuso S, Zeiler M, Waldherr K. Usability and engagement evaluation of an unguided online program for promoting a healthy lifestyle and reducing the risk for eating disorders and obesity in the school setting. Nutrients 2019 Mar 27;11(4):713 [FREE Full text] [doi: 10.3390/nu11040713] [Medline: 30934754]
100. Pope L, Garnett B, Dibble M. Engaging adolescents to inform the development of a mobile gaming app to incentivize physical activity. JMIR Res Protoc 2017 Aug 29;6(8):e161 [FREE Full text] [doi: 10.2196/resprot.8113] [Medline: 28851676]
101. Frontini R, Sousa P, Dixe MA, Ferreira R, Figueiredo MC. Designing a mobile app to promote healthy behaviors and prevent obesity: analysis of adolescents' preferences. Inform Health Soc Care 2020 Sep 01;45(3):327-341. [doi: 10.1080/17538157.2020.1725766] [Medline: 32237933]
102. Amresh A, Lyles A, Small L, Gary K. FitBit garden: a mobile game designed to increase physical activity in children. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH '17: International Conference on Digital Health; Jul 2 - 5, 2017; London United Kingdom. [doi: 10.1145/3079452.3079457]
103. Saksono H, Castaneda-Sceppa C, Hoffman J, Morris V, Seif EM, Parker A. Storywell: designing for family fitness app motivation by using social rewards and reflection. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu, HI, USA. [doi: 10.1145/3313831.3376686]
104. Militello LK, Hanna N, Nigg CR. Pokémon GO within the context of family health: retrospective study. JMIR Pediatr Parent 2018 Oct 03;1(2):e10679 [FREE Full text] [doi: 10.2196/10679] [Medline: 31518294]
105. Schwarz A, DeSmet A, Cardon G, Chastin S, Costa R, Grilo A, et al. Mobile exergaming in adolescents' everyday life-contextual design of where, when, with whom, and how: the SmartLife case. Int J Environ Res Public Health 2018 Apr 24;15(5):835 [FREE Full text] [doi: 10.3390/ijerph15050835] [Medline: 29695069]
106. Stanley K, Livingston I, Bandurka A, Kapiszka R, Mandryk R. PiNiZoRo: a GPS-based exercise game for families. In: Proceedings of the International Academic Conference on the Future of Game Design and Technology. 2010 Presented at: Futureplay '10: Futureplay '10 @ GDC Canada; May 6 - 7, 2010; Vancouver , British Columbia, Canada. [doi: 10.1145/1920778.1920817]

107. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
108. Dombrowski SU, O'Carroll RE, Williams B. Form of delivery as a key 'active ingredient' in behaviour change interventions. Br J Health Psychol 2016 Nov 06;21(4):733-740. [doi: 10.1111/bjhp.12203] [Medline: 27709824]
109. Fulton E, Kwah K, Wild S, Brown K. Lost in translation: transforming behaviour change techniques into engaging digital content and design for the StopApp. Healthcare (Basel) 2018 Jul 06;6(3):75 [FREE Full text] [doi: 10.3390/healthcare6030075] [Medline: 29986396]
110. Voorheis P, Zhao A, Kuluski K, Pham Q, Scott T, Sztur P, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. JMIR Mhealth Uhealth 2022 Mar 16;10(3):e35799 [FREE Full text] [doi: 10.2196/35799] [Medline: 35293871]
111. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Transl Behav Med 2018 Mar 01;8(2):212-224 [FREE Full text] [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
112. West R, Michie S. A Guide to Development and Evaluation of Digital Behaviour Change Interventions in Healthcare. United Kingdom: Silverback Publishing; 2016.
113. Nuijten R, Van Gorp P. SciModeler: A Metamodel and Graph Database for Consolidating Scientific Knowledge by Linking Empirical Data with Theoretical Constructs. Setúbal, Portugal: SCITEPRESS; 2021.
114. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
115. Kwasnicka D, Keller J, Perski O, Potthoff S, Ten Hoor GA, Ainsworth B, et al. White Paper: open digital health - accelerating transparent and scalable health promotion and treatment. Health Psychol Rev 2022 Dec;16(4):475-491 [FREE Full text] [doi: 10.1080/17437199.2022.2046482] [Medline: 35240931]
116. Nuijten R. Energize: Exploring the Impact of Gamification Strategies on Engagement With Mobile Health Apps. Eindhoven: Eindhoven University of Technology; 2022.
117. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
118. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May 13;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
119. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
120. Schroé H, Van Dyck D, De Paepe A, Poppe L, Loh WW, Verloigne M, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 07;17(1):127 [FREE Full text] [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
121. Nagler RH, Ramanadhan S, Minsky S, Viswanath K. Recruitment and retention for community-based eHealth interventions with populations of low socioeconomic position: strategies and challenges. J Commun 2013 Feb 01;63(1):201-220 [FREE Full text] [doi: 10.1111/jcom.12008] [Medline: 23439871]
122. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
123. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to ehealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
124. Schwarz A. User engagement in serious digital games for healthy lifestyle promotion in youth. Ghent University. 2019. URL: https://www.researchgate.net/publication/335401813_User_engagement_in_serious_digital_games_for_healthy_lifestyle_promotion_in_youth [accessed 2022-06-30]
125. Lieberman D. Designing serious games for learning and health in informal and formal settings. In: Serious Games Mechanisms and Effects. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2009.
126. De Grove F, Bourgonjon J, Van Looy J. Digital games in the classroom? A contextual approach to teachers' adoption intention of digital games in formal education. Comput Human Behav 2012 Nov;28(6):2023-2033. [doi: 10.1016/j.chb.2012.05.021]
127. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep 06;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]

128. Geuens J, Geurts L, Gerling K, De CR, Abeele V. A dyad of lenses for the motivational design of mHealth: bridging the gap between health theory and app design. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics (ICHI). 2019 Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); Jun 10-13 , 2019; Xi'an, China. [doi: 10.1109/ichi.2019.8904839]
129. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
130. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
131. Maps of digital desires: exploring the topography of gender and play in online games. In: Beyond Barbie and Mortal Kombat New Perspectives on Gender and Gaming. Cambridge, Massachusetts, United States: MIT Press; 2011.
132. Dixon R, Maddison R, Ni Mhurchu C, Jull A, Meagher-Lundberg P, Widdowson D. Parents' and children's perceptions of active video games: a focus group study. J Child Health Care 2010 Jun 04;14(2):189-199. [doi: 10.1177/1367493509359173] [Medline: 20203134]
133. Simons M, de Vet E, Hoornstra S, Brug J, Seidell J, Chinapaw M. Adolescents' views on active and non-active videogames: a focus group study. Games Health J 2012 Jun;1(3):211-218. [doi: 10.1089/g4h.2011.0032] [Medline: 26193439]
134. De Vet E, Simons M, Wesselman M. Dutch children and parents' views on active and non-active video gaming. Health Promot Int 2014 Jun 02;29(2):235-243. [doi: 10.1093/heapro/das064] [Medline: 23208151]
135. Thompson D, Cantu D, Rajendran M, Rajendran M, Bhargava T, Zhang Y, et al. Development of a teen-focused exergame. Games Health J 2016 Oct;5(5):342-356 [FREE Full text] [doi: 10.1089/g4h.2016.0035] [Medline: 27680385]
136. Kayali F, Silbernagl M, Peters K, Mateus-Berr R, Reithofer A, Martinek D, et al. Design considerations for a serious game for children after hematopoietic stem cell transplantation. Entertain Comput 2016 Jun;15:57-73. [doi: 10.1016/j.entcom.2016.04.002]
137. Schwarz A, Mertens L, Simons M, Spook JE, Thompson D, Cardon G, et al. Which game narratives do adolescents of different gameplay and sociodemographic backgrounds prefer? A mixed-methods analysis. Games Health J 2019 Jun;8(3):195-204. [doi: 10.1089/g4h.2017.0178] [Medline: 30411982]

Abbreviations

BCT: behavior change technique
mHealth: mobile health
MVPA: moderate- to vigorous-intensity physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
PSD: Persuasive Systems Design
SEP: socioeconomic position

Edited by L Buis; submitted 08.07.22; peer-reviewed by J Spook, E Lyons; comments to author 11.08.22; revised version received 02.11.22; accepted 21.12.22; published 06.03.23.

Please cite as:
Schwarz A, Winkens LHH, de Vet E, Ossendrijver D, Bouwsema K, Simons M
Design Features Associated With Engagement in Mobile Health Physical Activity Interventions Among Youth: Systematic Review of Qualitative and Quantitative Studies
JMIR Mhealth Uhealth 2023;11:e40898
URL: https://mhealth.jmir.org/2023/1/e40898
doi:10.2196/40898
PMID:36877551

Review

Smartphone and Mobile App Use Among Physicians in Clinical Practice: Scoping Review

Mauricette Lee[1], BSc; Abu Bakar Shakran Bin Mahmood[1]; Eng Sing Lee[2,3], PhD; Helen Elizabeth Smith[2], PhD; Lorainne Tudor Car[1,4], MSc, MD, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[3]National Health Group Polyclinics, Singapore, Singapore

[4]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MSc, MD, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: Health care professionals are increasingly using smartphones in clinical care. Smartphone use can affect patient quality of care and clinical outcomes.

Objective: This scoping review aimed to describe how physicians use smartphones and mobile apps in clinical settings.

Methods: We conducted a scoping review using the Joanna Briggs Institute methodology and reported the results according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. We used the following databases in our literature search: MEDLINE, Embase, Cochrane Library, Web of Science, Google Scholar, and gray literature for studies published since 2010. An additional search was also performed by scanning the reference lists of included studies. A narrative synthesis approach was used.

Results: A total of 10 studies, published between 2016 and 2021, were included in this review. Of these studies, 8 used surveys and 2 used surveys with focus group study designs to explore smartphone use, its adoption, experience of using it, and views on the use of smartphones among physicians. There were studies with only general practitioners (n=3), studies with only specialists (n=3), and studies with both general practitioners and specialists (n=4). Physicians use smartphones and mobile apps for communication (n=9), clinical decision-making (n=7), drug compendium (n=7), medical education and training (n=7), maintaining health records (n=4), managing time (n=4), and monitoring patients (n=2) in clinical practice. The Medscape medical app was frequently used for information gathering. WhatsApp, a nonmedical app, was commonly used for physician-patient communication. The commonly reported barriers were lack of regulatory oversight, privacy concerns, and limited Wi-Fi or internet access. The commonly reported facilitator was convenience and having access to evidence-based medicine, clinical decision-making support, and a wide array of apps.

Conclusions: Smartphones and mobile apps were used for communication, medical education and training, clinical decision-making, and drug compendia in most studies. Although the benefits of smartphones and mobile apps for physicians at work were promising, there were concerns about patient privacy and confidentiality. Legislation is urgently needed to protect the liability of health care professionals using smartphones.

(JMIR Mhealth Uhealth 2023;11:e44765) doi:10.2196/44765

KEYWORDS

evidence-based medicine; specialist; general practitioners; GP; primary care physicians; mobile apps; consultants; surgeons; pediatricians; clinical care; mobile phone

Introduction

Background

The use of smartphones has become increasingly indispensable [1]. This technology has revolutionized how people live, learn, work, communicate, and entertain themselves [2]. The use of smartphones among health care professionals is also widespread and affects the clinical care they provide [3]. Studies in various settings reported that most health care professionals use smartphones daily in their practice [4-8].

Smartphones and mobile apps offer an important and diverse set of clinical tools for health care professionals. They enable direct communication with colleagues and patients, instant access to medical knowledge, education, remote patient management, research, and digital diagnostics, to name a few [2]. However, the widespread adoption and use of smartphones in medical practice can affect the quality of care [9,10]. There are concerns about the impact of smartphones on professionalism, patient safety, and data confidentiality as well as the trustworthiness of sources accessed via smartphones [6-8,11-14].

The use of smartphones may vary between different groups of health care professionals and in different settings. For instance, some studies report that medical journal mobile apps are more commonly used by physicians than by nurses [15,16]. In contrast, medical calculators or drug compendium apps are used by both physicians and nurses [7,16]. Many medical mobile apps targeting health care professionals are available, and their number is growing. Studies have reported that the daily use of medical apps ranges from 1 to 20 minutes among physicians [7]. Knowing what types of apps are commonly used by various health care professionals can help discern their needs, guide the future evaluation of the quality of such apps, and inform the development of new apps.

Objectives

A growing number of studies are exploring the use of smartphones among health care professionals as well as their experiences and perceptions of the role of smartphones in clinical care [4,6,7,9-11,13-15,17-22]. However, to date, there are no existing scoping reviews, systematic reviews, and research syntheses available on this topic. Our objective was to collate and describe how smartphones and mobile apps were used by physicians, specifically, specialists and family physicians, within clinical settings. We presented the barriers, facilitators, and opinions of physicians regarding smartphones and mobile apps.

Methods

Overview

A scoping review was conducted using the Joanna Briggs Institute methodology and reported according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [23]. The scoping review methodology consisted of five key steps: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. The study protocol was registered in the Open Science Framework registries [24].

Step 1: Identifying the Research Question

This review aimed to collate and describe studies focusing on the use of smartphones among physicians. The overarching question for this scoping review was as follows: "How do physicians use smartphones in clinical practice?" More specifically, the research questions for this scoping review are as follows:

- What are the smartphone apps and features physicians access and why?
- How do physicians use their smartphones as an information source?
- What are physicians' opinions of the impact of smartphones on clinical care?

Step 2: Identifying Relevant Studies

We developed the MEDLINE (Ovid) search strategy collaboratively and iteratively with support from an experienced medical librarian. The search strategy was guided by relevant articles identified from previous manual searches, based on our research questions, and eligibility criteria (Multimedia Appendix 1). The same strategy was adopted to search for applicable studies in Embase, Cochrane Library, and Web of Science. Similar to previous studies, we also searched the reference lists of the included studies and gray literature in the first 10 pages of the search results in Google Scholar using the search terms in our search strategy and titles of the included studies [25,26]. Only studies published in the English language were included. Results were imported to EndNote 20 (Clarivate Analytics) [27].

The included studies in this review had to meet the inclusion and exclusion criteria presented in Textbox 1. We included studies published between January 2010 and January 2022 to capture data that aligned with the proliferation of smartphone ownership [28]. This review aimed to understand practicing physicians', defined as specialists, and family physicians' use of smartphones in clinical settings. Owing to the differences in training needs between medical trainees and nontrainees, medical trainees were excluded [29].

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Studies focusing on the use of smartphones among physicians defined as specialists and family physicians
- Studies exploring the use of mobile apps and the use of social media if this is done explicitly using smartphones (only if the motivation for use was physician driven)
- Studies focusing on personal smartphones, organizationally provided smartphones, or both
- Studies focusing on a mix of physicians if more than 50% of the physicians were specialists and family physicians
- Survey, observational, mixed methods, or qualitative studies
- Published between January 2010 and January 2022
- Printed in the English language

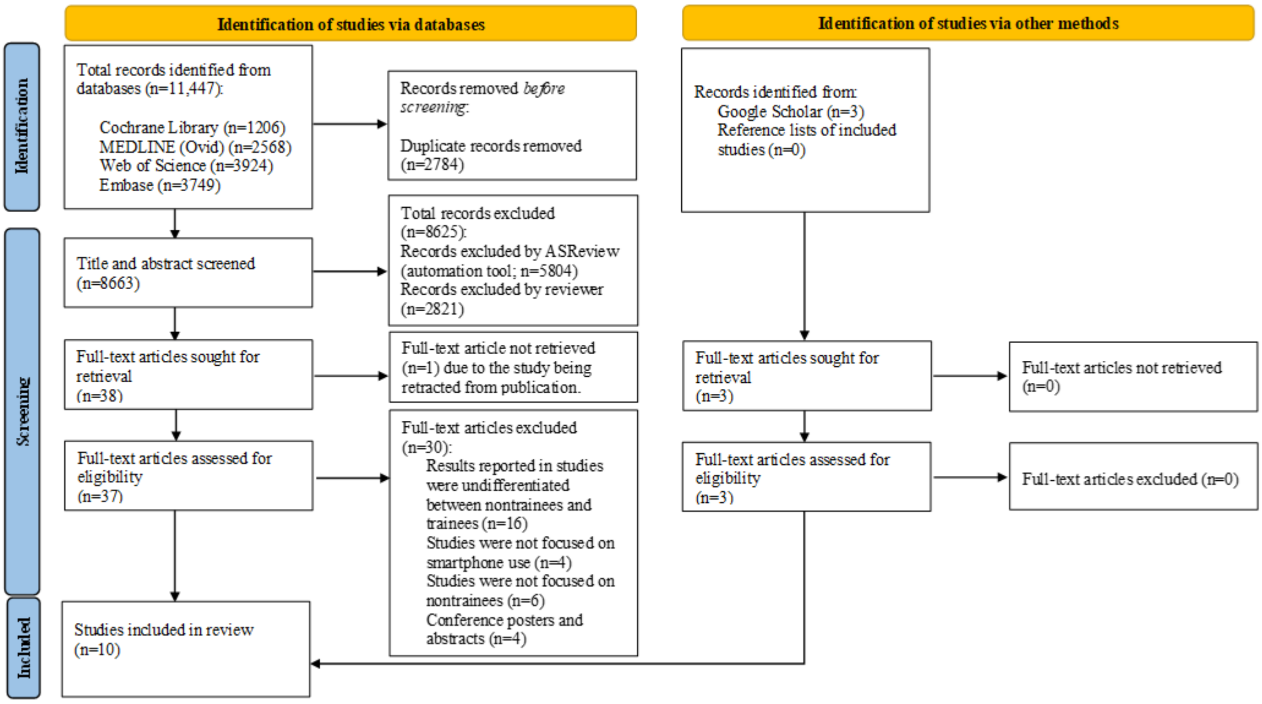
Exclusion criteria

- Studies that focused on patients, medical students, medical trainees, medical residents, house officers, health intervention, or medical education
- If the smartphones were implemented for research purposes
- Studies that focused on infection control of personal smartphones
- Editorials, opinion pieces, conference posters, and abstracts

Step 3: Study Selection

Studies were identified using the search criteria presented in Textbox 1. The search results from different electronic databases were combined in a single EndNote (Clarivate Analytics) library [27], and duplicate records were removed. The first reviewer (MLM) independently screened titles and abstracts on ASReview v1.0 (ASReview Lab) [30] to identify studies that potentially met the inclusion criteria. Only 33% of the titles and abstracts were screened. This was 1 rule that was predetermined and adhered to before screening commenced on ASReview [30]. This rule was set based on a study that found that 95% of eligible studies would be found after screening between 8% and 33% of studies on ASReview [31]. In parallel, a second reviewer (ABSBM) independently screened all titles and abstracts, including 33% of titles and abstracts that were screened by MLM on Covidence (Veritas Health Innovation) [32] to identify studies that potentially met the inclusion criteria. Disagreements on the included titles and abstracts were resolved through discussion between the first and second reviewers. Conflicts between the 2 reviewers were resolved through consensus, and when required, a third reviewer (Ahmad Ishqi Jabir) acted as an arbiter. In total, 2 reviewers (MLM and ABSBM) independently retrieved the full texts of the included titles and abstracts and read and assessed the studies against the eligibility criteria. Disagreements on the included full-text articles were resolved through discussion between the first and second reviewers. Conflicts between the 2 reviewers were resolved through consensus, and when required, a third reviewer (AIJ) acted as an arbiter. A total of 2 reviewers (MLM and ABSBM) independently extracted the data for each included study using a structured data extraction form. Figure 1 shows a flow diagram of the article selection process.

Figure 1. Flow diagram for the article selection process.



Step 4: Charting the Data

After the screening process was completed, EndNote Library [27] was set up to share articles between the reviewers. A data charting form was created using Microsoft Excel and used to extract data from the included studies. The data extracted from each study included the name of the first authors, year of publication, title, aims of the study, study design, study location, study time frame, sample size, participant characteristics, type of smartphone used, and study findings (Multimedia Appendix 2 [5,22,33-40]). Adapted from previous studies by Lee and De Jong [41,42], the data extracted from the included studies were categorized into functions, benefits, and challenges of smartphones and mobile apps for health care professionals. We used coding frames from these studies, as they captured data that were aligned with the proliferation of smartphone ownership. The data charting form was piloted by 2 reviewers (MLM and ABSBM), with 4 studies either different in study designs or population specialties to ensure consistent, reliable, and efficient data extraction. Conflicts between the 2 reviewers were resolved through consensus, and when required, a third reviewer (AIJ) served as an arbiter.

Step 5: Collating, Summarizing, and Reporting the Results

A comprehensive summary of the included studies, number of studies, study design, data collection methods, population types, and the aims of the study were presented. The use of smartphones was organized according to themes presented in a previous paper [41]. The framework we developed was aligned with all our research questions. The collated findings were ownership rates, type of mobile apps used, type of information sources used, type of websites accessed, use of smartphones for contact with colleagues and patients, and physicians' experiences of using smartphones. Relationships between population characteristics and the use of smartphones (differences in the use of smartphones among physicians in primary and secondary care) were identified. A narrative synthesis of the findings without the use of a qualitative analysis program was presented.

Results

Search Findings

Database searches yielded 11,447 records, and another 3 records were retrieved from the gray literature source. After removing 2784 duplicates, 8663 titles and abstracts were screened. Title and abstract screening led to the exclusion of 8625 records, resulting in 40 full texts that needed to be assessed for eligibility. Of these, 30 articles were excluded, resulting in 10 studies for the review (Figure 1). We used a sensitive search strategy that aimed to retrieve all relevant research in this novel area and as such had a high number of citations initially. We then screened citations in parallel and independently to ensure the reliability of our screening.

Study Characteristics

The 10 studies included in this review (Table 1 [5,22,33-40]) were conducted across 9 countries and published between 2016 and 2021. The study data collection time frames were reported in 4 studies [5,22,37,38], and data were collected between 2014 and 2019. Of the 10 included studies, 8 (80%) used surveys [5,22,33-38] and 2 (20%) used surveys with focus groups study designs [39,40]. A total of 3 studies recruited only general practitioners (GPs) [22,33,39], another 3 studies recruited only specialists [34,35,37], and 4 studies recruited both GPs and specialists [5,36,38,40]. The participants in the remaining studies were addressed as anesthetist consultants [34], pediatricians [35], specialists, or surgeons. Overall, 4 studies did not provide information on their clinical settings [5,33,35,36], 2 studies were conducted in hospitals [34,38], and the remaining studies were conducted in community health centers [22], large university surgical departments [37], rural practices [39], and health institutions [40]. The characteristics and details of the included studies on the use of smartphones and mobile apps by physicians are presented in Tables 1 and 2 (Multimedia Appendix 2). One study reported that physicians possessed more than 6 different work-related mobile apps on their smartphones [34]. Another study reported that most GPs had 1 to 3 medical apps, with very few owning more than 4 [22]. Only 1 study found that young GPs (aged <35 years) were more likely to own smartphones [22]. Another study found that younger physicians (aged ≤44 years) were less likely to allow their patients to communicate with them via the internet or phone, and they used medical apps more often [5]. One study reported that 10% (5/50) of physicians used organizationally provided smartphones, whereas the rest used personal smartphones for clinical use [38]. The results are presented in Tables 1 and 2 and are consistent with the PRISMA-ScR guidelines [43].

Table 1. Summary table of included studies (N=10).

Features	Study, n (%)
Country	
Ireland	2 (20)
Austria	1 (10)
Australia	1 (10)
Canada	1 (10)
China	1 (10)
Cyprus	1 (10)
Ecuador	1 (10)
Sudan	1 (10)
Turkey	1 (10)
Year of publication	
2021	1 (10)
2020	2 (20)
2019	1 (10)
2018	2 (20)
2017	2 (20)
2016	2 (20)
Population type	
Only GPs[a]	3 (30)
Only specialists	3 (30)
GPs and specialists	4 (40)
Age group (years)	
25-35	1 (10)
36-45	4 (40)
46-66	3 (30)
>66	0 (0)
Not applicable[b]	2 (20)
Sex[c]	
Mostly male	5 (50)
Mostly female	3 (30)
Intersex	0 (0)
Not applicable	2 (20)
Type of smartphone used[c]	
Mostly Android	1 (10)
Mostly iPhone	3 (30)
Others	0 (0)
Not applicable	6 (60)
Most commonly reported frequency of smartphone use	
Daily	2 (20)
Weekly	1 (10)
Monthly	0 (0)

Features	Study, n (%)
Sometimes	1 (10)
Rarely	0 (0)
Never	1 (10)
Not reported	5 (50)
Most commonly reported purpose of smartphone use	
Communication	9 (90)
Clinical decision-making	7 (70)
Medical education and training	7 (70)
Reference tools	7 (70)
Health record maintenance	4 (40)
Time management	4 (40)
Patient monitoring	2 (20)

[a]GP: general practitioner.

[b]This information was not reported in this study.

[c]The number of studies was calculated based on the majority reported under sex and type of smartphone used. For instance, if a study reported more males than females being recruited, we counted it as "Mostly males." Similarly, there are a number of studies on the type of smartphones used.

Table 2. Smartphones and mobile apps used by physicians.

Purpose of smartphone use	Examples of mobile apps and features used
Communication	WhatsApp [5,36,40], Google Hangout [40], Facebook [5,36], YouTube [36], SMS text messaging [36], email [22,36,37], voice calling [37], and instant messaging [33,37,39]
Information seeking and management	
Clinical decision-making	Medscape [35,36,40], UpToDate [35,40], Nature [40], MedCalc [5,40], Das28 [40], Diagnostic assistance tools [38], Prognosis [40], Dxsarus [40], Laborwerte [5], Labormedizin Pocket [5], and medical calculators [38]
Health record maintenance	Enil [40], Meddata [40], E-nabiz [40], PACSapp [40], Acibadem [40], and coding and billing [42]
Medical education and training	Twitter [34], Medscape [35,36,40], OrthoApp [40], Vcell [40], UpToDate [35,40], PubMed [36], and Nature [40]
Reference tools	UpToDate [35], Medscape [35,36], Cepilaç [40], Diagnosia [5], Embryotox [5], Antibiotika (Thalhammer) [5], Arzneimittel Pocket [5], Eponyms [40], and PubMed [36]
Clinical care	
Patient monitoring	Apple Health [37], Instant health rate [37], and Fitwell [37]
Time management	Google Calendar [40]

Communication

Of the 10 included studies, 7 (70%) reported the use of smartphones and mobile apps for communication [5,22,33,36,37,39,40]. One study reported that participants allowed their patients to contact them by phone and via web-based communication tools [5]. However, the study did not report on the smartphone features or apps used. Another study reported that smartphones allowed faster access to information, especially for communication among peers [40]. More than half of the physicians preferred using smartphones and mobile apps over other alternatives for communication with other physicians [36]. Of the 10 studies, 3 (30%) did not explicitly report how smartphones and mobile apps were used by physicians for communication [34,35,38]. One study reported using smartphones to contact other specialists for referrals or advice [34], another study reported using smartphones for communication [38], and the last study did not report on communication at all [35].

Information Seeking and Management

Of the 10 included studies, 7 (70%) reported the use of mobile apps for medical education and training [5,34-38,40]. Of the 7 studies, 3 did not provide examples of mobile apps used by GPs and specialists in medical education and training [5,37,38]. Of the 3 studies, 1 reported that most physicians perceived that they were provided with reliable clinical content and continuing medical education when using mobile apps for medical education and training [5]. Another study reported that most surgeons felt that texting improved the educational experience of their trainees [37]. Of the 10 studies, 7 (70%) reported the use of mobile apps for clinical decision-making [5,22,34-36,38,40]. Of the 7 studies, 1 reported that anesthetists used mobile apps for clinical algorithms, clinical planning, and assessment [34]. Another study reported that mobile apps for disease diagnosis were used by GPs [22], but the identity of the mobile apps used was not provided in those 2 studies.

Frequent use of drug compendium apps was reported in 7 studies [5,22,34-36,38,40]. A total of 4 studies found that the most frequent type of medical apps used by physicians were drug compendium apps [5,22,34,40]. The other most common types of reference tools were the literature search portals [5,22,35,36,38] and medical literature [5,22,34-36]. These were followed by medical journals [34-36], medical news [5,35,36], and medical textbooks [5,40]. Of the 7 studies, 1 reported that anesthetists used mobile apps for drug compendium, prescription, and dosing, and academic journals [34]. However, no examples of the mobile apps used were provided in this study. Another study reported that the Diagnosia and Embryotox medical apps were most often used [5]. Physicians were less familiar with medical apps such as Antibiotika (Thalhammer) from Germany, MedCalc (United States), Laborwerte (Germany), Arzneimittel Pocket (Germany), and Labormedizin Pocket (Germany) [5].

Of the 10 included studies, 4 (40%) reported the use of mobile apps for health record maintenance to access local appointment systems developed by their local health ministry [5,34,38,40]. Of the 4 studies, 1 reported the types of smartphone apps used by anesthetists, including apps for billing and accessing patient results and those that allowed access to hospital electronic medical records [34]. Another study reported that Enlil (Turkey), a national hospital management information system, was the most used app [40]. It was followed by other medical health recording systems from Turkey, such as Meddata, E-nabiz, PACSapp, and Acibadem [40].

Clinical Care

The use of mobile apps for time management was reported in 40% (4/10) of the included studies [22,34,35,40]. Only 1 study reported Google Calendar as the most used app among GPs, followed by the default mobile calendar and appointment app [40]. The other 3 studies did not provide examples of mobile apps used for time management [22,34,35].

Only 20% (2/10) of studies reported the use of mobile apps for patient monitoring [38,40]. According to Sezgin et al [40], patient-monitoring apps were the least prevalent among all mobile apps. Examples of the types of patient-monitoring smartphone apps provided in that study were pedometers, calorie trackers, and heart rate and information tracker tools (cardiograph) [40]. The study also reported that patients had been using the tracking apps and sharing them with their physicians [40]. Teferi et al [38] reported using an app to document the procedures for patient monitoring.

Smartphone Features Used and Web Access by Physicians

Of the 10 studies, only 1 (10%) reported the use of built-in features in the smartphone, such as torchlight, stopwatch, and camera [34]. The same study also reported the use of a smartphone to distract pediatric patients [34]. However, the study did not report on how the smartphone features were used by physicians.

A total of 2 studies reported that physicians used nonmedical apps more frequently than medical apps [22,42]. These nonmedical apps were used by physicians for calendar [22] and web access [22,42]. Overall, 4 studies reported on the use of smartphones for web access [22,35,38,40]. However, 3 of the 4 studies neither reported the reasons for physicians to access the internet nor stated the websites that were accessed [22,38,40]. Web browsers and search engines were also used for medical information searches, sometimes outperforming medical apps [40].

Barriers to the Use of Smartphones and Mobile Apps by Physicians

Overview

In addition to depicting the use of smartphones and mobile apps by physicians, many studies have described the factors that prevented physicians from using smartphones. This review found a wide variety of barriers to the use of smartphones and mobile apps by physicians, including the infringement of patient privacy and confidentiality, lack of regulatory oversight, negative impact on physician-patient and collegial relationships, quality concerns, limited Wi-Fi or internet access, lack of workplace integration, and lack of smartphone savviness.

Infringement of Patient Privacy and Confidentiality

The potential confidentiality breach of patient privacy was the most common barrier to the use of smartphones and mobile apps by physicians [33,36,37,39,40]. Only one study reported on privacy concerns, which included the fear of sending the message to the wrong person or number and the uncertainty around the receipt of the messages [33]. The lack of security and control over the apps' content were also perceived to be risks related to the infringement of patient privacy and confidentiality when using smartphones and mobile apps for communication at work [40]. However, the study did not report the names of these apps.

Lack of Regulatory Oversight

The included studies also addressed barriers to the regulation of smartphones and mobile apps used by physicians [33,37,39,40]. One study reported that only 27% of GPs have a written text policy for texting patients [33]. GPs who used texts always documented patient consent, and when texting medically sensitive information, they always obtained specific consent [33,39]. In the hospital setting, some physicians were unaware that they had any organizational policy on the use of smartphones [34] and sharing patient information via text messages [37]. Most surgeons in 1 study agreed that texting patient-related information should be regulated by a hospital policy (74%) or legislation (57%) [37].

Negative Impact on Physician-Patient and Collegial Relationships

Several studies in this review referenced barriers related to professional relationships. Hofer and Haluza [5] reported that employees were not allowed to use their smartphones at work, as it was found to be disruptive to the relationship with patients during consultation. It was also reported that, among consultant anesthetists with more than 3 years of experience as a consultant, up to 27% agreed that their smartphone was a distraction from their work [34]. However, none of those with less than 3 years in post believed that their smartphones were a distraction from their work [34]. In addition to distraction, GPs found that text messaging increased patient anxiety [39]. Some physicians also reported feeling uncomfortable using smartphones in front of patients [40].

Quality Concerns

Another reported barrier was concern about the quality of the information provided by medical apps [5,40]. For example, physicians expected professional organizations to inform evidence-based medical apps, including assessing the quality of medical information in the medical apps recommended by the organization [5].

Limited Wi-Fi or Internet Access

In total, 2 studies reported limited internet access as a barrier [5,36]. Physicians mentioned that they would use many more apps if smartphone reception was better in the hospital [5]. Hence, they suggested that the availability of an offline version of an app is important [5].

Lack of Workplace Integration

Sezgin et al [40] also raised the issue of the lack of extensive use of smartphones and mobile apps in the hospital system. This has been demonstrated by the lack of interoperability between the use of smartphones and other hospital devices [40].

Lack of Smartphone Savviness

Only one study reported the lack of advanced skills as a barrier to the use of smartphones and mobile apps [40]. For example, some physicians indicated that they were not aware and unsure of the appropriate apps that could be used to help them with their daily clinical tasks. Their lack of knowledge on smartphone use prevented them from using it in clinical settings.

Facilitators for the Use of Smartphones and Mobile Apps by Physicians

Overview

Numerous studies have reported on the facilitators for the use of smartphones and mobile apps by physicians. Facilitators included convenience and access to evidence-based medicine and clinical decision-making support. One study reported that smartphones and mobile apps were useful for conducting research. However, the authors did not elaborate further [36]. The user-friendliness of medical apps was perceived to facilitate the ease of use of mobile apps [5].

Convenience

Physicians used smartphones and mobile apps primarily for convenience [5,33,34,36,37,39,40]. For example, flexible communication channels [5,33,34,36,37,40] as well as a selection of powerful apps to accomplish a variety of tasks at work were readily available [5,36]. Portability [36,41], rapid access to information [37], and multimedia resources [5,36,41] were also examples of convenience.

Access to Evidence-Based Medicine and Clinical Decision-making Support

Access to various evidence-based and clinical decision-making support mobile apps was highlighted as a facilitator in this review [5,34,40]. The evidence-based medical mobile apps included apps for medical education and training and reference tools, as listed in Table 2.

Discussion

Summary of Key Findings

According to the studies included in this review, physicians primarily use smartphones and mobile apps for communication, medical education and training, clinical decision-making, and accessing the drug compendium. Medscape was frequently mentioned as a medical app used for information gathering. WhatsApp has been widely reported as a nonmedical app used for physician-physician and physician-patient communication. The most common barriers reported in the included studies were the risk of infringing on patient privacy and confidentiality, lack of regulatory oversight, limited Wi-Fi or internet access, the lack of extensive use of mobile apps in the hospital system, and the lack of smartphone savviness. The most common facilitators reported in the included studies were the availability of having flexible communication methods, easy access to evidence-based medicine, clinical decision-making support, availability of mobile app choices to accomplish many different purposes at work, and portability.

We found that physicians are more likely to use smartphones for work-related purposes because of the increasing availability of mobile apps. Prior studies showed that only 13% of physicians used their smartphones to watch web-based videos weekly for professional purposes, and continuing medical education activities were the most frequently viewed content [44]. However, this review found that studies frequently reported the daily and weekly use of smartphones and mobile apps by physicians [5,22,40]. In addition, smartphones and mobile apps were widely used not only for medical education and training but also for communication, clinical decision-making, and reference tools. We found mixed views regarding the use of smartphones for work-related purposes as a distraction for physicians [5,34]. For instance, physicians believe that using a smartphone during a consultation could negatively affect the patient-physician relationship [5]. This finding is consistent with a recent systematic review on the effect of web-based information-seeking behavior on the physician-patient relationship [45]. Another study found a correlation between the number of years of experience as a specialist and whether smartphones were perceived as a distraction [34]. Younger physicians tended to use smartphones more and were more likely to accept them in the workplace [34]. This was also found to be consistent with a recent systematic review of distraction with smartphones during nursing care [46]. Future research should conduct a review on the distraction of smartphones from physicians in the clinical setting and perhaps derive a precise estimate of the effect that smartphone distraction has on clinical care outcomes.

Our review identified some challenges to the use of smartphones and mobile apps by physicians. First, we found that physicians were unaware of their hospital's policy [41] on the use of smartphones at work. Only a minority of GPs had written a text policy for texting patients [32]. As a result, while our review found that most GPs who used text messaging always documented patient consent when texting medically sensitive information [33], there remains the potential for a breach in confidentiality. Although previous studies have suggested that the use of strong authentication mechanisms helps to mitigate the risks of a breach [47,48], we found evidence that not all physicians had their smartphones encrypted or password protected, and others were unsure whether their smartphones were encrypted [37].

Physicians are increasingly using instant messaging tools, such as WhatsApp, Facebook, and Google Hangout, for physician-physician and physician-patient communications. However, using social media at work may result in the mingling of personal and hospital data. Most social media tools mentioned in our review are not Health Insurance Portability and Accountability Act compliant, which aims to protect patient privacy and ensure the integrity of sensitive medical information [49]. Despite the absence of Health Insurance Portability and Accountability Act–specific regulations for smartphones and apps, some organizations have developed recommendations and guidelines for mobile security measures [49-53]. A previous study [47] suggested that educating health care professionals about the available hospital policy on the use of smartphones at work could be useful in implementing the policy. However, our review revealed a lack of direction for ideal smartphone use at work. Consequently, it might be helpful to have a policy or legislation that provides comprehensive guidance on authentication, access control, chain of responsibility, data ownership, allowed devices, acceptable use, training, and noncompliance with the use of smartphones and mobile apps [47,54]. Compliance with the legislation of smartphone use at work should be considered in the future during the appraisal process of health care professionals.

This review found that physicians use evidence-based medical apps because they provide instant access to evidence. One example of such an app is the evidence-based point-of-care information summaries [5,34,40]. Point-of-care information summaries are defined as medical compendia specifically designed to deliver predigested, rapidly accessible, comprehensive, periodically updated, and evidence-based information (and possibly guidance) to clinicians [55]. Our review found that Medscape and UpToDate were the most commonly reported evidence-based point-of-care information summaries apps. However, health care organizations lack information on the use of evidence-based medical apps [5]. They were unaware of the reliability of evidence-based information provided by medical apps [5]. To ensure quality and safety, the use of medical apps must undergo rigorous evaluation, validation, and development of best practice standards [40]. Therefore, as a means of mitigating the use of non–evidence-based information in clinical practice, future research should assess the quality of evidence within medical apps to support health care professionals to be more confident when using such apps for practice. In addition, the findings from such research may inform policy on the audit and regulation of medical apps.

There were some limitations when conducting this scoping review. As 8 (80%) of the 10 studies used quantitative methods such as surveys to gather data, deep descriptions and examples to provide an in-depth understanding of smartphones and mobile app use were limited [5,22,33-38]. In addition, only English-language studies were included in this narrative synthesis. Although our classification of data was determined through detailed analysis, team discussions, and consensus, there may be themes that we have overlooked. However, as our comprehensive analysis was based on a commonly used framework on smartphone use by health care professionals, missing out on themes may have been minimized [41].

Implication of the Findings

Physicians use smartphones and mobile apps for communication, clinical decision-making, drug compendium, medical education and training, maintaining health records, managing time, and monitoring patients in clinical practice. However, we found several gaps related to the use of smartphones and mobile apps by physicians at work. These gaps are the lack of regulatory oversight either at a hospital or at a government level, that is, the need to address concerns about the risks of infringement of patient privacy and confidentiality when using smartphones and mobile apps for communication of patient information. There is a need to identify medical apps that provide reliable clinical information and nonmedical apps that can be used for communication by physicians at work. We also found that the use of smartphones differs in different subgroups, such as participants of different ages, sexes, and work experience. Therefore, there may be possible implications on the association of the characteristics of participants with the use of smartphones and mobile apps. Future studies should explore the associations between smartphone use with clinical practice. Future research should also provide more information about smartphone use in clinical practice, including whether smartphones were used for work-related or personal purposes, how smartphone features and apps were used, and how health care professionals communicate using smartphones.

Conclusions

Our review found literature reporting on the use of smartphones and mobile apps for communication, medical education and training, clinical decision-making, and drug compendia. Challenges related to the use of smartphones and mobile apps include the lack of patient privacy and confidentiality and regulatory oversight. The benefits of smartphones and mobile apps for physicians at work include the availability of having flexible communication methods and mobile app choices to accomplish many different purposes at work, easy access to evidence-based medicine and clinical decision-making support, and portability. Physicians commonly use Medscape and WhatsApp mobile apps. Future research should address patient privacy issues, as well as legislation related to smartphone and mobile apps in clinical practice.

Acknowledgments

The authors would like to thank Ms Yasmin Lynda Munro from the Medical Library (Nanyang Technological University) for her help in developing the search strategy. The authors would also like to thank Ahmad Ishqi Jabir, a research associate from Lee Kong Chian School of Medicine (Nanyang Technological University), for his advice on the use of artificial intelligence for the study selection. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

LTC conceived the idea of the review. ML and ABSBM screened the studies and extracted data. ML synthesized the findings and wrote the manuscript. LTC, HES, ESL, and ABSBM provided insightful feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[DOCX File , 25 KB - mhealth_v11i1e44765_app1.docx]

Multimedia Appendix 2

Characteristics of the included studies on the use of smartphones and mobile apps for physicians.

[DOCX File , 41 KB - mhealth_v11i1e44765_app2.docx]

References

1. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb 05. URL: https://tinyurl.com/3erjjbjt [accessed 2022-09-23]
2. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
3. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
4. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht UV. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
5. Hofer F, Haluza D. Are Austrian practitioners ready to use medical apps? Results of a validation study. BMC Med Inform Decis Mak 2019 Apr 24;19(1):88 [FREE Full text] [doi: 10.1186/s12911-019-0811-2] [Medline: 31014333]
6. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 07;1(4):174-181 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000062]
7. Payne KF, Wharrad H, Watts K. Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
8. Franko OI, Tirrell TF. Smartphone app use among medical providers in ACGME training programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
9. Abolfotouh MA, BaniMustafa A, Salam M, Al-Assiri M, Aldebasi B, Bushnak I. Use of smartphone and perception towards the usefulness and practicality of its medical applications among healthcare workers in Saudi Arabia. BMC Health Serv Res 2019 Nov 12;19(1):826 [FREE Full text] [doi: 10.1186/s12913-019-4523-1] [Medline: 31718639]
10. Lo V, Wu RC, Morra D, Lee L, Reeves S. The use of smartphones in general and internal medicine units: a boon or a bane to the promotion of interprofessional collaboration? J Interprof Care 2012 Jul;26(4):276-282. [doi: 10.3109/13561820.2012.663013] [Medline: 22482742]
11. Wu R, Rossos P, Quan S, Reeves S, Lo V, Wong B, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. J Med Internet Res 2011 Aug 29;13(3):e59 [FREE Full text] [doi: 10.2196/jmir.1655] [Medline: 21875849]
12. Borshoff D. Is it time to switch to 'silent'? Anaesth Intensive Care 2011 Jan;39(1):135 [FREE Full text] [Medline: 21375108]
13. Katz-Sidlow RJ, Ludwig A, Miller S, Sidlow R. Smartphone use during inpatient attending rounds: prevalence, patterns and potential for distraction. J Hosp Med 2012 Oct;7(8):595-599. [doi: 10.1002/jhm.1950] [Medline: 22744793]

14. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
15. Smart NJ. A survey of smartphone and tablet computer use by colorectal surgeons in the UK and Continental Europe. Colorectal Dis 2012 Sep;14(9):e535-e538. [doi: 10.1111/j.1463-1318.2012.03161.x] [Medline: 22747977]
16. Mayer MA, Rodríguez Blanco O, Torrejon A. Use of health apps by nurses for professional purposes: web-based survey study. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e15195 [FREE Full text] [doi: 10.2196/15195] [Medline: 31682587]
17. Carey E, Payne KF, Ahmed N, Goodson A. The benefit of the smartphone in oral and maxillofacial surgery: smartphone use among maxillofacial surgery trainees and iPhone apps for the maxillofacial surgeon. J Maxillofac Oral Surg 2015 Jun;14(2):131-137 [FREE Full text] [doi: 10.1007/s12663-013-0550-3] [Medline: 26028826]
18. Choi JS, Yi B, Park JH, Choi K, Jung J, Park SW, et al. The uses of the smartphone for doctors: an empirical study from samsung medical center. Healthc Inform Res 2011 Jun;17(2):131-138 [FREE Full text] [doi: 10.4258/hir.2011.17.2.131] [Medline: 21886874]
19. El Hadidy TS, Alshafei AE, Mortell AE, Doherty EM. Smartphones in clinical practice: doctors' experience at two Dublin paediatric teaching hospitals. Ir J Med Sci 2018 Aug;187(3):565-573. [doi: 10.1007/s11845-017-1713-x] [Medline: 29170970]
20. Haroon M, Yasin F, Eckel R, Walker F. Perceptions and attitudes of hospital staff toward paging system and the use of mobile phones. Int J Technol Assess Health Care 2010 Oct;26(4):377-381. [doi: 10.1017/S0266462310001054] [Medline: 20923591]
21. Johnson AC, El Hajj SC, Perret JN, Caffery TS, Jones GN, Musso MW. Smartphones in medicine: emerging practices in an academic medical center. J Med Syst 2015 Jan;39(1):164. [doi: 10.1007/s10916-014-0164-4] [Medline: 25526706]
22. Liu Y, Ren W, Qiu Y, Liu J, Yin P, Ren J. The use of mobile phone and medical apps among general practitioners in Hangzhou City, eastern China. JMIR Mhealth Uhealth 2016 May 24;4(2):e64 [FREE Full text] [doi: 10.2196/mhealth.4508] [Medline: 27220417]
23. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
24. Lee M, Tudor Car L, Jabir AI. The use of mobile phones and its application by doctors: a scoping review protocol. Open Science Framework. 2022 Jul 29. URL: https://osf.io/ASZRB/ [accessed 2023-03-22]
25. Hagg E, Dahinten VS, Currie LM. The emerging use of social media for health-related purposes in low and middle-income countries: a scoping review. Int J Med Inform 2018 Jul;115:92-105 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.04.010] [Medline: 29779724]
26. Keren D, Lockyer J, Ellaway RH. Social studying and learning among medical students: a scoping review. Perspect Med Educ 2017 Oct;6(5):311-318 [FREE Full text] [doi: 10.1007/s40037-017-0358-9] [Medline: 28516340]
27. The EndNote Team. EndNote 20. Clarivate. Philadelphia, PA, USA: Clarivate; 2013. URL: https://support.clarivate.com/Endnote/s/article/Citing-the-EndNote-program-as-a-reference?language=en_US [accessed 2022-06-01]
28. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
29. Hardyman W, Bullock A, Brown A, Carter-Ingram S, Stacey M. Mobile technology supporting trainee doctors' workplace learning and patient care: an evaluation. BMC Med Educ 2013 Jan 21;13:6 [FREE Full text] [doi: 10.1186/1472-6920-13-6] [Medline: 23336964]
30. ASReview LAB developers. ASReview LAB software documentation (v1.0). Zenodo. 2022 Jun 08. URL: https://zenodo.org/record/6625881 [accessed 2022-07-01]
31. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 2021 Feb 01;3(2):125-133 [FREE Full text] [doi: 10.1038/s42256-020-00287-7]
32. Covidence: Better Systematic Review Management. Covidence. Melbourne, Australia: Veritas Health Innovation; 2022. URL: https://www.covidence.org/ [accessed 2022-07-01]
33. Quinlan D, Dahm M, Lyons A, Collins C. Patient texting in general practice: who, why, why not? A national survey of text messaging in Irish general practice. Ir Med J 2018 Apr 19;111(4):729. [Medline: 30465598]
34. Perkins EJ, Edelman DA, Brewster DJ. Smartphone use and perceptions of their benefit and detriment within Australian anaesthetic practice. Anaesth Intensive Care 2020 Sep;48(5):366-372. [doi: 10.1177/0310057X20947427] [Medline: 33021808]
35. Hadjipanayis A, Klonis A, Abdel-Mannan OA, Sutcliffe AG. Smartphone use in paediatric practice: a national survey. J Mob Technol Med 2016 Mar;5(1):3-8 [FREE Full text] [doi: 10.7309/jmtm.5.1.2]
36. Cherrez-Ojeda I, Vanegas E, Felix M, Mata VL, Jiménez FM, Sanchez M, et al. Frequency of use, perceptions and barriers of information and communication technologies among Latin American physicians: an Ecuadorian cross-sectional study. J Multidiscip Healthc 2020 Mar 12;13:259-269 [FREE Full text] [doi: 10.2147/JMDH.S246253] [Medline: 32214821]
37. Firdouse M, Devon K, Kayssi A, Goldfarb J, Rossos P, Cil TD. Using texting for clinical communication in surgery: a survey of academic staff surgeons. Surg Innov 2018 Jun;25(3):274-279. [doi: 10.1177/1553350618761980] [Medline: 29537349]

38. Teferi GH, Tilahun BC, Guadie HA, Amare AT. Smartphone medical app use & associated factors among physicians at referral hospitals in Amhara region, north Ethiopia, in 2019: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e19310 [FREE Full text] [doi: 10.2196/19310] [Medline: 33769303]
39. Leahy D, Lyons A, Dahm M, Quinlan D, Bradley C. Use of text messaging in general practice: a mixed methods investigation on GPs' and patients' views. Br J Gen Pract 2017 Nov;67(664):e744-e750 [FREE Full text] [doi: 10.3399/bjgp17X693065] [Medline: 28947620]
40. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: a mixed method study. Health Policy Technol 2017 Sep;6(3):251-267 [FREE Full text] [doi: 10.1016/j.hlpt.2017.07.007]
41. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
42. de Jong A, Donelle L, Kerr M. Nurses' use of personal smartphone technology in the workplace: scoping review. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e18774 [FREE Full text] [doi: 10.2196/18774] [Medline: 33242012]
43. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
44. Chase J. Healthcare professionals: the iPad and other drugs. Medical Marketing & Media: The Interactive Guide. 2013 Feb 1. URL: https://www.mmm-online.com/home/channel/features/healthcare-professionals-the-ipad-and-other-drugs/ [accessed 2022-07-01]
45. Luo A, Qin L, Yuan Y, Yang Z, Liu F, Huang P, et al. The effect of online health information seeking on physician-patient relationships: systematic review. J Med Internet Res 2022 Feb 10;24(2):e23354 [FREE Full text] [doi: 10.2196/23354] [Medline: 35142620]
46. Fiorinelli M, Di Mario S, Surace A, Mattei M, Russo C, Villa G, et al. Smartphone distraction during nursing care: systematic literature review. Appl Nurs Res 2021 Apr;58:151405. [doi: 10.1016/j.apnr.2021.151405] [Medline: 33745553]
47. Wani TA, Mendoza A, Gray K. Hospital bring-your-own-device security challenges and solutions: systematic review of gray literature. JMIR Mhealth Uhealth 2020 Jun 18;8(6):e18175 [FREE Full text] [doi: 10.2196/18175] [Medline: 32554388]
48. Redelmeier DA, Detsky AS. Pitfalls with smartphones in medicine. J Gen Intern Med 2013 Oct;28(10):1260-1263 [FREE Full text] [doi: 10.1007/s11606-013-2467-4] [Medline: 23625267]
49. Godard R. HIPAA compliance and cell phones: staying compliant while staying connected. Partners. 2022 Aug 16. URL: https://www.ispartnersllc.com/blog/hipaa-compliance-cell-phones/ [accessed 2022-09-28]
50. Mobile device security. National Cybersecurity Center of Excellence. URL: https://www.nccoe.nist.gov/mobile-device-security [accessed 2022-09-28]
51. Wani TA, Mendoza A, Gray K, Smolenaers F. Status of bring-your-own-device (BYOD) security practices in Australian hospitals - a national survey. Health Policy Technol 2020 Sep;11(3):100627 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100627]
52. Use of mobile devices by patients in hospitals. National Health Service 75 England. URL: https://tinyurl.com/mva2sym6 [accessed 2022-09-29]
53. Staff mobile device policy. National Health Service Trust. 2021 Sep 10. URL: https://www.leicspart.nhs.uk/wp-content/uploads/2021/09/Staff-Mobile-Device-Policy-exp-Sep-24.pdf [accessed 2022-09-29]
54. Olsen M, Lohning A, Campos M, Jones P, McKirdy S, Alghafri R, et al. Mobile phones of paediatric hospital staff are never cleaned and commonly used in toilets with implications for healthcare nosocomial diseases. Sci Rep 2021 Jun 21;11(1):12999 [FREE Full text] [doi: 10.1038/s41598-021-92360-3] [Medline: 34155278]
55. Kwag KH, González-Lorenzo M, Banzi R, Bonovas S, Moja L. Providing doctors with high-quality information: an updated evaluation of web-based point-of-care information summaries. J Med Internet Res 2016 Jan 19;18(1):e15 [FREE Full text] [doi: 10.2196/jmir.5234] [Medline: 26786976]

Abbreviations

GP: general practitioner
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by G Eysenbach; submitted 04.12.22; peer-reviewed by L Donelle, I Cherrez; comments to author 30.01.23; revised version received 08.02.23; accepted 12.03.23; published 31.03.23.

Please cite as:
Lee M, Bin Mahmood ABS, Lee ES, Smith HE, Tudor Car L
Smartphone and Mobile App Use Among Physicians in Clinical Practice: Scoping Review
JMIR Mhealth Uhealth 2023;11:e44765
URL: https://mhealth.jmir.org/2023/1/e44765
doi:[10.2196/44765](https://doi.org/10.2196/44765)
PMID:[37000498](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=37000498&dopt=Abstract)

Review

Health Monitoring Using Smart Home Technologies: Scoping Review

Plinio P Morita[1,2,3,4,5*], PEng, PhD; Kirti Sundar Sahu[1*], MPH, PhD; Arlene Oetomo[1*], BSc

[1]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Research Institute of Aging, University of Waterloo, Waterloo, ON, Canada

[4]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[5]Centre for Digital Therapeutics, University Health Network, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Plinio P Morita, PEng, PhD
School of Public Health Sciences, University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 5198884567 ext 31372
Fax: 1 5197466776
Email: plinio.morita@uwaterloo.ca

Abstract

Background: The Internet of Things (IoT) has become integrated into everyday life, with devices becoming permanent fixtures in many homes. As countries face increasing pressure on their health care systems, smart home technologies have the potential to support population health through continuous behavioral monitoring.

Objective: This scoping review aims to provide insight into this evolving field of research by surveying the current technologies and applications for in-home health monitoring.

Methods: Peer-reviewed papers from 2008 to 2021 related to smart home technologies for health care were extracted from 4 databases (PubMed, Scopus, ScienceDirect, and CINAHL); 49 papers met the inclusion criteria and were analyzed.

Results: Most of the studies were from Europe and North America. The largest proportion of the studies were proof of concept or pilot studies. Approximately 78% (38/49) of the studies used real human participants, most of whom were older females. Demographic data were often missing. Nearly 60% (29/49) of the studies reported on the health status of the participants. Results were primarily reported in engineering and technology journals. Almost 62% (30/49) of the studies used passive infrared sensors to report on motion detection where data were primarily binary. There were numerous data analysis, management, and machine learning techniques employed. The primary challenges reported by authors were differentiating between multiple participants in a single space, technology interoperability, and data security and privacy.

Conclusions: This scoping review synthesizes the current state of research on smart home technologies for health care. We were able to identify multiple trends and knowledge gaps—in particular, the lack of collaboration across disciplines. Technological development dominates over the human-centric part of the equation. During the preparation of this scoping review, we noted that the health care research papers lacked a concrete definition of a smart home, and based on the available evidence and the identified gaps, we propose a new definition for a smart home for health care. Smart home technology is growing rapidly, and interdisciplinary approaches will be needed to ensure integration into the health sector.

(JMIR Mhealth Uhealth 2023;11:e37347) doi:10.2196/37347

KEYWORDS

monitor; smart home; ambient assisted living; active assisted living; AAL; assisted living; review; internet of things; aging; gerontology; elder; older adult; older people; geriatric; digital health; eHealth; smart technology; older population; independent living; big data; machine learning; algorithm; deep learning

Introduction

Smart home technology is rapidly becoming a permanent fixture in our everyday lives. Globally, there are 175 million connected smart homes—a number projected to continue rising. Smart home technology employs the Internet of Things (IoT) concept to interconnect and share data among household devices across a Wi-Fi–based wireless network [1]. Through connection and automated action, smart homes provide convenience and comfort to homeowners [2-4]. Smart devices can include video monitors, motion sensors, alarms, smart planners or calendars, and thermostats. Data can be leveraged for a variety of purposes, including energy saving [5], security and safety [6], fall detection [7], light management [8], and fire detection [7]. However, the benefits of smart home technology run deeper than the superficial hype of comfort and convenience. These may be the solutions to our health care crisis.

The COVID-19 pandemic revealed what many health professionals already suspected: our health care system is overburdened. Our aging population places increased demand on the health care system. Many services are inaccessible to remote communities. Long-term care homes face high mortality and morbidity. To relieve an overwhelmed system, health care is turning to technology [9]—specifically, the application of smart home devices to support independent living. Through continuous behavioral monitoring, IoT devices can be harnessed to detect, diagnose, and monitor health conditions. At the community level, the collection and analysis of sensor data could inform public health initiatives. Interdisciplinary research teams are already working on the application of smart devices in health care. For example, smart wearable trackers, passive infrared sensors, and chair occupancy sensors deliver daily insights into the physical activity levels. Smart thermostats and bed occupancy sensors have been used to track sleep patterns. As physical activity and sleep are good overall health predictors, these can be powerful tools for motivating healthy behavioral changes [10]. The application of machine learning to these systems can be used for behavior change detection [11,12]. Applications can include monitoring the onset and progression of age-related diseases [10], detection of hazardous events (such as falls), and analyzing behavioral impacts following health interventions such as cancer treatments or physical therapy [12]. Information exchange with primary health care providers and caregivers will strengthen health care delivery. Public health authorities could also assess, in real time, the implications of COVID-19 lockdown policies at the population level. These data can be used to inform care delivery, support evidence-based policy making, and enhance care strategies in real time.

The main advantage of using IoT technologies is that they provide objective data in real time. Sensor data are collected passively without human effort; one can go about their day, forgetting about the device. The data are therefore less prone to performance and recall biases compared to the traditional data collection methods. As data are collected continuously and uploaded to the cloud storage, they are immediately available for analysis. The analysis can be conducted automatically, and the resulting insights can be shared immediately with users. The development and deployment of smart home technology for health care will require the concerted effort of an interdisciplinary research team: combining expertise in technology, engineering, and health care. Despite the potential of smart home solutions to health challenges, their real-world implementation continues to be scarce. There is a need to understand the current state of research in smart home technology for health care. Existing reviews on the application of smart home technology in health care are limited [2,3]. Here, we present a scoping review to address this need. The goal was to synthesize the literature on how smart home technologies are being used for health care within the home and community. This study also aims to identify gaps or opportunities in smart home technology to inform practice, policy making, and research. Our review was guided by the following research questions:

1. What smart home technologies are currently being used to monitor health care indicators in vulnerable populations at home or in the community?
2. What types of information are these sensors gathering?
3. What insights can be generated from these data sets?

Our extensive database search led to the identification of 49 peer-reviewed publications on smart home technology for health care, which met our inclusion criteria. We were able to identify multiple research trends and knowledge gaps and provide insight into the next steps needed to propel the field forward.

Methods

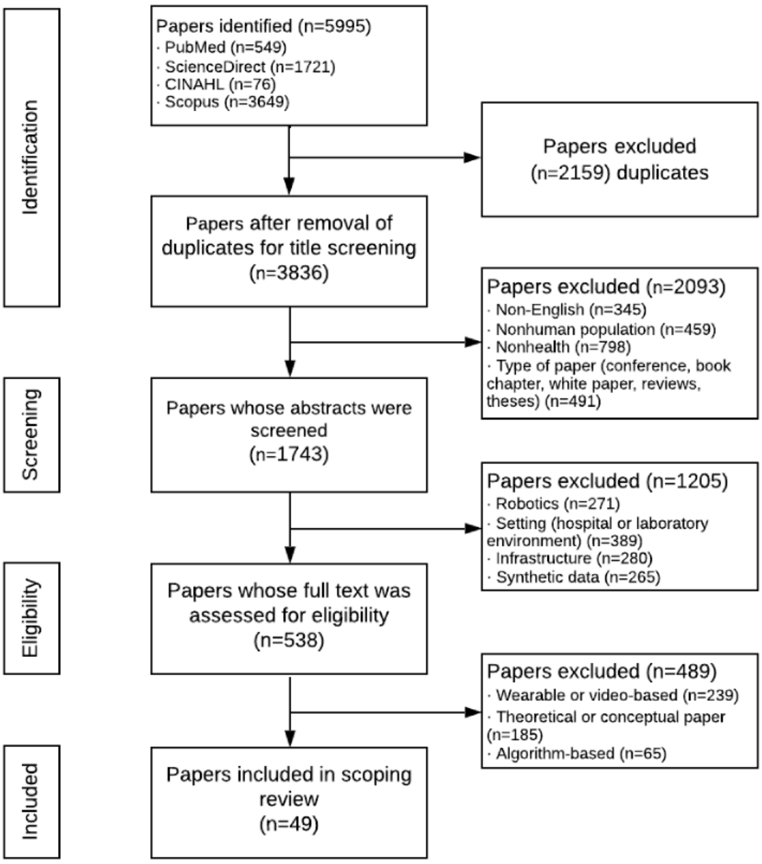
Data Sources and Search Strategy

This scoping review is based on the widely accepted framework by Arksey and O'Malley [13]. This framework was selected because it allows for the inclusion of various methodological designs across an interdisciplinary field. We searched for papers across 4 databases: PubMed, Scopus, ScienceDirect, and CINAHL. The search terms utilized are presented in Table S1 of Multimedia Appendix 1; they briefly encompassed the following search terms: health, monitor, smart home, ambient assisted living, active assisted living, and AAL. We limited our search to papers published between January 2008 and August 2021. Only peer-reviewed papers published in English were included. Of note, the term "surveillance" was not used in the search query, as its inclusion returned hundreds of results outside of the scope of this research project. A total of 5995 potential papers were identified using the search queries.

Paper Selection Process

Papers were organized into Mendeley and Zotero reference managers. Following the removal of 2159 duplicate papers, 3836 papers remained for title screening (Figure 1). Paper selection was further refined by ensuring that paper titles contained one of our keywords as mentioned above. AO and KSS each reviewed half of the papers. Papers not in English and those not related to humans were excluded: papers related to animal, agricultural, or biology research were excluded. Further, conference papers, book chapters, white papers, reviews, and theses were removed. Following title screening, 1743 papers were selected for abstract review by AO and KSS in Mendeley. AO and KSS screened the abstracts to ensure that the papers focused on remote sensor technology and its application in a home setting. Papers that used synthetic data or described infrastructure architecture or were in hospital or laboratory settings were excluded. The remaining 538 papers proceeded to full-text screening and were transitioned to Zotero for file management due to software issues in Mendeley. Studies using wearables or video-based technologies, theoretical or conceptual papers, and algorithm-based technologies were removed. Both authors independently and unanimously agreed on the inclusion of 29 papers with an additional 97 papers with conflicting votes. These papers were discussed on a case-by-case basis until a unanimous decision was reached. Of the 97 papers that had conflicting votes, 20 papers were included in this review. Thus, 49 papers were found to be eligible for the final scoping review. The selected papers were saved in a database, and a master chart was built by AO and KSS to summarize the key information for subsequent analysis.

Figure 1. Systematic study selection using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Results

Selection and Characterization of Studies on Smart Home Technologies

To gain an understanding of the types of smart home technologies being used and the information collected, we conducted a literature search across 4 databases (PubMed, Scopus, ScienceDirect, and CINAHL) between January 2008 and August 2021 by using the queries outlined in Table S1 of Multimedia Appendix 1. A total of 49 papers met the inclusion criteria for this scoping review (Table 1). Among the types of studies conducted, 31% (13/49) were pilot studies, 14% (7/49) were proof of concept, 12% (6/49) were algorithm evaluations, 10% (5/49) were proposals, 8% (4/49) were technical validations, 8% (4/49) were case studies, 6% (3/49) were method evaluations, 6% (3/49) were longitudinal studies, 4% (2/49) were platform evaluations, 2% (1/49) were randomized controlled trials, and 2% (1/49) were qualitative studies. When we examined the country of origin for each paper, we found that most of the studies were conducted in western societies, with 47% (23/49) of the papers originating from Europe and 35% (17/49) from North America. Few studies were conducted in Asia (6/49, 12%), Africa (2/49, 4%), and Oceania (1/49, 2%).

We observed an increase in the number of publications in recent years: 71% (35/49) of the papers were published within the last 5 years (2015-2020), while only 29% (14/49) of the papers were published before 2015. All the studies were either directly or indirectly associated with academic institutions. When classified based on a publication's domain, 64% (31/49) of the selected papers were published primarily in the fields of engineering and computer science, 18% (9/49) were published in biomedical engineering and health informatics journals, and 18% (9/49) were published in health-related journals (Figure 2 and Table S2 of Multimedia Appendix 1).

Table 1. Profile of the selected studies by type and human participation.

Type of study, reference	Sample size	Demographic profile of the participants (age [years], male/female)	Participant health profile
Pilot studies (n=13)			
Chen et al [14]	5	>45, 2 males, 3 females	Spinal cord injury, muscular dystrophy, multiple sclerosis, polio
Bock et al [15]	11	>18	Healthy
Fritz and Dermody [16]	10	>55	Chronic diseases
Skubic et al [17]	34	>70	Chronic diseases
Dawadi et al [18]	263	>18, 72 males, 191 females	Healthy
Choi et al [19]	37	>65, 7 males, 30 females	Chronic diseases
Clemente et al [20]	6	No data	No data
Pigini et al [21]	32	No data	Healthy and cardiac conditions
Monteriù et al [22]	13	>65	Healthy
Grgurić et al [23]	13	>65	No data
Dasios et al [24]	2	>70, 1 male, 1 female	Healthy
Marcelino et al [25]	23	>30, 11 males, 12 females	Healthy
Yu et al [26]	1	>65, 1 female	Chronic diseases
Proof of concept (n=7)			
Kim et al [27]	20	>65	Depression
Alberdi Aramendi et al [10]	29	>18	Healthy
Hassan et al [28]	0	N/A[a]	N/A
Shirali et al [29]	1	>65	No data
Jung [30]	22	>60, 10 males, 12 females	No data
Alsina-Pagès et al [31]	0	No data	N/A
Mahmoud et al [32]	1	No data	Healthy
Algorithm evaluation (n=6)			
Jakkula and Cook [33]	1	>18	Healthy
Rashidi et al [34]	40	>18	Healthy
Singla et al [35]	40	No data	Healthy
Damodaran et al [36]	0	N/A	N/A
Hamad et al [37]	19	No data	No data
Enshaeifar et al [38]	12	No data	Dementia
Proposals (n=5)			
Ros et al [39]	0	N/A	N/A
Navarro et al [40]	0	N/A	N/A
Gayathri et al [41]	0	N/A	N/A
Kwon et al [42]	150	>60, 23 males, 127 females	Healthy
Taiwo and Ezugwo [43]	0	N/A	N/A
Technical validation (n=4)			
Mora et al [44]	0	N/A	N/A
Bassoli et al [45]	0	N/A	N/A
Schlebusch [46]	10	>18, 7 males, 3 females	Healthy
Virone et al [47]	22	>45, 7 males, 15 females	Healthy

Type of study, reference	Sample size	Demographic profile of the participants (age [years], male/female)	Participant health profile
Case studies (n=4)			
Sprint et al [12]	3	>70, 3 females	Lung cancer, insomnia, leg pain
Lazarou et al [48]	4	>70, 1 male, 3 females	Amnestic, mild cognitive impairment, dementia
Hercog et al [49]	1	>60, 1 female	Healthy
Yang and Hsu [50]	0	N/A	N/A
Method evaluation (n=3)			
Yao et al [51]	0	N/A	N/A
Fleury et al [52]	13	>18	Healthy
Fiorini et al [53]	17	>18	Healthy
Longitudinal studies (n=3)			
Fritz et al [54]	11	>65	No data
Austin et al [55]	16	>70, 3 males, 13 females	Healthy
Lyons et al [56]	480	>70	No data
Platform evaluation (n=2)			
Junnila et al [57]	2	>70, 1 male, 1 female	Healthy and hip surgery rehabilitation
Lamprinakos et al [58]	207	>65	Frailty
Randomized controlled trial (n=1)			
Mora et al [1]	78	>18, 69 males, 9 females	Healthy
Qualitative study (n=1)			
Cahill et al [59]	200	No data	No data

[a]N/A: not applicable.

Figure 2. Journals of the published papers reviewed in this study.

Population Demographics

As it is common practice in computer science or engineering research to use simulated data for platform or algorithm evaluation, we first categorized the studies based on the source of their data. Approximately 78% (38/49) of the papers used data collected from human participants, and the remaining 22% (11/49) of the studies used simulated data (Table 1). The age of the participants ranged from 18 to 93 years. Of the 38 studies that utilized human participants, 63% (31/49) reported participant age, but only 33% (16/49) indicated the gender of the participants. Of those that did report gender, female participants were nearly 3 times more prevalent than male participants (425 females vs 145 males). Volunteer participants were typically students recruited from the researcher's institution or patients from memory care units and assisted-living residents. Of the papers on human participants, 79% (30/49) reported the health status of the participants.

Study Settings and Parameters

The 49 papers included in this review can be broadly divided into 2 groups: 41% (20/49) approached the use of IoT for health purposes and 59% (29/49) used IoT for technological validations. The primary research focus was recognizing human mobility patterns (Table 2; complete data in Table S3 of Multimedia Appendix 1). Study length ranged from a single day of data collection to 8 years. Data were collected primarily in real-world settings, including smart apartments or smart workplaces. One of the studies used simulated home environments [39]. If the study took place in an apartment, the number of rooms typically used was between 2 and 3. Typically, there was only a single occupant in the study location.

Table 2. Technical components of the selected studies with outcomes.

Type of study, reference	Primary focus	Outcome measure	Algorithm	Type of data
Pilot study				
Chen et al [14], Dasios et al [24], Yu et al [26]	Independent living for the older population who may or may not have chronic diseases	Activity, fall detection, indoor motion	Statistical analysis of the machine learning algorithm	Binary sensors: motion, light, temperature, humidity,
Marcelino et al [25]	e-Service provision	Physical, medical, social interaction by audio-visual communication with service providers	Qualitative and quantitative data analysis	Interview questionnaire
Proof of concept				
Alberdi Aramendi et al [10], Kim et al [27], Hassan et al [28], Shirali et al [29], Jung [30], Alsina-Pagès et al [31], Mahmoud et al [32]	From 2013 to 2020, the proof of concept improved from synthetic data to real-world data, single individual to multi-individual, but the objectives more or less—the same activity recognition, anomaly detection, pattern recognition to improve the quality of life of older individuals	Motion or presence data	Binary sensor data, machine learning algorithm-support vector machine as the typical model with many of the studies; the recent study used the parallel activity log inference algorithm	Sensor data
Algorithm evaluation				
Jakkula and Cook [33], Rashidi et al [34], Singla et al [35], Damodaran et al [36], Hamad et al [37], Enshaeifar et al [38]	All the studies tried to recognize normal activity patterns and anomaly detection	Motion or presence data, device-free solutions based on radio signals like home Wi-Fi 802.11 channel state information	Machine learning and deep learning algorithms	Passive infrared sensors
Proposal				
Ros et al [39], Navarro et al [40], Gayathri et al [41], Kwon et al [42], Taiwo and Ezugwo [43]	Activity recognition of the individual	Mobility pattern recognition	Machine learning, deep learning algorithms	Binary sensor and acoustic sensor data
Technical validation				
Mora et al [44], Bassoli et al [45], Schlebusch [46], Virone et al [47]	Active assisted living monitoring, intelligent toilet seat, differentiate regular patterns, and identify abnormalities in household activities	Passive infrared sensors, magnetic contact, bed occupancy, chair occupancy, toilet presence, fridge sensor, electrocardiogram and bioimpedance spectroscopy measurements, behavioral monitoring by presence data	Behavior explanatory models, sensor profiles, multivariate habits clusters, R-peak detection, software for automatic measurement of circadian activity deviation/circadian activity rhythms	Motion sensor data, electrocardiogram, bioimpedance spectroscopy, passive infrared sensor
Case studies				
Sprint et al [12], Lazarou et al [48], Hercog et al [49], Yang and Hsu [50]	Behavior change detection, home monitoring system, activity recognition, effective active home automation solution based on open-source home automation software, and wireless, custom-developed, Wi-Fi–based hardware	Activity change, sleep, physical activity, and activities of daily living, automatic classification of activities of daily living, system functionality	CASAS[a] middleware	Motion, light, temperature, door, motion, presence, utility usage sensors, passive infrared/current sensors
Method evalution				

Type of study, reference	Primary focus	Outcome measure	Algorithm	Type of data
Yao et al [51], Fleury et al [52], Fiorini et al [53]	Activity recognition	Automatic classification of activities of daily living	Support vector machine, unsupervised machine learning, rule-based reasoning method for activity recognition	Location, temperature, sound, postural transitions and walk periods, motion sensor, location, activity, motion
Longitudinal study				
Fritz et al [54], Austin et al [55], Lyons et al [56]	Remote monitoring of pain, loneliness	Recognize pain-associated behaviors	Machine learning algorithm, isolation forest (forest) anomaly detection algorithm, decision tree classifier, logistic regression classifier	Passive infrared–based sensor data, light, temperature, humidity
Platform evaluation				
Junnila et al [57], Lamprinakos et al [58]	Remote patient monitoring using home health or telehealth	Interoperability/adaptability, which can accommodate different types of sensors	Rule-based ontological framework, partial human monitoring is required	Passive infrared–based sensor data
Qualitative study				
Cahill et al [59]	Identify and validate the requirements for new technology enabling resident wellness and person-centered care delivery in a residential care environment	State of environment and state of care delivery, state of resident	Qualitative data analysis and machine learning algorithm	Sensor and interview data
Randomized controlled trial (secondary data analysis)				
Mora et al [1]	Internet of Things–based home monitoring for older patients with stroke	Behavioral aspects-bed/rest patterns, toilet usage, room presence, and many others	Regression framework and anomaly detection, unsupervised clustering techniques	Sensor data

[a]CASAS: Center for Advanced Studies in Adaptive Systems

Data Collection and Analysis

To determine which smart home technologies were being used, sensors were grouped into 16 main categories (Table 3): utilization of space (bed and chair occupancy, toilet, fridge, kitchen, or GPS), human vitals (blood pressure, electrocardiography, blood glucose, heart rate, or respiratory rate), and environmental sensors (light, air temperature, humidity, sound, airflow, smoke, carbon monoxide, gas, or flooding). Nearly 62% (30/49) of the studies used passive infrared sensors to report on motion detection. As motion detectors and object presence sensors primarily record binary (yes/no) data, it was unsurprising that this data type was the most reported in the studies examined. Quantitative data were reported in many papers. Audiovisual (sound, light), vital indicators (heart rate, respiratory, blood glucose, body temperature), and environmental conditions (room temperature, humidity) typically record quantitative data. Finally, several papers reported spatiotemporal data typical of GPS sensors.

As smart home data collection produces large quantities of data, data management software is frequently employed. Examining the papers, we found SQL [34,35,56,57] and MYSQL [24,25,55] were frequently used to organize the data. MATLAB and Python were used for data analysis and visualization by nearly all the studies. Various statistical methods were used for data analysis, including descriptive statistics, model building, machine learning, and deep learning. Descriptive statistics were primarily used to describe the demographic characteristics of the study participants, whereas multidomain approaches [52], longitudinal linear mixed-effect regression [55], and out-of-sample cross-validation methods [55] were used for statistical models.

As 41% (20/49) of the papers reported the use of machine learning algorithms, we sought to determine which algorithms were more commonly employed. Clustering in 5 studies [1,28,30,34,53] and Hidden Markov Model in 4 studies [23,30,34,39] were the most used in data analysis to identify a regular pattern and predict future patterns. The other algorithms used in the studies were decision tree emerging pattern [11,25,27], clustering conditional random field [37,51], context-aware reasoning [28,42], fuzzy logic [41,49], k-nearest neighbors [10,51], logistic regression classifier [51,55], AdaBoost [10], Bayes network [27], boosting model using ensemble [42], circadian activity rhythms [47], multi-Hidden Markov Model [34], multiple regression model [42], multivariate habits cluster [44], ontological modelling [41], software for automatic measurement of circadian activity deviation [47], and support vector machines [52].

Nearly 14% (7/49) of the papers used deep learning methods, which included artificial neural networks [40], activity recognition using the discontinuous varied-order sequential model [34], latent trajectory models [56], longitudinal linear mixed-effect regression recurrent neural networks [55], open pass neural networks [60], recurrent neural networks [32], and multilayer perceptron [10]. One study used mixed methods and included a thematic analysis of the quantitative data [25]. Another study used the activity discovery method [34], and yet another conducted qualitative data analysis by using a mixed methods approach [25]. Some studies used induction algorithms, behavioral monitoring systems, rapid iterative testing and evaluation [15], or QRS recognition [57] for electrocardiography.

Table 3. Types of sensors, data characteristics, and their association with health.

Sensor type	Data type	Health indicator/proxy
Motion: passive infrared sensors, radiofrequency identification, magnetic switches	Any movement within the room, door movement	Physical activity/speed/quality of physical health/sleep
Presence	Any movement within the room, indoor movement	Physical activity/gait speed/quality of physical health/sleep
Temperature	Temperature of room, temperature of stove/oven	Body temperature, health quality/activity-sleep/awake/sedentary
Light	Luminosity (lux)	Sleep/active
Sound/microphone	Noise	Sleep/active
Humidity	Indoor environment	Indoor environment
Biosensors	Fall detection	Activity/alert
Plug sensors	Appliance use: television, fridge, kitchen appliance, medicine dispenser	Activity
Body position sensors	Activity	Activity
Carbon monoxide	Indoor environment	Indoor environment
Flooding sensors	Water use/consumption	Indoor environment
Gas sensors	Use of gas in the kitchen	Indoor environment
Smoke detector	Indoor environment	Indoor environment
Pressure sensor/smart tiles/pressure pad	Bed movement, gait speed, chair movement	Sleep time/quality
Electrocardiogram patch	Heart health	Heart health
Airflow sensors	Room environment	Indoor environment
SpO_2	Oxygen saturation of blood	Heart health/lung health
Blood pressure	Heart health	Heart health
Heart rate	Heart health	Heart health
Respiratory rate	Lung health	Lung health
Blood glucose sensors	General health	Diabetes
Smart weighing scale	Body weight	Weight
Pedometer	Walking	Physical activity
Contact sensors	Usage of a phone book, cooking pot, medicine container	Activity analysis
GPS	Location	Location
Wi-Fi signal	Indoor activity	Location
Smart seismic sensor	Floor vibration	Activity analysis, including fall

Outcome Measures

All the studies reported that IoT improved the quality of care, increased participants' sense of comfort, enabled early detection, and increased participants' understanding of the impact of health events on overall health. The health indicators specifically measured through smart home technologies included fall detection [24], functional health decline/improvement [10], high-level activities of daily living/instrumental activities of daily living [34,35,48,50,59,61-63], leisure services [59], loneliness [55], medical services [17,21,30,64], patient health status [17,21,30,64], perception [58], physical activity [48], sedentary behaviors [24,62], medication adherence [62], movement patterns [29], sequence of gestures [61], sleep

[12,48,56], eating habits [10,24,57,62], situational awareness [30], social engagement [56], time spent outside the home [55], and overall well-being [24].

Limitations and Challenges in the Studies

To gain insight into future research needs in the field of smart home technologies, we extracted information pertaining to the challenges and limitations self-reported by researchers. In the 49 studies, the biggest challenge faced by the researchers was differentiating between multiple participants in a single space. The second challenge identified was the lack of technology interoperability and the ability to scale up. The third challenge identified was linked to data security and privacy. The additional challenges identified by the researchers included calibration of the sensors, cost of technology and data management, data streaming and integration, data velocity, data volume, difficulty differentiating activities, generalization of activities, and demographic discrepancies (data collected from young volunteers, while algorithms were designed for the older population). Heterogeneity, installation of the sensors, lack of patient motivation, large numbers of nodes, limited data bandwidth, limited indoor activities, malfunctioning sensors, privacy, sample size, security, service quality, user acceptance, and varying levels of data accuracy were also noted as challenges.

Discussion

Key Findings

Existing reviews on the application of smart home technology for health care are limited [2,3]. If at all present, they focus on a very specific specialty within health care, such as geriatric care [65], dementia [66,67], fall prevention [68], or pregnancy [69]. This scoping review aims to address this knowledge gap by elucidating how smart home technologies are being used for health care within the home and community. An extensive database search revealed 49 peer-reviewed publications, which met our inclusion criteria. A wide variety of sensors were used to meet the differing needs in each study. Passive infrared sensors, which report on motion detection, were the most studied smart home technology for health and report primarily binary data. Multiple studies quantified measurable health indicators (eg, heart rate, blood pressure, sleep, physical activity). Reported data were mostly organized using SQL or MYSQL. As expected, diverse data analyses and statistical methods, including machine learning and deep learning, were applied to big data analysis. Of note, although some studies were performed in home settings, none were unobtrusive or zero effort. There were often disruptions to daily routines or participants were required to log activities [70].

We recognize that there are several limitations to our study and that potentially relevant publications may have been overlooked due to the constraints in our search queries and inclusion criteria. As smart home technologies are often developed by the technology industry, not all work is likely published in peer-reviewed journals. Furthermore, our use of the query term “smart home” may have excluded relevant research settings in a community or an institution. For the purposes of this scoping review, database searches were conducted in August 2021. Due to the rapid nature of this field of research, new insights may have emerged since the initial search.

Defining a “Smart Home” for Health Care

During the preparation of this scoping review, we noted that the health care research papers lacked a concrete definition for a smart home. Based on the available evidence and the identified gaps, we propose the following definition for a smart home for health care.

A smart home for health care can be defined as a home equipped with smart sensors using Bluetooth, Wi-Fi, or similar technology, not restricted to IoT, to automate, regulate, and monitor home occupants’ physical health, mental health, and environments within the home. The focus must be on convenience, safety, and improvement of one’s quality of life, to address the needs of the individual, caregivers, and health professionals.

Sociodemographic Inequalities

The studies included in this review were predominantly performed in western societies. This bias could be due to our requirement that studies should be published in English. However, it is known that high-income nations dominate the field of smart home technology. This could be due to several factors. First, western countries are trending toward an aging population, and thus, the interest in assisted living technologies is higher [71]. Second, low- and middle-income countries are focused on reducing mortality and morbidity related to infectious diseases; therefore, their resources are not focused on the needs of an aging population [72-74]. To address global health and knowledge inequalities, researchers and funding bodies must ensure that low- and middle-income countries have the resources to benefit from health technologies. Future research should prioritize including study participants in nonwestern societies.

Computer science or engineering research often use simulated data due to budget, staffing, and time constraints. Traditional technical training does not consider health outcomes and overlooks the social determinants of health. Without health care experts as part of the research team, many are unaware of the importance of reporting the demographic characteristics of human study participants. This was reflected in our scoping review, as many of the included studies failed to report this information. Of those that did report demographics, we found that female participants were more prevalent, being nearly 3 times more likely to have been studied than male participants. This was unexpected, given that research is typically dominated by male participants [75,76]. Some potential reasons for this variance could be that women live longer [77], are more likely to live in assisted care units [78], are more likely to participate in studies [79], or have altruistic considerations [80]. Moreover, the use of simulated data despite the availability of actual data highlights the need for better access to high-quality data.

The Intersection of Health and Technology

Smart home technology is a rapidly growing interdisciplinary field at the intersection of health, information technology, and engineering [81]. Yet, our scoping review highlighted a strong bias toward publication within primarily engineering and information technology journals. Many of the papers included in this scoping review contained highly technical language, tools, and databases. However, the primary audience is the health care field. Although we acknowledge that much of the technology is in its early stages, with research focused on technical challenges (data handling, analysis, storage, security, and privacy), this finding highlights a lack of collaboration between health and technical fields. Future work must address this gap—fostering interdisciplinary research teams with a broad spectrum of skills and domain knowledge experts. The involvement of health professionals in smart home technology research will ensure that these tools are relevant and bolster their successful implementation.

Technological Challenges

Interoperability was a commonly noted challenge faced by researchers. Technology is constantly being upgraded and improved with new products continually hitting the market. As diverse companies compete to create the latest technology, interoperability becomes an issue. Because there are no standardized guidelines, companies develop their own unique protocols and architectures for handling data, which contribute to incompatibility across the IoT landscape. The result is a jungle of systems that are confusing and intimidating to navigate for many non–tech-savvy individuals. One must subscribe to a single system that may not meet all their needs, grapple with the inconvenience of systems that do not communicate seamlessly, or implement third-party software or hardware to bridge the gap. There is a need to continue to develop solutions that allow these systems to integrate and communicate with one another. Similarly, the other 2 challenges faced by the researchers were differentiating individuals within a multiparticipant household and data security and privacy. Health care technology brings a new layer of complexity due to risks associated with personally identified data, health data, privacy, data rights, and ethical considerations [82].

Data Quality

Some of the studies [18,27,55] examined here had insufficient data quality to make their research findings relevant in the health care field. In many cases, the number of study participants was minimal and lacked demographic information. The quality of many of the sensors used in a home setting is lesser than that of the instruments used in a clinical setting, often diminishing the value of the data. Additionally, some of the technologies were not diagnostic tools at all because the health indicators were not quantifiable (video or audio). Other health conditions such as loneliness or mental health cannot be quantified and thus must be measured through the integration of multiple proxy indicators. The challenges of data integration will likely be addressed with continued improvements in artificial intelligence. Here, we have highlighted the existing research on the application of smart home technology to improve health and revealed multiple gaps in our knowledge. The IoT has ushered in a period of ultraconnectivity [83], converting commercial, off-the-shelf sensors like smart Wi-Fi thermostats and wearable devices into vital sources of health data. With the collaborative efforts of technology experts and health care professionals, we have the potential to leverage these data to improve physical and mental health.

Conclusion

Smart home technology has the potential to improve the quality of life by monitoring health indicators in vulnerable persons. Despite their potential, there is still a lack of large-scale utilization of these technologies for health care. A scoping review of the existing literature enabled us to identify the types of sensors and the data being explored. The trends and knowledge gaps identified here will invite new progress in remote patient monitoring in public health. This kind of a care system can support and complement medical interventions to improve population health.

Acknowledgments

We would like to thank Adson Rocha for helping with screening, charting, and providing input throughout the process and the Ubiquitous Health Technology Lab (UbiLab) volunteers, namely, Clarisse Misola, Nadia Somani, Arjun Mehta, Chaeyoon Jeong, Thianna Edwards, Kunal Karhanis, and Harneet Dhillon, for their timely help whenever and wherever required. We would like to thank the Natural Sciences and Energy Research Council, ecobee, and The Mathematics of Information Technology and Complex Systems (MITACS) for supporting this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 30 KB - mhealth_v11i1e37347_app1.docx]

References

1. Mora N, Grossi F, Russo D, Barsocchi P, Hu R, Brunschwiler T, et al. IoT-Based Home Monitoring: Supporting Practitioners' Assessment by Behavioral Analysis. Sensors (Basel) 2019 Jul 23;19:14 [FREE Full text] [doi: 10.3390/s19143238] [Medline: 31340542]

2. Alam MR, Reaz MBI, Ali MAM. A Review of Smart Homes—Past, Present, and Future. IEEE Trans. Syst., Man, Cybern. C 2012 Nov;42(6):1190-1203. [doi: 10.1109/TSMCC.2012.2189204]
3. Alaa M, Zaidan A, Zaidan B, Talal M, Kiah M. A review of smart home applications based on Internet of Things. Journal of Network and Computer Applications 2017 Nov;97:48-65. [doi: 10.1016/j.jnca.2017.08.017]
4. Vaidya V, Vishwakarma P. A comparative analysis on smart home system to control, monitorsecure home, based on technologies like GSM, IOT. Bluetooth and PIC microcontroller with zigbee modulation. 2018 Presented at: International Conference on Smart City and Emerging Technology (ICSCET); January 5; Mumbai, India URL: https://ieeexplore.ieee.org/document/8537381 [doi: 10.1109/icscet.2018.8537381]
5. Tirado Herrero S, Nicholls L, Strengers Y. Smart home technologies in everyday life: do they address key energy challenges in households? Current Opinion in Environmental Sustainability 2018 Apr;31:65-70. [doi: 10.1016/j.cosust.2017.12.001]
6. Robles R, Kim TH. A review on security in smart home development. International Journal of Advanced Science and Technology. 2010 Feb. URL: https://www.researchgate.net/publication/228416463_A_Review_on_Security_in_Smart_Home_Development [accessed 2023-03-14]
7. Yu M, Rhuma A, Naqvi SM, Wang L, Chambers J. A posture recognition based fall detection system for monitoring an elderly person in a smart home environment. IEEE Trans Inf Technol Biomed 2012 Nov;16(6):1274-1286. [doi: 10.1109/TITB.2012.2214786] [Medline: 22922730]
8. Han J, Choi C, Park W, Lee I, Kim S. Smart home energy management system including renewable energy based on ZigBee and PLC. IEEE Trans. Consumer Electron 2014 May;60(2):198-202. [doi: 10.1109/tce.2014.6851994]
9. Gawrońska K, Lorkowski J. Smart homes for the older population: particularly important during the COVID-19 outbreak. Reumatologia 2021;59(1):41-46 [FREE Full text] [doi: 10.5114/reum.2021.103939] [Medline: 33707795]
10. Alberdi Aramendi A, Weakley A, Aztiria Goenaga A, Schmitter-Edgecombe M, Cook DJ. Automatic assessment of functional health decline in older adults based on smart home data. J Biomed Inform 2018 May;81:119-130 [FREE Full text] [doi: 10.1016/j.jbi.2018.03.009] [Medline: 29551743]
11. Sprint G, Cook D, Fritz R, Schmitter-Edgecombe M. Detecting health behavior change by analyzing smart home sensor data. 2016 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); May 18-20; St Louis, MO URL: https://ieeexplore.ieee.org/document/7501687 [doi: 10.1109/smartcomp.2016.7501687]
12. Sprint G, Cook DJ, Fritz R, Schmitter-Edgecombe M. Using Smart Homes to Detect and Analyze Health Events. Computer 2016 Nov;49(11):29-37. [doi: 10.1109/mc.2016.338]
13. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
14. Chen K, Harniss M, Patel S, Johnson K. Implementing technology-based embedded assessment in the home and community life of individuals aging with disabilities: a participatory research and development study. Disabil Rehabil Assist Technol 2014 Mar;9(2):112-120. [doi: 10.3109/17483107.2013.805824] [Medline: 23802128]
15. Bock C, Demiris G, Choi Y, Le T, Thompson HJ, Samuel A, et al. Engaging older adults in the visualization of sensor data facilitated by an open platform for connected devices. THC 2016 Jul 27;24(4):541-550. [doi: 10.3233/thc-161150]
16. Fritz RL, Dermody G. A nurse-driven method for developing artificial intelligence in "smart" homes for aging-in-place. Nurs Outlook 2019;67(2):140-153 [FREE Full text] [doi: 10.1016/j.outlook.2018.11.004] [Medline: 30551883]
17. Skubic M, Alexander G, Popescu M, Rantz M, Keller J. A smart home application to eldercare: Current status and lessons learned. THC 2009 Jul 23;17(3):183-201. [doi: 10.3233/thc-2009-0551]
18. Dawadi PN, Cook DJ, Schmitter-Edgecombe M, Parsey C. Automated assessment of cognitive health using smart home technologies. Technol Health Care 2013;21(4):323-343 [FREE Full text] [doi: 10.3233/THC-130734] [Medline: 23949177]
19. Choi YK, Thompson HJ, Demiris G. Use of an Internet-of-Things Smart Home System for Healthy Aging in Older Adults in Residential Settings: Pilot Feasibility Study. JMIR Aging 2020 Nov 10;3(2):e21964 [FREE Full text] [doi: 10.2196/21964] [Medline: 33170128]
20. Clemente J, Li F, Valero M, Song W. Smart Seismic Sensing for Indoor Fall Detection, Location, and Notification. IEEE J. Biomed. Health Inform 2020 Feb;24(2):524-532. [doi: 10.1109/jbhi.2019.2907498]
21. Pigini L, Bovi G, Panzarino C, Gower V, Ferratini M, Andreoni G, et al. Pilot Test of a New Personal Health System Integrating Environmental and Wearable Sensors for Telemonitoring and Care of Elderly People at Home (SMARTA Project). Gerontology 2017;63(3):281-286. [doi: 10.1159/000455168] [Medline: 28099965]
22. Monteriù A, Prist M, Frontoni E, Longhi S, Pietroni F, Casaccia S, et al. A Smart Sensing Architecture for Domestic Monitoring: Methodological Approach and Experimental Validation. Sensors (Basel) 2018 Jul 17;18(7):2310 [FREE Full text] [doi: 10.3390/s18072310] [Medline: 30018200]
23. Grgurić A, Mošmondor M, Huljenić D. The SmartHabits: An Intelligent Privacy-Aware Home Care Assistance System. Sensors (Basel) 2019 Feb 21;19(4):907 [FREE Full text] [doi: 10.3390/s19040907] [Medline: 30795587]
24. Dasios A, Gavalas D, Pantziou G, Konstantopoulos C. Hands-On Experiences in Deploying Cost-Effective Ambient-Assisted Living Systems. Sensors (Basel) 2015 Jun 18;15(6):14487-14512 [FREE Full text] [doi: 10.3390/s150614487] [Medline: 26094631]
25. Marcelino I, Laza R, Domingues P, Gómez-Meire S, Fdez-Riverola F, Pereira A. Active and Assisted Living Ecosystem for the Elderly. Sensors (Basel) 2018 Apr 17;18(4):1246 [FREE Full text] [doi: 10.3390/s18041246] [Medline: 29673234]

26. Yu J, An N, Hassan T, Kong Q. A Pilot Study on a Smart Home for Elders Based on Continuous In-Home Unobtrusive Monitoring Technology. HERD 2019 Jul;12(3):206-219. [doi: 10.1177/1937586719826059] [Medline: 30722699]
27. Kim J, Liu N, Tan H, Chu C. Unobtrusive Monitoring to Detect Depression for Elderly With Chronic Illnesses. IEEE Sensors J 2017 Sep;17(17):5694-5704. [doi: 10.1109/jsen.2017.2729594]
28. Hassan MK, El Desouky AI, Elghamrawy SM, Sarhan AM. Intelligent hybrid remote patient-monitoring model with cloud-based framework for knowledge discovery. Computers & Electrical Engineering 2018 Aug;70:1034-1048. [doi: 10.1016/j.compeleceng.2018.02.032]
29. Shirali M, Bayo-Monton J, Fernandez-Llatas C, Ghassemian M, Traver Salcedo V. Design and Evaluation of a Solo-Resident Smart Home Testbed for Mobility Pattern Monitoring and Behavioural Assessment. Sensors (Basel) 2020 Dec 14;20(24):1-25 [FREE Full text] [doi: 10.3390/s20247167] [Medline: 33327534]
30. Jung Y. Hybrid-Aware Model for Senior Wellness Service in Smart Home. Sensors (Basel) 2017 May 22;17(5):1182 [FREE Full text] [doi: 10.3390/s17051182] [Medline: 28531157]
31. Alsina-Pagès RM, Navarro J, Alías F, Hervás M. homeSound: Real-Time Audio Event Detection Based on High Performance Computing for Behaviour and Surveillance Remote Monitoring. Sensors (Basel) 2017 Apr 13;17(4):854 [FREE Full text] [doi: 10.3390/s17040854] [Medline: 28406459]
32. Mahmoud S, Lotfi A, Langensiepen C. Behavioural pattern identification and prediction in intelligent environments. Applied Soft Computing 2013 Apr;13(4):1813-1822. [doi: 10.1016/j.asoc.2012.12.012]
33. Jakkula V, Cook DJ. Anomaly detection using temporal data mining in a smart home environment. Methods Inf Med 2008;47(1):70-75. [doi: 10.3414/me9103] [Medline: 18213431]
34. Rashidi P, Cook DJ, Holder LB, Schmitter-Edgecombe M. Discovering Activities to Recognize and Track in a Smart Environment. IEEE Trans Knowl Data Eng 2011;23(4):527-539 [FREE Full text] [doi: 10.1109/TKDE.2010.148] [Medline: 21617742]
35. Singla G, Cook DJ, Schmitter-Edgecombe M. Recognizing independent and joint activities among multiple residents in smart environments. J Ambient Intell Humaniz Comput 2010 Mar 01;1(1):57-63 [FREE Full text] [doi: 10.1007/s12652-009-0007-1] [Medline: 20975986]
36. Damodaran N, Haruni E, Kokhkharova M, Schäfer J. Device free human activity and fall recognition using WiFi channel state information (CSI). CCF Trans. Pervasive Comp. Interact 2020 Jan 30;2(1):1-17. [doi: 10.1007/s42486-020-00027-1]
37. Hamad RA, Hidalgo AS, Bouguelia M, Estevez ME, Quero JM. Efficient Activity Recognition in Smart Homes Using Delayed Fuzzy Temporal Windows on Binary Sensors. IEEE J Biomed Health Inform 2020 Feb;24(2):387-395. [doi: 10.1109/JBHI.2019.2918412] [Medline: 31135373]
38. Enshaeifar S, Zoha A, Markides A, Skillman S, Acton ST, Elsaleh T, et al. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. PLoS One 2018;13(5):e0195605 [FREE Full text] [doi: 10.1371/journal.pone.0195605] [Medline: 29723236]
39. Ros M, Cuéllar M, Delgado M, Vila A. Online recognition of human activities and adaptation to habit changes by means of learning automata and fuzzy temporal windows. Information Sciences 2013 Jan;220:86-101. [doi: 10.1016/j.ins.2011.10.005]
40. Navarro J, Vidaña-Vila E, Alsina-Pagès RM, Hervás M. Real-Time Distributed Architecture for Remote Acoustic Elderly Monitoring in Residential-Scale Ambient Assisted Living Scenarios. Sensors (Basel) 2018 Aug 01;18(8):2492 [FREE Full text] [doi: 10.3390/s18082492] [Medline: 30071601]
41. Gayathri K, Easwarakumar K, Elias S. Fuzzy ontology based activity recognition for assistive health care using smart home. International Journal of Intelligent Information Technologies 2020;16(1):31. [doi: 10.4018/ijiit.2020010102]
42. Kwon O, Shim JM, Lim G. Single activity sensor-based ensemble analysis for health monitoring of solitary elderly people. Expert Systems with Applications 2012 Apr;39(5):5774-5783. [doi: 10.1016/j.eswa.2011.11.090]
43. Taiwo O, Ezugwu AE. Smart healthcare support for remote patient monitoring during covid-19 quarantine. Inform Med Unlocked 2020;20:100428 [FREE Full text] [doi: 10.1016/j.imu.2020.100428] [Medline: 32953970]
44. Mora N, Matrella G, Ciampolini P. Cloud-Based Behavioral Monitoring in Smart Homes. Sensors (Basel) 2018 Jun 15;18(6):1951 [FREE Full text] [doi: 10.3390/s18061951] [Medline: 29914127]
45. Bassoli M, Bianchi V, De Munari I, Ciampolini P. An IoT Approach for an AAL Wi-Fi-Based Monitoring System. IEEE Trans. Instrum. Meas 2017 Dec;66(12):3200-3209. [doi: 10.1109/tim.2017.2753458]
46. Schlebusch T. Unobtrusive Health Screening on an Intelligent Toilet Seat. Acta Polytech 2011 Jan 05;51:5. [doi: 10.14311/1454]
47. Virone G, Alwan M, Dalal S, Kell S, Turner B, Stankovic J, et al. Behavioral Patterns of Older Adults in Assisted Living. IEEE Trans. Inform. Technol. Biomed 2008 May;12(3):387-398. [doi: 10.1109/TITB.2007.904157]
48. Lazarou I, Karakostas A, Stavropoulos TG, Tsompanidis T, Meditskos G, Kompatsiaris I, et al. A Novel and Intelligent Home Monitoring System for Care Support of Elders with Cognitive Impairment. J Alzheimers Dis 2016 Oct 18;54(4):1561-1591. [doi: 10.3233/JAD-160348] [Medline: 27636843]
49. Hercog D, Sedonja D, Recek B, Truntič M, Gergič B. Smart Home Solutions Using Wi-Fi-based Hardware. Tehnički vjesnik 2020 Aug 15;27(4):1351-1358 [FREE Full text] [doi: 10.17559/tv-20190730075217]

50. Yang C, Hsu Y. Remote monitoring and assessment of daily activities in the home environment. Journal of Clinical Gerontology and Geriatrics 2012 Sep;3(3):97-104. [doi: 10.1016/j.jcgg.2012.06.002]
51. Yao L, Sheng Q, Benatallah B, Dustdar S, Wang X, Shemshadi A, et al. WITS: an IoT-endowed computational framework for activity recognition in personalized smart homes. Computing 2018 Mar 6;100(4):369-385. [doi: 10.1007/s00607-018-0603-z]
52. Fleury A, Vacher M, Noury N. SVM-based multimodal classification of activities of daily living in Health Smart Homes: sensors, algorithms, and first experimental results. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):274-283. [doi: 10.1109/TITB.2009.2037317] [Medline: 20007037]
53. Fiorini L, Cavallo F, Dario P, Eavis A, Caleb-Solly P. Unsupervised Machine Learning for Developing Personalised Behaviour Models Using Activity Data. Sensors (Basel) 2017 May 04;17(5):1034 [FREE Full text] [doi: 10.3390/s17051034] [Medline: 28471405]
54. Fritz RL, Wilson M, Dermody G, Schmitter-Edgecombe M, Cook DJ. Automated Smart Home Assessment to Support Pain Management: Multiple Methods Analysis. J Med Internet Res 2020 Nov 06;22(11):e23943 [FREE Full text] [doi: 10.2196/23943] [Medline: 33105099]
55. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A Smart-Home System to Unobtrusively and Continuously Assess Loneliness in Older Adults. IEEE J Transl Eng Health Med 2016;4:2800311 [FREE Full text] [doi: 10.1109/JTEHM.2016.2579638] [Medline: 27574577]
56. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive Computing Technologies to Continuously Assess Alzheimer's Disease Progression and Intervention Efficacy. Front Aging Neurosci 2015;7:102 [FREE Full text] [doi: 10.3389/fnagi.2015.00102] [Medline: 26113819]
57. Junnila S, Kailanto H, Merilahti J, Vainio A, Vehkaoja A, Zakrzewski M, et al. Wireless, multipurpose in-home health monitoring platform: two case trials. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):447-455. [doi: 10.1109/TITB.2009.2037615] [Medline: 20007034]
58. Lamprinakos G, Asanin S, Broden T, Prestileo A, Fursse J, Papadopoulos K, et al. An integrated remote monitoring platform towards Telehealth and Telecare services interoperability. Information Sciences 2015 Jul;308:23-37. [doi: 10.1016/j.ins.2015.02.032]
59. Cahill J, Portales R, McLoughin S, Nagan N, Henrichs B, Wetherall S. IoT/Sensor-Based Infrastructures Promoting a Sense of Home, Independent Living, Comfort and Wellness. Sensors (Basel) 2019 Jan 24;19(3):485 [FREE Full text] [doi: 10.3390/s19030485] [Medline: 30682864]
60. Amiribesheli M, Benmansour A, Bouchachia A. A review of smart homes in healthcare. J Ambient Intell Human Comput 2015 Mar 14;6(4):495-517. [doi: 10.1007/s12652-015-0270-2]
61. Zhu H, Chen H, Brown R. A sequence-to-sequence model-based deep learning approach for recognizing activity of daily living for senior care. J Biomed Inform 2018 Aug;84:148-158 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.006] [Medline: 30004019]
62. Bouchard K, Maitre J, Bertuglia C, Gaboury S. Activity Recognition in Smart Homes using UWB Radars. Procedia Computer Science 2020;170:10-17 [FREE Full text] [doi: 10.1016/j.procs.2020.03.004]
63. Fang H, Tang P, Si H. Feature Selections Using Minimal Redundancy Maximal Relevance Algorithm for Human Activity Recognition in Smart Home Environments. Journal of Healthcare Engineering 2020 Nov 27;2020:1-13. [doi: 10.1155/2020/8876782]
64. Hassan MK, El Desouky AI, Elghamrawy SM, Sarhan AM. Intelligent hybrid remote patient-monitoring model with cloud-based framework for knowledge discovery. Computers & Electrical Engineering 2018 Aug;70:1034-1048. [doi: 10.1016/j.compeleceng.2018.02.032]
65. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: A systematic review. Int J Med Inform 2016 Jul;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
66. Moyle W, Murfield J, Lion K. The effectiveness of smart home technologies to support the health outcomes of community-dwelling older adults living with dementia: A scoping review. Int J Med Inform 2021 Sep;153:104513. [doi: 10.1016/j.ijmedinf.2021.104513] [Medline: 34116363]
67. Husebo BS, Heintz HL, Berge LI, Owoyemi P, Rahman AT, Vahia IV. Sensing Technology to Monitor Behavioral and Psychological Symptoms and to Assess Treatment Response in People With Dementia. A Systematic Review. Front Pharmacol 2019;10:1699 [FREE Full text] [doi: 10.3389/fphar.2019.01699] [Medline: 32116687]
68. Chan JKY, Klainin-Yobas P, Chi Y, Gan JKE, Chow G, Wu XV. The effectiveness of e-interventions on fall, neuromuscular functions and quality of life in community-dwelling older adults: A systematic review and meta-analysis. Int J Nurs Stud 2021 Jan;113:103784. [doi: 10.1016/j.ijnurstu.2020.103784] [Medline: 33120138]
69. Bossung V, Kast K. [Smart sensors in pregnancy: Narrative review on the use of smart home technology in routine prenatal care]. Z Evid Fortbild Qual Gesundhwes 2021 Aug;164:35-43. [doi: 10.1016/j.zefq.2021.05.004] [Medline: 34215532]
70. Boger J, Young V, Hoey J, Jiancaro T, Mihailidis A. Zero-Effort Technologies: Considerations, Challenges, and Use in Health, Wellness, and Rehabilitation, Second Edition. Kentfield: Morgan & Claypool Publishers; 2018.

71. Marshall BL, Dalmer NK, Katz S, Loos E, López Gómez D, Peine A. Digitization of Aging-in-Place: An International Comparison of the Value-Framing of New Technologies. Societies 2022 Feb 25;12(2):35. [doi: 10.3390/soc12020035]
72. Santosa A, Wall S, Fottrell E, Högberg U, Byass P. The development and experience of epidemiological transition theory over four decades: a systematic review. Glob Health Action 2014;7:23574 [FREE Full text] [doi: 10.3402/gha.v7.23574] [Medline: 24848657]
73. Kuate Defo B. Demographic, epidemiological, and health transitions: are they relevant to population health patterns in Africa? Glob Health Action 2014;7:22443 [FREE Full text] [doi: 10.3402/gha.v7.22443] [Medline: 24848648]
74. Dyson T. Population and Development: The Demographic Transition. London: Zed Books; 2010.
75. Holdcroft A. Gender bias in research: how does it affect evidence based medicine? J R Soc Med 2007 Jan;100(1):2-3 [FREE Full text] [doi: 10.1177/014107680710000102] [Medline: 17197669]
76. Lamon S, Knowles O. Why are males still the default subjects in medical research? The Conversation. 2021 Oct 4. URL: http://theconversation.com/why-are-males-still-the-default-subjects-in-medical-research-167545 [accessed 2022-12-17]
77. Eskes T, Haanen C. Why do women live longer than men? Eur J Obstet Gynecol Reprod Biol 2007 Aug;133(2):126-133. [doi: 10.1016/j.ejogrb.2007.01.006] [Medline: 17324494]
78. Gruneir A, Forrester J, Camacho X, Gill SS, Bronskill SE. Gender differences in home care clients and admission to long-term care in Ontario, Canada: a population-based retrospective cohort study. BMC Geriatr 2013 May 16;13:48 [FREE Full text] [doi: 10.1186/1471-2318-13-48] [Medline: 23678949]
79. Liu KA, Mager NAD. Women's involvement in clinical trials: historical perspective and future implications. Pharm Pract (Granada) 2016;14(1):708 [FREE Full text] [doi: 10.18549/PharmPract.2016.01.708] [Medline: 27011778]
80. Lobato L, Bethony JM, Pereira FB, Grahek SL, Diemert D, Gazzinelli MF. Impact of gender on the decision to participate in a clinical trial: a cross-sectional study. BMC Public Health 2014 Nov 06;14:1156 [FREE Full text] [doi: 10.1186/1471-2458-14-1156] [Medline: 25377601]
81. Marikyan D, Papagiannidis S, Alamanos E. A systematic review of the smart home literature: A user perspective. Technological Forecasting and Social Change 2019 Jan;138:139-154. [doi: 10.1016/j.techfore.2018.08.015]
82. Guhr N, Werth O, Blacha PPH, Breitner MH. Privacy concerns in the smart home context. SN Appl. Sci 2020 Jan 21;2(2):247. [doi: 10.1007/s42452-020-2025-8]
83. IoT market size worldwide 2017-2025. Statista. URL: https://www.statista.com/statistics/976313/global-iot-market-size/ [accessed 2022-12-17]

Abbreviations

IoT: Internet of Things

Edited by L Buis; submitted 16.02.22; peer-reviewed by S Fritz, T Deserno, J Chan; comments to author 04.04.22; revised version received 29.07.22; accepted 21.02.23; published 13.04.23.

Please cite as:
Morita PP, Sahu KS, Oetomo A
Health Monitoring Using Smart Home Technologies: Scoping Review
JMIR Mhealth Uhealth 2023;11:e37347
URL: https://mhealth.jmir.org/2023/1/e37347
doi:10.2196/37347
PMID:37052984

Review

Mobile Health Self-management Support for Spinal Cord Injury: Systematic Literature Review

Renaldo M Bernard[1], PhD; Vanessa Seijas[1,2,3], MD; Micheal Davis[1,2], MSc; Anel Volkova[1,2], BSc; Nicola Diviani[1,2], PhD; Janina Lüscher[1], PhD; Carla Sabariego[1,2,3], PhD

[1]Swiss Paraplegic Research, Nottwil, Switzerland

[2]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[3]Center for Rehabilitation in Global Health Systems, World Health Organization Collaborating Center, University of Lucerne, Lucerne, Switzerland

Corresponding Author:
Renaldo M Bernard, PhD
Swiss Paraplegic Research
Guido A. Zäch-Strasse 4
Nottwil, 6207
Switzerland
Phone: 41 419396654
Email: renaldo.bernard@paraplegie.ch

Abstract

Background: Self-management plays a critical role in maintaining and improving the health of persons with spinal cord injury (SCI). Despite their potential, existing mobile health (mHealth) self-management support (SMS) tools for SCI have not been comprehensively described in terms of their characteristics and approaches. It is important to have an overview of these tools to know how best to select, further develop, and improve them.

Objective: The objective of this systematic literature review was to identify mHealth SMS tools for SCI and summarize their characteristics and approaches to offering SMS.

Methods: A systematic review of the literature published between January 2010 and March 2022 was conducted across 8 bibliographic databases. The data synthesis was guided by the self-management task taxonomy by Corbin and Strauss, the self-management skill taxonomy by Lorig and Holman, and the Practical Reviews in Self-Management Support taxonomy. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards guided the reporting.

Results: A total of 24 publications reporting on 19 mHealth SMS tools for SCI were included. These tools were introduced from 2015 onward and used various mHealth technologies and multimedia formats to provide SMS using 9 methods identified by the Practical Reviews in Self-Management Support taxonomy (eg, social support and lifestyle advice and support). The identified tools focused on common SCI self-management areas (eg, bowel, bladder, and pain management) and overlooked areas such as sexual dysfunction problems and environmental problems, including barriers in the built environment. Most tools (12/19, 63%) unexpectedly supported a single self-management task instead of all 3 tasks (ie, medical, role, and emotional management), and emotional management tasks had very little support. All self-management skills (eg, problem-solving, decision-making, and action planning) had coverage, but a single tool addressed resource use. The identified mHealth SMS tools were similar in terms of number, introduction period, geographical distribution, and technical sophistication compared with SMS tools for other chronic conditions.

Conclusions: This systematic literature review provides one of the first descriptions of mHealth SMS tools for SCI in terms of their characteristics and approaches to offering SMS. This study's findings highlight a need for increased coverage of key SMS for SCI components; adopting comparable usability, user experience, and accessibility evaluation methods; and related research to provide more detailed reporting. Future research should consider other data sources such as app stores and technology-centric bibliographic databases to complement this compilation by identifying other possibly overlooked mHealth SMS tools. A consideration of this study's findings is expected to support the selection, development, and improvement of mHealth SMS tools for SCI.

(JMIR Mhealth Uhealth 2023;11:e42679) doi:10.2196/42679

KEYWORDS

mobile phone; mobile health; mHealth; eHealth; telemedicine; telehealth; spinal cord injury; self-management; internet-based intervention; World Wide Web; systematic review

Introduction

Background

Spinal cord injury (SCI) is a complex chronic health condition that carries a high health, economic, and social burden for those affected and their families. SCI can be traumatic or nontraumatic in nature and is characterized by the loss or impairment of motor, sensory, or autonomic functions below the level of the injury [1]. Frequent health complications include pressure injury, urinary tract infections, bowel dysfunction, mental health conditions, pulmonary complications, pain, and sexual dysfunction [2,3]. The limitations in functioning caused by SCIs are largely dependent on the neurological level and severity of the injury, associated comorbidities and complications, the age of onset, available health and social care resources, and the presence of barriers or facilitators in the person's environment [1]. Similarly, wider participation in society is also made difficult without a concerted effort to pursue further education or sustainable employment, sufficient financial support, and the alleviation of comorbidities [4].

Self-management plays a critical role in maintaining and improving the health of persons with SCI [5]. It is widely understood as the ability of an individual to manage the symptoms, treatment, biopsychosocial consequences, and lifestyle changes inherent to living with a chronic health condition [6]. Corbin and Strauss [7] introduced 3 tasks, namely, medical, role, and emotional management, that describe how people with chronic health conditions manage their health. Lorig and Holman [8] described 6 key skills that support the execution of these tasks: problem-solving, decision-making, resource use, forming patient-provider partnerships, action planning, and self-tailoring. Pearce et al [9] argued that self-management is nonetheless not the sole responsibility of persons affected by chronic health conditions and proposed the Practical Reviews in Self-Management Support (PRISMS) taxonomy to highlight 14 self-management support (SMS) activities such as the provision of social support and equipment. SMS is often provided in the form of traditional institutional and paper-based options [6]. However, technology-based SMS options help overcome traditional barriers of distance, time, and high economic costs and are increasingly becoming available for SCI [10,11].

The use of technology-based SMS for chronic conditions has expanded with the widespread adoption of mobile health (mHealth) technology [10]. Compared with early desktop computer–based technologies, mHealth provides more person-centered, available, accessible, and scalable tools [12]. It introduces the use of mobile and wireless information and communications technologies, including geospatial services, movement, light and proximity sensors, and Bluetooth technology, bundled into mobile devices, apps, and wearable technologies, among other similar products, to support meeting health needs [13]. In the context of SMS, this could involve using a mobile device to receive visual, auditory, and tactile-based reminders to perform a health behavior (eg, taking medication), self-monitor health status (eg, recording vital signs), learn from web-based informational resources, and secure social support from online peer groups [9,14]. mHealth is well positioned to benefit from the high adoption rates among persons with SCI. Over 87% of participants with traumatic SCI in a 2018 study indicated that they were mobile internet users, which represented a 35% increase from 2012 [15] and a 12% higher rate than the global mobile internet subscription rate in 2019 [16]. An increase in the global user base has also been attributed to the recent pandemic [17], which is also expected to have a similar impact among persons with SCI in the last 2 years.

Nonetheless, to the best of our knowledge, the available mHealth SMS for SCI has not been comprehensively compiled. Reviews on the closest related topics have focused on accessing telerehabilitation [10], telehealth care [18], and telecounseling [19-21] outside clinical settings but have not adequately considered SMS and, with the exception of one study [10], mHealth. The latest review was also completed in early 2016, which does not account for the expected rapid increase in the development of mHealth options over the last 6 years. Therefore, it is important to have an overview of available mHealth SMS options for SCI.

Objectives

The objective of this systematic literature review was to identify and summarize the mHealth SMS tools developed for SCI. It aimed to describe their volume, features, evidence base, and reporting and recommend future directions for the development, evaluation, and reporting of these tools. Articulating data on effectiveness, gaps in coverage, usability shortcomings, and impact is expected to help patients and clinicians with selecting tools and support researchers and developers in optimizing existing tools or deciding and planning the development of new ones.

Methods

Overview

A systematic review was conducted to identify and summarize the mHealth SMS tools for SCI. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [22] and the extension for searching [23] were used to guide reporting (Multimedia Appendix 1). The inclusion of an assessment of methodological quality for other types of observational studies overlooked by the study protocol [24] was the single protocol deviation observed.

Search Strategy

MEDLINE, Academic Search Premier, LISTA, Business Source Premier, Scopus, CINAHL Complete, PsycINFO, and Web of Science Core Collection were searched using keywords for SCI and mHealth (Multimedia Appendix 2). The reference lists of included articles were also hand searched.

Eligibility Criteria

Publications were eligible for inclusion if they described an mHealth [13] SMS tool [6] for SCI. Eligible mHealth SMS tools were optimized for access from mobile devices to help accommodate the accessibility needs of people with SCI and were intended for use outside a clinical setting or not dependent on assistance from others to obtain benefits. Publications including primary research studies, books, and gray literature (eg, conference proceedings, theses, government documents, and professional publications) made available in the English language between January 2010 and March 2022 were considered. Gray literature such as commentaries and letters to the editor that were unlikely to discuss mHealth SMS tools for SCI in sufficient detail were not considered.

Eligibility Assessment

In total, 3 researchers (AV, MD, and RMB), including a health scientist, psychologist, and health technologist, were involved in screening. They attended a training workshop to help ensure consistency in screening using the web-based service Rayyan (Rayyan Systems Inc) [25] without its artificial intelligence–based features. The screeners completed a training set of 100 publications. Conflicting screening decisions (ie, *include*, *maybe*, or *exclude*) were discussed to clarify any misunderstandings. A total of 2 screeners were then randomly assigned a screening set of titles and abstracts. A third screener afterward performed a second screening of 29% (29/100) of the publications. In total, 2 screeners (AV and RMB) conducted eligibility checks on the full texts. Screening was independently conducted to reduce the risk of reviewer bias [26], and conflicting screening decisions were resolved collaboratively.

Risk-of-Bias Assessment

The same researchers who conducted the screening (RMB and AV), along with a rehabilitation physician (VS), independently evaluated the risk of bias for the included studies based on recommendations from Ma et al [27] and according to the assessment strategy shown in Textbox 1. Disagreements in evaluations were resolved collaboratively.

Textbox 1. Strategy for risk-of-bias assessment.

Experimental studies

- Revised version of the Cochrane Collaboration tool for assessing risk of bias in randomized trials [28]

Mixed methods studies

- Mixed Methods Appraisal Tool for systematic mixed studies reviews [29]

Other observational studies

- Joanna Briggs Institute Checklist for Analytical Cross-Sectional Studies [30]

Qualitative studies

- Joanna Briggs Institute Checklist for Qualitative Research [31]

Quasi-experimental studies

- Joanna Briggs Institute Checklist for Quasi-Experimental Studies [32]

Data Extraction and Synthesis

MD and RMB completed the data extraction. These researchers attended a training workshop to help ensure consistency and reliability using a web-based data extraction form. This form was discussed and modified for increased clarity. One researcher extracted data from the included publications, another reviewed and verified the extracted data, and discrepancies were resolved collaboratively. The extracted data were collated and summarized by 2 researchers (RMB and AV) using a descriptive synthesis. The analysis of frequencies, except for publication characteristics, only considered unique data extracted from publications focusing on the same mHealth option. The synthesis of evaluative information considered usability [33] and user experience [34]. Data extraction and synthesis were also guided by frameworks for self-management tasks [7] and skills [8], as detailed in Textbox 2, and support activities [9]. Aspects of SMS for SCI that were targeted by the included mHealth tools [35] were described using emergent themes.

Textbox 2. Self-management task and skill frameworks.

Self-management tasks [7]

- Medical management
  - Making health-related appointments, following treatment plans, tracking symptoms, and taking medication as directed
- Role management
  - Organizing and coordinating the various everyday roles and responsibilities related to work, family, community, and self-care and adapting these roles as needed
- Emotional management
  - Regulating and coping with emotions resulting from living with a condition in a healthy and effective manner

Self-management skills [8]

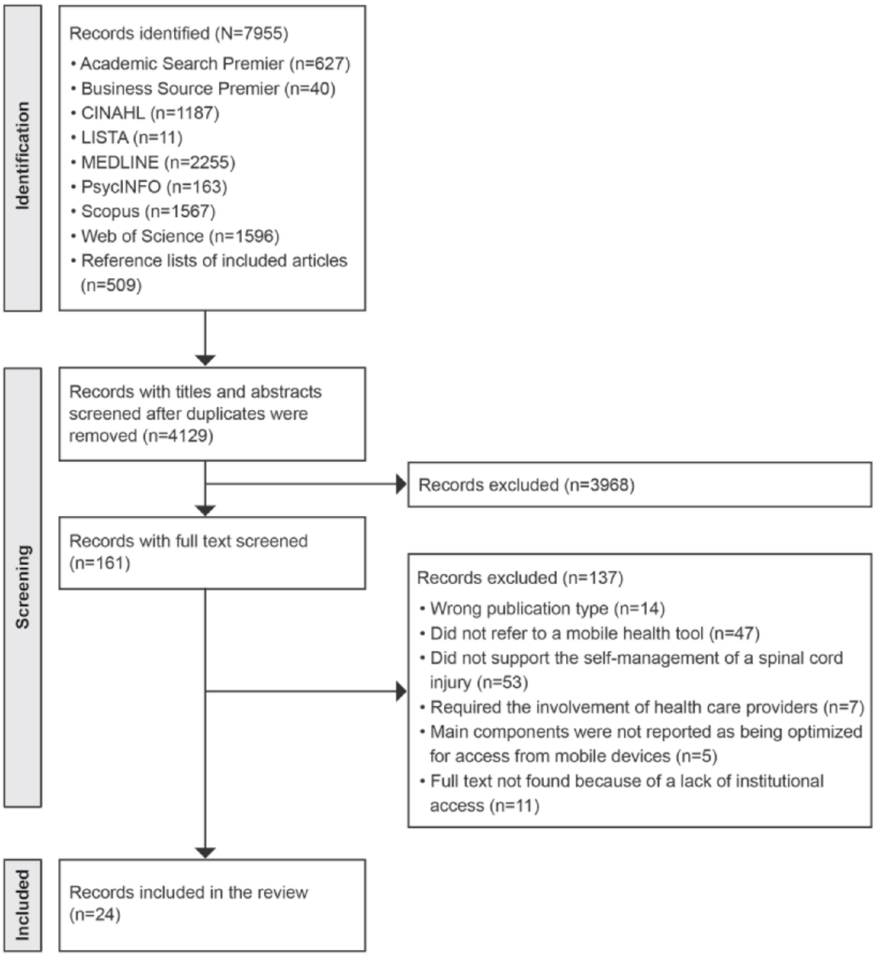
- Problem-solving
  - Identifying problems and finding, implementing, and evaluating solutions
- Decision-making
  - Weighing options and choosing the best course of action in response to changes in their condition
- Resource use
  - Finding and effectively using resources
- Forming patient-provider partnerships
  - Learning from and partnering with health care professionals to understand the patterns experienced with a condition, make informed decisions, and discuss related issues
- Action planning
  - Developing a realistic action plan that can be confidently used to achieve a set goal
- Self-tailoring
  - Developing and implementing personalized self-management strategies as needed

Results

Overview

A total of 24 publications [36-59] were included, and Figure 1 details the methodological process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the review search, selection, and inclusion process.



Characteristics of the Included Publications

The 24 included publications comprised 20 (83%) studies [36,37,40,42-57,59], 3 (12%) reports [38,39,58], and 1 (4%) protocol paper [41] (Table S1 Multimedia Appendix 3). The included publications primarily aimed to describe and develop mHealth tools (10/24, 42%) [38,39,43,45,46,48,54,56-58], evaluate implementation factors (9/24, 38%) [37,40,45,46,49,50,54,55,57], evaluate usability and user experience (7/24, 29%) [45-48,52-54] and effectiveness (6/24, 25%) [36,41,42,44,51,54], and describe stakeholder perspectives (1/24, 4%). The included publications were published between 2015 and 2022, and most (18/24, 75%) were published from 2018 onward [41-57,59]. The research teams were mainly based in North America (15/24, 62%) [38,40-42,44-47,50,52,53,55,56,58,59], followed by Europe (6/24, 25%) [36,37,48,49,54,57], Asia (2/24, 8%) [43,51], and Europe and Asia (1/24, 4%) [39].

Of the 20 included studies, 7 (35%) were quasi-experimental [37,42,43,50,51,53,57], 5 (25%) were mixed methods [46-48,54], 4 (20%) were qualitative [49,52,56,59], 2 (10%) were observational [45,55], and 2 (10%) were experimental [36,44]. The risk of bias was deemed low in just over half (11/20, 55%) of the included studies [44,45,47,49,51,52,55-57,59] (Multimedia Appendix 4 [36-59]). No studies were excluded based on the risk assessment. Study participants experienced various limitations in body functions, including musculoskeletal and movement-related functions [37,39,43], sensory functions and pain [36,47,51,59], urination and defecation [40,45,47,50,58,59], skin [38,45-50,52,53,59], sleep [54], and functions that help manage the psychological and social demands of daily life [44,45,47,54,59]. The sample sizes of the 20 included studies ranged from 4 to 75 participants where reported [36,37,40,42-57,59]. The age range of the study participants was 18 to 81 years, and most study participants were male (232/396, 59%) where reported (19/20, 95%) [36,37,40,42,44-57,59].

Characteristics of the Underlying mHealth Technology

A total of 19 mHealth tools were identified (Table 1 and Table S2 Multimedia Appendix 3). In total, 4% (1/24) of the publications focused on 2 tools [37], 12% (3/24) focused on 1 tool [41,47,50], and 2 sets of 2 focused on 1 tool equally (2/24, 8%) [40,48,49,58]. In total, 4 tools were unnamed (4/24, 17%) [38,40,43,48,49]. The included publications documented mHealth tools mainly at their testing stage (10/19, 53%) [36,40,43-45,47,50-52,54,55,58], followed by the developmental (6/19, 32%) [38,39,41,46,48,56,59], proof-of-concept (3/19, 16%) [37,42,49], proposal (1/19, 5%) [57], and launch (1/19, 5%) [53] stages.

The design and development of the included mHealth tools followed phases largely characteristic of user-centered design as the process was iterative and sought to understand users, the relevant tasks they needed to perform, and the environment of use (10/19, 53%) [36,39,40,44,46,47,50-52,54,57,59]. Participatory design, in which stakeholders are encouraged to make substantial contributions to design decisions, was used to a lesser extent (3/19, 16%) [45,48,49,56]. A total of 32% (6/19) of the publications [37,38,42,43,53,55] did not describe the design process adopted.

The primary technologies used were apps (16/19, 84%) [37-39,41-55,57,59], mobile-optimized websites (2/19, 11%) [40,56,58], and a glove (ie, wearable; 1/19, 5%) [36]. These technologies were mainly connected to mobile phones (10/19, 53%) [39,42-44,46,48,49,51,53,54,57], followed by tablets (7/19, 37%) [37,39,41,45,47,50,52,54,55], unspecified mobile devices (4/19, 21%) [36,38,40,56,58], pressure mats (2/19, 11%) [46,53], smartwatches (ie, wearable; 2/19, 11%) [42,55], smart garments (ie, wearable; 1/19, 5%) [39], and Raspberry Pi (1/19, 5%) [46].

The Android mobile operating system was the most frequently chosen (8/19, 42%) [39,41-44,48,50,53,57], closely followed by iOS (7/19, 37%) [37,38,45-47,52,53]. A total of 11% (2/19) of the tools used both operating systems [41,47,50,53], 11% (2/19) were operating system–agnostic [40,56,58], and 21% (4/19) did not report this information [36,51,54,55]. Further requirements regarding the device and operating system version and full language and country availability were largely vague or absent and could not be extracted.

When reported, devices required a display (18/19, 95%) [37-48,50-57,59], internet connectivity (12/19, 63%) [38,40,43-46,48,51,53-56], audio (6/19, 32%) [36,39,45,48,55,57], camera (5/19, 26%) [37,43-45,48], Bluetooth (5/19, 26%) [36,39,42,46,53], reminder features (5/19, 26%) [38,43,44,48,54], accelerometer sensor (4/19, 21%) [38,42,43,57], notification features (4/19, 21%) [38,39,42,46], messaging (3/19, 16%) [44,45,48], and cloud storage (1/19, 5%) [38]. Table 2 summarizes each requirement of the included mHealth tools. Multimedia Appendix 5 [36-59] organizes each requirement by self-management tasks, skills, and support components and tasks.

Table 1. Number of mobile health (mHealth) tools introduced per year (n=19).

Year of introduction	mHealth tools, n (%)
2015	3 (16)
2016	2 (11)
2017	0 (0)
2018	0 (0)
2019	7 (37)
2020	3 (16)
2021	3 (16)
2022	1 (5)

Table 2. Device requirements of the included mobile health tools (n=19).

Device requirement and citation	Description of use	Frequency, n (%)
Display [37-59]	Used for presenting the mobile device's user interface in visual and tactile form	18 (95)
Internet connectivity [38,40,43-46,48,49,51,53-56,58]	Used for accessing web-based information, having voice and video calls, and sending and storing information via the web	12 (63)
Audio [36,39,45,48,49,55,57]	Used for listening to multimedia content with sound, creating audio messages, and having voice calls	6 (32)
Bluetooth [36,39,42,46,53]	Used for exchanging data over short distances between mobile devices and paired technologies	5 (26)
Camera [37,43-45,48,49]	Used for having video calls and capturing still images	5 (26)
Reminders [38,43,44,48,49,54]	Used for alerting users to participate in a planned activity	5 (26)
Accelerometer [38,42,43,57]	Embedded in a smartphone or wearable (eg, smartwatch) and used for motion sensing	4 (21)
Notifications [38,39,42,46]	Used for informing users of available mobile technology information updates via audio, visual, and tactile indicators	4 (21)
Messaging [44,45,48,49]	Used for multimedia communication via the internet	3 (16)
Cloud storage [38]	Used for data storage	1 (5)

Characteristics of Approaches Providing SMS for SCI

The mHealth tools supported the completion of all self-management tasks (Table 3). Emotional management had little support (3/19, 16%) [44,47,54,59] compared with medical (14/19, 74%) [36-39,41,43,44,46-50,53-56,58,59] and role (12/19, 63%) [38,40-42,45,47-52,54,56-59] management tasks. Most mHealth tools supported 1 self-management task (12/19, 63%) [36,37,39,42,43,45,46,51-53,55,57], followed by 26% (5/19) supporting 2 self-management tasks [38,40,44,48,49,56,58] and 11% (2/19) supporting 3 self-management tasks [41,47,50,54,59].

The mHealth tools supported the practice of all self-management skills (Table 4). The top 4 self-management skills were supported more than the average number of times. These 4 represented 84% (31/37) of the total number of times that self-management skills were supported. Most mHealth tools (7/19, 37%) supported 1 self-management skill [41,44,47-51,56,57,59], followed by 32% (6/19) supporting 2 self-management tasks [38,40,43,45,54,55,58] and 3 self-management tasks [36,37,39,42,46,52,53].

The mHealth tools incorporated 64% (9/14) of the PRISMS support components (Table 5). The top 4 components were incorporated more than the average number of times. These 4 represented 74% (35/47) of the total number of times that the components were incorporated. Most mHealth tools (8/19, 42%) incorporated 1 support component [36,37,39,42,46,52,53,55], followed by 21% (4/19) incorporating 4 support components [40,44,51,54,58], 16% (3/19) incorporating 3 support components [38,45,48,49], and 11% (2/19) incorporating 2 support components [43,57] and 5 support components [41,47,50,56,59]. These 10 components largely focused on supporting pressure injury prevention, physical activity promotion, and bladder management (Table 3). The lowest focus was placed on spasticity management, autonomic dysreflexia management, sleep management, and shoulder posture monitoring (Table 6).

The adopted self-management approaches were individualized only or combined with a group-based approach. Individualized approaches included multimedia educational content (eg, audio, text, images, and video), real-time behavioral visualizations or illustrations, textual or haptic feedback, personalized physical movement plans, games, 2-way messaging with health care professionals, content requiring active end-user engagement (eg, diary), and progress-tracking features (16/19, 84%) [36-39,41-53,55,57,59]. Combined approaches included forums and progress-tracking leaderboards (3/19, 16%) [40,54,56,58].

The included mHealth tools were intended mostly for use in nonclinical (ie, home and community environment) settings (17/19, 89%) [37-39,42-46,48,51-58], and only 11% (2/19) were also intended for use in clinical settings [36,41,47,50,59]. The adopted approaches largely relied on research (14/19, 74%) [36,42,44-48,51,52,54,55,58], followed by theory (5/19, 26%) [38,39,41,43,56] and expertise (1/19, 5%) [57]. Approaches targeting therapeutic exercise for legs [43] and shoulder posture monitoring [39] solely relied on theory. The provision of training and practice for everyday activities that targeted therapeutic exercise for the hands was the only approach that was without an app [36].

Table 3. Characteristics of approaches providing self-management support for spinal cord injury (SCI; n=19).

mHealth tool name, citation, and country availability	Self-management focus areas	Relevant self-management tasks	Relevant self-management skills	Relevant self-management support components
AW-Shift[a] [53], United States	Pressure injury management	Medical management	Decision-making	Monitoring of the condition with feedback
Ball Strike, Pop Flux [37], Italy	Therapeutic exercise for hands, legs, or trunk	Medical management	Action planning	Training and rehearsal for practical self-management activities
CMAP[c] [46], United States	Pressure injury management	Medical management	Problem-solving	Monitoring of the condition with feedback
Fisiofriend [57], Italy	Physical activity promotion	Role management	Maintaining patient-provider partnership, action planning, and self-tailoring	Training and rehearsal for practical self-management activities and monitoring of the condition with feedback
iMHere[d] [44], United States	Bladder management, pressure injury management, and psychosocial support	Medical and emotional management	Maintaining patient-provider partnership, self-tailoring, and decision-making	Practical support with adherence (medication or behavioral), information about the condition or its management, provision of easy access to advice or support when needed, and monitoring of the condition with feedback
MMT[e] [36], Turkey	Therapeutic exercise for hands, legs, or trunk	Medical management	Problem-solving	Training and rehearsal for everyday activities
M2M[f] [55], United States	Physical activity promotion	Medical management	Self-tailoring and problem-solving	Information about the condition or its management
NR [38], United States	Pressure injury management	Medical and role management	Decision-making and resource use	Practical support with adherence (medication or behavioral), information about the condition or its management, and monitoring of the condition with feedback
NR [40,58], United States	Bladder management	Role and medical management	Maintaining patient-provider partnership and problem-solving	DPractical support with adherence (medication or behavioral), information about the condition or its management, provision of easy access to advice or support when needed, and social support
NR [43], Thailand	Therapeutic exercise for hands, legs, or trunk	Medical management	Self-tailoring and decision-making	Practical support with adherence (medication or behavioral) and training and rehearsal for practical self-management activities
NR [48,49], Switzerland	Pressure injury management	Medical and role management	Maintaining patient-provider partnership, self-tailoring, and problem-solving	Practical support with adherence (medication or behavioral), information about the condition or its management, and provision of easy access to advice or support when needed
PHOENIX[h] [45], United States	Pressure injury, bladder, and bowel management	Role management	Action planning and problem-solving	Information about the condition or its management, lifestyle advice and support, and social support
PHIRE[i] [42], United States	Physical activity promotion	Role management	Self-tailoring	Monitoring of the condition with feedback

mHealth tool name, citation, and country availability	Self-management focus areas	Relevant self-management tasks	Relevant self-management skills	Relevant self-management support components
PUT[j] [52], Canada	Pressure injury management	Role management	Problem-solving	Information about the condition or its management
Punsook [51], Thailand	Bladder and pain management	Role management	Maintaining patient-provider partnership, action planning, and problem-solving	Practical support with adherence (medication or behavioral), information about the condition or its management, monitoring of the condition with feedback, and provision of easy access to advice or support when needed
SCI Health Storylines [41,47,50,59], Canada	Bladder management, bowel management, pressure injury management, spasticity management, autonomic dysreflexia management, physical activity promotion, pain management, psychosocial support, medicating and dieting, sensation of pain, handling stress and other psychological demands, and looking after one's health	Medical, role, and emotional management	Action planning, decision-making, and self-tailoring	Practical support with adherence (medication or behavioral), information about the condition or its management, training and rehearsal for practical self-management activities, monitoring of the condition with feedback, and training and rehearsal for psychological strategies
WHEELS[k] [54], the Netherlands	Physical activity promotion, psychosocial support, sleep management, and medicating and dieting	Medical, role, and emotional management	Problem-solving and action planning	Practical support with adherence (medication or behavioral), information about the condition or its management, social support, and training and rehearsal for practical self-management activities
WOWii[l] [56], United States	Physical activity promotion	Medical and role management	Problem-solving, action planning, and decision-making	Practical support with adherence (medication or behavioral), information about the condition or its management, social support, lifestyle advice and support, and training and rehearsal for practical self-management activities

[a]mHealth: mobile health.

[b]AW-Shift: Assisted Weight Shift.

[c]CMAP: Comprehensive Mobile Assessment of Pressure.

[d]iMHere: Interactive Mobile Health and Rehabilitation.

[e]MMT: Mobile Music Touch.

[f]M2M: Movement-to-Music.

[g]NR: not reported.

[h]PHOENIX: Peer-Supported Health Outreach, Education, and Information Exchange.

[i]PHIRE: Personal Health Informatics and Rehabilitation Engineering.

[j]PUT: Pressure Ulcer Target.

[k]WHEELS: Wheelchair Exercise and Lifestyle Study.

[l]WOWii: Workout on Wheels internet intervention.

Table 4. Supported self-management skills (n=19).

Self-management skills	Frequency, n (%)
Problem-solving [36,40,45, 46, 48, 49, 51, 52, 54-56, 58]	10 (53)
Decision-making [38,39,41, 43, 44, 50, 53, 56, 59]	7 (37)
Self-tailoring [42-44,47-49, 55, 57]	7 (37)
Action planning [37,41,45,47,50, 51, 54, 56, 57, 59]	7 (37)
Maintaining patient-provider partnership [40,44,48,49,51, 57, 58]	5 (26)
Resource use [38]	1 (5)

Table 5. Incorporated self-management support components (n=19).

Self-management support components	Frequency, n (%)
Information about the condition, its management, or both [38,40,41,44,45,47-49,51,52,54-56,58,59]	11 (58)
Practical support with adherence (medication or behavioral) [38-41,43,44,47-49,51,54,56,58,59]	10 (53)
Monitoring of the condition with feedback [38,41,42,44,46,47,51,53,57,59]	8 (42)
Training or rehearsal for practical self-management activities [37,41,43,47,50,54,56,57,59]	6 (32)
Provision of easy access to advice or support when needed [40,44,48,49,51,58]	4 (21)
Social support [40,45,54,56,58]	4 (21)
Lifestyle advice and support [45,56]	2 (11)
Training or rehearsal for everyday activities [36]	1 (5)
Training or rehearsal for psychological strategies [47,59]	1 (5)

Table 6. Targeted self-management focus areas (n=19).

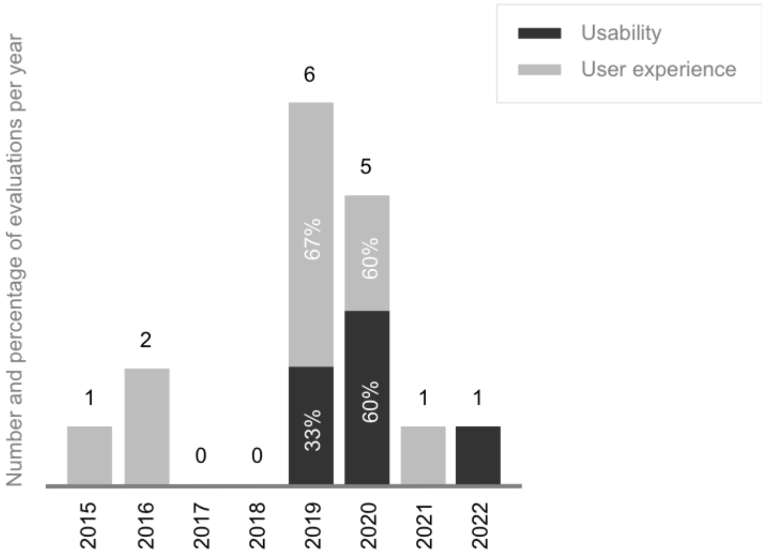
Self-management focus areas	Frequency, n (%)
Pressure injury management [38,41,44-50,52,53,59]	8 (42)
Physical activity promotion [41,42,50,54-57]	6 (32)
Bladder management [40,41,44,45,47,50,51,58,59]	5 (26)
Psychosocial support [44,47,54,59]	3 (16)
Therapeutic exercise for hands, legs, or trunk [36,37,43]	3 (16)
Bowel management [41,45,47,50,59]	2 (11)
Pain management [47,50,51,59]	2 (11)
Medicating and dieting [47,54,59]	2 (11)
Spasticity management [41,47]	1 (5)
Autonomic dysreflexia management [41]	1 (5)
Sleep management [54]	1 (5)
Shoulder posture monitoring [39]	1 (5)

Evaluation of mHealth SMS for SCI

The included studies reported a significant change in trunk control [37], urinary tract infections [44], hand sensory functions [36], self-management for neurogenic bladder dysfunction [40], and bowel management confidence [50] (Table 1). No significant changes in urinary tract leakage and infections or pain [51] and psychosocial-related outcomes [44] were observed. None of the included studies published before 2019 (6/20, 30%) had the primary aim of evaluating usability or user experience. The included studies that conducted these evaluations increased 4-fold during the last 4 years (15/18, 83%) [45-48,51,54,56] compared with the previous period (3/18, 17%) [37,39,40] (Figure 2). When evaluated by 58% (14/24) of the included publications [37,40,43,45-49,51,52,54,56,57,59], interviews, focus groups, surveys, and field studies were used. A widely adopted instrument (eg, the System Usability Scale; 4/14, 29%) [46-48,54] was seldom used for evaluating usability or user experience (Table 3). All evaluations relied on empirical methods involving participants with SCI. The results from the usability evaluations were largely *very good* (6/12, 50%) [37,40,45-48,51], followed by those indicating *good* (3/12, 25%)

[46-48] and *poor* (1/12, 8%) [54] usability. The user experience evaluations were mostly *good* (3/4, 75%) [43,46,57] and *very good* (1/4, 25%) [51]. Other evaluations of usability [56,59] and user experience [49,52] generated 12 change requests from study participants regarding design, content, and functionality. No accessibility evaluations were reported.

Figure 2. Temporal distribution of mobile health tool evaluation.



Discussion

Principal Findings and Comparison With Prior Work

The 24 included publications introduced 19 mHealth SMS tools for SCI since 2015 using various mHealth technologies and multimedia formats. The findings support the notion that the adoption of mHealth SMS tools for SCI is a growing area of interest [10]. The findings are similar to those of reviews of a comparable period identifying 23 heart failure– [60]; 21 cardiovascular disease– [61]; 23 HIV-, AIDS-, or HIV and AIDS– [62]; and 17 Parkinson disease–related [63] mHealth SMS tools that were introduced from 2012 onward. The geographical distribution of the tools in the included publications is also similar to that in these reviews, with the large majority of tools being introduced in North America and Europe except in the case of HIV or AIDS [62], where no tools introduced in Europe were reported.

mHealth Technologies Underlying SMS for SCI

A review of mHealth SMS tools for heart failure [60] reported that 48% of the identified tools benefited from a participatory design approach compared with 16% (3/19) in this review. However, as evidenced in the study by Allin et al [64], adopting participatory design for the development of mHealth SMS tools for SCI is instrumental in highlighting accessibility, design, and information quality concerns and developing potential solutions in response. Nonetheless, research also highlights that a participatory design approach does not guarantee sustained engagement [64]. The remaining reviews did not report the approach adopted for designing the identified tools, which is similar to 32% (6/19) of the tools identified in this review. Unlike comparable reviews [61,62], our review found a nearly equal operating system share between Android and iOS despite the former being the most used by a large margin [65]. Unlike comparable reviews [61,63] except for that by Mehraeen et al [62], widely used wearable devices (eg, smartwatches) were not identified in this review despite their demonstrated potential for health self-management. This likely results partially from the difficulty to accurately measure physical activity in persons with SCI using wearables [66]. The identified device requirements are also common features found in many mobile devices, which allowed for the easy adoption of these tools and which were also present in and similarly used by the tools identified in comparable reviews.

Approaches to SMS for SCI

The lack of support for all 3 self-management tasks (ie, medical, role, and emotional) was unexpected as SMS tools should aim to include content that addresses all of them [8]. The lack of support for emotional management was unexpectedly pronounced given that managing the psychological demands of SCI and other chronic conditions is a core task for those affected [67]. Compared with the 10 most common problems reported by persons with SCI [68,69], the included mHealth tools largely addressed similar problems but prioritized them differently. For example, pressure injury was ranked much lower by people with SCI than the high level of coverage this complication had in the included mHealth tools. Moreover, the included mHealth tools did not address environmental problems such as barriers in the built environment and sexual dysfunction problems. Compared with non–mHealth-based self-management interventions identified in a recent scoping review [70], the identified mHealth tools reflected a very similar level of focus on SCI symptoms. Except for sexual functions, the identified mHealth tools considered many additional symptoms and health self-management options in comparison. The interventions and mHealth tools identified in this review similarly ascertained the provision of *information about the condition and/or its management* as being the most common PRISMS support component offered. However, the findings from the recent scoping review and this review differ in their coverage of the remaining PRISMS components and self-management skills. This is likely partially due to mHealth being more suited for offering *practical support with adherence*, *monitoring of the condition with feedback*, and *provision of easy access to advice or support when needed*, for example, compared with alternative methods (eg, paper- and desktop computer–based options). The recent scoping review [70] also found that most self-management interventions following an SCI were individualized or combined with a group-based approach. The identified mHealth tools used more sophisticated but comparable formats with alternative self-management interventions [70].

Evaluation of mHealth SMS Tools for SCI

mHealth tools require a high level of usability to ensure that they can be easily used over time without expending unwarranted effort. Comparable reviews have reported a lower percentage of usability evaluations for hypertension (2/21, 10%) [71], diabetes (14/31, 45%) [72], and heart failure (9/18, 50%) [73] than that reported by the included studies. Although the usability of the included tools was generally ranked positively, the failure to use standardized measurement instruments makes it difficult to ascertain what exactly was measured and compare with findings from similar studies. Comparable reviews have reported a slightly higher adoption of standardized instruments for usability evaluations for diabetes (6/31, 19%) [72] and heart failure (4/18, 22%) [73]. Comparable studies investigating user experience of self-management tools were few [74,75], similarly revealed positive results [74], did not adopt widely used assessment instruments [74], and benefited from qualitative methods to gain insights into improvements [74,75].

Implications for Future Practice and Research

More effort is needed to develop mHealth SMS tools for SCI with consideration for incorporating all self-management tasks and undersupported self-management skills and support components. New approaches that can bridge the observed fragmentation of SMS provided by mHealth tools for SCI should be pursued. For example, mHealth SMS tools for chronic health conditions share several common features, and a reference architecture could be of benefit for the efficient and cost-effective development of mHealth SMS tools for SCI, other chronic health conditions, or a combination of these. These technologies are shaped by their underlying technical frameworks as much as by their features. Decisions regarding the design, development, and implementation of mHealth tools need to be reported in detail and investigated to inform future decision-making regarding mHealth tools. Usability and user experience evaluations should use commonly adopted instruments, including the System Usability Scale [76]; the Usefulness, Satisfaction, and Ease of Use Questionnaire [77]; and the Post-Study System Usability Questionnaire [78], to enhance the validity of evaluations and comparability of findings. Furthermore, empirical methods such as usability testing with users should be complemented by other methods [79], including expert inspections and automated evaluations, to improve the validity of these evaluations. Considering and reporting the supported level of functioning by an mHealth tool is essential given the considerable accessibility needs of people with SCI (eg, difficulties associated with sensory and motor impairments). Similar reviews should include more technology-centric databases, for example, the one from the Institute of Electrical and Electronics Engineers, in their search strategy. A systematic search of the most used app stores can complement this review's findings by identifying and evaluating SMS apps for SCI that are available to the public.

Limitations

The included publications were unlikely to account for all available mHealth SMS tools for SCI. Furthermore, one of the identified apps was retired from the Apple App Store (ie, Assisted Weight Shift) [53], another was retired from the Google Play Store (ie, Pop Flux) [37], and a single app was available from both digital distribution platforms (eg, Interactive Mobile Health and Rehabilitation) [44]. Nonetheless, this systematic literature review is necessary to comprehensively account for these tools. The mHealth tools were also insufficiently described by the included studies, and this prevented a deeper evaluation. For example, despite notable differences in the cost and features of mobile devices using the Android and iOS operating systems, it was difficult to understand how the operating system was chosen without a rationale being provided, especially when their adoption rates were almost the same. Information about the intervention, such as its name; details about primary and secondary users, including lesion type and injury etiology; the design process followed; and minimum hardware and software requirements, was vaguely reported or absent and could have provided valuable insight. For example, it might have indicated a fuller coverage of self-management tasks. This inadequate reporting might also reflect publication restrictions regarding word limits and alternative focus topics where authors instead strategically prioritize other details. Despite these shortcomings in reporting, the included studies still provided more relevant details than tools identified via other means, such as app store descriptions. The publication year restriction could have excluded otherwise eligible mHealth tools, but the findings from this study and the latest review on a related topic [10] strongly suggest that very few or no tools would have been missed as a result. Only considering mobile-optimized web-based services for inclusion likely reduced the number of web-based mHealth tools included, but it is an essential feature given the accessibility needs of people with SCIs. Usability and user experience evaluations were limited as they relied on empirical evaluations, which typically focus on testing select system tasks with users instead of all possible tasks. However, the focus is often on essential tasks, and the practice reduces costs such as time, money, and effort to conduct the evaluation [79].

Conclusions

This systematic literature review provides one of the first overviews of mHealth SMS tools for SCI and represents one of the first steps in a wider research agenda aiming to comprehensively account for these tools. This review identified 19 mHealth tools reported across the 24 included publications and an increasing development trend. A synthesis of these findings highlighted the need for mHealth to support key underserviced SMS components for SCI, more standardized or commonly used evaluation methods for usability and user experience, and more detailed reporting that includes key technical details and decisions that shape the mHealth tool. Future research is encouraged to consider other sources for the identification of mHealth SMS tools for SCI, such as app stores and more technology-centric bibliographic databases, to complement this compilation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 26 KB - mhealth_v11i1e42679_app1.docx]

Multimedia Appendix 2
Search concepts and terms and search strategies for the queried bibliographic databases.
[DOCX File , 55 KB - mhealth_v11i1e42679_app2.docx]

Multimedia Appendix 3
Characteristics of the included publications (N=24) and included mobile health (mHealth) tools (n=19).
[DOCX File , 40 KB - mhealth_v11i1e42679_app3.docx]

Multimedia Appendix 4
Risk-of-bias assessment of the included studies.
[XLSX File (Microsoft Excel File), 102 KB - mhealth_v11i1e42679_app4.xlsx]

Multimedia Appendix 5
Device requirements by self-management area, skills, and support methods.
[DOCX File , 358 KB - mhealth_v11i1e42679_app5.docx]

References

1. International Perspectives on Spinal Cord Injury. Geneva: World Health Organization; 2013.
2. Pilusa SI, Myezwa H, Potterton J. Experiences of secondary health conditions amongst people with spinal cord injury in South Africa: a qualitative study. S Afr J Physiother 2021 Apr 06;77(1):1530 [FREE Full text] [doi: 10.4102/sajp.v77i1.1530] [Medline: 33937547]
3. Wahman K, Nilsson Wikmar L, Chlaidze G, Joseph C. Secondary medical complications after traumatic spinal cord injury in Stockholm, Sweden: towards developing prevention strategies. J Rehabil Med 2019 Jul 08;51(7):513-517 [FREE Full text] [doi: 10.2340/16501977-2568] [Medline: 31243468]
4. Gross-Hemmi MH, Fekete C, Post MW, Scheel-Sailer A, Schwegler U, Brinkhof MW. Detecting subgroups in social participation among individuals living with spinal cord injury: a longitudinal analysis of community survey data. Spinal Cord 2021 Apr;59(4):419-428. [doi: 10.1038/s41393-020-00576-3] [Medline: 33446936]
5. Conti A, Clari M, Kangasniemi M, Martin B, Borraccino A, Campagna S. What self-care behaviours are essential for people with spinal cord injury? A systematic review and meta-synthesis. Disabil Rehabil 2022 Apr;44(7):991-1006. [doi: 10.1080/09638288.2020.1783703] [Medline: 32602384]
6. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
7. Corbin A, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. San Francisco, California: Jossey-Bass; 1988.
8. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
9. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CL, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82 [FREE Full text] [doi: 10.1177/1355819615602725] [Medline: 26377727]
10. Wellbeloved-Stone CA, Weppner JL, Valdez RS. A systematic review of telerehabilitation and mHealth interventions for spinal cord injury. Curr Phys Med Rehabil Rep 2016 Nov 30;4(4):295-311. [doi: 10.1007/s40141-016-0138-1]
11. Tung JY, Stead B, Mann W, Ba'Pham, Popovic MR. Assistive technologies for self-managed pressure ulcer prevention in spinal cord injury: a scoping review. J Rehabil Res Dev 2015;52(2):131-146 [FREE Full text] [doi: 10.1682/JRRD.2014.02.0064] [Medline: 26237111]
12. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2022-03-17]

13. MHealth New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
14. World Health Organization. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/260480 [accessed 2022-03-17]
15. Rigot SK, Worobey LA, Boninger ML, Robinson-Whelen S, Roach MJ, Heinemann AW, et al. Changes in internet use over time among individuals with traumatic spinal cord injury. Arch Phys Med Rehabil 2022 Apr;103(4):832-9.e2. [doi: 10.1016/j.apmr.2021.04.021] [Medline: 34062118]
16. Digital trends in Europe 2021 - ICT trends and developments in Europe, 2017-2020. International Telecommunication Union. 2021. URL: https://www.itu.int/pub/D-IND-DIG_TRENDS_EUR.01-2021 [accessed 2022-03-17]
17. 2.9 billion people still offline New data from ITU suggest ‘COVID connectivity boost’ – but world’s poorest being left far behind. International Telecommunication Union. 2021. URL: https://www.itu.int/en/mediacentre/Pages/PR-2021-11-29-FactsFigures.aspx [accessed 2022-05-12]
18. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]
19. Dorstyn D, Mathias J, Denson L. Applications of telecounselling in spinal cord injury rehabilitation: a systematic review with effect sizes. Clin Rehabil 2013 Dec;27(12):1072-1083. [doi: 10.1177/0269215513488001] [Medline: 23823709]
20. Dorstyn DS, Mathias JL, Denson LA. Psychosocial outcomes of telephone-based counseling for adults with an acquired physical disability: a meta-analysis. Rehabil Psychol 2011 Feb;56(1):1-14. [doi: 10.1037/a0022249] [Medline: 21401281]
21. Dorstyn D. Effectiveness of cognitive behaviour therapy and telecounselling for the treatment of psychological problems following spinal cord injury. University of Adelaide. 2012. URL: https://digital.library.adelaide.edu.au/dspace/bitstream/2440/75356/8/02whole.pdf [accessed 2022-03-17]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
24. Bernard R, Davis M, Volkova A, Sabariego C. Mobile health technologies for self-management in spinal cord injuries: a systematic literature review protocol. OSF Registries. 2022 Apr 22. URL: https://osf.io/nwh63 [accessed 2022-04-22]
25. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
26. Denison HJ, Dodds RM, Ntani G, Cooper R, Cooper C, Sayer AA, et al. How to get started with a systematic review in epidemiology: an introductory guide for early career researchers. Arch Public Health 2013 Aug 07;71(1):21 [FREE Full text] [doi: 10.1186/0778-7367-71-21] [Medline: 23919540]
27. Ma L, Wang Y, Yang Z, Huang D, Weng H, Zeng X. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020 Feb 29;7(1):7 [FREE Full text] [doi: 10.1186/s40779-020-00238-8] [Medline: 32111253]
28. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
29. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inform 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
30. Checklist for analytical cross sectional studies. Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf [accessed 2022-03-17]
31. Checklist for qualitative research. Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf [accessed 2022-03-17]
32. Checklist for quasi-experimental studies (non-randomized experimental studies). Joanna Briggs Institute. 2017. URL: https://jbi.global/sites/default/files/2019-05/JBI_Quasi-Experimental_Appraisal_Tool2017_0.pdf [accessed 2022-03-17]
33. BS EN ISO 9241-11:2018 Ergonomics of human-system interaction. Usability: Definitions and concepts. British Standards Institution. 2018. URL: https://landingpage.bsigroup.com/LandingPage/Standard?UPI=000000000030319071 [accessed 2022-03-17]
34. British Standards Document BS EN ISO 9241-210 Ergonomics of human-system interaction. Human-centred design for interactive systems. British Standards Institution. 2019. URL: https://landingpage.bsigroup.com/LandingPage/Undated?UPI=000000000030388991 [accessed 2022-03-17]
35. Boldt C, Brach M, Grill E, Berthou A, Meister K, Scheuringer M, et al. The ICF categories identified in nursing interventions administered to neurological patients with post-acute rehabilitation needs. Disabil Rehabil 2005;27(7-8):431-436. [doi: 10.1080/09638280400014071] [Medline: 16040546]

36. Estes LT, Backus D, Starner T. A wearable vibration glove for improving hand sensation in persons with spinal cord injury using passive haptic rehabilitation. In: Proceedings of the 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth). 2015 Presented at: 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); May 20-23, 2015; Istanbul, Turkey. [doi: 10.4108/icst.pervasivehealth.2015.259137]
37. Fizzotti G, Rognoni C, Imarisio A, Meneghini A, Pistarini C, Quaglini S. Tablet technology for rehabilitation after spinal cord injury: a proof-of-concept. Stud Health Technol Inform 2015;210:479-483. [Medline: 25991193]
38. Khan A, Phung N. Undergraduate research in assistive technology: design and development of a preventive weight shifting app to reduce the risk of pressure ulcers in wheelchair bound patients with spinal cord injuries (phase 1) 1. J Med Device 2016;10(2):020927. [doi: 10.1115/1.4033277]
39. Wang Q, Toeters M, Chen W, Timmermans A, Markopoulos P. Zishi: a smart garment for posture monitoring. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose, California, USA.
40. Wilde MH, McMahon JM, Fairbanks E, Brasch J, Parshall R, Zhang F, et al. Feasibility of a web-based self-management intervention for intermittent urinary catheter users with spinal cord injury. J Wound Ostomy Continence Nurs 2016;43(5):529-538 [FREE Full text] [doi: 10.1097/WON.0000000000000256] [Medline: 27488740]
41. Mortenson WB, Mills PB, Adams J, Singh G, MacGillivray M, Sawatzky B. Improving self-management skills among people with spinal cord injury: protocol for a mixed-methods study. JMIR Res Protoc 2018 Nov 14;7(11):e11069 [FREE Full text] [doi: 10.2196/11069] [Medline: 30429114]
42. Hiremath SV, Amiri AM, Thapa-Chhetry B, Snethen G, Schmidt-Read M, Ramos-Lamboy M, et al. Mobile health-based physical activity intervention for individuals with spinal cord injury in the community: a pilot study. PLoS One 2019;14(10):e0223762 [FREE Full text] [doi: 10.1371/journal.pone.0223762] [Medline: 31613909]
43. Kongcharoen J, Pruitikanee S, Puttinaovarat S, Tubtiang Y, Chankeaw P. Gamification smartphone application for leg physical therapy. Int J Onl Eng 2019 May 14;15(08):31. [doi: 10.3991/ijoe.v15i08.10488]
44. Kryger MA, Crytzer TM, Fairman A, Quinby EJ, Karavolis M, Pramana G, et al. The effect of the interactive mobile health and rehabilitation system on health and psychosocial outcomes in spinal cord injury: randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e14305 [FREE Full text] [doi: 10.2196/14305] [Medline: 31464189]
45. Newman SD, Toatley SL, Rodgers MD. Translating a spinal cord injury self-management intervention for online and telehealth delivery: a community-engaged research approach. J Spinal Cord Med 2019 Sep;42(5):595-605 [FREE Full text] [doi: 10.1080/10790268.2018.1518123] [Medline: 30199344]
46. Olney CM, Vos-Draper T, Egginton J, Ferguson J, Goldish G, Eddy B, et al. Development of a comprehensive mobile assessment of pressure (CMAP) system for pressure injury prevention for veterans with spinal cord injury. J Spinal Cord Med 2019 Nov;42(6):685-694 [FREE Full text] [doi: 10.1080/10790268.2019.1570437] [Medline: 30702395]
47. Singh G, MacGillivray M, Mills P, Adams J, Sawatzky B, Mortenson WB. Patients' perspectives on the usability of a mobile app for self-management following spinal cord injury. J Med Syst 2019 Dec 11;44(1):26. [doi: 10.1007/s10916-019-1487-y] [Medline: 31828440]
48. Amann J, Fiordelli M, Brach M, Bertschy S, Scheel-Sailer A, Rubinelli S. Co-designing a self-management app prototype to support people with spinal cord injury in the prevention of pressure injuries: mixed methods study. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e18018 [FREE Full text] [doi: 10.2196/18018] [Medline: 32673241]
49. Amann J, Fiordelli M, Scheel-Sailer A, Brach M, Rubinelli S. Opportunities and challenges of a self-management app to support people with spinal cord injury in the prevention of pressure injuries: qualitative study. JMIR Mhealth Uhealth 2020 Dec 09;8(12):e22452 [FREE Full text] [doi: 10.2196/22452] [Medline: 33295876]
50. MacGillivray MK, Sadeghi M, Mills PB, Adams J, Sawatzky BJ, Mortenson WB. Implementing a self-management mobile app for spinal cord injury during inpatient rehabilitation and following community discharge: a feasibility study. J Spinal Cord Med 2020 Sep;43(5):676-684 [FREE Full text] [doi: 10.1080/10790268.2019.1614343] [Medline: 31091160]
51. Harnphadungkit K. Development and effectiveness testing of “Punsook”: a smartphone application for intermittent urinary catheter users with spinal cord injury. Siriraj Med J 2021 Feb 15;72(2):99-107. [doi: 10.33192/smj.2021.14]
52. Shirai T, Bulandres P, Choi J, D'Ortenzio D, Moon N, Musselman K, et al. The use of a mobile educational tool on pressure injury education for individuals living with spinal cord injury/disease: a qualitative research study. Disabil Rehabil 2022 Feb;44(3):468-477. [doi: 10.1080/09638288.2020.1771780] [Medline: 32493089]
53. Goodwin BM, Olney CM, Ferguson JE, Hansen AH, Eddy B, Goldish G, et al. Visualization of user interactions with a pressure mapping mobile application for wheelchair users at risk for pressure injuries. Assist Technol 2022 Jul 04;34(4):444-453. [doi: 10.1080/10400435.2020.1862938] [Medline: 33395558]
54. Hoevenaars D, Holla JF, Te Loo L, Koedijker JM, Dankers S, Houdijk H, WHEELS Study Group. Mobile app (WHEELS) to promote a healthy lifestyle in wheelchair users with spinal cord injury or lower limb amputation: usability and feasibility study. JMIR Form Res 2021 Aug 09;5(8):e24909 [FREE Full text] [doi: 10.2196/24909] [Medline: 34379056]
55. Wilroy JD, Lai B, Davlyatov G, Mehta T, Thirumalai M, Rimmer JH. Correlates of adherence in a home-based, self-managed exercise program tailored to wheelchair users with spinal cord injury. Spinal Cord 2021 Jan;59(1):55-62 [FREE Full text] [doi: 10.1038/s41393-020-0497-4] [Medline: 32541883]

56. Cole M, Froehlich-Grobe K, Driver S, Shegog R, McLaughlin J. Website redesign of a 16-week exercise intervention for people with spinal cord injury by using participatory action research. JMIR Rehabil Assist Technol 2019 Dec 17;6(2):e13441 [FREE Full text] [doi: 10.2196/13441] [Medline: 31845902]
57. Bizzarini E, Chittaro L, Frezza M, Polo M, Malisan C, Menosso R, et al. A mobile app for home-based exercise in spinal cord injured persons: proposal and pilot study. Digit Health 2022;8:20552076211070724 [FREE Full text] [doi: 10.1177/20552076211070724] [Medline: 35140978]
58. Wilde MH, Fairbanks E, Parshall R, Zhang F, Miner S, Thayer D, et al. A web-based self-management intervention for intermittent catheter users. Urol Nurs 2015;35(3):127-33, 138 [FREE Full text] [Medline: 26298947]
59. Mortenson WB, Singh G, MacGillivray M, Sadeghi M, Mills P, Adams J, et al. Development of a self-management app for people with spinal cord injury. J Med Syst 2019 Apr 22;43(6):145. [doi: 10.1007/s10916-019-1273-x] [Medline: 31011881]
60. Bezerra Giordan L, Tong HL, Atherton JJ, Ronto R, Chau J, Kaye D, et al. The use of mobile apps for heart failure self-management: systematic review of experimental and qualitative studies. JMIR Cardio 2022 Mar 31;6(1):e33839 [FREE Full text] [doi: 10.2196/33839] [Medline: 35357311]
61. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. mHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare (Basel) 2022 Feb 08;10(2):322 [FREE Full text] [doi: 10.3390/healthcare10020322] [Medline: 35206936]
62. Mehraeen E, SeyedAlinaghi S, Pashaei Z, Mirzapour P, Barzegary A, Vahedi F, et al. Mobile applications in HIV self-management: a systematic review of scientific literature. AIDS Rev 2022 Mar 01;24(1):24-31. [doi: 10.24875/AIDSRev.21000025] [Medline: 34723447]
63. Lee J, Yeom I, Chung ML, Kim Y, Yoo S, Kim E. Use of mobile apps for self-care in people with parkinson disease: systematic review. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e33944 [FREE Full text] [doi: 10.2196/33944] [Medline: 35060910]
64. Allin S, Shepherd J, Tomasone J, Munce S, Linassi G, Hossain SN, et al. Participatory design of an online self-management tool for users with spinal cord injury: qualitative study. JMIR Rehabil Assist Technol 2018 Mar 21;5(1):e6 [FREE Full text] [doi: 10.2196/rehab.8158] [Medline: 29563075]
65. Android OS market share of smartphone sales to end users from 2009 to 2021. Statista. URL: https://www.statista.com/statistics/216420/global-market-share-forecast-of-smartphone-operating-systems/ [accessed 2022-06-09]
66. Benning N, Knaup P, Rupp R. Measurement performance of activity measurements with newer generation of apple watch in wheelchair users with spinal cord injury. Methods Inf Med 2021 Dec;60(S 02):e103-e110 [FREE Full text] [doi: 10.1055/s-0041-1740236] [Medline: 34856623]
67. Cieza A, Kirchberger I, Biering-Sørensen F, Baumberger M, Charlifue S, Post MW, et al. ICF Core Sets for individuals with spinal cord injury in the long-term context. Spinal Cord 2010 Apr;48(4):305-312. [doi: 10.1038/sc.2009.183] [Medline: 20065984]
68. Rubinelli S, Glässel A, Brach M. From the person's perspective: perceived problems in functioning among individuals with spinal cord injury in Switzerland. J Rehabil Med 2016 Feb;48(2):235-243 [FREE Full text] [doi: 10.2340/16501977-2060] [Medline: 26926924]
69. Brinkhof MW, Al-Khodairy A, Eriks-Hoogland I, Fekete C, Hinrichs T, Hund-Georgiadis M, SwiSCI Study Group. Health conditions in people with spinal cord injury: contemporary evidence from a population-based community survey in Switzerland. J Rehabil Med 2016 Feb;48(2):197-209 [FREE Full text] [doi: 10.2340/16501977-2039] [Medline: 26931074]
70. McIntyre A, Marrocco SL, McRae SA, Sleeth L, Hitzig S, Jaglal S, et al. A scoping review of self-management interventions following spinal cord injury. Top Spinal Cord Inj Rehabil 2020;26(1):36-63 [FREE Full text] [doi: 10.1310/sci2601-36] [Medline: 32095066]
71. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
72. Larbi D, Randine P, Årsand E, Antypas K, Bradway M, Gabarron E. Methods and evaluation criteria for apps and digital interventions for diabetes self-management: systematic review. J Med Internet Res 2020 Jul 06;22(7):e18480 [FREE Full text] [doi: 10.2196/18480] [Medline: 32628125]
73. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]
74. Mayer G, Hummel S, Oetjen N, Gronewold N, Bubolz S, Blankenhagel K, et al. User experience and acceptance of patients and healthy adults testing a personalized self-management app for depression: a non-randomized mixed-methods feasibility study. Digit Health 2022;8:20552076221091353 [FREE Full text] [doi: 10.1177/20552076221091353] [Medline: 35425641]
75. Khan AI, Gill A, Cott C, Hans PK, Steele Gray C. mHealth tools for the self-management of patients with multimorbidity in primary care settings: pilot study to explore user experience. JMIR Mhealth Uhealth 2018 Aug 28;6(8):e171 [FREE Full text] [doi: 10.2196/mhealth.8593] [Medline: 30154073]
76. Brooke J. Sus-a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Taylor & Francis; 1996.

77. Lund AM. Measuring usability with the use questionnaire. Usability Interface 2001;8(2):3-6.
78. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Human Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
79. Nielsen J. Usability Engineering. Amsterdam, Netherlands: Elsevier Science; 1993.

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMS: Practical Reviews in Self-Management Support
SCI: spinal cord injury
SMS: self-management support

Edited by L Buis; submitted 14.09.22; peer-reviewed by Y Kim, S Newman; comments to author 29.11.22; revised version received 14.02.23; accepted 01.03.23; published 26.04.23.

Please cite as:
Bernard RM, Seijas V, Davis M, Volkova A, Diviani N, Lüscher J, Sabariego C
Mobile Health Self-management Support for Spinal Cord Injury: Systematic Literature Review
JMIR Mhealth Uhealth 2023;11:e42679
URL: https://mhealth.jmir.org/2023/1/e42679
doi:10.2196/42679
PMID:37099372

Review

Assessing the Pragmatic Nature of Mobile Health Interventions Promoting Physical Activity: Systematic Review and Meta-analysis

Chad Stecher[1], PhD; Bjorn Pfisterer[2], MSc; Samantha M Harden[3], PhD; Dana Epstein[1], PhD; Jakob M Hirschmann[4], MA; Kathrin Wunsch[2], PhD; Matthew P Buman[1], PhD

[1]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[2]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[3]Department of Human Nutrition, Foods, and Exercise, Virginia Tech, Blacksburg, VA, United States

[4]Institute of Sport Sciences, Goethe University, Frankfurt, Germany

Corresponding Author:
Chad Stecher, PhD
College of Health Solutions
Arizona State University
500 N 3rd Street
Room 438
Phoenix, AZ, 85004
United States
Phone: 1 6024960957
Email: chad.stecher@asu.edu

Abstract

Background: Mobile health (mHealth) apps can promote physical activity; however, the pragmatic nature (ie, how well research translates into real-world settings) of these studies is unknown. The impact of study design choices, for example, intervention duration, on intervention effect sizes is also understudied.

Objective: This review and meta-analysis aims to describe the pragmatic nature of recent mHealth interventions for promoting physical activity and examine the associations between study effect size and pragmatic study design choices.

Methods: The PubMed, Scopus, Web of Science, and PsycINFO databases were searched until April 2020. Studies were eligible if they incorporated apps as the primary intervention, were conducted in health promotion or preventive care settings, included a device-based physical activity outcome, and used randomized study designs. Studies were assessed using the Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) and Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) frameworks. Study effect sizes were summarized using random effect models, and meta-regression was used to examine treatment effect heterogeneity by study characteristics.

Results: Overall, 3555 participants were included across 22 interventions, with sample sizes ranging from 27 to 833 (mean 161.6, SD 193.9, median 93) participants. The study populations' mean age ranged from 10.6 to 61.5 (mean 39.6, SD 6.5) years, and the proportion of males included across all studies was 42.8% (1521/3555). Additionally, intervention lengths varied from 2 weeks to 6 months (mean 60.9, SD 34.9 days). The primary app- or device-based physical activity outcome differed among interventions: most interventions (17/22, 77%) used activity monitors or fitness trackers, whereas the rest (5/22, 23%) used app-based accelerometry measures. Data reporting across the RE-AIM framework was low (5.64/31, 18%) and varied within specific dimensions (Reach=44%; Effectiveness=52%; Adoption=3%; Implementation=10%; Maintenance=12.4%). PRECIS-2 results indicated that most study designs (14/22, 63%) were *equally explanatory and pragmatic*, with an overall PRECIS-2 score across all interventions of 2.93/5 (SD 0.54). The most pragmatic dimension was *flexibility (adherence)*, with an average score of 3.73 (SD 0.92), whereas *follow-up*, *organization*, and *flexibility (delivery)* appeared more explanatory with means of 2.18 (SD 0.75), 2.36 (SD 1.07), and 2.41 (SD 0.72), respectively. An overall positive treatment effect was observed (Cohen *d*=0.29, 95% CI 0.13-0.46). Meta-regression analyses revealed that more pragmatic studies (−0.81, 95% CI −1.36 to −0.25) were associated with smaller increases in physical activity. Treatment effect sizes were homogenous across study duration, participants' age and gender, and RE-AIM scores.

Conclusions: App-based mHealth physical activity studies continue to underreport several key study characteristics and have limited pragmatic use and generalizability. In addition, more pragmatic interventions observe smaller treatment effects, whereas study duration appears to be unrelated to the effect size. Future app-based studies should more comprehensively report real-world applicability, and more pragmatic approaches are needed for maximal population health impacts.

Trial Registration: PROSPERO CRD42020169102; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=169102

(JMIR Mhealth Uhealth 2023;11:e43162) doi:10.2196/43162

KEYWORDS

physical activity; mobile health; mHealth; Reach, Effectiveness, Adoption, Implementation, Maintenance; RE-AIM; Pragmatic-Explanatory Continuum Indicator Summary-2; PRECIS-2; systematic review; meta-analysis; digital health; mobile phone

Introduction

Background

Regular physical activity can combat numerous chronic conditions and is associated with reduced premature mortality [1,2]. Despite these benefits, behavioral interventions and public policy have been largely unsuccessful in promoting higher physical activity among the general population. Worldwide, 28% of individuals are currently classified as insufficiently active [3], and physical inactivity has an estimated annual health care cost of >US $50 billion globally [4]. Thus, increasing physical activity across the world is an important economic and public health objective that requires scalable and pragmatic strategies [5].

Mobile health (mHealth) tools are one promising approach for improving health care delivery and scaling behavioral interventions worldwide [6,7]. Mobile app–based platforms can be particularly effective at increasing intervention accessibility and cost-effectiveness, and they offer the ability to tailor intervention methods to individuals' unique needs [8-10]. Accordingly, the use of app-based mHealth tools in health care has rapidly increased since 2008 [10,11], and several review papers have recently highlighted the important potential role of app-based interventions for improving global physical activity levels [12-14]. In addition, app-based interventions saw a large relative increase in publications compared with SMS text messaging, telehealth, or web-based interventions [14], making app-based interventions one of the most popular new clinical tools [15] and an important intervention approach to review to inform current and future researchers, as well as health care providers (eg, general practitioners).

Despite the growth of research using app-based tools to promote physical activity, there is limited evidence that app-based interventions for increasing physical activity have been widely adopted by policy makers or integrated into clinical or other practice settings [16,17]. One potential explanation for this lack of real-world application is that this research has generally centered on internal validity (ie, reliability or accuracy of the outcomes) over external validity (ie, generalizability or applicability of results) [18,19]. In other words, the existing research has emphasized explanatory approaches rather than more pragmatic study designs [20]. Explanatory studies measure whether an intervention has a beneficial effect under ideal and thoroughly controlled circumstances and, therefore, substantially differ from real-world conditions (eg, restrictive selection of study sample and control of intervention delivery). Pragmatic study designs can determine the effect of an intervention under more realistic conditions by maximizing external validity (eg, broad and inclusive eligibility criteria and flexibility in intervention delivery) [20-23]. Studies are not strictly dichotomous in their design; instead, they are situated along the explanatory-pragmatic continuum [21,22,24]. Essentially, the challenge is to strike a balance between a highly effective program and whether it can be integrated into practice settings. mHealth interventions have the unique advantages of leveraging automation, data-informed decision-making, and other technological components that might aid in adherence to the core elements (eg, key ingredients or mechanism of change) while scaling out [25].

Existing systematic reviews of mHealth studies have broadly called for increased pragmatism [18,26,27]; however, only one research review has specifically explored the generalizability and applicability of app-based physical activity interventions [16]. However, the results were limited by the insufficient reporting of external validity factors within the included studies. Thus, the review authors were not able to determine the generalizability of the findings and recommended that future mHealth researchers better report all study characteristics [16]. Specific study design characteristics, such as the study sample's demographics (eg, average age and gender) and the duration of the intervention, are important dimensions to evaluate when determining the generalizability of a study's findings to the full population.

Given the continued growth of app-based physical activity interventions [14] and the lack of clarity surrounding the pragmatic nature of these approaches, we conducted a systematic review and meta-analysis of mHealth apps for physical activity promotion.

Objective

Our primary aim was to analyze the degree to which these interventions reported the study characteristics necessary to inform generalizability and applicability and to assess the explanatory versus pragmatic nature of these studies. Our secondary aim was to explore the association between study design characteristics (eg, explanatory vs pragmatic, intervention duration, and participant demographics) and the observed effect sizes on participants' physical activity.

Methods

Protocol and Registration

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [28,29].

Search Strategy and Study Selection

We conducted a systematic search in 4 electronic databases on April 4, 2020: PubMed, Scopus, Web of Science, and PsycINFO. The search combined synonyms and keywords related to an app-based mHealth intervention for promoting physical activity (Table 1; Multimedia Appendix 2). We attempted to control for language bias by using a search strategy without language restriction (ie, no selective inclusion of trials published in English) [30]. In addition to these databases, the list of papers discussed by relevant systematic reviews [8,31-40] was examined to identify any further eligible studies.

Table 1. Search strategy used in PubMed on April 4, 2020.

Search category	Search term
mHealth[a]	mHealth OR mobile health OR m-health OR activity tracker OR fitness tracker OR wearable OR tablet OR personal digital assistant OR pda OR short message service OR sms OR text message OR android OR iphone OR iOS OR mobile phone OR cellphone OR cell phone OR cellular phone OR cellular telephone OR mobile telephone OR smart-phone OR smartphone OR mobile application OR mobile app
Physical activity	physical activity OR leisure activity OR active living OR exercise OR sport OR fitness OR motor activity OR sedentary behavior OR sedentary lifestyle OR sitting OR physical inactivity
Intervention	Intervention OR trial OR program
Study design	clinical trial OR controlled trial OR controlled study OR double blind OR RCT[b] OR pragmatic trial OR practical trial OR PCT[c] OR ecological trial OR dynamic trial OR real-world OR real world
Combined	mHealth AND Physical activity AND Intervention AND Study design

[a]mHealth: mobile health.

[b]RCT: randomized controlled trial.

[c]PCT: practical clinical trial.

The included studies were limited to app-based physical activity interventions that were published in a peer-reviewed journal between January 2012 and April 2020 that primarily targeted physical activity and at most one other behavioral outcome and that presented quantitative outcome data. We further restricted our review to studies that collected device-based physical activity measures, as opposed to self-reported measures because device-based measures are frequently observed to be more reliable [41,42] and the use of physical activity–monitoring devices has become more commonplace in the real world [43], demonstrating the feasibility, acceptability, and pragmatism of these intervention tools. A complete list of the eligibility criteria is presented in Table 2. We obtained additional data sources (when available) such as the study protocol, the CONSORT (Consolidated Standards of Reporting Trials) checklist, or any other publicly available information from the corresponding authors provided via an email invitation to assess the Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) framework for internal and external validity factors [44,45] and the Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) tool for evaluating interventions' pragmatism [24]. Specifically, this email contained a brief description of our study, and then asked, "In order to comprehensively evaluate the reporting of RE-AIM and PRECIS-2 criteria, we are also extracting data from study protocols and companion articles (eg, qualitative or quantitative methods measuring implementation). Would you be willing to help us by providing these additional resources?"

All records from the databases and supplementary searches were managed using the Microsoft EndNote X9 (Clarivate) reference manager software. After removing duplicates, we exported the records to Abstrackr (Brown University) for semiautomatic citation screening [46]. The relevance of the titles and abstracts was independently assessed by 2 authors (BP and JMH). Each eligible full text was independently reviewed by 2 researchers (SMH and MPB). Discrepancies were resolved through discussion between the screening authors. Any remaining conflicts were discussed among the other authors (CS, DE, KW, and BP) until consensus was reached.

Table 2. Eligibility criteria.

Data type	Eligibility criteria
Population	Participants of any age participating in physical activity programs in the context of health promotion or preventive care settings were included. Studies focusing on special populations (eg, pregnant women) or studies including participants with physical or psychological morbidities preventing them from participating in physical activity were excluded.
Intervention	Stand-alone mobile apps and web apps exclusively designed for mobile devices; multicomponent interventions (eg, supported through brief counseling sessions or paired with other mHealth[a] technologies) were included as long as the app was the primary component to the intervention; interventions that targeted ≥2 health behaviors in addition to physical activity (eg, diet, sleep, and SB[b]) were excluded; apps solely used for data collection purposes or as an appointment reminder service only were not eligible.
Comparator	Active or inactive comparator arms were included; single-subject design trials were excluded.
Outcome	Device-based measures of physical activity.
Study design	RCTs[c] and randomized ecologically valid research designs (ie, practical clinical trials, RCTs); randomized pilot and feasibility studies were included.

[a]mHealth: mobile health.

[b]SB: sedentary behavior.

[c]RCT: randomized controlled trial.

Data Collection Process

General Study Characteristics

We adapted an existing extraction template [32] to collect and summarize the general study characteristics. Specifically, we collected information about the study setting and design, study population, intervention components, outcome measures, key findings, and statistical analyses performed (Multimedia Appendix 3). Two authors (BP and JMH) separately extracted additional quantitative data for the meta-analyses; discrepancies were resolved through discussion and consultation with a third author (SMH).

RE-AIM Evaluation and PRECIS-2 Assessment

We used the RE-AIM framework to describe the degree of reporting of study characteristics across 5 dimensions (ie, reach, effectiveness, adoption, implementation, and maintenance). The evaluation was assisted by a 31-item RE-AIM coding system used in a previous study [47]. We then applied the PRECIS-2 tool to compare the interventions with usual care and to identify the pragmatic versus explanatory nature of each study. Following the guidance of Loudon et al [24] and the PRECIS-2 toolkit published on the web, usual care was defined as the primary care that patients usually received for medical advice and treatment. The PRECIS-2 tool comprises 9 domains (ie, eligibility criteria, recruitment, setting, flexibility [delivery], flexibility [adherence], follow-up, primary outcome, and primary analysis), each of which is assigned a score from 1 to 5 (1 is *very explanatory* and 5 is *very pragmatic*) [24]. In accordance with previous research [47], mean scores of >3.5 were deemed *primarily pragmatic*. Values between 2.5 and 3.5 were considered *equally pragmatic and explanatory*, and scores <2.5 were rated as *primarily explanatory*.

Although both frameworks can be applied regardless of the study setting, additional modifications to these frameworks are recommended for a given setting [48]. Thus, we adapted the RE-AIM and PRECIS-2 coding sheets [49] for our setting (Multimedia Appendix 4 presents these adapted coding sheets). The final scoring by the study is presented in Multimedia Appendix 5.

Quality Assessment

For each study, we also assessed quality of the study using the revised Cochrane risk-of-bias (RoB 2.0) tool for randomized controlled trials [50]. Two authors (BP and JMH) independently performed these assessments, and any disagreements were resolved through discussion with a third author (SMH, DE, and MPB). The studies were classified as having a *low risk of bias* if all the 5 assessment domains were considered low risk. Otherwise, the studies were classified as having s*ome concerns* when concerns were raised in at least 1 of the 5 domains, or they were classified as having *high risk of bias* when at least one of the domains was judged to be at high risk. These categories were drawn from the original Cochrane RoB 2.0 tool [50].

Statistical Analyses

We used counts and percentages to summarize the general study characteristics and RE-AIM and PRECIS-2 scores for each study.

Meta-analyses were performed by using *meta* commands in Stata 16 (StataCorp) [51]. We used the standardized average treatment effect in each study's primary app- or device-based physical activity outcome (ie, minutes of moderate to vigorous physical activity or step count) to compare treatment effects across studies with different outcomes. The standardized average treatment effect (or Cohen *d*) was calculated as the difference in the mean change in primary physical activity outcome between the intervention group and the control group divided by the pooled SD of the physical activity outcome in both the intervention and control groups, with a priori interpretations [52] of trivial (<0.2), small (0.2-0.5), moderate (0.5-0.8), and large (>0.8) effects.

In addition, we tested for heterogeneous treatment effects using random-effects models estimated through restricted maximum likelihood. All the following moderating variables were log transformed to better compare the effect sizes: baseline physical activity, sample size, participants' age, participants' gender, intervention duration, RoB score, RE-AIM score, and PRECIS-2 score. Bubble plots were used to graphically examine the relationships between treatment effect size and the continuous moderating variables.

We assessed the statistical significance of treatment effect heterogeneity by using Cochran Q test and calculating the Higgins I^2 statistic [53]. The following thresholds for the interpretation of the I^2 statistic were used:0%-40%, 30%-60%, 50%-90%, or 75%-100%; these were interpreted as *not likely important*, *moderate*, *substantial*, and *considerable* heterogeneity, respectively [53].

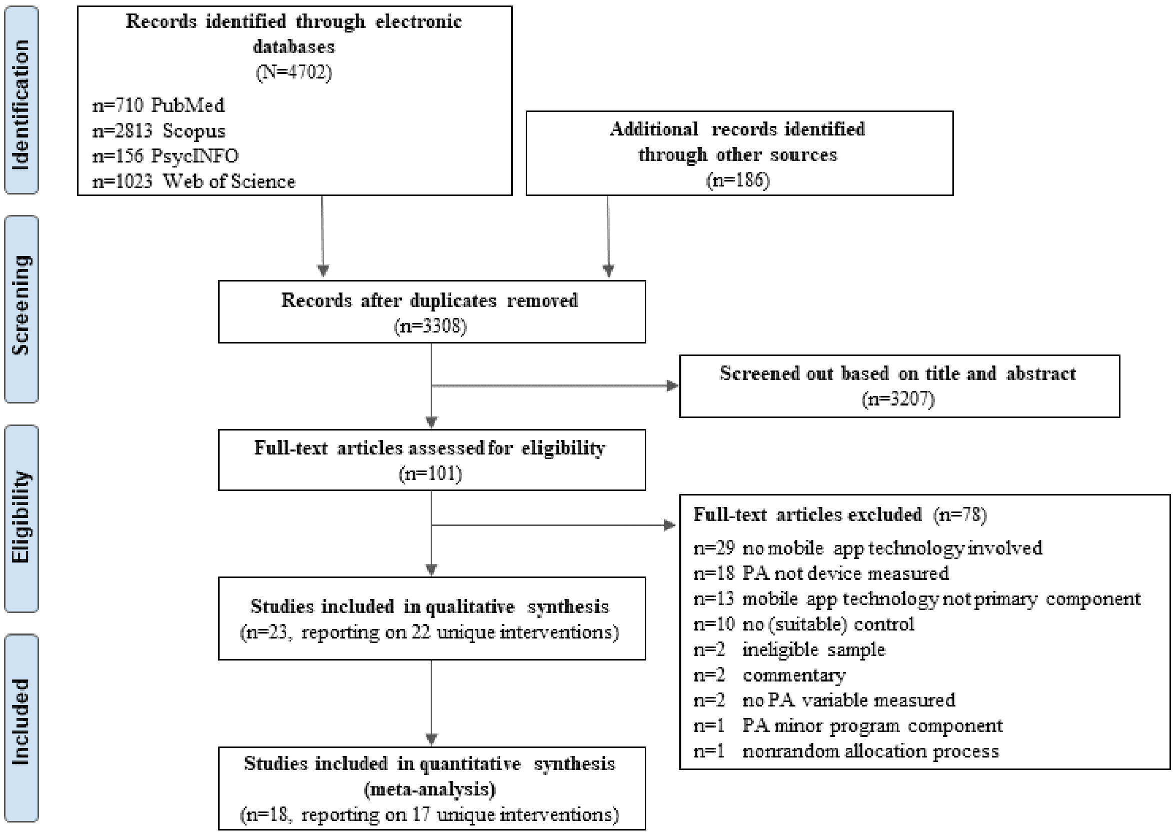
Finally, the combined impact of small-study effects and publication bias was assessed by using the trim-and-fill method and performing the Egger test using the *metafor* package [54] in R (version 3.6.3; R Foundation for Statistical Computing) [55]. The results are reported with 95% CI, and a *P* value of <.05 was considered statistically significant.

Results

Study Selection

The search yielded 3308 unique studies after duplicates were removed. Of the 3308 studies, we screened 3207 (96.95%) studies based on title and abstract, leaving 101 (3.05%) potentially relevant studies. After additional content reviews, 23 studies reporting 22 unique interventions met the eligibility criteria for inclusion in the RE-AIM and PRECIS-2 analyses. We emailed the corresponding authors of all 23 studies to request additional study information. We received responses from 52% (12/23) of the studies, and these responses either contained more information on the study (7/12, 58%) or simply stated that there was no additional information available (5/12, 42%). In total, only 74% (17/23) of these studies presented sufficient quantitative detail for inclusion in the meta-analyses. The detailed study selection process is visualized in the PRISMA flowchart (Figure 1).

Figure 1. Flowchart of study selection. PA: physical activity.



Study Characteristics

All interventions were published in English between 2012 and 2020 and were conducted in 10 countries, with most interventions (10/22, 45%) having based in the United States [56-65]. Of the 22 interventions, 21 (95%) used a randomized controlled trial design, of which 19 (90%) interventions randomized participants on an individual level and 3 (14%) interventions were randomized in clusters [66-68]. One study explicitly used a pragmatic study design [69]; 6 studies identified their trials as pilot studies [56,61,62,64,70,71], and 1 was classified as a feasibility study [72]. One study used a factorial design between multiple intervention components as part of a multiphase optimization strategy [57]. An overview of these study characteristics for each study is presented in detail in Multimedia Appendix 6.

A total of 3555 participants were included across all 22 interventions, with sample sizes ranging from 27 to 833 (mean 161.6, SD 193.9, median 93) participants. All studies were conducted in a health promotion or preventive care setting, and the most common study settings were the local community (10/22, 45%), a university or other type of school (7/22, 32%), or a clinical care setting (3/22, 14%). In addition, 10 interventions exclusively targeted insufficiently active individuals. Study populations varied in age and gender, with mean ages ranging from 10.6 to 61.5 (mean 39.6, SD 6.5) years, and the proportion of males included across all studies was 42.8% (1521/3555). Moreover, 2 studies exclusively targeted men, and 2 studies included women only.

Intervention length varied from 2 weeks to 6 months (mean 60.9, SD 34.9 days). The primary app- or device-based physical activity outcomes differed between interventions, with most interventions (17/22, 77%) using activity monitors or fitness trackers and the rest (5/22, 23%) using app-based accelerometry measures. All studies reported either moderate to vigorous physical activity, daily steps, or both measures. The comparator groups received either no intervention (10/22, 45%); a minimal intervention such as generic physical activity information (6/22, 27%); a basic app version targeting physical activity (3/22, 14%); a control app unrelated to physical activity (1/22, 5%); or a wearable activity monitor with access to its corresponding generic tracking app (2/22, 9%).

A total of 27% (6/22) of studies targeted physical activity, and 5% (1/22) of studies targeted additional health behavior outcome (ie, diet or sedentary behavior). With regard to the physical activity intervention strategies used in all studies, 27% (6/22) of studies provided brief in-person expert consultations (eg, goal setting or generic physical activity information), and 5% (1/22) of interventions included weekly telephone counseling. Most studies (19/22, 83%) also used emails and text messages as physical activity reminders or to provide participants with an activity summary.

The interventions' apps varied greatly between the studies and consisted of both commercial products and apps designed solely for research purposes. The apps included features such as physical activity tracking and self-monitoring, feedback, goal setting, social interaction, and gamification features (Multimedia Appendix 6 provides the full list of app features by intervention).

RoB Assessment

Table 3 shows the RoB in the included studies. Overall, 17% (4/23) of studies showed a *low risk*; 43% (10/23) of studies raised *some concerns*; and 39% (9/23) of studies were rated *high risk*. A lack of balance across randomized study groups in terms of baseline physical activity and gender contributed to a *high risk of bias* classification for 3 studies, and 2 other studies were considered to have a *high risk of bias* for deviating from their intended intervention design, which the authors attributed to a lack of participant engagement with the intervention's physical activity app and the intended intervention. In addition, most studies (14/22, 64%) did not provide enough information to determine whether the data were analyzed according to their prespecified data analysis plan, which resulted in them being classified as having *some concerns*.

Table 3. Risk-of-bias (RoB) assessment based on the revised Cochrane RoB tool for randomized trials (RoB 2.0).[a]

Study, year	Randomization bias[b]	Deviation bias[c]	Missing data bias[d]	Measurement bias[e]	Selection bias[f]	Overall
Direito et al [69], 2015	+	+	+	+	+	+
Edney et al [66], 2020	+	+	+	+	+	+
Fanning et al [57], 2017	?	?	?	+	?	−
Fukuoka et al [58], 2019	+	+	+	+	+	+
Garcia-Ortiz et al [73], 2018	+	?	?	+	+	?
Garde et al [74], 2018	+	?	?	+	?	−
Glynn et al [75], 2014	?	+	+	+	+	?
Gremaud et al [59], 2018	+	?	+	+	?	?
Harries et al [76], 2016	?	?	+	+	?	−
Hurkmans et al [77], 2018	−	?	+	+	+	−
King et al [60], 2016	+	+	+	+	?	?
Kitagawa et al [70], 2020	?	+	+	+	?	?
Leinonen et al [72], 2017	+	−	−	+	+	−
Lyons et al [61], 2017	+	+	+	+	+	+
Martin et al [56], 2015	+	+	+	+	?	?
Pope and Gao [62], 2020	−	?	+	+	?	−
Recio-Rodriguez et al [78], 2016	+	?	?	+	+	?
Robertson et al [67], 2018	+	−	−	+	?	−
Schade et al [63], 2020	?	?	−	+	?	−
Simons et al [68], 2018	−	?	+	+	?	−
Walsh et al [71], 2016	+	+	+	+	?	?
Zhang, and Jemmott [64], 2019	+	?	+	+	?	?
Zhou et al [65], 2018	+	+	+	+	?	?

[a]+ = low risk of bias; ?=some concerns; −=high risk of bias.
[b]Bias arising from the randomization process.
[c]Bias because of deviations from the intended intervention.
[d]Bias because of missing outcome data.
[e]Bias because of measurement tools used to collect outcome data.
[f]Bias in selection of the reported result.

RE-AIM Evaluation

Overview

The overall rating of sufficiently reported individual RE-AIM items across all interventions was 18% (5.64/31, SD 2.30%; Table 4). Reporting ranged from 2 to 11 of the 31 RE-AIM items. The most commonly reported items were those in the Effectiveness (2.6/5, 52%) and Reach (1.8/4, 45%) dimensions. Reported data within the Maintenance categories were observed in only 12% (1.1/9) of the interventions, and the reporting of items in the Adoption and the Implementation dimensions were found in 4% (0.3/8) and 10% (0.5/5) of the interventions, respectively. A summary of the key findings of the factors within each dimension is presented in the subsequent section.

Table 4. Inclusion of Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) items across all interventions (N=22).[a,b]

RE-AIM dimension and items	Values, n (%)
Reach (44.3%)	
Exclusion criteria	17 (77)
Participation rate	16 (73)
Representativeness	6 (27)
Use of qualitative methods to understand reach and recruitment	0 (0)
Effectiveness (52.7%)	
Measure of primary outcome	22 (100)
Measure of broader outcomes (ie, QoL[c], negative outcomes)	11 (50)
Measure of robustness across subgroups	4 (18)
Measure of short-term attrition	14 (64)
Use of qualitative methods or data to understand outcomes	7 (32)
Adoption-setting (3.4%)	
Setting exclusions	2 (9)
Setting adoption rate	1 (4)
Setting representativeness	0 (0)
Use of qualitative methods to understand adoption at setting level	0 (0)
Adoption-staff (0%)	
Staff exclusions	0 (0)
Staff participation rate	0 (0)
Staff representativeness	0 (0)
Use of qualitative methods to understand staff participation	0 (0)
Implementation (10%)	
Delivered as intended	5 (23)
Adaptations to intervention	4 (18)
Cost of intervention (time or money)	0 (0)
Consistency of implementation across staff or time or settings subgroups	2 (9)
Use of qualitative methods to understand implementation	0 (0)
Maintenance-individual (9%)	
Measure of primary outcome at ≥6-mo follow-up	3 (14)
Measure of broader outcomes (ie, QoL, negative outcomes) at follow-up	2 (9)
Measure of long-term robustness across subgroups	2 (9)
Measure of long-term attrition	3 (14)
Use of qualitative methods to understand long-term effects	0 (0)
Maintenance-setting (3.4%)	
Program ongoing (≥6-mo poststudy funding)	1 (4)
Long-term program adaptations	2 (9.1)
Some discussion of sustainability of business model	0 (0)
Use of qualitative methods to understand setting-level institutionalization	0 (0)

[a]The table formatting was adapted from Burke et al [47].

[b]Overall RE-AIM was 18.2%.

[c]QoL: quality of life.

Reach

Exclusion criteria commonly included health contraindications for participating in physical activity or comprised mHealth-specific requirements (eg, specifications around technical devices). Most studies provided accurate information (ie, either n and valid denominator or percentage) on the *participation rate* (16/22, 73%) [56-60,62-66,68,69,71,72,75,78]; however, only a few (3/22, 14%) reported the sample size in relation to the total number exposed to recruitment [65,68,72], and the remaining trials reported only on the relation of the sample size to potentially eligible participants [56-60,62-64,66,69,71,75,78]. A few interventions (6/22, 27%) adequately reported the *representativeness* of the study sample. One intervention compared their sample to eligible individuals who declined participation [72], and 5 compared their sample and their target audience [58,62,66,70,71]. Comparisons were made on physical activity variables and anthropometry and fitness measures.

Effectiveness

All studies (23/23, 100%) reported a *measure of primary outcome* related to physical activity (per review eligibility criteria), and half of the interventions (11/22, 50%) addressed a *measure of broader outcomes* [56,57,60,61,65-67,69,70,72,75]. Moreover, 45% (10/22) of studies compared their physical activity–related findings to a public health goal (ie, physical activity guidelines) [56,58,62-64,71,74-76]; however, only a few studies (4/22, 18%) analyzed the *robustness across study subgroups* (eg, gender and age groups) [56,58,64,76]. Potential explanations for physical activity–related findings were explored using *qualitative research methods* in several interventions (7/22, 32%) [57,62,67-69,72,76].

Adoption

Both nonresearch and research staff participation were considered, and more participation of either nonresearch or research staff would result in a study being less pragmatic if it exceeded the usual standard of care. However, no items were reported within the dimension "Adoption-staff." Regarding "Adoption-setting," 2 studies specified *settingexclusions* (eg, unqualified staff and irregular physical education classes) [67,68]. One intervention presented a valid *setting adoption rate* [68].

Implementation

The *delivered as intended* and the *adaptations to intervention* items were infrequently addressed and were mainly of technical nature (eg, app bug or app appearance). None of the studies sufficiently reported the *cost of* intervention, meaning that costs were not addressed across all levels of the intervention or were not detailed enough (eg, app development, technical equipment, and support). The *consistency of implementation* was outlined in 2 trials (eg, fidelity checks) [58,78].

Maintenance

A few interventions (3/22, 14%) assessed a *≥6-month follow-up* measure; 2 studies reported a 6-month follow-up phase [58,66]; 1 implemented a 9-month follow-up measure [73]; and all these studies reported an accurate *long-term attrition rate*. Two studies analyzed the *long-term robustness* (eg, age and weight status) [58,73]. A *measure of broader outcomes* was reported in 2 interventions, assessing the quality of life using the 12-Item Short-Form Health Survey [58,66].

Items within the Maintenance-Setting dimension were only addressed by 3 interventions, including potential *long-term adaptations* (eg, implementing an educational app component) [56,72,74]. The sustainability of the program in the RE-AIM context was not discussed at all.

PRECIS-2 Assessment

The overall PRECIS-2 score across all interventions was 2.93/5 (SD 0.54). Of the 22 assessed interventions, 14 (64%) interventions were categorized as *equally pragmaticand explanatory* (range 2.56-3.44) [57,59,62–67,69–71,73,74,76]; 5 (23%) studies were identified as being *primarily explanatory* (range 2.00-2.44) [58,60,61,68,77]; and 3 (14%) studies were *primarily pragmatic* (range 3.56-4.44) [56,72,74].

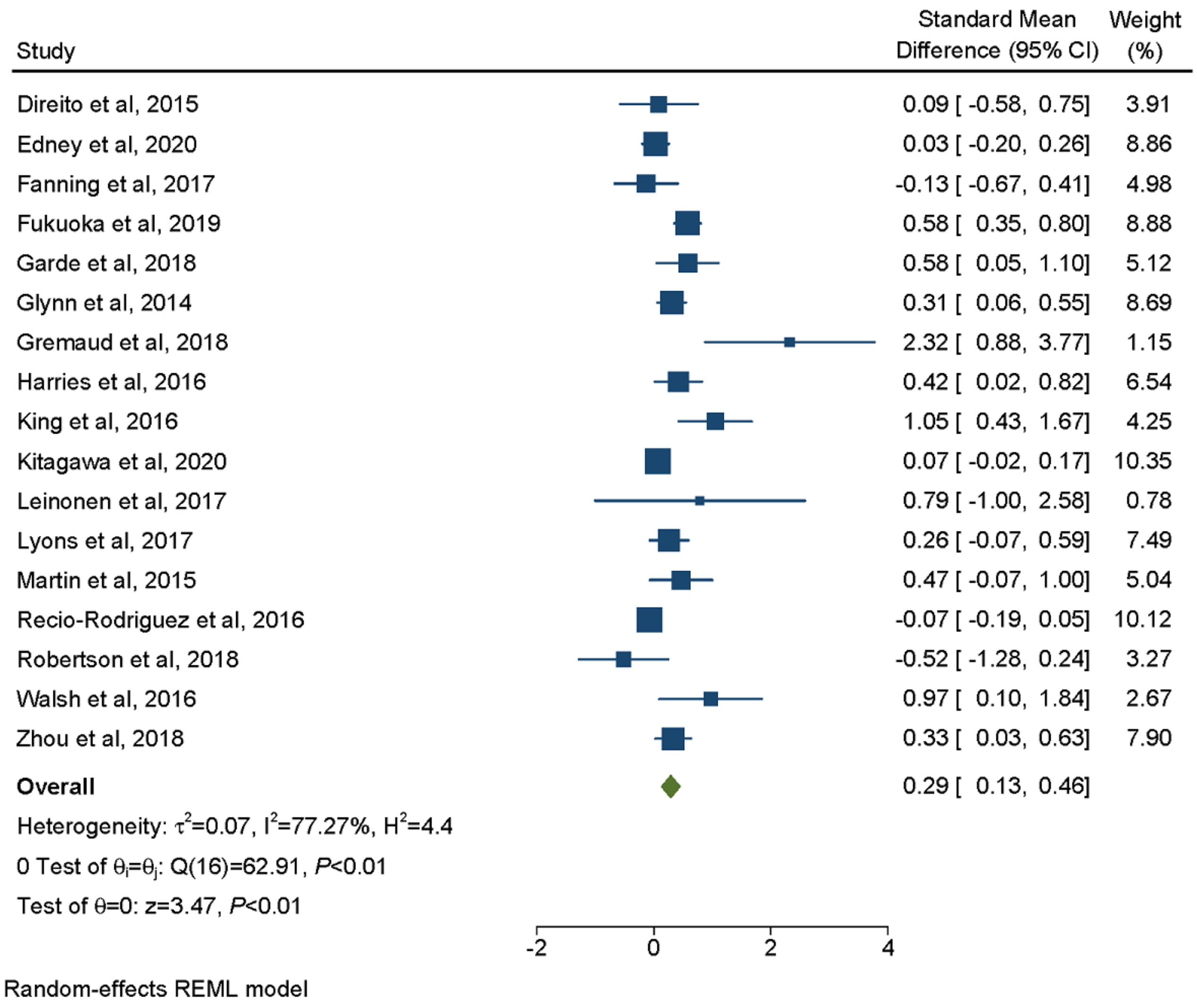
The most pragmatic dimension across all interventions was *flexibility (adherence)*, with an average score of 3.73 (SD 0.92), as demonstrated by letting the participants use the app at their convenience or lacking any measures to improve adherence. *Follow-up*, *organization*, and *flexibility (delivery)* appeared to be more explanatory, with means of 2.18 (SD 0.75), 2.36 (SD 1.07), and 2.41 (SD 0.72), respectively. For example, delivery flexibility was considered more explanatory based on in-person requirements, clinician oversight, or specific app use or compliance requirements. Domains considered equally explanatory and pragmatic were *eligibility criteria*, *recruitment*, *setting*, *primary outcome*, and *primary analysis* (range 2.95-3.45). Overall, the studies in this review were equally pragmatic and explanatory in terms of the eligibility criteria.

Meta-analysis

Overall Treatment Effect

Data from only 17 interventions were extracted for this meta-analysis because 5 interventions did not present complete outcome data (ie, they did not report SE or 95% CI). Overall, these 17 mHealth interventions significantly improved the participants' physical activity (Cohen *d*=0.29, 95% CI 0.13-0.46; Figure 2).

Figure 2. Forest plot of standardized treatment effects on physical activity with studies weighted by the inverse of the SE of the estimated treatment effect. REML: restricted maximum likelihood.



Meta-regression Analyses

Meta-regression analyses revealed a statistically significant negative association between the standardized treatment effect and the study's sample size (*P*=.01), PRECIS-2 score (*P*<.001), and study participants' baseline physical activity (*P*<.001; Table 5), that is, a larger sample size, higher PRECIS-2 score (ie, more pragmatic), and higher observed baseline physical activity levels were all associated with smaller treatment effect sizes on participants' physical activity. None of the other covariate measures (ie, intervention duration, participants' age, participants' gender, and RE-AIM score) were significantly related to changes in participants' physical activity.

To graphically depict the interaction between the treatment effect size and the continuous measure of a study's PRECIS-2 score, we created a bubble plot with studies represented by circles sized by the inverse of the SE of the estimated treatment effect (Figure 3). The plot also shows the weighted linear relationship between these study characteristics and the 95% CI for this estimated relationship.

Table 5. Meta-regression results showing the interaction between study characteristics and the standardized treatment effect on physical activity.[a]

Covariate	Standardized mean difference (95% CI)	*P* value
Log (intervention duration [days])	0.0171 (−0.0338 to 0.0680)	.51
Log (participant mean age [years])	−0.00296 (−0.224 to 0.218)	.98
Log (sample size)	−0.0616[b] (−0.111 to −0.0123)	.01
Log (percentage male)	−0.0615 (−0.266 to 0.143)	.56
Log (baseline step count)	−0.420[c] (−0.637 to −0.202)	<.001
Log (baseline MVPA[d] [minutes])	−0.199[c] (−0.288 to −0.109)	<.001
Log (PRECIS-2[e] score)	−0.805[f] (−1.361 to −0.249)	<.001
Log (RE-AIM[g] score)	−0.0277 (−0.177 to 0.122)	.72
Log (risk-of-bias score)	−0.199 (−0.406 to 0.0690)	.06

[a]All covariates were log transformed; therefore, the coefficients measure the associated change in the standardized treatment effect size from a 1% increase in the indicated variable.

[b]*P*<.05.

[c]*P*<.001.

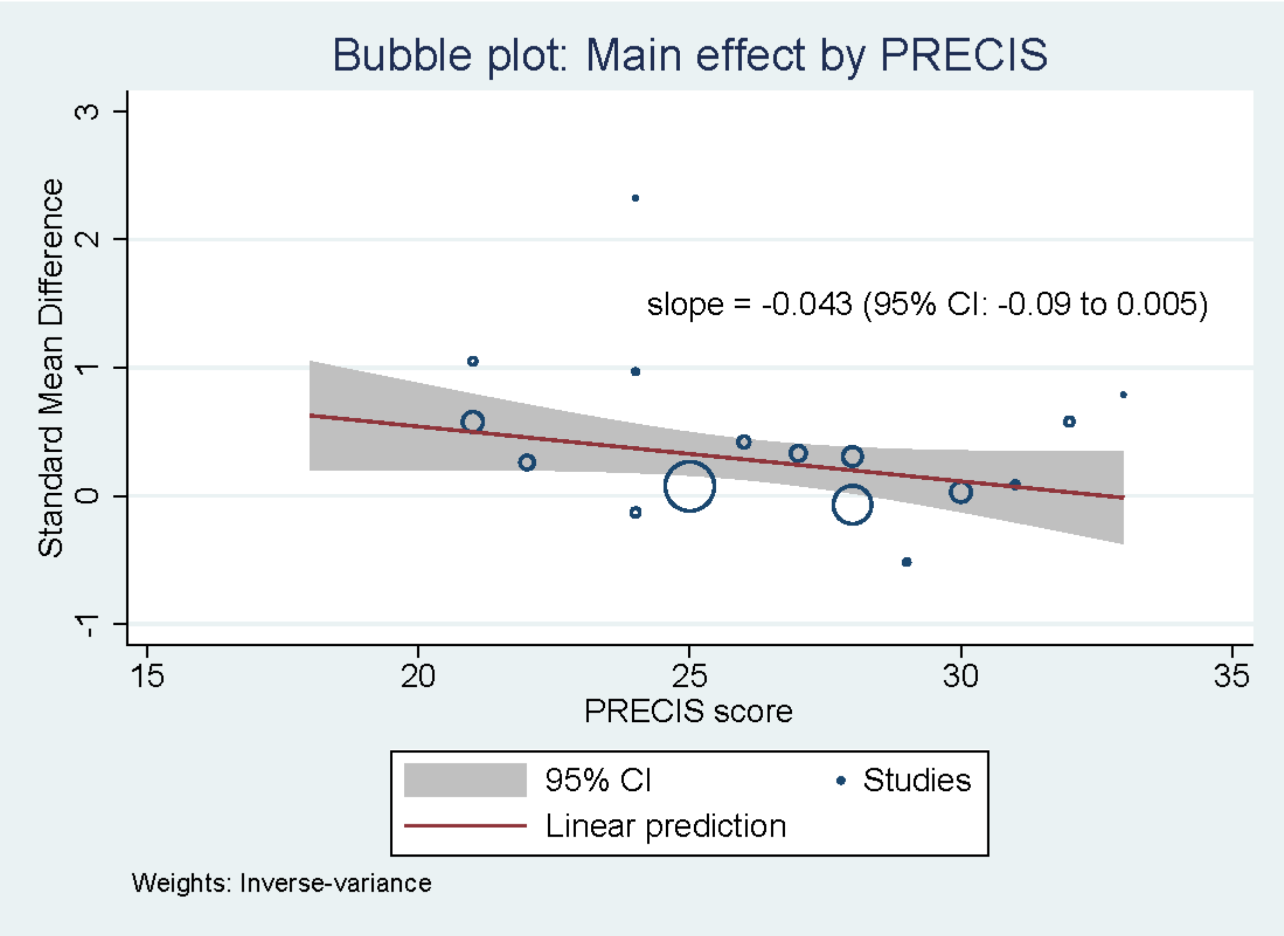
[d]MVPA: moderate to vigorous physical activity.

[e]RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance.

[f]*P*<.01.

[g]PRECIS-2: Pragmatic-Explanatory Continuum Indicator Summary-2.

Figure 3. Bubble plot of standardized treatment effect on Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) score (a single outlier was removed).



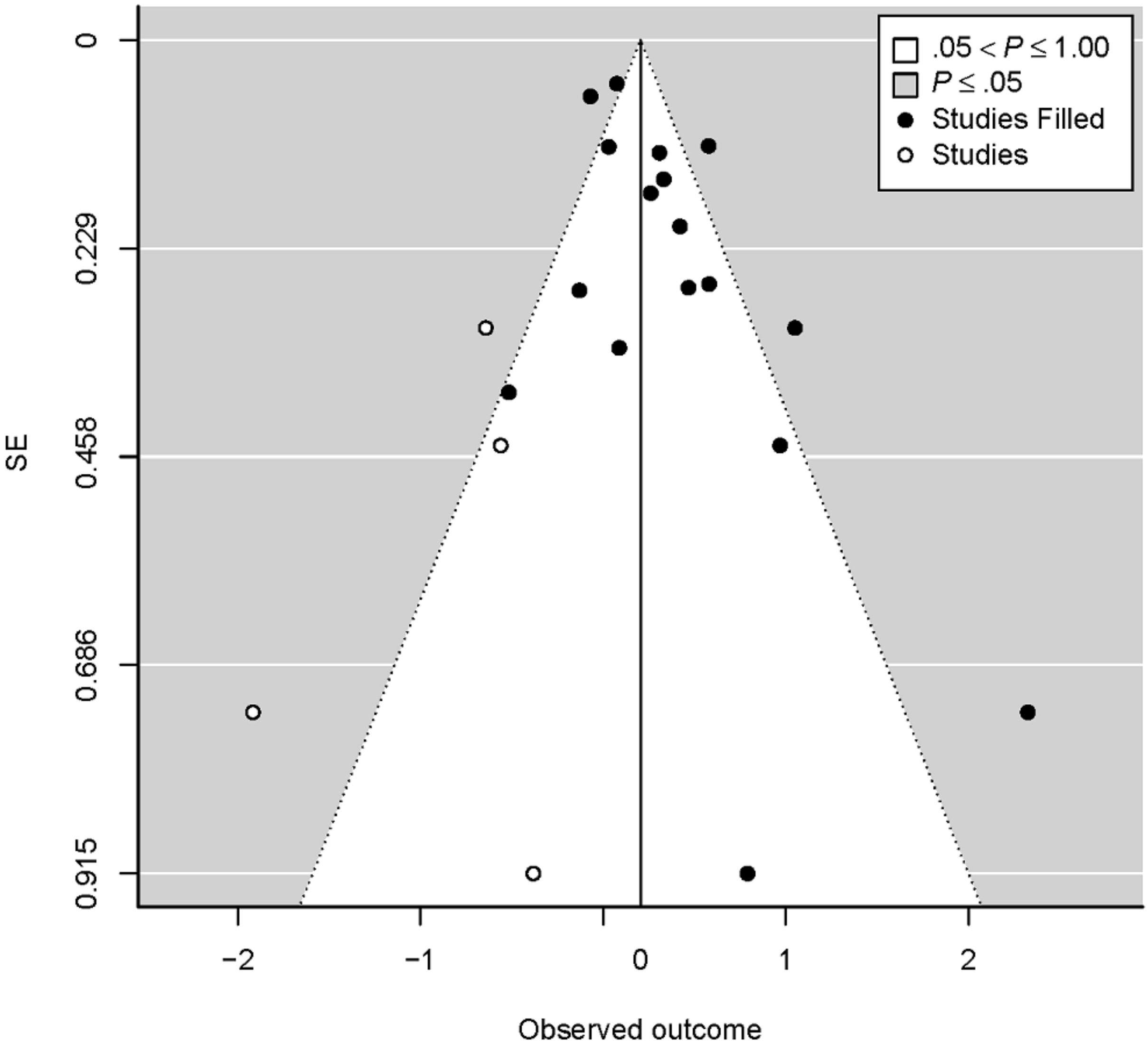
Overall Treatment Effect Heterogeneity

The meta-analysis showed considerable heterogeneity between the studies, with an I^2 value of 77.27%. The I^2 value represents the estimated percentage of variability in the results because of heterogeneity rather than chance [53]. Cochran Q test for treatment effect heterogeneity across these studies was Q_{16}=62.91, which demonstrates a statistically significant degree of heterogeneity (P<.001).

Analysis of Publication Bias and Small-Study Effects

We used the trim-and-fill method to explore the potential impact of publication bias in this literature, which estimated the number of studies missing from this literature to be 4 (SE 2.80; Figure 4). After imputing these missing studies, the overall standardized treatment effect size was slightly reduced from 0.29 (95% CI 0.13-0.46) to 0.20 (95% CI 0.01-0.40) but remained statistically significant. A high I^2 value of 83.8% indicated that the heterogeneity between studies remained at a considerable level after imputing these potentially missing studies. We then carried out the Egger test for small-study effects, which reached statistical significance under most specifications (Multimedia Appendix 7).

Figure 4. Trim-and-fill funnel plot for included studies in this meta-analysis.



Discussion

Principal Findings

Among recent studies using app-based interventions to promote physical activity, we observed a significant degree of underreporting on several RE-AIM dimensions, which limits researchers' and policy makers' ability to assess the generalizability of the research results. In addition, the interventions in this literature, in general, had more explanatory rather than pragmatic designs, which further limits our ability to forecast how successful these interventions would be in promoting physical activity if implemented among the general population. Finally, the aggregate study results showed a small but significant improvement in participants' physical activity. However, treatment effect sizes varied according to the PRECIS-2 classification, where the more pragmatic trials produced smaller treatment effects on physical activity. Taken together, these findings suggest that app-based physical activity interventions would have limited efficacy in promoting physical activity if more widely scaled and adopted among the general population, suggesting that more pragmatic study designs are needed to increase the transferability from research to practice. The recommendations provided by Blackman et al [16] should be used more widely by researchers in this literature when designing and reporting study findings.

RE-AIM Evaluation and PRECIS-2 Assessment

RE-AIM Evaluation

Our findings build on a prior review of mHealth physical activity interventions that also observed a lack of reporting on study characteristics and research findings in this literature [16]. Without sufficient information on these important study dimensions, the previous review was unable to determine the generalizability of the research findings at that time. Our more detailed and updated review demonstrates that only small improvements in transparency and the reporting of study characteristics have been achieved in mHealth physical activity research since then.

Our finding that recent mHealth physical activity studies lack transparency builds on similar observations reported in reviews of physical activity interventions using both mHealth and other intervention tools [47,49,79]. Specifically, the review by Blackman et al [16] on the mHealth physical activity literature found that few studies reported on the maintenance of intervention effects and the degree of implementation fidelity. In addition, the review by Harden et al [49] on group-based physical activity interventions showed that external validity factors were consistently underreported, and the review by Burke et al [47] on physical activity interventions for adults with spinal cord injuries found that several items within the Adoption and Maintenance dimensions of RE-AIM were not reported in any study, limiting the generalizability of these studies.

Two specific areas of underreporting in the mHealth physical activity studies that we reviewed were in the Adoption and Maintenance dimensions. The lack of reported information on the ability of health care providers to adopt these app-based physical activity intervention tool or tools significantly limits the willingness of clinicians and organizations to implement these new intervention approaches [16,47,49]. More pragmatic study designs with greater reporting of the Adoption and Maintenance dimensions are needed to increase the implementation of these mHealth tools in real-world settings. In addition, none of the studies reported sufficiently the cost of the intervention (in terms of either time or money), making it difficult to assess the benefit versus cost of these tools. Rubin et al [80] noted that prior complications experienced when integrating mHealth technologies into clinical practice have likely increased providers' hesitancy to adopt new mHealth strategies. Therefore, we believe that increased reporting of interventions' organizational requirements and costs (eg, required staff qualifications, equipment for delivery and analysis, cost of acquiring the intervention tools, and maintenance) would increase the applicability of this research.

PRECIS-2 Assessment

With regard to the PRECIS-2 results, our domain-specific assessments suggest that these recent studies testing app-based physical activity interventions tend to be primarily explanatory in nature. To combat a lack of app engagement, many studies used additional text messages or email reminders to reengage participants with the interventions' app. These additional intervention components lowered our assessment of pragmatism, as it is not clear how well these methods can be widely implemented in usual care practices. Although the apps were considered relatively pragmatic in terms of their ease of accessibility, many studies also used frequent assessments and in-person intervention components or brought participants into research-specific facilities, limiting their overall level of pragmatism. Importantly, the use of device-based physical activity measures did not influence PRECIS-2 scores, as these devices are increasingly available and integrated into usual care.

Challenges and Adaptations of RE-AIM and PRECIS-2

To address the underreporting of study characteristics, we combined the main intervention report with additional documents available on the web but found few additional study details through these additional sources; thus, we want to emphasize that a greater "consensus around the use of frameworks and checklists across scientific fields and journals" is still needed [47]. We also expanded the original RE-AIM framework to include a third scoring category (*inadequately/insufficiently reported)* but found that assessing this added nuance in reporting adds substantially more work to the review process. Therefore, we refer readers and future reviewers to the ongoing creation of domain-specific review tools [81], which will hopefully be able to strike a better balance between researcher burden and improved accuracy.

Meta-analysis

Overall, these recent app-based physical activity interventions produced small but significant increases in participants' physical activity. This finding is in line with the results of previous reviews that also found a small and significant effect of app-based interventions on promoting physical activity [31-33,82]. In addition, our meta-analysis found that study effect sizes were not significantly different between interventions with durations longer than 8 weeks compared with those with shorter durations (Multimedia Appendix 7), which suggests that duration alone is not a predictor of a successful physical activity intervention and that additional approaches and intervention tools are still needed to change and maintain physical activity increases. Finally, a few of these studies were able to demonstrate, or even assess, the maintenance of physical activity after the interventions were withdrawn. This finding emphasizes the need for an improved understanding of physical activity habits and the maintenance of initial behavioral change.

The lack of evidence for an optimal physical activity intervention duration and for the maintenance of physical activity increases has been noted in previous reviews of the mHealth literature. Contrary to our findings, Romeo et al [32] found that the most effective physical activity interventions had durations longer than 8 weeks. In addition, the review by

Schoeppe et al [33] on app-based health interventions showed the greatest effects among interventions for up to 3 months in duration. The discrepancy between our results and those of previous studies demonstrates the need for more evidence on the optimal intervention duration. With regard to the maintenance of intervention effects, a recent systematic review by Pradal-Cano et al [82] described the need for longer-term studies to observe the maintenance of intervention effects after the intervention components are withdrawn. Among the studies reviewed by Pradal-Cano et al [58,66,73], only 3 reported on the maintenance of intervention effects at least 6 months after the intervention was withdrawn, and there were fixed findings on maintenance among these studies.

Strengths and Limitations

Adapting 2 complementary implementation science tools to better understand the generalizability and applicability of app-based physical activity intervention findings is a key strength of this review; however, this review is not without limitations. First, our literature search identified a relatively small number of unique interventions, which limited the power of our statistical methods. Second, the included studies significantly varied in terms of design parameters (eg, sampling frame and intervention components) and methodological parameters (eg, outcome measures). This considerable heterogeneity was identified in the meta-analyses and indicated the difficulties in synthesizing this literature. Although we focused only on app-based physical activity interventions, most interventions incorporated additional intervention components, precluding us from isolating the individual effect of the app on physical activity. Third, our literature search was performed in 2020, and more studies using mobile apps to increase physical activity have been published since then [83-85]. Although it is beyond the scope of this paper to incorporate these studies into our complete analyses, they all provide additional evidence that mobile apps can improve physical activity. In addition, one of these recent studies reported long-term behavioral maintenance outcomes [85], which is an important step in the mHealth app literature. Another important limitation is that most included studies targeted adults (19/22, 86%), which limits the generalizability of our findings to physical activity interventions among younger and older populations. Fourth, the significant degree of underreporting of study characteristics limited our ability to assess treatment moderation by individual RE-AIM dimensions, which is an important area for future research. Finally, our statistical analyses indicated the presence of a publication bias, potentially compromising the robustness of our findings. However, subsequent trim-and-fill analyses suggested that the overall treatment effect was only slightly reduced when attempting to account for these missing studies.

Conclusions

This review highlights important limitations in the mHealth literature that uses app-based interventions to promote physical activity. Specifically, studies continue to underreport several key study characteristics that are necessary to determine the generalizability and scalability of these intervention approaches. Importantly, more pragmatic study designs are needed to help researchers and policy makers confidently implement app-based tools in standard care practice. In addition, studies with different intervention durations were equally effective in increasing physical activity, suggesting that additional intervention methods and approaches are necessary to improve the maintenance and growth of initial physical activity improvements.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.
[DOC File , 68 KB - mhealth_v11i1e43162_app1.doc]

Multimedia Appendix 2
Search strategy for all electronic databases.
[PDF File (Adobe PDF File), 108 KB - mhealth_v11i1e43162_app2.pdf]

Multimedia Appendix 3
Data extraction form (general study characteristics).
[PDF File (Adobe PDF File), 76 KB - mhealth_v11i1e43162_app3.pdf]

Multimedia Appendix 4
Combined coding sheet (adapted Reach, Effectiveness, Adoption, Implementation, Maintenance [RE-AIM] and Pragmatic-Explanatory Continuum Indicator Summary-2 [PRECIS-2]).
[PDF File (Adobe PDF File), 158 KB - mhealth_v11i1e43162_app4.pdf]

Multimedia Appendix 5
Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) and Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) scoring.

[PDF File (Adobe PDF File), 815 KB - mhealth_v11i1e43162_app5.pdf]

Multimedia Appendix 6
Extracted study characteristics.
[DOCX File , 331 KB - mhealth_v11i1e43162_app6.docx]

Multimedia Appendix 7
Additional meta-analysis results.
[XLSX File (Microsoft Excel File), 260 KB - mhealth_v11i1e43162_app7.xlsx]

References

1. Warburton DE, Bredin SS. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
2. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Global Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
4. World Health Organization. Global Action Plan on Physical Activity 2018-2030. Geneva: World Health Organization; 2019.
5. Aarons GA, Sklar M, Mustanski B, Benbow N, Brown CH. "Scaling-out" evidence-based interventions to new populations or new health care delivery systems. Implement Sci 2017 Sep 06;12(1):111 [FREE Full text] [doi: 10.1186/s13012-017-0640-6] [Medline: 28877746]
6. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838 [FREE Full text] [doi: 10.3390/ijerph15122838] [Medline: 30551555]
7. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
8. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
9. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
10. Roess A. The promise, growth, and reality of mobile health - another data-free zone. N Engl J Med 2017 Nov 23;377(21):2010-2011. [doi: 10.1056/NEJMp1713180] [Medline: 29116869]
11. Clement J. Number of apps available in leading app stores as of 2nd quarter 2020. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2023-04-28]
12. The growing value of digital health evidence and impact on human health and the healthcare system. IQVIA. 2017 Nov 7. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2023-04-28]
13. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
14. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
15. Lee J. Future of the smartphone for patients and healthcare providers. Healthc Inform Res 2016 Jan;22(1):1-2 [FREE Full text] [doi: 10.4258/hir.2016.22.1.1] [Medline: 26893944]
16. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, et al. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. J Med Internet Res 2013 Oct 04;15(10):e224 [FREE Full text] [doi: 10.2196/jmir.2745] [Medline: 24095951]
17. Bull S. Beyond acceptability and feasibility: moving mHealth into impact. Mhealth 2016 Dec 19;2:45 [FREE Full text] [doi: 10.21037/mhealth.2016.12.02] [Medline: 28293615]
18. Huebschmann AG, Leavitt IM, Glasgow RE. Making health research matter: a call to increase attention to external validity. Annu Rev Public Health 2019 Apr 01;40:45-63. [doi: 10.1146/annurev-publhealth-040218-043945] [Medline: 30664836]
19. Estabrooks P, Gyurcsik N. Evaluating the impact of behavioral interventions that target physical activity: issues of generalizability and public health. Psychol Sport Exercise 2003 Jan;4(1):41-55. [doi: 10.1016/s1469-0292(02)00016-x]
20. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. Health Educ Behav 2013 Jun 24;40(3):257-265. [doi: 10.1177/1090198113486805] [Medline: 23709579]

21. Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. J Orthop Sports Phys Ther 2003 Apr;33(4):163-165. [doi: 10.2519/jospt.2003.33.4.163] [Medline: 12723672]
22. Gartlehner G, Hansen R, Nissman D, Lohr K, Carey T. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.
23. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003 Dec 22;3(1):28 [FREE Full text] [doi: 10.1186/1471-2288-3-28] [Medline: 14690550]
24. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015 May 08;350:h2147 [FREE Full text] [doi: 10.1136/bmj.h2147] [Medline: 25956159]
25. Arvey SR, Fernandez ME. Identifying the core elements of effective community health worker programs: a research agenda. Am J Public Health 2012 Sep;102(9):1633-1637 [FREE Full text] [doi: 10.2105/AJPH.2012.300649] [Medline: 22813092]
26. Estabrooks PA, Brownson RC, Pronk NP. Dissemination and implementation science for public health professionals: an overview and call to action. Prev Chronic Dis 2018 Dec 20;15:E162 [FREE Full text] [doi: 10.5888/pcd15.180525] [Medline: 30576272]
27. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
29. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
30. Rothstein H, Sutton A, Borenstein M. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chichester, UK: John Wiley & Sons, Ltd; 2005.
31. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
32. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
33. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
34. Coughlin SS, Stewart J. Use of consumer wearable devices to promote physical activity: a review of health intervention studies. J Environ Health Sci 2016 Nov;2(6):1-6 [FREE Full text] [doi: 10.15436/2378-6841.16.1123] [Medline: 28428979]
35. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):021 [FREE Full text] [Medline: 27034992]
36. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
37. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
38. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
39. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep 26;22(3):451-469 [FREE Full text] [doi: 10.1177/1460458214567004] [Medline: 25649783]
40. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
41. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
42. Haskell W. Physical activity by self-report: a brief history and future issues. J Phys Act Health 2012 Jan;9 Suppl 1:S5-10 [FREE Full text] [doi: 10.1123/jpah.9.s1.s5] [Medline: 22287448]
43. Matthews C, Hagströmer M, Pober D, Bowles H. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]

44. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126 [FREE Full text] [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
45. Gaglio B, Glasgow R. Evaluation approaches for dissemination and implementation research. In: Dissemination and Implementation Research in Health: Translating Science to Practice, 2nd edition. New York: Oxford University Press; 2017.
46. Wallace B, Small K, Brodley C, Lau J, Trikalinos T. Deploying an interactive machine learning system in an evidence-based practice center. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. 2012 Presented at: IHI '12: ACM International Health Informatics Symposium; Jan 28 - 30, 2012; Miami Florida USA. [doi: 10.1145/2110363.2110464]
47. Burke SM, Tomasone JR, Scime NV, Ma JK, Harden SM, Wolfe DL. Physical activity self-management interventions for adults with spinal cord injury: part 2 – Exploring the generalizability of findings from research to practice. Psychol Sport Exercise 2018 Jul;37:286-295. [doi: 10.1016/j.psychsport.2018.02.013]
48. Gaglio B, Phillips SM, Heurtin-Roberts S, Sanchez MA, Glasgow RE. How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. Implement Sci 2014 Aug 28;9:96 [FREE Full text] [doi: 10.1186/s13012-014-0096-x] [Medline: 25163664]
49. Harden SM, Burke SM, Haile AM, Estabrooks PA. Generalizing the findings from group dynamics-based physical activity research to practice settings: what do we know? Eval Health Prof 2015 Mar 28;38(1):3-14. [doi: 10.1177/0163278713488117] [Medline: 23716732]
50. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
51. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Cambridge, Massachusetts, United States: Academic Press; 1969.
53. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
54. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
55. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
56. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
57. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone "app"-delivered randomized factorial trial targeting physical activity in adults. J Behav Med 2017 Oct;40(5):712-729. [doi: 10.1007/s10865-017-9838-y] [Medline: 28255750]
58. Fukuoka Y, Haskell W, Lin F, Vittinghoff E. Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. JAMA Netw Open 2019 May 03;2(5):e194281 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.4281] [Medline: 31125101]
59. Gremaud AL, Carr LJ, Simmering JE, Evans NJ, Cremer JF, Segre AM, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. JAHA 2018 Jul 03;7(13):e007735. [doi: 10.1161/jaha.117.007735]
60. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. PLoS One 2016;11(6):e0156370 [FREE Full text] [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
61. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
62. Pope ZC, Gao Z. Feasibility of smartphone application- and social media-based intervention on college students' health outcomes: A pilot randomized trial. J Am Coll Health 2022 Jan 09;70(1):89-98. [doi: 10.1080/07448481.2020.1726925] [Medline: 32150514]
63. Schade S, Mahoney J, Spotts A, Greenauer N, Veerabhadrappa P. Pokémon Go did not increase step count or distance travelled among college students. Human Movement 2020;21(2):64-70. [doi: 10.5114/hm.2020.89916]
64. Zhang J, Jemmott Iii JB. Mobile app-based small-group physical activity intervention for young African American women: a pilot randomized controlled trial. Prev Sci 2019 Aug;20(6):863-872. [doi: 10.1007/s11121-019-01006-4] [Medline: 30788692]
65. Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28 [FREE Full text] [doi: 10.2196/mhealth.9117] [Medline: 29371177]

66. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
67. Robertson J, Macvean A, Fawkner S, Baker G, Jepson RG. Savouring our mistakes: learning from the FitQuest project. Int J Child Comput Interact 2018 Jun;16:55-67. [doi: 10.1016/j.ijcci.2017.12.003]
68. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and process evaluation of a smartphone app to promote an active lifestyle in lower educated working young adults: cluster randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [FREE Full text] [doi: 10.2196/10003] [Medline: 30143477]
69. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
70. Kitagawa T, Higuchi Y, Todo E, Ueda T, Ando S, Murakami T. Tailored feedback reduced prolonged sitting time and improved the health of housewives: a single-blind randomized controlled pilot study. Women Health 2020 Feb 21;60(2):212-223. [doi: 10.1080/03630242.2019.1616043] [Medline: 31113310]
71. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
72. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of gamified mobile service aimed at physical activation in young men: population-based randomized controlled study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
73. Garcia-Ortiz L, Recio-Rodriguez JI, Agudo-Conde C, Patino-Alonso MC, Maderuelo-Fernandez J, Repiso Gento I, EVIDENT Investigators Group, Mobilizing Minds Research Group. Long-term effectiveness of a smartphone app for improving healthy lifestyles in general population in primary care: randomized controlled trial (Evident II study). JMIR Mhealth Uhealth 2018 Apr 27;6(4):e107 [FREE Full text] [doi: 10.2196/mhealth.9218] [Medline: 29702473]
74. Garde A, Chowdhury M, Rollinson AU, Johnson M, Prescod P, Chanoine JP, et al. A multi-week assessment of a mobile exergame intervention in an elementary school. Games Health J 2018 Feb;7(1):1-8. [doi: 10.1089/g4h.2017.0023] [Medline: 29394109]
75. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
76. Harries T, Eslambolchilar P, Rettie R, Stride C, Walton S, van Woerden HC. Effectiveness of a smartphone app in increasing physical activity amongst male adults: a randomised controlled trial. BMC Public Health 2016 Sep 02;16(1):925 [FREE Full text] [doi: 10.1186/s12889-016-3593-9] [Medline: 27590255]
77. Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J. Face-to-face versus mobile versus blended weight loss program: randomized clinical trial. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e14 [FREE Full text] [doi: 10.2196/mhealth.7713] [Medline: 29326093]
78. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MD, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
79. Galaviz KI, Harden SM, Smith E, Blackman KC, Berrey LM, Mama SK, et al. Physical activity promotion in Latin American populations: a systematic review on issues of internal and external validity. Int J Behav Nutr Phys Act 2014 Jun 17;11(1):77 [FREE Full text] [doi: 10.1186/1479-5868-11-77] [Medline: 24938641]
80. Rubin DS, Severin R, Arena R, Bond S. Leveraging technology to move more and sit less. Prog Cardiovasc Dis 2021 Jan;64:55-63. [doi: 10.1016/j.pcad.2020.10.007] [Medline: 33129794]
81. Norton WE, Loudon K, Chambers DA, Zwarenstein M. Designing provider-focused implementation trials with purpose and intent: introducing the PRECIS-2-PS tool. Implement Sci 2021 Jan 07;16(1):7 [FREE Full text] [doi: 10.1186/s13012-020-01075-y] [Medline: 33413489]
82. Pradal-Cano L, Lozano-Ruiz C, Pereyra-Rodríguez JJ, Saigí-Rubió F, Bach-Faig A, Esquius L, et al. Using mobile applications to increase physical activity: a systematic review. Int J Environ Res Public Health 2020 Nov 07;17(21):8238 [FREE Full text] [doi: 10.3390/ijerph17218238] [Medline: 33171871]
83. Vetrovsky T, Borowiec A, Juřík R, Wahlich C, Śmigielski W, Steffl M, et al. Do physical activity interventions combining self-monitoring with other components provide an additional benefit compared with self-monitoring alone? A systematic review and meta-analysis. Br J Sports Med 2022 Dec 07;56(23):1366-1374 [FREE Full text] [doi: 10.1136/bjsports-2021-105198] [Medline: 36396151]
84. Lau E, Mitchell M, Faulkner G. Long-term usage of a commercial mHealth app: a "multiple-lives" perspective. Front Public Health 2022;10:914433 [FREE Full text] [doi: 10.3389/fpubh.2022.914433] [Medline: 36438245]

85. Yoshimura E, Tajiri E, Michiwaki R, Matsumoto N, Hatamoto Y, Tanaka S. Long-term effects of the use of a step count-specific smartphone app on physical activity and weight loss: randomized controlled clinical trial. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35628 [FREE Full text] [doi: 10.2196/35628] [Medline: 36279159]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
PRECIS-2: Pragmatic-Explanatory Continuum Indicator Summary-2
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance
RoB: risk-of-bias

Edited by L Buis; submitted 01.10.22; peer-reviewed by M Kapsetaki, D Liebovitz, R Gal; comments to author 20.12.22; revised version received 20.02.23; accepted 14.03.23; published 04.05.23.

Please cite as:
Stecher C, Pfisterer B, Harden SM, Epstein D, Hirschmann JM, Wunsch K, Buman MP
Assessing the Pragmatic Nature of Mobile Health Interventions Promoting Physical Activity: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2023;11:e43162
URL: https://mhealth.jmir.org/2023/1/e43162
doi:10.2196/43162
PMID:37140972

Review

Participant Engagement in Microrandomized Trials of mHealth Interventions: Scoping Review

Utek Leong[1], BSocSci; Bibhas Chakraborty[2,3,4], PhD

[1]Department of Psychology, National University of Singapore, Singapore, Singapore

[2]Centre for Quantitative Medicine and Program in Health Services and Systems Research, Duke-NUS Medical School, National University of Singapore, Singapore, Singapore

[3]Department of Statistics and Data Science, National University of Singapore, Singapore, Singapore

[4]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

Corresponding Author:
Bibhas Chakraborty, PhD
Centre for Quantitative Medicine and Program in Health Services and Systems Research
Duke-NUS Medical School
National University of Singapore
8 College Road, #06-31
Singapore, 169857
Singapore
Phone: 65 66016502
Email: bibhas.chakraborty@duke-nus.edu.sg

Abstract

Background: Microrandomized trials (MRTs) have emerged as the gold standard for the development and evaluation of multicomponent, adaptive mobile health (mHealth) interventions. However, not much is known about the state of participant engagement measurement in MRTs of mHealth interventions.

Objective: In this scoping review, we aimed to quantify the proportion of existing or planned MRTs of mHealth interventions to date that have assessed (or have planned to assess) engagement. In addition, for the trials that have explicitly assessed (or have planned to assess) engagement, we aimed to investigate how engagement has been operationalized and to identify the factors that have been studied as determinants of engagement in MRTs of mHealth interventions.

Methods: We conducted a broad search for MRTs of mHealth interventions in 5 databases and manually searched preprint servers and trial registries. Study characteristics of each included evidence source were extracted. We coded and categorized these data to identify how engagement has been operationalized and which determinants, moderators, and covariates have been assessed in existing MRTs.

Results: Our database and manual search yielded 22 eligible evidence sources. Most of these studies (14/22, 64%) were designed to evaluate the effects of intervention components. The median sample size of the included MRTs was 110.5. At least 1 explicit measure of engagement was included in 91% (20/22) of the included MRTs. We found that objective measures such as system usage data (16/20, 80%) and sensor data (7/20, 35%) are the most common methods of measuring engagement. All studies included at least 1 measure of the physical facet of engagement, but the affective and cognitive facets of engagement have largely been neglected (only measured by 1 study each). Most studies measured engagement with the mHealth intervention (Little e) and not with the health behavior of interest (Big E). Only 6 (30%) of the 20 studies that measured engagement assessed the determinants of engagement in MRTs of mHealth interventions; notification-related variables were the most common determinants of engagement assessed (4/6, 67% studies). Of the 6 studies, 3 (50%) examined the moderators of participant engagement—2 studies investigated time-related moderators exclusively, and 1 study planned to investigate a comprehensive set of physiological and psychosocial moderators in addition to time-related moderators.

Conclusions: Although the measurement of participant engagement in MRTs of mHealth interventions is prevalent, there is a need for future trials to diversify the measurement of engagement. There is also a need for researchers to address the lack of attention to how engagement is determined and moderated. We hope that by mapping the state of engagement measurement in existing MRTs of mHealth interventions, this review will encourage researchers to pay more attention to these issues when planning for engagement measurement in future trials.

(JMIR Mhealth Uhealth 2023;11:e44685) doi:10.2196/44685

KEYWORDS

microrandomized trials; engagement; adherence; mobile health; mHealth interventions; mobile phone

Introduction

Background

In the past decade, digital solutions that leverage mobile technologies to improve health and well-being have become increasingly popular and have emerged as promising adjuncts to traditional health care provision [1]. These so-called mobile health (mHealth) interventions generally involve the use of mobile technologies such as mobile apps, SMS text messaging, and wearable devices to improve patient health outcomes by delivering health-related intervention content. Mounting evidence suggests that mHealth interventions are largely effective for treating chronic health conditions [2,3] and for preventing unhealthy behaviors [4]. Effectiveness aside, it is not difficult to see why mHealth interventions are so popular; mHealth interventions are highly scalable and cost-efficient [1]. High rates of mobile ownership worldwide also signal the potential for mHealth interventions to reach a diverse audience, including the underserved; however, we must acknowledge that there are barriers to access (such as the lack of internet access) that prevent mHealth interventions from being truly equitable [5].

Recently, more sophisticated mHealth interventions have been proposed to take advantage of the technological advances in mobile technology. These novel interventions (such as just-in-time adaptive interventions) tend to be multicomponent, that is, they tend to involve the manipulation of ≥2 components hypothesized to have a treatment effect. They also tend to be adaptive, in the sense that components of the intervention (eg, its content and timing of delivery) can change in response to some input data provided by the user (tailoring data collected from surveys or sensors). To make this concrete, let us consider a hypothetical mHealth intervention designed to reduce the severity of depression symptoms by sending daily motivational messages via SMS text messaging. The intervention is said to be multicomponent if both message content and timing of SMS delivery are thought to be *active* ingredients that can influence depression symptom severity. Such an intervention could be made adaptive if daily message content is tailored to the participant's mood the night before such that if a given participant had high negative mood the night before, a more strongly worded motivational message would be sent the next day. Unfortunately, conventional randomized controlled trials (RCTs) cannot be used to develop and optimize these interventions because they do not allow researchers to separate the treatment effect of individual treatment components from the overall treatment effect. In addition, RCTs do not allow researchers to investigate time-varying effects, which is of interest when the goal is to identify the optimal time to administer an intervention component [6]. Therefore, if the RCT design is used to study the aforementioned hypothetical mHealth intervention, we will only be able to estimate the *overall* treatment effect of sending motivational messages on depression symptom severity and not the specific treatment effect of message content and timing of SMS delivery on the severity of depressive symptoms.

To address these limitations of the RCT design, several cutting-edge trial designs have been proposed in recent years. The microrandomized trial (MRT) design in particular has gained considerable traction as a way to optimize multicomponent and adaptive mHealth interventions (including but not limited to just-in-time adaptive interventions) [6-9]. Essentially, the MRT design involves the repeated random assignment of participants to different intervention options of a single or multiple intervention components; therefore, an MRT of our hypothetical multicomponent motivational SMS text messaging intervention would entail repeatedly randomizing participants to receive *different types of motivational messages* at *different times daily*. This repeated random assignment then facilitates the estimation of the time-varying causal effects of each specific treatment component [6], that is, we can estimate the treatment effect of message content and timing of SMS text message delivery on the severity of depressive symptoms. Therefore, unlike RCTs, MRTs allow researchers to investigate the effectiveness of specific components of mHealth interventions, which could be informative for theory, future research, and intervention optimization. Notably, RCTs and MRTs are not mutually exclusive. One additional benefit of the MRT design is that it can be easily embedded within the treatment arm of a conventional RCT; therefore, the overall treatment effect and the effect of specific intervention components can be tested simultaneously.

Regardless of the trial design used, the measurement of participant engagement is integral to understanding the feasibility of mHealth interventions. This is because engagement with the constituent digital or nondigital intervention stimuli and tasks of an mHealth intervention is necessary for the individual to experience the intended distal health outcomes of the intervention [10,11]. The measurement of engagement, however, is not straightforward. Engagement, like many other psychological constructs, is an abstract and fuzzy concept that is not directly measurable (unlike, for example, the measurement of height). To measure engagement, researchers must first operationalize engagement, that is, define engagement in measurable terms [12]. To unpack how exactly engagement with mHealth interventions can be operationalized, it is instructive to consider *how* engagement can be measured, *which* kinds of engagement can be measured, and *what* levels of engagement can be measured.

Measures of Engagement

According to Yardley et al [13] and then Short et al [14], there are 7 methods of engagement measurement that researchers can use to obtain a sense of participant engagement in their digital interventions: self-report questionnaires, ecological momentary assessments (EMAs), qualitative methods, system usage data, sensor data, social media data, and psychophysiological measures. The measurement of engagement via self-report questionnaires and EMAs involves directly asking participants to report (via single items or questionnaires) their subjective experience of using the digital intervention or their use of the intervention. Qualitative methods of engagement, by contrast, involve the inference of engagement from qualitative sources (such as written responses and semistructured interviews). Measuring engagement via system usage data involves the quantification of how the digital intervention is used through metrics including, but not limited to, the number of log-ins, time spent on the intervention, and number of modules viewed. Engagement can also be measured by analyzing passively collected social media and sensor data if social media and sensors (eg, pedometers and heart rate sensors) are a feature of the intervention. Finally, psychophysiological measures of engagement involve the use of measures such as electroencephalography, eye tracking, or functional magnetic resonance imaging to infer engagement from neural and physiological activity.

Facets of Engagement

Engagement is thought to be a multifaceted construct composed of 3 distinct facets—physical, affective, and cognitive [11,14]. The physical facet of engagement refers to the "actual performance of an activity or task" [11]. The affective facet by contrast is thought to capture "a wide range of positive affective reactions to a task or activity, from feeling pride, enthusiasm, and satisfaction, to affective states that may underlie more enduring experiences of attachment, identification, and commitment" [11]. Finally, the cognitive facet of engagement is thought to refer to "selective attention and processing of information related to a task or activity" [11]. These facets represent distinct *kinds* of engagement that can be measured in mHealth interventions.

Levels of Engagement

When discussing the measurement of engagement in digital interventions, it is crucial to ask the question, "engagement with what?" [11]. This is because engagement measures can either be measures of engagement with the features and the active ingredients of the intervention or engagement with the health behavior of interest. Formally, Cole-Lewis et al [15] termed engagement with the mHealth intervention as "Little e" and engagement with the health behavior of interest as "Big E"; elsewhere, the terms microengagement and macroengagement are used instead [13]. In essence, Little e and Big E represent 2 distinct levels of engagement, where the 7 methods of engagement outlined in the *Measures of Engagement* section can be applied to measure participant engagement in the mHealth intervention context.

This Study

Given the importance of engagement to mHealth interventions, researchers have endeavored to understand how engagement has been conceptualized and operationalized in studies evaluating mHealth interventions. For instance, Pham et al [16] recently reviewed how engagement has been defined and measured in mHealth apps for chronic conditions. Perski et al [10], by contrast, reviewed how engagement was conceptualized in digital behavior change interventions (their review was not limited to mHealth interventions; it included other digital interventions). Other recent reviews evaluated the measurement of engagement in mHealth interventions designed for specific health conditions [17,18]. However, none of these reviews examined mHealth interventions evaluated by MRTs, perhaps owing to the relative infancy of the trial design. Thus, not much is known about the state of participant engagement measurement in MRTs of mHealth interventions. Furthermore, it is not yet known what kinds of factors have been studied as determinants of engagement in these MRTs.

Therefore, we conducted a scoping review to map this relatively new research area. We chose to conduct a scoping review as we expected that only a handful of mHealth intervention MRTs have been conducted to date—too few to be meaningfully synthesized with a systematic review. This scoping review aimed to address 3 review questions:

1. What proportion of existing (or planned) MRTs of mHealth interventions to date have assessed (or have planned to assess) engagement?
2. How has engagement been operationalized in existing (or planned) MRTs of mHealth interventions that have assessed (or have planned to assess) engagement?
3. In existing (or planned) MRTs of mHealth interventions that have assessed (or have planned to assess) engagement, what kind of factors have been studied as determinants of engagement?

Methods

Protocol and Registration

The protocol for this scoping review was developed using the Joanna Briggs Institute Manual for Evidence Synthesis [19] and was designed to ensure adherence to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews) guidelines [20]. The protocol and its appendices were prospectively registered with the Open Science Framework (OSF) on June 30, 2022 [21].

Eligibility Criteria

We prioritized the inclusion of papers published in peer-reviewed journals. We included preprints, trial protocols, and dissertations (this was mistakenly left out of the "Types of Sources" section of our protocol [21]) only if no corresponding peer-reviewed journal articles were available. Conference abstracts were excluded from this scoping review.

All papers fulfilling these criteria to date were considered for inclusion if they were written in English and if they reported MRTs of mHealth interventions. We also included any secondary analyses of mHealth intervention engagement data collected from an MRT if the primary analysis (if available) did not report the assessment of engagement in detail. We defined mHealth interventions as any intervention designed to improve health outcomes through (though not limited to) the modification of health behavior (such as physical activity or treatment adherence), the improvement of patient knowledge, health monitoring, and the reduction of psychological distress via mobile technology such as SMS text messaging; mobile phone apps; or devices (including but not limited to smartwatches, wearables, and sensors) [1].

As the review's objectives concerned the assessment of engagement in MRTs of mHealth interventions, we included all studies in which authors *explicitly* attempted or claimed to quantitatively or qualitatively measure the participation in or use of mHealth interventions directly (by measuring participation in or performance of mHealth intervention activities or components) or indirectly (using measurements derived from non–intervention-related activities or components as a proxy), regardless of how they actually defined and measured engagement (eg, if they use alternative terms like adherence).

Information Sources and Search Strategy

We conducted a broad search for all published MRTs of mHealth interventions to date (the search was initially conducted on July 13, 2022, and again on September 28, 2022) by searching the following 5 bibliographic databases: MEDLINE (via PubMed), Embase, PsycINFO, CINAHL, and Cochrane Library. The search strategy was originally developed for MEDLINE, and we consulted an academic librarian from the National University of Singapore to ensure that the search strategy was comprehensive and sound. This search strategy was then translated for the 4 other databases (only syntax was changed to accommodate differences in search engines; keywords remained the same). Although only 1 broad search was eventually performed, it must be noted that we registered 2 separate searches in our protocol—1 for all published MRTs of mHealth interventions to date and 1 fine-grained search for MRTs of mHealth interventions that have assessed (or have planned to assess) engagement. During our search process, we realized that the latter search was redundant as it was nested within the former (because we used the Boolean operator AND between the mHealth intervention search terms and the engagement-related search terms). Therefore, we condensed the 2 planned searches into 1 by using the Boolean operator OR instead, such that our database searches indexed any MRTs that mentioned mHealth interventions or engagement-related terms. The comprehensive search strategies for all 5 databases (and their respective previous iterations) can be found on OSF [21].

To search for gray literature and unpublished studies, we searched the reference lists of included studies for any additional sources not indexed by our database search. We also posted an open call for unpublished MRTs of mHealth interventions on Twitter and contacted known experts of the MRT design to request unpublished and file-drawered studies. Finally, we performed a search (similarly, this search was initially conducted on July 13, 2022, and again on September 28, 2022) of MRTs of mHealth intervention on 2 preprint servers (PsyArXiv and medRxiv; we added this search during our search process to ensure the comprehensiveness of our gray literature search) and on 2 clinical trial registries, ClinicalTrials.gov (as detailed in our protocol) and the International Clinical Trials Registry Platform (this was added during the search process as well). The following search terms were used: "microrandomised," "microrandomized," "micro-randomised," and "micro-randomized."

Selection of Sources of Evidence

The results of the searches described in the previous section were imported into EndNote (version 20; Clarivate; we did not use Zotero as planned because of technical difficulties) for source selection and screening. The titles and abstracts of all potential evidence sources were first screened for eligibility. Eligible sources were then subjected to a full-text screening. Before the 2 screening stages, both authors discussed a subset of the search results (5 titles and abstracts and 4 full-text articles) to calibrate the selection of evidence sources. UL performed the screening using the eligibility criteria, and BC verified the screening at both stages. Any disagreements were resolved by consensus.

Data Charting Process and Data Items

As described in our protocol [21], we developed an initial data extraction form (a Microsoft Excel [Microsoft Corporation] spreadsheet) to chart the data from eligible evidence sources to obtain the information necessary to answer our review questions. Both authors (UL and BC) piloted this initial data extraction form with 4 included articles to calibrate the charting process and to ensure that relevant data items were captured by the form. This form was continuously updated during the charting process through the discussion of the extracted results. UL performed data charting, and BC verified the charted data for all eligible evidence sources. Any disagreements were resolved by consensus.

The initial data collection form was designed to abstract the following information from each paper: whether the paper described a primary or secondary analysis of MRT data, type of paper, sample size of the MRT, sample characteristics, purpose of the study, type of mHealth intervention assessed, mode of delivery for the mHealth intervention, if engagement was or will be assessed, how engagement was operationalized (if assessed), if determinants of engagement were or will be assessed, and (if any) what determinants of engagement were or will be assessed; for comprehensiveness, we also charted any moderating variables and control variables (covariates) assessed.

After piloting the form and during the charting process, we included additional data items to capture the following information: primary and secondary (if any) outcomes of the study, randomization design of the MRT, frequency of microrandomization, and the overall duration of the MRT. The final version of the data extraction form is available on OSF [21].

Synthesis of Results

To quantify the proportion of existing and planned MRTs of mHealth interventions to date that have assessed (or have planned to assess) engagement, we tabulated the number of evidence sources charted to have assessed or planned to assess engagement. The included evidence sources were grouped by their purpose and presented in a tabular format. The mHealth interventions of each included evidence source were categorized based on their target. We used the following categories: mental health promotion, smoking cessation, physical activity promotion, sleep improvement, dietary lapse prevention or weight management behavior promotion, gambling reduction, and alcohol use reduction.

To understand how engagement has been operationalized in MRTs of mHealth interventions, we sought to determine *how* included evidence sources measured engagement, *which* kinds of engagement they measured, and *what* levels of engagement they measured. To determine *how* engagement has been measured, we classified *explicit* measures of engagement from each included source according to the methods of engagement measurement outlined by Short et al [14] described in the *Introduction* section. We combined the self-report questionnaires and EMA categories for parsimony, as they are largely similar methods of measuring engagement. To determine *which* kinds of engagement have been measured, we classified *explicit* measures of engagement by the facets (physical, affective, or cognitive) of engagement they appear to measure [11]. Finally, to determine *what* levels of engagement have been measured, we classified the *explicit* measures of engagement from each included source as Little e or Big E measures [15].

To identify the factors that have been studied as determinants of engagement in MRTs of mHealth interventions, we extracted the variables of interest, moderators, and covariates from each model (with a measure of engagement as the dependent variable) tested in each included source. We then organized these variables into the following categories: notification related (eg, type of prompt sent), time related (eg, days since the start of the intervention or day of the week), psychological, societal, health behavior related (eg, alcohol use), contextual (eg, location data), physiological (heart rate), demographic, anthropometric (eg, weight change), or task related (eg, intervention-related activities).

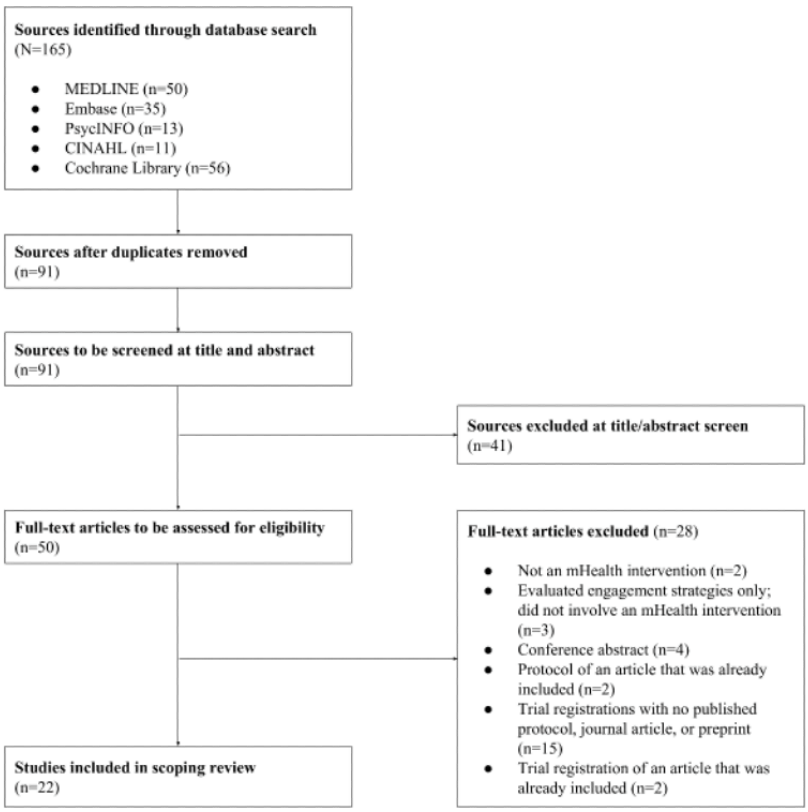
Results

Selection of Sources of Evidence

A total of 165 evidence sources were retrieved by our database search. After removing duplicates, 91 evidence sources were retained for further screening. During the title and abstract screening, 41 sources were excluded. Of the remaining 50 evidence sources, 28 were excluded at the full-text screening (Figure 1).

Notably, 17 of these sources excluded at full-text screening were trial registrations (a total of 19 trial registrations were retrieved by our database search of the Cochrane Library). A total of 15 (88%) of these 17 sources had no published protocol, journal article, or preprint; we performed a manual Google search of their respective trial identification numbers to confirm this. In total, 2 (12%) of these 17 sources were duplicate trial registrations, that is, a corresponding protocol, journal, article, or preprint for each registration was already indexed by our database search. Therefore, only 22 evidence sources identified by our database search were considered eligible for this scoping review. No additional studies were identified and included from our planned searches of gray literature and unpublished studies.

Figure 1. Evidence source selection flow diagram.



Characteristics of Sources of Evidence

All charted data described in the preceding section are available on OSF [21] and Multimedia Appendix 1 [22-43]. We present a subset of the charted data that are pertinent to our review questions.

Table 1 details the characteristics of each included evidence source. Of the 22 included sources, 12 (54%) were published journal articles, 8 (36%) were trial protocols, 1 (5%) was a preprint, and 1 (5%) was a dissertation. Only 1 evidence source was a secondary analysis of MRT data [22]. All included sources were published between 2018 and 2022. More than half of the included sources (14/22, 64%) were designed to evaluate the effect of intervention components. Physical activity promotion was the most common target of the mHealth interventions (8/22, 36%). Interventions were largely delivered via smartphone apps. The median sample size of the included MRTs was 110.5.

Table 1. Characteristics of included evidence sources.

Source and intervention type		Mode of delivery	Engagement assessed?
Evaluate effect of intervention components			
Aguilera et al [23], 2021	Mental health promotion	SMS	Yes
Battalio et al [24], 2021	Smoking cessation	App	Yes
Figueroa et al [25], 2022	Physical activity promotion	SMS, app	No
Goldstein et al [26], 2021	Dietary lapse prevention or weight management behavior promotion	App	Yes
Klasnja et al [27], 2021	Physical activity promotion	SMS	Yes
Klasnja et al [28], 2019	Physical activity promotion	App	Yes
Kramer et al [29], 2020	Physical activity promotion	App	Yes
Latham [30], 2021[a]	Sleep improvement	App	Yes
Jeganathan et al [31], 2022	Physical activity promotion	SMS[b]	Yes
NeCamp et al [32], 2020	Physical activity promotion, mental health promotion, and sleep improvement	App	No
Spruijt-Metz et al [33], 2022	Physical activity promotion	App	Yes
Wang et al [34], 2022	Physical activity promotion and sleep improvement	App	Yes
Dowling et al [35], 2022	Gambling reduction	App	Yes
Rodda et al [36], 2022	Gambling reduction	App	Yes
Evaluate strategies to improve engagement			
Bell et al [22], 2020	Alcohol use reduction	App	Yes
Bidargaddi et al [37], 2018	Mental health promotion	App	Yes
Nahum-Shani et al [38], 2021	Smoking cessation	App	Yes
Nordby et al [39], 2022	Mental health promotion	SMS	Yes
Evaluate feasibility and acceptability of intervention			
Militello et al [40], 2022	Mental health promotion	App	Yes
Yang et al [41], 2022	Smoking cessation	App	Yes
Describing engagement			
Hoel et al [42], 2022	Mental health promotion	App	Yes
Valle et al [43], 2020	Dietary lapse prevention or weight management behavior promotion	App	Yes

[a]This study was also designed to evaluate the feasibility and acceptability of its mobile health intervention.

[b]SMS text messages were delivered as smartphone and smartwatch notifications.

Synthesis of Results

Operationalization of Engagement

Overview

Of the 22 included sources, 20 (91%) explicitly included at least 1 measure of engagement; 2 (9%) studies did not claim to measure engagement at all [25,32]; NeCamp et al [32] did not do so because of technical limitations. Though we did not chart the different terms used to refer to participant engagement, we noticed during our full-text screening that some studies did indeed use alternative terms in place of the term "engagement," such as adherence [27] and investment [30].

Measures of Engagement

Table 2 summarizes the measures of engagement used in each study. Across all included studies, system usage data were by far the most frequently used measure of engagement. Sixteen (80%) out of the 20 studies that explicitly measured engagement included at least 1 measure of this category. Generally, researchers used 2 types of system usage data: (1) responsiveness to self-reports, logs, or EMAs [23,24,26,27,29,30,33,35-37,41,42] and (2) access or use of interventions [22,26,33,35,36,39-41,43].

Table 2. Measures of engagement used in microrandomized trials of mobile health (mHealth) interventions.

Source	SR[a] or EMA[b]	SU[c]	Sensor data	Qualitative methods	SM[d]	PP[e]
Evaluate effect of intervention components						
Aguilera et al [23], 2021		✓				
Battalio et al [24], 2021		✓	✓			
Goldstein et al [26], 2021		✓				
Klasnja et al [27], 2021		✓	✓			
Klasnja et al [28], 2019			✓			
Kramer et al [29], 2020		✓				
Latham [30], 2021[f]	✓	✓				
Jeganathan et al [31], 2022			✓			
Spruijt-Metz et al [33], 2022		✓	✓			
Wang et al [34], 2022			✓			
Dowling et al [35], 2022		✓				
Rodda et al [36], 2022		✓				
Evaluate strategies to improve engagement						
Bell et al [22], 2020		✓				
Bidargaddi et al [37], 2018		✓				
Nahum-Shani et al [38], 2021	✓					
Nordby et al [39], 2022	✓	✓				
Evaluate feasibility and acceptability of intervention						
Militello et al [40], 2022	✓	✓				
Yang et al [41], 2022		✓	✓			
Describing engagement						
Hoel et al [42], 2022		✓		✓		
Valle et al [43], 2020		✓				

[a]SR: self-report data.

[b]EMA: ecological momentary assessment.

[c]SU: system usage data.

[d]SM: social media data.

[e]PP: psychophysiological data.

[f]This study was also designed to evaluate the feasibility and acceptability of its mHealth intervention.

Sensor data were the second most common measure of engagement. Overall, 35% (7/20) of the studies that explicitly measured engagement included at least 1 measure of this category [24,27,28,31,33,34,41]. Wang et al [34], for example, measured the proportion of days in a week that participants wore the study's FitBit smartwatch to track their step counts and sleep duration.

Engagement was measured via self-reports or EMAs in 20% (4/20) of the studies that explicitly measured engagement [30,38-40]. Latham [30] evaluated a sleep intervention designed to improve the regularity of wake times in college students via prompts. One measure of engagement in this study was participants' self-reported adherence to the sleep-related suggestions included in the prompt. Nahum-Shani et al [38] proposed to study how prompts to engage in self-regulatory strategies increased engagement in self-regulatory activities; researchers planned to measure engagement as self-reported engagement in self-regulatory activities during the hour after receiving a prompt. In their evaluation of a web-based intervention delivered via SMS text messaging, Nordby et al [39] measured engagement as the self-reported frequency of practicing the coping strategies taught in the web-based intervention. Militello et al [40] assessed the feasibility and acceptability of intervention prompts to encourage engagement in mindfulness activities guided by a mindfulness mobile app. Here, engagement was measured as self-reported performance of a mindfulness activity or exercise in the 24 hours after receiving an intervention prompt.

Only 1 study measured engagement with qualitative methods. In this study, researchers sought to describe engagement with an Acceptance and Commitment Therapy (ACT)–based mobile app in a clinical and a nonclinical sample [42]. The researchers inferred participant engagement by assessing whether participant responses reflected an understanding of the ACT intervention content. The following 3 indicators were used: the identification of the function of behavior, process alignment (whether the content of a given participant's response is congruent with the core ACT process underlying the intervention prompt received), and the qualitative content of responses.

Only 8 (40%) out of the 20 studies that explicitly measured engagement used >1 method to measure engagement. Interestingly, no study used >2 methods. No studies measured engagement with social media data or psychophysiological measures.

Facets of Engagement

Table 3 summarizes the facets of engagement measured by each included study. The physical facet of engagement was the most frequently measured facet of engagement; all 20 studies that explicitly measured engagement included at least 1 measure of this facet [22-24,26-31,33-43]. Multimedia Appendix 2 [22-24,26-31,33-43] provides examples of how this facet of engagement was measured in each included study.

Only 1 study included a measure of the affective facet of engagement [30]. Recall that the affective facet of engagement "captures a wide range of positive affective reactions to a task or activity," including the "the affective states that may underlie more enduring experiences of attachment, identification, and commitment" [11]. By asking participants how likely they were to complete the intervention (ie, their commitment to the intervention), it could be argued that Latham [30] measured this facet of engagement.

Similarly, only 1 study assessed the cognitive facet of engagement—recall that this involves the "selective attention and processing of information related to a task or activity" [11]. This processing of information related to a task was comprehensively measured by Hoel et al [42] using the qualitative measures described in *Measures of Engagement* section.

Table 3. Facets of engagement measured in microrandomized trials of mobile health (mHealth) interventions.

Source	Physical	Affective	Cognitive
Evaluate effect of intervention components			
Aguilera et al [23], 2021	✓		
Battalio et al [24], 2021	✓		
Goldstein et al [26], 2021	✓		
Klasnja et al [27], 2021	✓		
Klasnja et al [28], 2019	✓		
Kramer et al [29], 2020	✓		
Latham [30], 2021[a]	✓	✓	
Jeganathan et al [31], 2022	✓		
Spruijt-Metz et al [33], 2022	✓		
Wang et al [34], 2022	✓		
Dowling et al [35], 2022	✓		
Rodda et al [36], 2022	✓		
Evaluate strategies to improve engagement			
Bell et al [22], 2020	✓		
Bidargaddi et al [37], 2018	✓		
Nahum-Shani et al [38], 2021	✓		
Nordby et al [39], 2022	✓		
Evaluate feasibility and acceptability of intervention			
Militello et al [40], 2022	✓		
Yang et al [41], 2022	✓		
Describing engagement			
Hoel et al [42], 2022	✓		✓
Valle et al [43], 2020	✓		

[a]This study was also designed to evaluate the feasibility and acceptability of its mHealth intervention.

Levels of Engagement

Table 4 summarizes the levels of engagement measured in each included study. Of the 20 studies that explicitly measured engagement, 14 (70%) studies measured Little e only, 2 (10%) studies measured Big E only, and 4 (20%) studies measured both Little e and Big E. Clearly, measures of engagement in MRTs of mHealth interventions are most often Little e measures.

Table 4. Levels of engagement measured in microrandomized trials of mobile health (mHealth) interventions.

Source	Little e		Big E	
	Yes or no	Example	Yes or no	Example
Evaluate effect of intervention components				
Aguilera et al [23], 2021	Yes	Response rates to daily mood rating SMS	No	N/A[a]
Battalio et al [24], 2021	Yes	If end-of-day logs for smoking are completed	No	N/A
Goldstein et al [26], 2021	Yes	Percentage of interventions accessed	No	N/A
Klasnja et al [27], 2021	Yes	Adherence to wearing the FitBit	No	N/A
Klasnja et al [28], 2019	Yes	Adherence to activity tracker	No	N/A
Kramer et al [29], 2020	Yes	Whether participants responded to first message of the chatbot in an intervention conversation	No	N/A
Latham [30], 2021[b]	Yes	Percentage of sleep diaries completed	Yes	Self-reported adherence to intervention prompt's suggestion
Jeganathan et al [31], 2022	Yes	Nonadherence with recommendations for watch wear time	No	N/A
Spruijt-Metz et al [33], 2022	Yes	Time since FitBit was last worn	No	N/A
Wang et al [34], 2022	Yes	Proportion of days that daily step/sleep minutes were provided within a week	No	N/A
Dowling et al [35], 2022	Yes	EMA[c] compliance	No	N/A
Rodda et al [36], 2022	Yes	EMA compliance	No	N/A
Evaluate strategies to improve engagement				
Bell et al [22], 2020	Yes	Whether participants opened the intervention app in the hour after microrandomization	No	N/A
Bidargaddi et al [37], 2018	No	N/A	Yes	Whether participants performed the self-monitoring intervention activity
Nahum-Shani et al [38], 2021	No	N/A	Yes	Whether participants engaged in self-regulatory activities 1 h after randomization
Nordby et al [39], 2022	Yes	Minutes spent on the intervention	Yes	Self-reported frequency of practicing coping strategies taught
Evaluate feasibility and acceptability of intervention				
Militello et al [40], 2022	Yes	Opening the application	Yes	Self-reported engagement with mindfulness exercises 24 hours after randomization
Yang et al [41], 2022	Yes	Percentage of EMAs completed	Yes	Percentage of prompted strategies completed
Describing engagement				
Hoel et al [42], 2022	Yes	Proportion of submitted and nonblank logs	No	N/A
Valle et al [43], 2020	Yes	Proportion of intervention messages viewed before end of day	No	N/A

[a]N/A: not applicable.

[b]This study was also designed to evaluate the feasibility and acceptability of its mHealth intervention.

[c]EMA: ecological momentary assessment.

Determinants of Engagement

Table 5 presents the determinants, moderators, and covariates of engagement studied (if any) in MRTs that assessed or planned to assess engagement. Of the 20 included studies that measured engagement explicitly, 6 (30%) investigated the determinants of participant engagement. Of the 6 studies, 4 (67%) studies were designed to evaluate strategies to improve engagement and investigated the influence of notification-related variables on participant engagement as variables of interest [22,37-39]. The remaining 2 (33%) of the 6 studies were designed to evaluate the effect of intervention components on health outcomes or to describe engagement. The former study assessed a time-based variable as its variable of interest—the causal effect of being in an intervention week on participant engagement [34]. The latter study assessed task-related variables (lapses in self-monitoring and behavioral goal attainment) and an anthropometric variable (weight change) as determinants of participant engagement [43].

Of the 6 studies, only 3 (50%) studies designed to evaluate strategies to improve engagement investigated how the determinants of engagement were moderated. Two of these studies exclusively examined the moderating effect of time-related variables [22,37]. Concretely, Bell et al [22] investigated how the causal effect of sending a push notification (vs not sending it) on engagement was moderated by the number of days in the study. Bidargaddi et al [37], by contrast, investigated if the causal effect of sending (vs not sending) a push notification on engagement was moderated by the number of weeks in the study or by the day of the week (sent on a weekday or a weekend). The third study of this trio planned to study the moderating effect of a comprehensive set of physiological and psychosocial moderators representing vulnerability and receptivity, in addition to time-related moderators [38].

Table 5. Determinants, moderators, and covariates of engagement assessed in microrandomized trials of mobile health (mHealth) interventions.

Source	Determinants	Moderators	Covariates
Evaluate effect of intervention components			
Wang et al [34], 2022	Time related	N/A[a]	N/A
Evaluate strategies to improve engagement			
Bell et al [22], 2020	Notification related	Time related	Demographic, time related, and health behavior related
Bidargaddi et al [37], 2018	Notification related	Time related	Time related, notification related, and task related
Nahum-Shani et al [38], 2021	Notification related	Psychological, societal, health behavior related, contextual, time related, physiological, and demographic	Demographic and time related
Nordby et al [39], 2022	Notification related	N/A	N/A
Describing engagement			
Valle et al [43], 2020	Task related, anthropometric	N/A	Time related, notification related, and anthropometric

[a]N/A: not applicable.

Discussion

Principal Findings

In this scoping review, we aimed to better understand the state of participant engagement measurement in MRTs of mHealth interventions. To do so, we quantified the proportion of existing and planned studies that have explicitly assessed engagement and investigated how engagement has been operationalized in these MRTs. Of the 22 eligible studies indexed by our search, 20 (91%) studies included at least 1 explicit measure of engagement. Overall, our findings suggest that MRTs of mHealth interventions have operationalized engagement in overly narrow terms. We also sought to identify the factors that have been studied as determinants of engagement in MRTs of mHealth interventions. We found that out of the 20 studies that measured engagement explicitly, only 6 (30%) studies investigated the determinants of engagement. Even fewer attempts had been made to investigate the moderators of engagement.

Operationalization of Engagement

Measures of Engagement

Objective measures of engagement—in particular, system usage data (16/20, 80%) and sensor data (7/20, 35%)—were the most common methods of measuring engagement in MRTs of mHealth interventions. The relative popularity of measuring engagement with objective measures, especially system usage data, in MRTs of mHealth interventions is not surprising. System usage has been a central focus in the extant mHealth intervention literature [16]. In fact, it is one of the most common measures of engagement in mHealth interventions [10,44,45]. Subjective measures of engagement, by contrast, were far less common: self-report or EMA (4/20, 20%) and qualitative methods (1/20, 5%). Unfortunately, the lack of attention to the subjective experiences of participants in engagement measurement is not unique to MRTs of mHealth interventions [14,17]. Surprisingly, only 8 (40%) out of the 20 studies measured engagement using >1 method (no study used >2 methods). Of these 8 studies, only half (4/8, 50%) used both subjective and objective measures of engagement.

Taken together, these findings highlight a pressing need for future MRTs of mHealth interventions to diversify the methods of engagement used; the aforementioned lack of diversity does not seem limited to mHealth interventions evaluated using MRTs [14]. Researchers should keep in mind that subjective and objective methods are complementary, not competing, methods to measure engagement—subjective methods provide unique information about participant engagement that objective methods do not capture and vice versa [13,14]. Let us consider the distinction between qualitative and sensor data measures of engagement. Using qualitative methods, we may glean interesting insights about how a participant feels about an intervention or how cognitively invested they are in the intervention. This is certainly not possible for sensor data extracted from a pedometer. However, with said sensor data, it is possible to obtain detailed information (unobtrusively) about health behavior participation and how it fluctuates over time. We recommend that future MRTs of mHealth interventions adopt a multimethod approach to engagement measurement [13] such that engagement data from several subjective and objective measures are collected and interpreted.

Facets of Engagement

In this review, we found that the physical facet of engagement was the dominant kind of engagement measured in MRTs of mHealth interventions. Indeed, all 20 studies included at least 1 explicit measure of this facet. Surprisingly, the affective and cognitive facets of engagement were only measured by 1 study each. Clearly, our findings suggest an imbalance in the *kinds* of engagement measured and that researchers' conceptualizations of engagement, and consequently their operationalizations of engagement, are largely constrained to intervention-related task or activity performance. Given that self-report and qualitative measures of engagement are best suited to measure the affective and cognitive facets of engagement, we cannot rule out that this imbalance is a product of the lack of diversity in methods of measuring engagement described in *Measures of Engagement* subsection in the *Discussion* section.

From the theoretical position that engagement is a multidimensional latent construct composed of physical, affective, and cognitive facets, this imbalance is particularly worrying because it signals that the construct of engagement is not being adequately measured in MRTs of mHealth interventions. Scholars who adopt this position generally agree that no facet of engagement alone can constitute engagement. Instead, they concur that engagement involves the physical, emotional, and cognitive energies of a person working in concert [11]. Therefore, without measuring all 3 facets of engagement, it is not possible to accurately identify how engaged participants are with a task. We hope that this review will draw attention to this gap in engagement measurement and encourage future MRTs of mHealth interventions to incorporate more measures of the affective and cognitive facets of engagement.

On a related note, although an assessment of the quality of engagement measurement in MRTs of mHealth interventions is beyond the scope of this review, we did observe that many included studies relied on single items to measure engagement. Estimates of reliability were also rarely (if ever) reported. As single items have a bad reputation for being unreliable measures of psychological constructs [46], we encourage researchers to clearly report estimates of reliability (such as test-retest reliability) so that readers can evaluate for themselves how much variation in "engagement scores" can be attributed to measurement error.

Levels of Engagement

The distinction between Little e and Big E is an important consideration when studying engagement in digital health interventions. Recall that Little e and Big E can be construed as 2 distinct answers to the question "Engagement with what?" [11]. Our findings suggest that most explicit measures of engagement in MRTs of mHealth interventions are Little e measures (measures of engagement with the mHealth intervention) and that only a handful of studies have measured engagement with the health behavior of interest (or Big E).

Although this review focuses on explicit claims of engagement measurement, a careful analysis of the outcome measures used in all 22 studies makes it clear that many of these outcomes qualify as Big E measures, even though they were not explicitly conceptualized as such [15]. This was observed in 12 studies [24-29,31-36]. All 12 studies were designed to evaluate the effects of intervention components. Most of these studies measured the physical aspect of engagement using sensor data. If we account for such studies, we may conclude that all 22 studies of mHealth interventions included in this review included at least one measure of engagement and that out of the 22 MRTs of mHealth interventions included here, 4 (18%) studies measured Little e only, 4 (18%) studies measured Big E only, and 14 (64%) studies measured both Little e and Big E (Multimedia Appendix 3 [22-43]). It was difficult for us to decide whether the outcome measures of these 12 studies should be deemed measures of engagement in this review. Our concern stems from the fact that the inclusion of these outcomes as measures of engagement hinges on our use of the Little e and Big E distinction to understand how engagement has been operationalized. If this distinction was not invoked, there would be no clear evidence from these 12 studies to suggest that these outcome measures are measures of engagement or that the authors themselves considered them to be measures of engagement. Let us consider the engagement-related information extracted from Goldstein et al [26], which is one of the 12 studies. The outcome measure of this study, whether a dietary lapse was experienced since the last EMA, is a clear-cut measure of Big E. However, it was not included in the authors' own list of engagement measures stated in the paper. If the authors themselves do not conceptualize these outcomes as measures of engagement, would it be appropriate to include these outcomes as measures of engagement in this scoping review? Even if we were to include this outcome as a measure of engagement, can we assume that the underlying motivations of Goldstein et al [26]—in terms of modeling decisions and decisions about the study design—are similar to those of researchers who explicitly frame health behavior outcomes as measures of engagement? This is important because we cannot rule out the possibility that researchers' choice of causal effects, moderators, and control variables are at least partly influenced by how they conceptualize outcome measures. On the basis of these considerations, we decided not to consider the outcome measures of these 12 studies as measures of engagement in this scoping review.

Nevertheless, our findings clearly suggest the need for future MRTs of mHealth interventions to strike a balance between Little e and Big E measurement or at least be more intentional and explicit with Big E measurement (especially when using sensor data as an outcome measure). As the field begins to recognize that sustained engagement is not always required for participants to experience the intended health outcomes of an intervention [13], we encourage researchers to find this balance so that they can gain a sense of effective engagement in the interventions they develop—the sufficient amount of engagement needed to attain the intended outcome of the intervention [11,14].

Determinants of Engagement

We found that very few studies investigated the determinants of engagement (6/20, 30% of the studies that measured engagement). In studies that did assess the determinants of engagement, notification-related causal effects were most common. This is likely attributable to the fact that most of these studies were designed to evaluate strategies to improve engagement [22,37-39]. Even fewer studies (3/6, 50%) assessed the moderators of engagement. Although all 3 studies assessed time-related (time-variant) moderators such as the number of days in the study or the day of the week, only 1 study [38] planned to investigate time-invariant moderators (such as psychological or social variables) in addition to the time-related moderators. These findings suggest that there is a striking lack of attention to how engagement is determined and to the effect of time-invariant psychosocial moderators on engagement in existing MRTs of mHealth interventions. To advance our understanding of engagement in the multicomponent and adaptive mHealth interventions tested by MRTs, it is necessary for future MRTs to address this research gap.

To begin addressing this research gap, we recommend that researchers adopt existing theoretical frameworks to guide their selection of the determinants and moderators of participant engagement in MRTs of mHealth interventions. If widely adopted, this approach should ensure some semblance of parity in the kinds of determinants and moderators of engagement studied across MRTs and provide researchers with a common taxonomy (or at least a common language) to guide their inquiry. With this, researchers can compare and synthesize results from different MRTs to better understand how engagement is modulated across mHealth interventions tested with MRTs.

Researchers can consider studying the determinants and moderators of engagement through the lens of participant engagement frameworks. Recently, Nahum-Shani et al [11] proposed the affect-integration-motivation and attention-context-translation framework for participant engagement. In this paper, they outlined 3 areas, namely, attention, contextual influences, and the translation of motivation to behavior (attention-context-translation), that might influence the neural-based process (affect-integration-motivation) of how engagement with a task (eg, walking) is realized through engagement with a stimulus (eg, a prompt to take a walk). It would be interesting for future MRTs to examine how constructs from each of these 3 areas contribute to participant engagement. Alternatively, researchers can consider selecting theoretically relevant determinants and moderators from the Big Five Personality trait framework [47], which is composed of trait openness, conscientiousness, extraversion, agreeableness, and neuroticism. This approach might be a good first step toward clarifying the role of individual differences in participant engagement, considering the lack of attention given to the psychological characteristics of participants in the extant MRT literature and the relevance of personality to health behaviors and outcomes [48]. Researchers should pay particular attention to the role of conscientiousness as it seems to be the most relevant to mHealth engagement [49] and it has been consistently linked to positive health behaviors [50,51]. These 2 frameworks are by no means exhaustive. We encourage researchers interested in understanding the determinants and moderators of engagement to seek out other appropriate frameworks to advance this line of research.

Limitations

There are 3 notable limitations of this scoping review. First, at the time of conducting our database searches, there was no available Medical Subject Heading in PubMed for MRTs (or equivalent controlled vocabularies for other databases). Therefore, our database searches might not have picked up papers and protocols that did not use the phrase "micro-randomised trial" or "micro-randomized trial" as a keyword or in their title and abstract. Nevertheless, we believe that the main findings of this scoping review still hold true, as our database and manual searches would have indexed most mHealth intervention MRTs planned and conducted to date. Second, we did not use existing frameworks such as the Frequency, Intensity, Time, and Type principle [14] to further categorize engagement measured using system usage data. This has been done in previous scoping reviews [17] and is necessary to obtain a nuanced understanding of engagement measurement in mHealth interventions. Unfortunately, we were not able to do so, as some studies and protocols did not clearly operationalize their measurement of engagement in exact terms. Finally, it must be noted that because of the inclusion and exclusion criteria, we were not able to include several well-designed MRTs in this review because they were not strictly evaluations of mHealth interventions—they were designed either to evaluate digital but not mHealth interventions [52,53] or to evaluate engagement strategies only [54-56]. To fully understand the extent of engagement measurement in digital health interventions evaluated by MRTs, we encourage future reviews to broaden their inclusion and exclusion criteria to include these 2 types of evidence sources.

Conclusions

In this scoping review, we demonstrate that although most MRTs of mHealth interventions have measured engagement explicitly, they have operationalized engagement in overly narrow terms; there is an overemphasis on using objective measurements of engagement, measuring the physical facet of engagement, and measuring engagement with the mHealth intervention (as opposed to engagement with the health behavior of interest). There is also a lack of attention to how engagement is determined and moderated in these existing trials. We hope that by mapping the state of engagement measurement, this review will encourage researchers to pay more attention to these issues when planning engagement measurement in future MRTs. Although these issues are by no means unique to mHealth interventions evaluated with MRTs, the relative infancy of the MRT design suggests that there is still time and opportunity for the field to course correct and establish best practices for the measurement of engagement in MRTs of mHealth interventions.

Acknowledgments

The authors would like to thank Loh Mee Lan from the National University of Singapore Libraries for her guidance in formulating the database search strategies used in this review. This scoping review was supported by the Khoo Bridge Funding Award, the Academic Research Fund Tier 2 grant (Ministry of Education T2EP20122-0013) from the Ministry of Education, Singapore, and the start-up funding from Duke-NUS Medical School awarded to author BC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

All charted data for the included evidence sources.

[XLSX File (Microsoft Excel File), 29 KB - mhealth_v11i1e44685_app1.xlsx]

Multimedia Appendix 2

Engagement measures of the included evidence sources organized according to the method of engagement measurement used, the facets of engagement measured, and the levels of engagement measured.

[XLSX File (Microsoft Excel File), 35 KB - mhealth_v11i1e44685_app2.xlsx]

Multimedia Appendix 3

Levels of engagement measured in microrandomized trials of mobile health (mHealth) interventions if 12 outcome measures are considered measures of Big E.

[PDF File (Adobe PDF File), 128 KB - mhealth_v11i1e44685_app3.pdf]

References

1. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
2. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017 Mar 01;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
3. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
4. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
5. Campbell BR, Ingersoll KS, Flickinger TE, Dillingham R. Bridging the digital health divide: toward equitable global access to mobile health interventions for people living with HIV. Expert Rev Anti Infect Ther 2019 Mar;17(3):141-144 [FREE Full text] [doi: 10.1080/14787210.2019.1578649] [Medline: 30721103]
6. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
7. Liu X, Deliu N, Chakraborty B. Microrandomized trials: developing just-in-time adaptive interventions for better public health. Am J Public Health 2023 Jan;113(1):60-69. [doi: 10.2105/AJPH.2022.307150] [Medline: 36413704]
8. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]
9. Coppersmith DD, Dempsey W, Kleiman EM, Bentley KH, Murphy SA, Nock MK. Just-in-time adaptive interventions for suicide prevention: promise, challenges, and future directions. Psychiatry 2022;85(4):317-333. [doi: 10.1080/00332747.2022.2092828] [Medline: 35848800]
10. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

11. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
12. Vessonen E. Respectful operationalism. Theory Psychol 2021 Feb;31(1):84-105 [FREE Full text] [doi: 10.1177/0959354320945036]
13. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
14. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
15. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
16. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
17. Madujibeya I, Lennie T, Aroh A, Chung ML, Moser D. Measures of engagement with mHealth interventions in patients with heart failure: scoping review. JMIR Mhealth Uhealth 2022 Aug 22;10(8):e35657 [FREE Full text] [doi: 10.2196/35657] [Medline: 35994345]
18. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Sep 11;26:100454 [FREE Full text] [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
19. Peters M, Godfrey C, McInerney P, Munn Z, Trico A, Khalil H. Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2020:406-451.
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Leong U, Chakraborty B. Participant engagement in micro-randomized trials of mHealth interventions: a scoping review. Open Science Framework. 2022 Jun 30. URL: https://osf.io/xdnr8/ [accessed 2022-06-30]
22. Bell L, Garnett C, Qian T, Perski O, Potts HW, Williamson E. Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial. JMIR Res Protoc 2020 Aug 07;9(8):e18690 [FREE Full text] [doi: 10.2196/18690] [Medline: 32763878]
23. Aguilera A, Hernandez-Ramos R, Haro-Ramos AY, Boone CE, Luo TC, Xu J, et al. A text messaging intervention (StayWell at home) to counteract depression and anxiety during COVID-19 social distancing: pre-post study. JMIR Ment Health 2021 Nov 01;8(11):e25298 [FREE Full text] [doi: 10.2196/25298] [Medline: 34543230]
24. Battalio SL, Conroy DE, Dempsey W, Liao P, Menictas M, Murphy S, et al. Sense2Stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. Contemp Clin Trials 2021 Oct;109:106534 [FREE Full text] [doi: 10.1016/j.cct.2021.106534] [Medline: 34375749]
25. Figueroa CA, Deliu N, Chakraborty B, Modiri A, Xu J, Aggarwal J, et al. Daily motivational text messages to promote physical activity in university students: results from a microrandomized trial. Ann Behav Med 2022 Feb 11;56(2):212-218. [doi: 10.1093/abm/kaab028] [Medline: 33871015]
26. Goldstein SP, Zhang F, Klasnja P, Hoover A, Wing RR, Thomas JG. Optimizing a just-in-time adaptive intervention to improve dietary adherence in behavioral obesity treatment: protocol for a microrandomized trial. JMIR Res Protoc 2021 Dec 06;10(12):e33568 [FREE Full text] [doi: 10.2196/33568] [Medline: 34874892]
27. Klasnja P, Rosenberg DE, Zhou J, Anau J, Gupta A, Arterburn DE. A quality-improvement optimization pilot of BariFit, a mobile health intervention to promote physical activity after bariatric surgery. Transl Behav Med 2021 Mar 16;11(2):530-539. [doi: 10.1093/tbm/ibaa040] [Medline: 32421187]
28. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
29. Kramer JN, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
30. Latham MD. A micro-randomized trial to improve college students' wake time regularity. University of Oregon. URL: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc18&NEWS=N&AN=2020-79976-228 [accessed 2022-07-13]
31. Jeganathan VS, Golbus JR, Gupta K, Luff E, Dempsey W, Boyden T, et al. Virtual application-supported environment to increase exercise (VALENTINE) during cardiac rehabilitation study: rationale and design. Am Heart J 2022 Jun;248:53-62. [doi: 10.1016/j.ahj.2022.02.012] [Medline: 35235834]

32. NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, et al. Assessing real-time moderation for developing adaptive mobile health interventions for medical interns: micro-randomized trial. J Med Internet Res 2020 Mar 31;22(3):e15033 [FREE Full text] [doi: 10.2196/15033] [Medline: 32229469]
33. Spruijt-Metz D, Marlin BM, Pavel M, Rivera DE, Hekler E, De La Torre S, et al. Advancing behavioral intervention and theory development for mobile health: the HeartSteps II protocol. Int J Environ Res Public Health 2022 Feb 17;19(4):2267 [FREE Full text] [doi: 10.3390/ijerph19042267] [Medline: 35206455]
34. Wang J, Fang Y, Frank E, Walton MA, Burmeister M, Tewari A, et al. Effectiveness of gamified team competition in the context of mHealth intervention for medical interns: a micro-randomized trial. medRxiv 2022 Mar 14 [FREE Full text] [doi: 10.1101/2022.03.11.22272278]
35. Dowling NA, Merkouris SS, Youssef GJ, Lubman DI, Bagot KL, Hawker CO, et al. A gambling just-in-time adaptive intervention (GamblingLess: in-the-moment): protocol for a microrandomized trial. JMIR Res Protoc 2022 Aug 23;11(8):e38958 [FREE Full text] [doi: 10.2196/38958] [Medline: 35998018]
36. Rodda SN, Bagot KL, Merkouris SS, Youssef G, Lubman DI, Thomas AC, et al. Smartphone app delivery of a just-in-time adaptive intervention for adult gamblers (gambling habit hacker): protocol for a microrandomized trial. JMIR Res Protoc 2022 Jul 26;11(7):e38919 [FREE Full text] [doi: 10.2196/38919] [Medline: 35881441]
37. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
38. Nahum-Shani I, Potter LN, Lam CY, Yap J, Moreno A, Stoffel R, et al. The mobile assistance for regulating smoking (MARS) micro-randomized trial design protocol. Contemp Clin Trials 2021 Nov;110:106513 [FREE Full text] [doi: 10.1016/j.cct.2021.106513] [Medline: 34314855]
39. Nordby ES, Gjestad R, Kenter RM, Guribye F, Mukhiya SK, Lundervold AJ, et al. The effect of SMS reminders on adherence in a self-guided internet-delivered intervention for adults with ADHD. Front Digit Health 2022 May 16;4:821031 [FREE Full text] [doi: 10.3389/fdgth.2022.821031] [Medline: 35651537]
40. Militello L, Sobolev M, Okeke F, Adler DA, Nahum-Shani I. Digital prompts to increase engagement with the headspace app and for stress regulation among parents: feasibility study. JMIR Form Res 2022 Mar 21;6(3):e30606 [FREE Full text] [doi: 10.2196/30606] [Medline: 35311675]
41. Yang MJ, Sutton SK, Hernandez LM, Jones SR, Wetter DW, Kumar S, et al. A just-in-time adaptive intervention (JITAI) for smoking cessation: feasibility and acceptability findings. Addict Behav 2023 Jan;136:107467. [doi: 10.1016/j.addbeh.2022.107467] [Medline: 36037610]
42. Hoel S, Victory A, Sagorac Gruichich T, Stowe ZN, McInnis MG, Cochran A, et al. A mixed-methods analysis of mobile ACT responses from two cohorts. Front Digit Health 2022 May 12;4:869143 [FREE Full text] [doi: 10.3389/fdgth.2022.869143] [Medline: 35633737]
43. Valle CG, Nezami BT, Tate DF. Designing in-app messages to nudge behavior change: lessons learned from a weight management app for young adults. Organ Behav Hum Decis Process 2020 Nov;161:95-101 [FREE Full text] [doi: 10.1016/j.obhdp.2020.10.004]
44. Tuvesson H, Eriksén S, Fagerström C. mHealth and engagement concerning persons with chronic somatic health conditions: integrative literature review. JMIR Mhealth Uhealth 2020 Jul 24;8(7):e14315 [FREE Full text] [doi: 10.2196/14315] [Medline: 32706686]
45. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
46. Allen MS, Iliescu D, Greiff S. Single item measures in psychological science. Eur J Psychol Assess 2022;38(1):1-5 [FREE Full text] [doi: 10.1027/1015-5759/a000699]
47. John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues. In: John OP, Robins RW, editors. Handbook of personality: Theory and research. New York, NY, USA: The Guilford Press; 2008:114-158.
48. Ferguson E. Personality is of central concern to understand health: towards a theoretical model for health psychology. Health Psychol Rev 2013 May;7(Suppl 1):S32-S70 [FREE Full text] [doi: 10.1080/17437199.2010.547985] [Medline: 23772230]
49. Aziz M, Erbad A, Belhaouari SB, Almourad MB, Altuwairiqi M, Ali R. Who uses mHealth apps? Identifying user archetypes of mHealth apps. Digit Health 2023 Jan 22;9:20552076231152175 [FREE Full text] [doi: 10.1177/20552076231152175] [Medline: 36714545]
50. Bogg T, Roberts BW. Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. Psychol Bull 2004 Nov;130(6):887-919. [doi: 10.1037/0033-2909.130.6.887] [Medline: 15535742]
51. Molloy GJ, O'Carroll RE, Ferguson E. Conscientiousness and medication adherence: a meta-analysis. Ann Behav Med 2014 Feb;47(1):92-101. [doi: 10.1007/s12160-013-9524-4] [Medline: 23783830]

52. Beinema T, Op den Akker H, Hurmuz M, Jansen-Kosterink S, Hermens H. Automatic topic selection for long-term interaction with embodied conversational agents in health coaching: a micro-randomized trial. Internet Interv 2022 Mar;27:100502 [FREE Full text] [doi: 10.1016/j.invent.2022.100502] [Medline: 35198412]
53. Carpenter SM, Yap J, Patrick ME, Morrell N, Dziak JJ, Almirall D, et al. Self-relevant appeals to engage in self-monitoring of alcohol use: a microrandomized trial. Psychol Addict Behav 2022 Jul 14:adb0000855. [doi: 10.1037/adb0000855] [Medline: 35834200]
54. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]
55. Psihogios AM, Rabbi M, Ahmed A, McKelvey ER, Li Y, Laurenceau JP, et al. Understanding adolescent and young adult 6-mercaptopurine adherence and mHealth engagement during cancer treatment: protocol for ecological momentary assessment. JMIR Res Protoc 2021 Oct 22;10(10):e32789 [FREE Full text] [doi: 10.2196/32789] [Medline: 34677129]
56. Rabbi M, Philyaw Kotov M, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, et al. Toward increasing engagement in substance use data collection: development of the substance abuse research assistant app and protocol for a microrandomized trial using adolescents and emerging adults. JMIR Res Protoc 2018 Jul 18;7(7):e166 [FREE Full text] [doi: 10.2196/resprot.9850] [Medline: 30021714]

Abbreviations

ACT: Acceptance and Commitment Therapy
EMA: ecological momentary assessment
mHealth: mobile health
MRT: microrandomized trial
OSF: Open Science Framework
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 29.11.22; peer-reviewed by J Yap, J Golbus, J Rooksby; comments to author 23.01.23; revised version received 20.02.23; accepted 31.03.23; published 22.05.23.

Please cite as:
Leong U, Chakraborty B
Participant Engagement in Microrandomized Trials of mHealth Interventions: Scoping Review
JMIR Mhealth Uhealth 2023;11:e44685
URL: https://mhealth.jmir.org/2023/1/e44685
doi:10.2196/44685
PMID:37213178

Review

Conversational Agents and Avatars for Cardiometabolic Risk Factors and Lifestyle-Related Behaviors: Scoping Review

Lynnette Nathalie Lyzwinski[1,2], MPhil, PhD; Mohamed Elgendi[3], PhD; Carlo Menon[1,3], PhD

[1]Menrva Research Group, Schools of Mechatronic Systems Engineering and Engineering Science, Simon Fraser University, Metro Vancouver, BC, Canada

[2]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[3]Biomedical and Mobile Health Technology Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

Corresponding Author:
Lynnette Nathalie Lyzwinski, MPhil, PhD
Menrva Research Group
Schools of Mechatronic Systems Engineering and Engineering Science
Simon Fraser University
250-13450 102 Avenue
Metro Vancouver, BC, V3T 0A3
Canada
Phone: 1 778 302 1151
Email: lnl2@sfu.ca

Abstract

Background: In recent years, there has been a rise in the use of conversational agents for lifestyle medicine, in particular for weight-related behaviors and cardiometabolic risk factors. Little is known about the effectiveness and acceptability of and engagement with conversational and virtual agents as well as the applicability of these agents for metabolic syndrome risk factors such as an unhealthy dietary intake, physical inactivity, diabetes, and hypertension.

Objective: This review aimed to get a greater understanding of the virtual agents that have been developed for cardiometabolic risk factors and to review their effectiveness.

Methods: A systematic review of PubMed and MEDLINE was conducted to review conversational agents for cardiometabolic risk factors, including chatbots and embodied avatars.

Results: A total of 50 studies were identified. Overall, chatbots and avatars appear to have the potential to improve weight-related behaviors such as dietary intake and physical activity. There were limited studies on hypertension and diabetes. Patients seemed interested in using chatbots and avatars for modifying cardiometabolic risk factors, and adherence was acceptable across the studies, except for studies of virtual agents for diabetes. However, there is a need for randomized controlled trials to confirm this finding. As there were only a few clinical trials, more research is needed to confirm whether conversational coaches may assist with cardiovascular disease and diabetes, and physical activity.

Conclusions: Conversational coaches may regulate cardiometabolic risk factors; however, quality trials are needed to expand the evidence base. A future chatbot could be tailored to metabolic syndrome specifically, targeting all the areas covered in the literature, which would be novel.

(JMIR Mhealth Uhealth 2023;11:e39649) doi:10.2196/39649

KEYWORDS

chatbots; avatars; conversational coach; diet; physical activity; cardiovascular disease; hypertension; cardiometabolic; behavior change; hypertension diabetes; metabolic syndrome; mobile phone

Introduction

Background

Metabolic syndrome (MetS) is a highly prevalent condition that affects up to approximately 30% of adults aged >65 years worldwide [1]. It consists of multiple symptoms, namely abdominal obesity, glucose intolerance, hypertension, and high cholesterol as well as low high-density lipoprotein [2]. It is associated with a substantially increased risk of premature morbidity and mortality from diabetes and cardiovascular disease (CVD) [2]. Low levels of physical activity (PA) are strongly associated with MetS, including obesity and overweight [3], high blood pressure [4], and insulin intolerance [5]. Furthermore, low levels of activity are significantly associated with increased risk of complications of MetS, including diabetes and CVD [5,6]. In addition, research has found that losing weight by approximately 5% to 10% results in significantly reduced MetS-associated markers [1] in patients with existing disease, highlighting that MetS may be modifiable through lifestyle-related weight interventions. Dietary modifications, including reduced sodium, sugar, and fat intake, are also highly beneficial for reducing the risk of the syndrome and its complications [7].

In recent years, mobile health (mHealth) has increasingly been used to support behavior changes related to weight loss, including improving dietary intake and physical activity [8]. Research on the use of mHealth interventions has found support for a moderate effect size for assisting with weight loss [8]. This includes the use of SMS text messaging for behavior change and mHealth apps that target weight loss using a range of behavior change techniques (BCTs) [9], including self-monitoring, feedback, goal setting, education, tips, personal tailoring, reminders, encouragement, and social and professional support [8]. mHealth is a form of health care that enables timely accessibility, portability, and personalized medicine tailored to the needs of the user [10,11]. It includes smartphones, PDAs, MP3 players, iPads (Apple Inc), smart clothing, and smart watches [10,11].

Emerging research in the mHealth field has focused on developing conversational agents that can simulate human professional interactions for managing different health problems [12], including weight issues [13]. Furthermore, avatars have also been developed to display a conversational coach in addition to written conversational text, simulating real-life interactions with a professional, such as a live fitness coach [14,15]. Having a conversational coach complement or replace metabolic-related health advice from professionals may increase accessibility and enable more timely health monitoring and diagnosis of health conditions [15] such as MetS if physicians also gain access to patient data. Given that technology in the field is advancing, it is time to determine whether these conversational agents are effective for assisting with MetS-associated risk factors, including overweight, obesity, physical inactivity, and unhealthy dietary intake. There is also a need to better understand what types of weight-related and MetS-related studies have been undertaken using conversational agents and to identify challenges with the technology and future areas of research.

Aims

This review aimed to better understand the evidence surrounding the use of conversational coaches for metabolic-related risk factors and biomarkers. Furthermore, this review aimed to determine whether conversational coaches are effective for improving weight-related behaviors and metabolic indicators and whether conversational agents are acceptable for consumers as agents of behavior change.

Research Questions

- Research question (RQ) 1: How effective are conversational agents (chatbots and avatars) for weight-related behaviors, including diet and exercise?
- RQ 2: How effective are conversational agents for improving metabolic risk factors, including blood pressure, cholesterol, abdominal obesity, and glucose (diabetes management)?
- RQ 3: What are consumers' perspectives on the use of chatbots?

Methods

A systematic review of PubMed and MEDLINE was conducted in December 2021 for all relevant studies on conversational coaches for metabolic risk factors published over the last 10 years. Google Scholar was also searched for any additional papers along with manual hand searching.

Inclusion and Exclusion Criteria

This review included studies on chatbots or avatar conservational agents that acted as coaches for improving metabolic health behaviors, including dietary intake (sodium and sugar intake), PA, and weight (including abdominal obesity). Studies that evaluated one or more physiological indicators of metabolic health or risk factors for MetS, such as diabetes, glucose intolerance, hypertension, cholesterol, and serum triglycerides, were also included. Studies must have been published in the English language to be included. Chatbots that were used for survey reasons but not primarily for targeting weight-related or metabolic risk factors were excluded. Studies whose primary focus was not on conversational coaches were excluded (including those that had an avatar element but did not primarily focus on evaluating it). Studies on wearables that did not include avatars or chatbots were excluded. Studies in pregnant women were excluded.

Search

The keywords included word variations for "chatbot," "virtual assistant," "virtual coach," or "avatar"; weight-related behaviors, including "diet," "exercise," or "weight"; and metabolic risk factors, including "hypertension," "cholesterol," or "diabetes." The search strategy is shown in Textbox 1.

Textbox 1. PubMed search strategy example.

1. Cardiometabolic risk factors

- Weight

"obesity"[MeSH Terms] OR "obese"[tiab] OR "obesity"[tiab] OR "overweight"[tiab] OR "overweight"[tiab] OR "BMI"[tiab] OR "Body mass index"[tiab] OR "Body mass index"[MeSH Terms] OR "physical activity"[Tiab] OR adiposity [tiab] OR weight gain[tiab] OR body weight[tiab] OR "abdominal visceral fat"[Tiab] OR "adipose tissue"[MeSH Terms] "weight loss"[Mesh] OR "weight loss"[tiab] or "metabolic syndrome"

- Diet and physical activity

diets[tiab] OR "diet"[mesh] OR diet[tiab] OR "energy intake"[tiab] OR nutrition[tiab] OR "diet, food, and nutrition"[MeSH Terms] OR diets[tiab] OR Caloric restriction[tiab]OR "physical activity"[tiab]

- Hypertension

hypertension[tiab] OR "Blood Pressure"[tiab] OR Prehypertension[tiab] OR BP[tiab] OR "Systolic blood pressure"[tiab] OR SBP[tiab] OR "Diastolic blood pressure"[tiab] OR DBP[tiab] OR cardiovascular[tiab] OR hypotensive[tiab] OR "Hypertension"[MeSH] OR "Blood Pressure"[MeSH] OR "Prehypertension"[MeSH]

- Cholesterol

"cholesterol"[MeSH Terms] OR cholesterol[tiab]

- Diabetes

"Diabetes Mellitus"[MeSH] or diabetes[tiab] or diabetic[tiab] or prediabetes[tiab] or pre-diabetes[tiab] OR "glucose"[MeSH Terms] OR "glucose"[tiab]

AND

2. Technology

chatbot*[tiab] OR chat bot[tiab] OR chat-bot[tiab] OR chatter bot[tiab] OR chat bots[tiab] OR chat-bots[tiab] OR chatter bots[tiab] OR chatterbot*[tiab] OR smart bot[tiab] OR smartbot[tiab] OR smart bots[tiab] OR smartbots[tiab] OR smart-bot*[tiab] OR virtual agent*[tiab] OR virtual character*[tiab] OR virtual coach*[tiab] OR virtual human[tiab] OR avatar*[tiab] OR embodied agent*[tiab] OR relational agent*[tiab] OR animated character*[tiab])

1 AND 2

Screening and Data Extraction

Titles were screened for relevance to the RQs, followed by abstract and full-text retrieval of eligible studies that met the inclusion criteria. A second reviewer (LL) screened the abstracts and full texts against the inclusion and exclusion criteria to ensure agreement. Quantitative and qualitative data were extracted and summarized in a tabular format, including study characteristics, measures, outcomes, and intervention details.

Results

General Description

LL and ME screened the final selected papers individually. A total of 52 full texts were selected [13,14,16-65]; however, after double peer screening, 1 protocol and 1 dated technology were removed. The final number included 50 papers [13,14,16-59,61-63,66]. Details of the search process and reasons for exclusion are illustrated in Figure 1 [67].

Most of the studies were feasibility and usability studies. A few studies were qualitative and explored consumer perspectives on conversational agents for weight-related behaviors [14,19]. The countries where the studies were conducted included Australia, the United States, Italy, Spain, and Taiwan [13,14,16-29]. Most of the studies explored virtual agents for diet and exercise, with only 2 (4%) exploring chatbots for hypertension management [17,19]. The majority were conducted among adults, but 3 (6%) were conducted among teenagers and preteens [14,26,29]. The study characteristics and results are summarized in Table 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the search and screening process. MetS: metabolic syndrome.

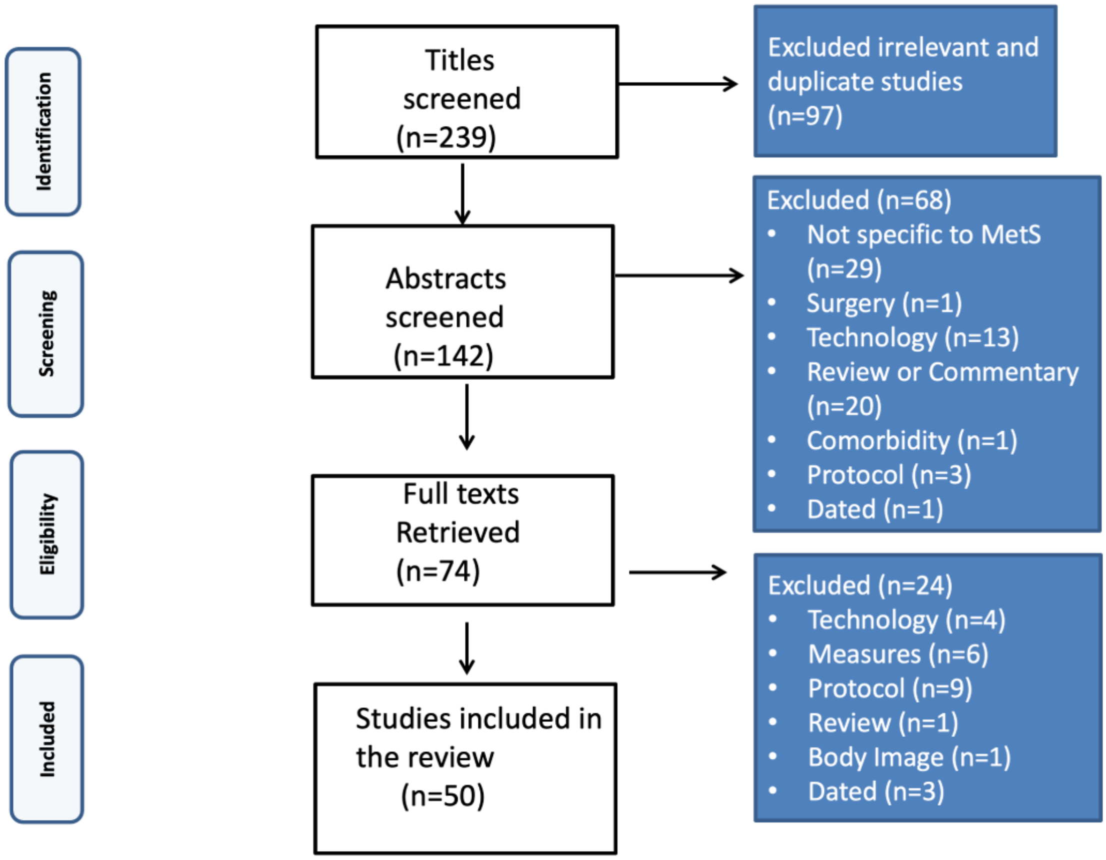


Table 1. Study characteristics.

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
Echeazarra et al [17], 2021	• Location: Spain • N=112 • Design: 2-year RCT[a]	Female: 42	Mean 52.1	BP[b]	• Tensiobot (telegram app) • Reminders to check BP • Education on how to properly check BP using videos • Warnings and graphic feedback on BP • GPs[c] can connect with the app to access patient data • Advice offered 24/7	• No significant differences in adherence between groups • Bot group had higher levels of knowledge on good practice skills for BP (t=2.11; *df*=82.3; 95% CI 0.39-12.6; *P*<.05) • Measurements (*P*<.05) • Bot found to be acceptable/likable • Adherence after intervention: 85%
Griffin et al [19], 2021	• Location: United States • N=15 • Design: mixed methods questionnaires with semistructured interviews qualitative	Female: 53	Mean 59 (SD 11)	BP	• Theoretical discussion around chatbots for hypertension medication management	• Most patients were interested in and open to trying a chatbot for hypertension medication management and reminders • Privacy concerns and usability with mobile phones
Larbi et al [20], 2021	• Location: Switzerland • N=30 • Design: feasibility study	Female: 50	Range 18-69	PA[d]	• MYA social media chatbot	• Perceptions of usefulness and informativeness: 53% • User friendly: 83% • Failed to understand user input: 63.3% • Potential confusion with using the technology 43.3%
Lin et al [27], 2021	• Location: Taiwan • N=96 • Design: factorial experimental study with 4 arms	Female: 53	Mean 21.5; range 18-42	PA (core muscle exercise)	• VR[e] avatar	• Increase in PA (vector movement) of 986.7 (SD 1.03) points in in normal realistic avatar relative to muscular avatar • Higher self-efficacy for core muscle exercise in normal avatars vs muscular avatars in female participants (+0.66, SD 0.1 points) and higher levels than in male participants (+0.9, SD 0.2 points) • *P*<.05
Dol et al [37], 2021	• Location: The Netherlands • N=71 • Design: qualitative study	Female: 100	Mean 44.4 (SD 12.86); range 19-70	Emotional eating	• Conversational coach for emotional eating	• The design of the conversational coach should integrated dialectal behavioral coaching strategies, as preferred by participants with emotional eating behavior

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
Lin et al [27], 2021	• Location: Taiwan • N=104 • Design: experimental design study	Female: 50	Mean 70.39 (SD 6.51); range 60-88	PA perceived exertion Self-efficacy	• Assigned to either age-matched or young avatars for PA Theory: Proteus effect of avatar embodiment • Watched videos in a digital gym where they exercised • Wore a head-mounted display	• Older male participants assigned to young avatars had higher perceived exertion than counterparts assigned to older ones (+1.56, SD 0.31 points; male participants only) • Female participants assigned to young avatars had higher self-efficacy for future exercise than counterparts (+0.45 points) and male participants • $P<.05$
Maher et al [13], 2021	• Location: Australia • N=31 • Design: proof-of-concept study	Female: 67	Range 45-75	PA, Mediterranean diet, and weight	• AI[f] Paola chatbot teaches users about exercise and uses BCTs[g], including goal setting, self-monitoring, and feedback	• Mean increase in diet score: 5.7 (95% CI 4.2-7.3) • Mean PA increase: 109.8 min (95% CI 1.9-217.9; $P<.01$) • Mean weight loss: 1.3 kg (95% CI −2.5 to −0.7; $P<.05$) • $P<.01$ • No significant changes in blood pressure
Hickman et al [40], 2021	• Location: United States • N=109 • Design: 2-arm RCT	Female: 59	Mean 52 (SD 11)	Hypertension, quality of the physician-patient interaction	• Avatar intervention or video on hypertension management	• Scores for the quality of the patient-provider interaction were better over time (F_3=5.25; $P<.01$) in the within-subjects analysis along with a time by experimental condition interaction (F_3=2.91; $P<.05$) • Between-subject effects per treatment were insignificant • No significant changes in blood pressure
Napolitano et al [49], 2021	• Location: United States • N=136 • Design: feasibility study (12 weeks)	Female: 100	Mean 27.8 (SD 5.4)	Weight, diet, and PA; exercise self-efficacy	• Conversational coach gave lessons on health behaviors	• No significant results were found for differences in weight, PA, or consumption of fast food between the intervention arm and control groups • High attrition 44% • Goal achievement for nutrition <10%
Santini et al [55], 2021	• Location: Austria, Italy, and Netherlands • N=60 (2 waves) • Design: qualitative study with focus groups and phone interviews	Female: 53.3% wave 1; 51.6% wave 2	Mean 61.9	Health behaviors, diet, and PA	• Embodied coach for diet and PA	• Desire for the avatar to motivate older adults to exercise • Supportive tone and language that is not authoritarian or patronizing
Krishnakumar et al [44], 2021	• Location: India • N=102 • Design: pre-post intervention (1 arm) • 16 weeks	Female: 31.4	Mean 50.8	Diabetes (blood sugar), diet, PA, and weight (logged)	• Wellthy CARE mobile app	

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
						• The use of the Wellthy CARE digital therapeutic for patients with T2D[h] showed a significant reduction in the mean levels of HbA_{1c}[i] −1.16% (95% CI −1.40 to −0.92; *P*<.01); FBG[j] (−11 mg/dL), and PPBG[k] (−22 mg/dL); *P*<.05 • Weight decreased by 1.32 kg (95% CI −0.63 to −2.01 kg) after 16 weeks
Dhinagaran et al [36], 2021	• Location: Singapore • N=60 • Design: one arm web-based feasibility stud	Female: 62	Mean 33.7	Diet, PA, sleep, and stress	• Chatbot for diabetes prevention, diet, exercise delivered over Facebook Messenger (Meta Platforms Inc)	• Engagement: 50% • Retention: 93% • Satisfaction: high at 92% • 50% agreed that the chatbot was acceptable and usable • No significant changes in health behaviors including PA • Minimal improvement in diet: increase in fruit intake (3 portions) by 4% and vegetables once per day by 2%
To et al [61], 2021	• Location: Australia • N=116 • Design: quasi-experimental study (6 weeks)	Female: 81.9	Mean 49.1 (SD 9.3)	PA	• Fitbit plus a chatbot in the Messenger app	• Usability score: 89.4% • Desire to continue using: 35.4% • Helped them: 53% • Mean PA increase: 154.2 min/week (95% CI 2.28-5.63) • OR for meeting PA guidelines: 6.37 (95% CI 3.31 to 12.27) • Mean steps/day increase: 627 (95% CI 219 to 1035)
Mitchell et al [48], 2021	• Location: United States • N=158 • Design: mixed methods survey with qualitative interviews study	Female: 100	Mean 56 (SD 11) intervention; 57 (SD 11) control	Diabetes	• Avatar for diabetes self-management	• Avatars provide support for diabetes self-management via 3 areas (self, social, and physical) that are linked with engagement
Strombotne et al [58], 2021	• Location: United States • N=590 • Design: quasi-experimental study	Female: 11	Mean treatment=58.1; control=57.7	Diabetes and risk factors	• Conversational coach and ketogenic diet	• BP decrease (systolic): 1.4 mm Hg (95% CI −2.72 to 0.14) • Diastolic BP levels decreased: −1.43 (95% CI −2.72 to −0.14) mm Hg • HbA_{1c} decreased: −0.69 (95% CI −1.02 to 0.36) • Diabetes medication fills: −0.38 (95% CI −0.49 to −0.26) • BMI: −1.07 (95% CI −1.95 to −0.19) kg/m^2
Alves Da Cruz [31], 2020		Female 48.1	Mean 63.4 (SD 12.7)	HR[l], BP, and RR[m]		

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
	• Location: Brazil • N=27 • Design: cluster randomized crossover trial				• Avatars with exergames for PA in patients undergoing cardiovascular rehabilitation	• Increase in HR (z=82.8; *P*<.01) and RR (z=12.9; *P*<.01) during and (5 min) after exergame • Changes in systolic BP but not diastolic with differences within moments z=11.26 (*P*<.01) • With no statistical significance between groups
Kowalska et al [43], 2020	• Location: Poland • N=249 • Design: cross-sectional study	Female: 36.5	Mean 65.3 (SD 13.8)	CVD[n]	• Telehealth voice technology with health professionals and voice conversational agent	• High desirability for telehealth consultations with a cardiologist combined with a conversational agent • Desirability for telemonitoring of vitals: 67.5% • 70.7% wanted a consultation with a cardiologist remotely
Piao et al [51], 2020	• Location: South Korea • N=106 • Design: Exploratory Randomized Controlled Trial • 12 weeks	Female: 56 intervention; 57 control	Range 20-59	Health behaviors (diet and exercise); SRHI[o]	• Lifestyle coaching chatbot • Informed by habit formation • Cues and goals	• Significant improvement in health behavior • The intervention group had higher scores on the SRHI of 7.12 (SD 5.57) with *P*<.05 at 4 weeks; no significant differences between groups at 12 weeks, PA remained higher after 12 weeks (*P*<.05)
Naylor et al [50], 2020	• Location: United States • N=20 • Design: pilot study	N/A[p]	Mean 8.4 (SD 1.3)	VO_2 (mL × kg^{-1} × min^{-1}) using indirect calorimetry questionnaire on liking and motivation	• Children played tennis with their friend and an avatar	• Increased VO2 during game play in both cooperative (3.8 + 1.8 mL × kg^{-1} × min^{-1}) and competitive play (4.4 +1.8 mL × kg^{-1} × min^{-1}) compared with resting condition (*P*<.01) • Children liked exercising more in cooperative games than in competitive games (*P*<.01) • No differences between game styles in motivation for PA (*P*>.05)
Hahn et al [39], 2020	• Location: United States • N=42 (child and parent dyads [n=40 completed baseline and follow-up measures]) • Design: pilot intervention	Female (children): 55.2	Treatment: mean 8.06 (SD 1.10); control: mean 7.5 (SD 1.38)	PA using Fitbit and self-report on motivation for PA	Children wore Fitbit with a personalized dog avatar for socializing and support (digital fitness kiosk); theory informed (social cognitive theory)	• Completion rate: 81.63% • Mean number of PA goals reached: 3.28 • Mean time playing with pets: 20.35 min • Mean number of active min: 66 min (no statistical significance was found)
Navarro et al [24], 2020	• Location: United States • N=305 • Design: 3-arm RCT	Female: N/A	Mean 20.0 (SD 2.2); range 18-37	Cardiac frequency, step counts, accelerometer, and HR monitor	• Randomly assigned to avatars embodying them (same face) or different from them (strangers) • Avatars wore normal clothes or gym clothes	

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
						• Higher cardiac output (frequency) from 6 to 12 min in users of avatars that had a similar appearance (face) • Higher output in users with avatars that additionally wore sports clothing at 6-7 and 10-minute periods • Support for the Proteus effect hypothesis • No changes in step count
Davis et al [16], 2020	• Location: Australia • N=28 • Design: pilot single-arm study	Female: 68	Mean 56.2 (SD 8); range 45-75	Diet: Mediterranean diet adherence tool. Weekly log for diet and step count; activity tracked using a wrist worn tracker (Garmin) that syncs with Paola. Minutes of moderate to vigorous PA assessed with Active Australia Survey	• Conversational assistant Paola for diet and PA consisted of educational modules, weekly check-ins, and 24/7 availability for PA and diet questions • 12-week pilot	• Assisted with increasing PA (step goal achieved 59% of the time) • Adherence to diet: 91%
Navarro et al [23], 2020	• Location: Spain • N=42 • Design: 3 arms—2 avatars vs control	Female: 100	Mean 31.9 (SD 11.7); range 19-61	PA, IPAQ[q], self-efficacy to regulate exercise, and PA enjoyment scale (PACES[r])	• Avatar: ideal (perfect body) or normal (matching the participant) and web-based intervention without the avatar	• Increased PA in all groups ($F_{1,39}$=15.8; P<.01; web-based intervention effects) • No effects of time by avatar assignment, ie, interaction
Balsa et al [32], 2020	• Location: Portugal • N=20 • Design: usability study with qualitative interviews	Female: experts 88.9%; end users 27.3%	Mean 62.62; mean end users 70.9; mean experts 54.3	Usability of the app for diabetes medication adherence and improving lifestyle behaviors, diet, and PA	• The conversational coach resembles a human • Integrated BCTs: goal setting, self-monitoring, feedback, and social support/counseling	• Usability score: 73.75 (SD 13.31) (indicates high usability of the coach)
Chin et al [35], 2020	• Location: United States • N=15 • Design: feasibility study	Female: 60%	Mean 67 (SD 5.84)	PA	• Health coach for PA • As part of a PA program using a Google Home device (Google LLC)	• Usability was high • 80% of the participants did not experience challenges when interacting with the conversational coach
Fadhil et al [18], 2019	• Location: Italy • N=19 • Design: validation study (4 weeks)	Female: 42	Mean 28.5; range 19-53	Diet and PA questionnaires via chatbot and motivation (HAPA[s])	• CoachAI text based conversational agent • Tailored coaching for habits	• Participants were satisfied with the agent • High trust to share personal information to the coach
Ahn et al [30], 2019	• Location: United States • N=67 • Design: field study (3 days)	Female: 61.19	Mean 11.24 (SD 0.85); range 9-13	PA and basic psychological needs	• Use of a digital dog, with and without a points-based reward system	• Higher levels of PA in the rewards points group briefly versus control ($F_{1,58}$=5.32; P<.05) • No significant effects on PA over time
Stephens et al [26], 2019	• Location: United States • N=23 • Design: feasibility study	Female: 57	Mean 15.2; range 9.7-18.5	Weight management; pre-diabetes		• Usefulness rate: 96% • Progress toward goals frequency: 81%

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
					• Tess text-based chatbot counsellor for healthy behavior change usability assessed with progress toward goals and engagement	
Srivastana et al [57], 2019	• Location: United States • N=10 • Design: usability study	Female: 70	Range 44-67	Prediabetes	• The web-based module used to support diabetes prevention education and a mobile app that is an electronic diary and a coach	• Success of modules 60% as they meet weight loss of 5% • Compliance with dietary recommendations: 59%-87% • Compliance with PA: 52%-93%
Thompson et al [59], 2019	• Location: United States • N=27 • Design: pilot feasibility study	Female: 73 (teens)	Range 10-15	Diabetes	• Conversational agent with human features • Conversations around diabetes education	• Attrition: low (<10%) • High satisfaction: >80% • Technical issues<10% • Teens and families had a positive experience
Thompson et al [29], 2018	• Location: United States • N=48 • Design: laboratory-based study	Female: 50	Range 12-14	PA	• PA exergame with an avatar coach	• Completion: 87.5%; teens enjoyed the game (mean enjoyment score 68%) • Vigorous PA during 74.9% of the game
Duncan-Carnesciali et al [38], 2018	• Location: United States • N=198 • Design: cross-sectional, survey-based design using quantitative and qualitative paradigms	Female: 97.5	Range 26-76	Diabetes	• Avatar for diabetes management	• Ethnicity including Arab or Middle Eastern, Asian, and White or European descents as well as age were significantly associated with an excellent rating of the video with *P*<.05
Klaassen et al [42], 2018	• Location: N/A • N=21 • Design: usability study	Female: 52	Mean 13.9	Diabetes	• Conversational coach game with feedback • Integrates BCTs including information on consequences	• Usability index of 44.18 (SD 21.18; low)
Sinoo et al [56], 2018	• Location: Netherlands • N=21 • Design: experimental study	Female: 37	Mean 9.2 (SD 1.1)	Diabetes self-management	• Avatar for gameplay and diabetes self-management vs robot	• Preference for the robot (mean friendship score 4.0, SD 0.6) over the avatar (mean friendship score 2.9, SD 0.7) as a companion • Usability moderate: 58.7 (SD 24.5) • Similarity of avatar to robot led to greater friendship (*P*<.01)
Tongpeth et al [62], 2018	• Location: Australia • N=22 (development of the application) • N=10 (feasibility testing) • Design: pilot feasibility	Female: 10	Mean 52.2 (SD 10.4)	Cardiovascular: acute coronary syndrome management	• An interactive, avatar-based education application for improving patients' knowledge of, and response to, acute coronary syndrome symptoms	• Symptom recognition increased: 24% • Satisfaction: 87.3% • Knowledge increase: 15.7%

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
Friedrichs et al [63], 2014	• Location: Netherlands • N=958 • Design: 3-arm RCT	Female: 60.4	Mean 42.9 (SD 14.5)	PA; Dutch Short questionnaire	• Avatar with a web intervention or a digital web-based text condition versus control	• Significant increases in PA in the intervention arms versus control with B=0.39 in the avatar arm and B=0.44 in the text arms (*P*<.05) • No differences between the text arm or the avatar arm for PA
Stein et al [25], 2017	• Location: United States • N=70 • Design: longitudinal observational study	Female: 74.5	Mean 47 (SD 1.8); range 18-76	Weight and dietary intake	• Lark Weight Loss Health Coach (participants were a part of a diabetes prevention weight loss program) • Advice on dietary intake and PA • BCTs used include motivation, encouragement, reminders, and education	• 31% increase in healthy eating • Mean weight change: −2.4 kg (SE 0.82; 95% CI −4.03 to −0.77)
Thompson [66], 2016	• Location: United States • N=48 (round 1) • N=43 (round 2) • Design: mixed methods survey with qualitative interviews	Female: 50	Range 12-14	Preferences for a PA intervention	• Exergame with a self-representation avatar	• Desired gameplay with the avatar that could be controlled by eliciting the desired action: 62.5% male and 58.3% female • Personalized avatar: 41.7% • Most common avatar features to be customized: • Body: 95.8% • Clothing: 93.8% • Hair color: 87.5%
Behm-Morawitz et al [33], 2016	• Location: United States • N=90, female=100% (the 2 male participants were excluded) • Design: qualitative research and RCT	Female: 100	Range 18-61	Weight and PA self-efficacy	• Avatar (embodied) and video game to promote PA	• Findings support the use of the avatar for weight management t_{18}=2.15 (*P*<.05) with the intervention losing 1.75 lbs versus 0.91 lbs in the control • No effects on dietary self-efficacy • Strong correlation with avatar sense of self-presence and confidence in meeting health goals (r=0.95; *P*<.01) • Themes: avatar benefits include motivation and assisting with self-efficacy for PA • Barrier: games are not for everyone
Kuo et al [45], 2016	• Location: Taiwan • N=76 • Design: 2-arm intervention in laboratory	Female: 63.15	Mean 21.2	Eating behavior observed in laboratory	• Avatar that embodied the participants or was a weight-reduced (thinner) version of them	• Avatars that embodied a thinner version of the participants shaped eating behaviors more compared with identical self-avatars; including selecting less ice cream (Cohen d=0.35; F1,73=7.8; *P*<.01) and opted for sugar free drinks (Cohen d=0.29; F1,73=6.0; *P*<.01)

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
Ruiz et al [54], 2016	• Location: United States • N=41 • Design: laboratory study	Female: 0	Mean 64 (SD 7)	Cardiovascular behavioral risk factors (diet and exercise)	• Avatar vs a voice (nonanimated) for behavior change linked with CVD	• Avatar increased intentions (+1.56 points) to improve lifestyle behaviors relative to controls (Cohen d=0.77 $P<.01$; $t_{36}=2.48$) • Differences in confidence to change risk of heart disease was nonsignificant
LeRouge et al [14], 2015	• Location: United States • N=41 • Design: user-centered design, 3 phases with focus group and interviews	N/A	Teenagers: 12-17	Perceptions of the avatar for diet and exercise	• Interactive avatar coach	• Desire for a fun human-like interaction • Desire for a lifestyle coach and personal embodiment avatar and an authoritarian one • Desire for customization of the avatar • Advice on activity on the go and meals when eating at home • Goal setting • Technical issues could be a barrier including the internet
Thomas et al [28], 2015	• Location: United States • N=37 • Design: feasibility and usability study with pre-post test	Female: 100	Mean 55.0 (SD 8.2)	Weight-related eating behaviors	• Conversational coach for weight (focuses on dietary intake and managing eating behaviors)	• The coach assisted with perceptions of increased self-control over eating (confidence to control eating: +1 point (SD 0.2; $P<.01$) and skills for controlling eating +0.7 points (SD 0.1; $P<.01$)
Ruiz et al [53], 2014	• Location: United States • N=150 • Design: RCT	Male: 100	Mean 62 (SD 7.9)	Diabetes (knowledge)	• Computer program with an avatar to increase diabetes knowledge and medication (adherence)	• There were no significant differences between the intervention group and control group in terms of knowledge, with $P=.95$ • Satisfaction levels were higher in the digital intervention group (F4=3.11; $P<.01$)
Li et al [46], 2014	• Location: Singapore • N=140 • Design: factorial design experiment	Female: 41	Range 9-12	PA attitudes, motivation, and game performance	• Assigned to varying avatars (normal and overweight)	• Healthy weight avatars linked with greater scores in motivation for Nintendo exercise ($F_{1,134}=5.49$; $P<.05$ [boys]) attitude, and performance ($F_{1,134}=2.27$; $P<.05$ [girls])
Napolitano et al [22], 2013	• Location: United States • N=128 (phase I) N=8 (phase 2) • Design: mixed methods (pilot usability testing) study with interviews	Female: 100	Mean 34.1 (SD 13.0); range 18-60 (phase 1)	Weight, PA [14], and weight self-efficacy; satisfaction; preferences survey and interviews	• Avatar for diet and exercise • Informed by social cognitive theory • Behavioral modeling • Targeted self-efficacy • 4 weeks	• The avatar helpful: 87.5% • Mean weight loss after 4 weeks: 1.6 (SD 1.7) kg • All women found that it helped with their diet and exercise • Most were interested in the avatar
Bickmore et al [34], 2013	• Location: United States • N=122 • Design: 4-arm RCT (2 months)	Female: 61	Mean 33.0 (SD 12.6); range 21-69	Diet (NIH[t]/NCI[u] fruit and vegetable scan) and PA (IPAQ)	• Animated counselor • for diet and PA (separate and combined)	

Study and year	Location, N, and design	Sex (%)	Age (years)	Health targets and measures	Technology and procedures	Outcomes
						• No significant differences between groups in PA after adjustment • Fruit and vegetable servings significantly increased in the diet arm (F3,103=4.52; *P*<.01) • No significant differences in weight or PA between groups • Likability: Karen was perceived as nice by 35% of the participants • 50% of the participants found Karen helpful
Johnoson-Glenberg et al [41], 2013	• Location: United States • N=19 • Design: pilot feasibility study (pre-post study)	N/A	Grades 4-12 (ages 9-18)	Diet (nutrition and food choice test and knowledge)	• Diet and exercise game (exergame) with an alien interactive coach	• Differences in dietary knowledge of nutrition pre and post intervention (t=4.13; *P*<.01) and knowledge of the My Plate content in the study (t=3.29; *P*<.01)
Ruiz et al [52], 2012	• Location: United States • N=30 • Design: comparative pilot with three arms (with randomization) intervention	N/A	N/A	PA	• 3D avatar-based VR intervention	• Participants completing a 3D VR intervention mediated by avatars resembling the participants showed significant improvement in PA (*P*<.05) • No significant effects of the intervention on obese or overweight participants
Mestre et al [47], 2011	• Location: France • N=6 • Design: laboratory experimental study	N/A	Range 19-25	PA enjoyment	• Digital coach paced participants in a VR bicycling setting	• The VR coach and VR cycling were associated with higher levels of PA enjoyment (F2,10=13.24; *P*<.001) in the feedback group

[a]RCT: randomized controlled trial.
[b]BP: blood pressure.
[c]GP: general practitioner.
[d]PA: physical activity.
[e]VR: virtual reality.
[f]AI: artificial intelligence.
[g]BCT: behavior change technique.
[h]T2D: type 2 diabetes.
[i]HbA_{1c}: hemoglobin A_{1c}.
[j]FBG: fasting blood glucose.
[k]PPBG: postprandial blood glucose.
[l]HR: heart rate.
[m]RR: respiratory rate.
[n]CVD: cardiovascular disease.
[o]SRHI: Self-Report Habit Index.
[p]N/A: not applicable.
[q]IPAQ: International Physical Activity Questionnaire.
[r]PACES: physical activity enjoyment scale.
[s]HAPA: Health Action Process Approach.
[t]NIH: National Institutes of Health.
[u]NCI: National Cancer Institute.

Weight

A few studies evaluated the effects of conversational assistants for weight loss [13,22-24,44]. The study by Maher et al [13] in Australia found that the conversational assistant (chatbot) Paola assisted with a weight loss of 1.3 kg at 12 weeks follow-up (95% CI –0.1 to –2.5). In addition, there was a mean waist circumference reduction of 2.5 cm at follow-up compared with baseline (95% CI –3.5 to –0.7). The chatbot used a range of BCTs, including goal setting, self-monitoring, education, social support, and feedback to users on PA and the Mediterranean diet [13]. A study in the United States found that the Lark Weight Loss Coach, an artificial intelligence–powered bot, assisted participants with a weight loss of 2.38% (95% CI –3.75 to 1.0) with a mean use of 15 weeks [25]. The conversational agent was informed by cognitive behavioral therapy and used a range of BCTs, including education, encouragement, and reminders surrounding dietary and PA targets [25]. The determinants of weight loss included the duration of using the artificial intelligence program and engaging with it, logging meals, and the number of counseling sessions completed [25]. A large study in the United States examining the use of an avatar coach that targeted self-efficacy and modelled vicarious experiences for diet and PA (4 weeks) found that women lost an average of 1.6 (SD 1.7) kg at follow-up [22]. A study in India found that an avatar coaching app with calls from health professionals assisted with a weight loss of 1.39 kg (95% CI –0.63 to –2.01; P<.01) at 16 weeks [44]. A randomized controlled trial (RCT) with a qualitative component found that avatars increase motivation and PA self-efficacy linked with weight loss [33]. However, some studies did not report any significant weight loss [34,49].

Diet

A few studies evaluated the effects of conversational coaches (chatbots and avatars) on dietary intake and found that overall, the coaches assisted with ameliorating dietary habits and goals [13,16,25,28,34,45,49]. A study in the United States found that healthy dietary intake improved in 30% of participants who were using a conversational weight loss coach [25]. Another study found that eating behaviors improved in users of a conversational eating coach, which included increases in the mean scores for the perceptions of skills to eat healthily and self-control over their eating habits (0.7 increase in points) as well as confidence to control food consumption in social situations (1.0 increase in points; P<.01) [28]. The Paola chatbot study found a mean increase in the Mediterranean diet score [68] of 5.8 points at 12 weeks follow-up [13]. Similarly, a study of Karen, an animated counselor, found significant increases ($F_{3,103}$=4.5; P<.01) in fruit and vegetable intake in the diet intervention arm relative to the control group [34]. A further study found that eating behaviors were shaped by the appearance of the avatar, with healthier eating behavioral patterns in participants who had thinner avatars including reduced portions of ice cream and opting for healthier sugar-free drink alternatives [45].

Physical Activity

A few conversational assistant PA coaches, including chatbots and avatars, were evaluated, and overall, they assisted with increasing PA [13,16,21,23,24,27,55,63]. Most of them involved exergames with the avatar. However, one of the studies did not find any improvements in PA among the 2 avatars, attributing improvements only to the web-based part of the intervention [23], and another study did not find a difference between the web-based intervention and the chatbot (only when considering a standard control) [63]. A preliminary usability study in Australia found that step count goals increased 59% of the time in users of the chatbot that targeted PA and that participants had a preference for personalization and greater knowledge-based content [16]. Another pilot study of Paola, the chatbot in Australia, found that it assisted with increasing mean step count by 109 minutes per week at 12 weeks follow-up (95% CI 1.9-217.7) [13]. A study involving an exergame that used a PA avatar coach in teens found that 75% of the time, participants engaged in 15.88 (SD 5.8) minutes of vigorous PA throughout the game [29]. Participants also wanted the avatar to have a supportive and nonpatronizing or nondisparaging tone in interactions regarding PA and found that it could motivate older adults when adequately personalized [55]. Similarly, a study in children also found that they desired the option to personalize the avatar, including controlling and customizing its physical appearance during game play when exercising [66].

Proteus Effect

The Proteus effect is a phenomenon wherein individuals embody and emulate the behaviors of their virtual characters such as avatars [69,70]. A few studies demonstrated support for the Proteus effect when it came to PA behaviors, although the type of avatar varied. A study in Taiwan found that younger looking avatars were associated with higher levels of PA than older looking avatars but only in women. Male participants had higher levels of PA than female participants who used an older looking avatar, highlighting differences between sexes [27]. A further study found a higher cardiac output resulting from increased intensity of PA in adult users of an avatar that resembled them and wore gym clothes when compared with avatars that appeared unfamiliar like strangers in regular clothing, which reduced heart rate [24]. Similarly, a study in Taiwan found increases in physical activity assessed in movements (986.7 points higher) in users of a “normal avatar”, more closely resembling them than the most muscular avatar [21]. They also found that self-efficacy was higher (0.66 points) for core muscle exercises in female participants assigned to normal avatars relative to their muscular counterparts and male participants assigned to the same standard avatar (0.9 points higher), with P<.05 [21]. Similarly, dietary behavior was also shaped by thinner embodied avatars in another study [45].

Diabetes

Most diabetes studies were feasibility studies. The results of diabetes conversational coaches were mixed. A few studies did not have positive findings concerning the applications with avatars for diabetes [42,53]. However, one study reported a usability score of 73, which is relatively high. Notably, the study integrated a range of BCTs, including goal setting, feedback, self-monitoring, social support, and counseling [32]. Low usability scores were reported in a few studies, including one that reported an overall score of 44.58 (SD 21.18) [42].

Similarly, an RCT of a diabetes coaching avatar did not find that knowledge increased relative to controls, but intervention participants in the computer-based programmed dynamic avatar had higher satisfaction levels (F_4=3.11; P=.01) [53]. Another study in the United States in participants with prediabetes found that 60% of patients had successfully completed the modules and met weight targets during 6 months of use (60% success rate) [57]. Engagement was also moderately high (50%) in a study in Singapore involving a chatbot, although usability was high along with retention (93%) [36]. In a study in teenagers, attrition was also low, and 80% of the participants were satisfied with the conversational diabetes coach [59]. A study that evaluated a coaching application for diabetes found an improvement of –11 mg/dL in fasting blood glucose levels [44]. However, the intervention also involved phone calls from health professionals [44]. Similarly, an avatar application with a ketogenic diet program assisted with a reduction in hemoglobin A_{1c} levels of 0.69% (SE 0.168%) [58]. Qualitative research found that avatars created an environment of social presence that facilitated social support and coherence for patients with diabetes [48]. In another study of avatars combined with robots, children preferred robots over avatars, but their friendship increased if the two had a greater similarity, which impacted usability [56].

CVD and Associated Risk Factors

A few studies evaluated the use of conversational coaches for CVD. One of them was a pilot study of the Tensiobot chatbot [17], a coaching application that teaches users how to properly check their blood pressure using recommended practice guidelines and provides users with graphic feedback and reminders. The study found that the chatbot group did not differ from the control group in terms of adhering to blood pressure measurement recommendations. However, there were significantly higher levels of knowledge (+6.53 points) with regard to checking blood pressure in the chatbot group than in the control group (P<.05) [17]. Blood pressure (diastolic) was significantly reduced, that is, by 1.43 mm Hg (SE 0.65; 95% CI –2.72 to –0.14; P<.01), in users of an avatar application that also involved a ketogenic diet [58]. In addition, a mixed methods study with a qualitative component found that users in general were interested in trying a hypertension chatbot for medication management as well as for health communication and self-care [19]. In addition to these studies, a general diet and PA chatbot study evaluated changes in blood pressure, but these changes were nonsignificant [13]. A study in Poland found high desirability for a CVD voice technology coach, in addition to accessing phone-based telemedical services by health professionals [43]. A further study in Brazil evaluated avatars for cardiovascular rehabilitation and found that an avatar with an exergame influenced heart rate, systolic blood pressure, and respiratory rate during the intervention and up to 5 minutes after its completion [31]. Furthermore, a study found that the avatar intervention increased the intent to improve lifestyle behavioral risk factors in patients relative to controls (P=.01), although confidence did not change [54]. Finally, a study evaluated a cardiovascular educational avatar application and found that it increased symptom recognition by 24% and knowledge of CVD by 15%, with a high satisfaction rate of 87% among patients [62].

User Perceptions

Several studies found that users were interested in using conversational coaches for lifestyle behaviors [14,19,22]. Overall, participants enjoyed using the chatbots and avatars or found them helpful for diet, exercise, and hypertension management [17,22,25,26,29]. User-friendliness was reported by 83% of the participants in a study that evaluated a PA social media chatbot [20]. Similarly, 87.5% of women in a weight loss avatar intervention found it helpful [22]. With the exception of studies on diabetes conversational coaches, adherence or completion of tasks was high across studies on lifestyle (diet and PA) conversational coaches, ranging from 85% to 90% [13,16,17,29]. The qualitative study themes were related to the desirability for a conversational coach for hypertension and weight-related behaviors, especially for one that simulates human interaction closely, provides advice and goals for meals when cooking, and provides educational support [14] including for hypertension management [14,19].

Technological Challenges

A few tech challenges were brought up across the studies. Although users found that the conversational coach answered basic questions correctly, failure to understand and respond to more complex or spontaneous questions was reported in the studies. The percentage of failure for spontaneous or complex questions was 79% in one study [16], and participants in another study gave a high ranking for the chatbot's failure to recognize their input [20]. Paola chatbot correctly answered spontaneous questions on diet in 4 out of 20 attempts, with a success of 20%, while the percentage of correctly answered simple and predetermined questions and responses was 96% and 97% [16].

Discussion

Principal Findings

This review aimed to better understand the effectiveness of virtual coaches for managing metabolic health and weight-related risk factors. It appears that virtual coaches hold potential for assisting patients with improving their dietary intake and PA behaviors, leading to subsequent weight loss. However, more studies that are larger and sufficiently powered RCTs are needed to establish a stronger evidence base. RCTs are the gold standard of evidence but are often costly and time-consuming [71,72]. Most of the studies were limited, as they were pilot studies. Ideally, it would be of interest to research long-term weight changes and cardiometabolic risk factor modifications over longer periods.

It appears that PA interventions may benefit from using avatars that embody the participant. The Proteus effect is based on the hypothesis that users adjust their behavior by modeling the virtual character with which they interact [73]. Thus, it seems that incorporating an avatar may enhance mHealth chatbot interventions, as it adds the element of user interaction and promotes the modeling of behavior through embodiment [73]. However, 1 (2%) study did not find that the avatars enhanced the effects of the web-based intervention [23].

We also found that consumers seemed to be interested in and enthusiastic about trying virtual coaches for managing their weight-related behaviors and blood pressure. Adherence to the intervention was also high throughout the studies, which indicates that this technology is acceptable and usable for patients. However, there is a need to undertake qualitative research on developing a MetS coach to further understand consumer perspectives. The main barrier to consider when developing future virtual agents is that the virtual agents did not always answer correctly to spontaneous responses. As consumers want personalized and tailored mHealth for weight-related behaviors [74], future applications should ensure that the virtual agents are sufficiently advanced to be able to interact with users in a natural and personalized manner.

It appears that diabetes virtual coaches should be improved to maximize engagement and adherence, as not all studies found that they were helpful. Although outside the scope of this review, we note that some studies used BCTs, which could suggest that future applications may benefit from integrating BCTs [13,20,22,25,32,42,51]. In addition, we identified some studies on blood pressure and CVD management, which demonstrated preliminary improvement in patients with hypertension as well as knowledge of CVD. However, we did not identify any virtual coaches for managing MetS. Therefore, there is a need to develop virtual coaches specifically tailored to this syndrome and its associated risk factors. Such virtual coaches could be integrated into a combined synchronized application that involves diabetes and CVD education and monitoring.

MetS is linked with high blood pressure, which is one of the main hallmarks of the disease. The theoretical mechanisms underpinning the development of hypertension in patients with MetS have included a combination of endothelial dysfunction, systemic inflammation, adiposity, and oxidative stress [75]. Dysfunction in the renin-angiotensin system has also been theorized to be a determinant [75]. Obesity itself has also been identified as a risk factor for high blood pressure in MetS [76]. Blood pressure is modifiable to some extent through lifestyle changes previously described, including dietary sodium restriction, PA, stress reduction [77], and medication [78]. Future virtual coaches may target hypertension as part of a MetS intervention, and this review found that patients are willing to try chatbots for managing their blood pressure.

MetS is also associated with high glucose levels of at least 100 mg/dL when patients are fasting [78], which indicates that they are in the prediabetes stage, as diabetes begins at fasting glucose levels of 126 mg/dL [79]. In a recent longitudinal study, patients who reduced their fasting blood glucose levels decreased their overall risk of diabetes by 54% when compared with their counterparts who did not improve their blood sugar levels (95% CIs exclude 1) [80]. A recent study found that individuals who consumed high amounts of sugar were 32% more likely to have MetS than their counterparts [81]. Thus, a future MetS virtual coach could target blood glucose monitoring and offer personalized advice on optimum sugar intake.

In addition to targeting dietary intake, PA is integral to managing this syndrome. A meta-analysis found that the risk of cardiovascular events was reduced by 30% in physically active individuals compared with those who were inactive [6]. A longitudinal study in middle-aged women found that increasing step counts significantly reduced, by 30%, the risk of MetS in this population and that they had clinically improved levels of the protective cholesterol high-density lipoprotein, whereas their serum triglycerides had significantly decreased [82]. A review found that walking on a daily basis reduced the risk of type 2 diabetes by nearly half [5]. Furthermore, recent research suggests that sedentary behavior, including sitting time, is an independent and significant risk factor for MetS syndrome [83]. Thus, PA chatbots and avatars, which were found to increase PA time, steps, and self-efficacy in this review, could be integrated into a comprehensive future MetS interventions.

Given that chatbots and avatars hold potential for increasing PA and reducing sedentary behavior, as well as improving dietary intake, studies are needed to evaluate their effectiveness for managing the symptoms and risk factors associated with MetS specifically.

In addition, stress is often an underlying determinant of maladaptive weight-related behaviors, including binge eating, emotional eating, and an unhealthy dietary intake as well as weight gain [84-88]. Future avatar and chatbot interventions for cardiometabolic factors could also consider integrating psychological supportive interventions such as mindfulness-based stress reduction, which assists with weight and stress [89-93], as an element.

Conclusions

In summary, we found that virtual coaches hold promise for regulating diet, PA, weight, and possibly hypertension. However, studies on virtual coaches are few in number; therefore, more research, including RCTs, is needed to confirm the effectiveness of virtual coaches. Overall, most participants in the reviewed studies were interested in using virtual coaches, including chatbots and avatars, for regulating their weight-related behaviors, and study adherence was good. Future interventions could be ameliorated to reduce technical challenges associated with these conversational agents and ensure that they respond correctly to complex and spontaneous questions. Furthermore, future research could involve developing a comprehensive conversational agent for MetS, such as a health coach that simultaneously targets diet (sodium, sugar, and fat intake), exercise, weight (including abdominal obesity), blood pressure, and diabetes, and evaluating it. This would include a health coach that simultaneously targets diet (sodium, sugar, and fat intake), exercise, weight (including abdominal obesity), blood pressure, and diabetes.

Conflicts of Interest

None declared.

References

1. Han TS, Lean ME. A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovasc Dis 2016;5:2048004016633371 [FREE Full text] [doi: 10.1177/2048004016633371] [Medline: 26998259]
2. About metabolic syndrome. American Heart Association. 2021. URL: https://wwwheartorg/en/health-topics/metabolic-syndrome/about-metabolic-syndrome [accessed 2022-01-10]
3. Wiklund P. The role of physical activity and exercise in obesity and weight management: time for critical appraisal. J Sport Health Sci 2016 Jun;5(2):151-154 [FREE Full text] [doi: 10.1016/j.jshs.2016.04.001] [Medline: 30356545]
4. Hegde SM, Solomon SD. Influence of physical activity on hypertension and cardiac structure and function. Curr Hypertens Rep 2015 Oct;17(10):77 [FREE Full text] [doi: 10.1007/s11906-015-0588-3] [Medline: 26277725]
5. Hamasaki H. Daily physical activity and type 2 diabetes: a review. World J Diabetes 2016 Jun 25;7(12):243-251 [FREE Full text] [doi: 10.4239/wjd.v7.i12.243] [Medline: 27350847]
6. Li J, Siegrist J. Physical activity and risk of cardiovascular disease--a meta-analysis of prospective cohort studies. Int J Environ Res Public Health 2012 Feb;9(2):391-407 [FREE Full text] [doi: 10.3390/ijerph9020391] [Medline: 22470299]
7. Hoyas I, Leon-Sanz M. Nutritional challenges in metabolic syndrome. J Clin Med 2019 Aug 24;8(9):1301 [FREE Full text] [doi: 10.3390/jcm8091301] [Medline: 31450565]
8. Lyzwinski LN. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014 Jun 30;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]
9. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
10. mHealth: new horizons for health through mobile technologie. World Health Organization Global Observatory for eHealth. 2011. URL: http://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1 [accessed 2022-01-10]
11. Lefebvre C. Integrating cell phones and mobile technologies into public health practice: a social marketing perspective. Health Promot Pract 2009 Oct;10(4):490-494. [doi: 10.1177/1524839909342849] [Medline: 19809002]
12. Griffin AC, Xing Z, Khairat S, Wang Y, Bailey S, Arguello J, et al. Conversational agents for chronic disease self-management: a systematic review. AMIA Annu Symp Proc 2020;2020:504-513 [FREE Full text] [Medline: 33936424]
13. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
14. LeRouge C, Dickhut K, Lisetti C, Sangameswaran S, Malasanos T. Engaging adolescents in a computer-based weight management program: avatars and virtual coaches could help. J Am Med Inform Assoc 2016 Jan;23(1):19-28 [FREE Full text] [doi: 10.1093/jamia/ocv078] [Medline: 26159466]
15. Virtual trainer. Charamel. 2021. URL: https://wwwcharamelcom/en/case-studies/virtual-trainer [accessed 2022-09-01]
16. Davis CR, Murphy KJ, Curtis RG, Maher CA. A process evaluation examining the performance, adherence, and acceptability of a physical activity and diet artificial intelligence virtual health assistant. Int J Environ Res Public Health 2020 Dec 07;17(23):9137 [FREE Full text] [doi: 10.3390/ijerph17239137] [Medline: 33297456]
17. Echeazarra L, Pereira J, Saracho R. TensioBot: a chatbot assistant for self-managed in-house blood pressure checking. J Med Syst 2021 Mar 15;45(4):54. [doi: 10.1007/s10916-021-01730-x] [Medline: 33723721]
18. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
19. Griffin AC, Xing Z, Mikles SP, Bailey S, Khairat S, Arguello J, et al. Information needs and perceptions of chatbots for hypertension medication self-management: a mixed methods study. JAMIA Open 2021 Apr;4(2):ooab021 [FREE Full text] [doi: 10.1093/jamiaopen/ooab021] [Medline: 33898936]
20. Larbi D, Gabarron E, Denecke K. Social media chatbot for increasing physical activity: usability study. Stud Health Technol Inform 2021 Oct 27;285:227-232. [doi: 10.3233/SHTI210604] [Medline: 34734878]
21. Lin JT, Wu D, Yang J. Exercising with a six pack in virtual reality: examining the proteus effect of avatar body shape and sex on self-efficacy for core-muscle exercise, self-concept of body shape, and actual physical activity. Front Psychol 2021;12:693543 [FREE Full text] [doi: 10.3389/fpsyg.2021.693543] [Medline: 34690859]
22. Napolitano MA, Hayes S, Russo G, Muresu D, Giordano A, Foster GD. Using avatars to model weight loss behaviors: participant attitudes and technology development. J Diabetes Sci Technol 2013 Jul 01;7(4):1057-1065 [FREE Full text] [doi: 10.1177/193229681300700430] [Medline: 23911189]
23. Navarro J, Cebolla A, Llorens R, Borrego A, Baños RM. Manipulating self-avatar body dimensions in virtual worlds to complement an internet-delivered intervention to increase physical activity in overweight women. Int J Environ Res Public Health 2020 Jun 05;17(11):4045 [FREE Full text] [doi: 10.3390/ijerph17114045] [Medline: 32517117]
24. Navarro J, Peña J, Cebolla A, Baños R. Can avatar appearance influence physical activity? User-avatar similarity and proteus effects on cardiac frequency and step counts. Health Commun 2022 Feb;37(2):222-229. [doi: 10.1080/10410236.2020.1834194] [Medline: 33054371]

25. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
26. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
27. Tammy Lin JH, Wu DY. Exercising with embodied young avatars: how young vs. older avatars in virtual reality affect perceived exertion and physical activity among male and female elderly individuals. Front Psychol 2021 Oct 25;12:693545 [FREE Full text] [doi: 10.3389/fpsyg.2021.693545] [Medline: 34759856]
28. Thomas JG, Spitalnick JS, Hadley W, Bond DS, Wing RR. Development of and feedback on a fully automated virtual reality system for online training in weight management skills. J Diabetes Sci Technol 2015 Jan;9(1):145-148 [FREE Full text] [doi: 10.1177/1932296814557326] [Medline: 25367014]
29. Thompson DI, Cantu D, Callender C, Liu Y, Rajendran M, Rajendran M, et al. Photorealistic avatar and teen physical activity: feasibility and preliminary efficacy. Games Health J 2018 Apr;7(2):143-150 [FREE Full text] [doi: 10.1089/g4h.2017.0103] [Medline: 29406774]
30. Ahn SJ, Johnsen K, Ball C. Points-based reward systems in gamification impact children's physical activity strategies and psychological needs. Health Educ Behav 2019 Jun;46(3):417-425 [FREE Full text] [doi: 10.1177/1090198118818241] [Medline: 30678507]
31. Alves da Cruz MM, Ricci-Vitor AL, Bonini Borges GL, Fernanda da Silva P, Ribeiro F, Marques Vanderlei LC. Acute hemodynamic effects of virtual reality-based therapy in patients of cardiovascular rehabilitation: a cluster randomized crossover trial. Arch Phys Med Rehabil 2020 Apr;101(4):642-649. [doi: 10.1016/j.apmr.2019.12.006] [Medline: 31926142]
32. Balsa J, Félix I, Cláudio AP, Carmo MB, Silva IC, Guerreiro A, et al. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. J Med Syst 2020 Jun 13;44(7):130. [doi: 10.1007/s10916-020-01583-w] [Medline: 32533367]
33. Behm-Morawitz E, Lewallen J, Choi G. A second chance at health: how a 3D virtual world can improve health self-efficacy for weight loss management among adults. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):74-79. [doi: 10.1089/cyber.2015.0317] [Medline: 26882324]
34. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
35. Chin J, Quinn K, Muramatsu N, Marquez D. A user study on the feasibility and acceptance of delivering physical activity programs to older adults through conversational agents. Proc Hum Factors Ergon Soc Annu Meet 2020 Dec;64(1):33-37 [FREE Full text] [doi: 10.1177/1071181320641010] [Medline: 34177216]
36. Dhinagaran DA, Sathish T, Soong A, Theng Y, Best J, Tudor Car L. Conversational agent for healthy lifestyle behavior change: web-based feasibility study. JMIR Form Res 2021 Dec 03;5(12):e27956 [FREE Full text] [doi: 10.2196/27956] [Medline: 34870611]
37. Dol A, Bode C, Velthuijsen H, van Strien T, van Gemert-Pijnen L. Application of three different coaching strategies through a virtual coach for people with emotional eating: a vignette study. J Eat Disord 2021 Jan 14;9(1):13 [FREE Full text] [doi: 10.1186/s40337-020-00367-4] [Medline: 33446275]
38. Duncan-Carnesciali J, Wallace BC, Odlum M. An evaluation of a diabetes self-management education (DSME) intervention delivered using avatar-based technology: certified diabetes educators' ratings and perceptions. Diabetes Educ 2018 Jun;44(3):216-224. [doi: 10.1177/0145721718769474] [Medline: 29644931]
39. Hahn L, Rathbun SL, Schmidt MD, Johnsen K, Annesi JJ, Ahn SJ. Using virtual agents and activity monitors to autonomously track and assess self-determined physical activity among young children: a 6-week feasibility field study. Cyberpsychol Behav Soc Netw 2020 Jul;23(7):471-478 [FREE Full text] [doi: 10.1089/cyber.2019.0491] [Medline: 32379498]
40. Hickman RL, Clochesy JM, Alaamri M. Effects of an eHealth intervention on patient-provider interaction and functional health literacy in adults with hypertension. SAGE Open Nurs 2021;7:23779608211005863 [FREE Full text] [doi: 10.1177/23779608211005863] [Medline: 33997280]
41. Johnson-Glenberg MC, Hekler EB. "Alien health game": an embodied exergame to instruct in nutrition and MyPlate. Games Health J 2013 Dec;2(6):354-361. [doi: 10.1089/g4h.2013.0057] [Medline: 26197077]
42. Klaassen R, Bul KC, Op den Akker R, van der Burg GJ, Kato PM, Di Bitonto P. Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. Sensors (Basel) 2018 Jan 30;18(2):402 [FREE Full text] [doi: 10.3390/s18020402] [Medline: 29385750]
43. Kowalska M, Gładyś A, Kalańska-Łukasik B, Gruz-Kwapisz M, Wojakowski W, Jadczyk T. Readiness for voice technology in patients with cardiovascular diseases: cross-sectional study. J Med Internet Res 2020 Dec 17;22(12):e20456 [FREE Full text] [doi: 10.2196/20456] [Medline: 33331824]
44. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating glycemic control in patients of south Asian origin with type 2 diabetes using a digital therapeutic platform: analysis of real-world data. J Med Internet Res 2021 Mar 25;23(3):e17908 [FREE Full text] [doi: 10.2196/17908] [Medline: 33764306]

45. Kuo H, Lee C, Chiou W. The power of the virtual ideal self in weight control: weight-reduced avatars can enhance the tendency to delay gratification and regulate dietary practices. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):80-85. [doi: 10.1089/cyber.2015.0203] [Medline: 26544532]
46. Li BJ, Lwin MO, Jung Y. Wii, myself, and size: the influence of proteus effect and stereotype threat on overweight children's exercise motivation and behavior in exergames. Games Health J 2014 Feb;3(1):40-48. [doi: 10.1089/g4h.2013.0081] [Medline: 26197254]
47. Mestre DR, Ewald M, Maiano C. Virtual reality and exercise: behavioral and psychological effects of visual feedback. Stud Health Technol Inform 2011;167:122-127. [Medline: 21685653]
48. Mitchell S, Bragg A, Gardiner P, De La Cruz B, Laird L. Patient engagement and presence in a virtual world world diabetes self-management education intervention for minority women. Patient Educ Couns 2022 Apr;105(4):797-804 [FREE Full text] [doi: 10.1016/j.pec.2021.06.033] [Medline: 34226067]
49. Napolitano MA, Harrington CB, Patchen L, Ellis LP, Ma T, Chang K, et al. Feasibility of a digital intervention to promote healthy weight management among postpartum African American/Black women. Int J Environ Res Public Health 2021 Feb 23;18(4):2178 [FREE Full text] [doi: 10.3390/ijerph18042178] [Medline: 33672229]
50. Naylor JB, Patton BJ, Barkley JE. VO2, liking, and relative reinforcing value of cooperative and competitive exergame play in young children. Int J Exerc Sci 2020;13(5):1501-1511 [FREE Full text] [Medline: 33414882]
51. Piao M, Ryu H, Lee H, Kim J. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
52. Ruiz JG, Andrade AD, Anam R, Aguiar R, Sun H, Roos BA. Using anthropomorphic avatars resembling sedentary older individuals as models to enhance self-efficacy and adherence to physical activity: psychophysiological correlates. Stud Health Technol Inform 2012;173:405-411. [Medline: 22357026]
53. Ruiz JG, Andrade AD, Anam R, Lisigurski M, Karanam C, Sharit J. Computer-based programmed instruction did not improve the knowledge retention of medication instructions of individuals with type 2 diabetes mellitus. Diabetes Educ 2014;40(1):77-88. [doi: 10.1177/0145721713508825] [Medline: 24168837]
54. Ruiz JG, Andrade AD, Karanam C, Krishnamurthy D, Niño L, Anam R, et al. The communication of global cardiovascular risk by avatars. Stud Health Technol Inform 2016;220:341-344. [Medline: 27046602]
55. Santini S, Stara V, Galassi F, Merizzi A, Schneider C, Schwammer S, et al. User requirements analysis of an embodied conversational agent for coaching older adults to choose active and healthy ageing behaviors during the transition to retirement: a cross-national user centered design study. Int J Environ Res Public Health 2021 Sep 14;18(18):9681 [FREE Full text] [doi: 10.3390/ijerph18189681] [Medline: 34574615]
56. Sinoo C, van der Pal S, Blanson Henkemans OA, Keizer A, Bierman BP, Looije R, et al. Friendship with a robot: children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. Patient Educ Couns 2018 Jul;101(7):1248-1255. [doi: 10.1016/j.pec.2018.02.008] [Medline: 29548599]
57. Srivastava P, Verma A, Geronimo C, Button TM. Behavior stages of a physician- and coach-supported cloud-based diabetes prevention program for people with prediabetes. SAGE Open Med 2019;7:2050312119841986 [FREE Full text] [doi: 10.1177/2050312119841986] [Medline: 31105938]
58. Strombotne KL, Lum J, Ndugga NJ, Utech AE, Pizer SD, Frakt AB, et al. Effectiveness of a ketogenic diet and virtual coaching intervention for patients with diabetes: a difference-in-differences analysis. Diabetes Obes Metab 2021 Dec;23(12):2643-2650 [FREE Full text] [doi: 10.1111/dom.14515] [Medline: 34351035]
59. Thompson D, Callender C, Gonynor C, Cullen KW, Redondo MJ, Butler A, et al. Using relational agents to promote family communication around type 1 diabetes self-management in the diabetes family teamwork online intervention: longitudinal pilot study. J Med Internet Res 2019 Sep 13;21(9):e15318 [FREE Full text] [doi: 10.2196/15318] [Medline: 31538940]
60. Thompson D, Cullen KW, Redondo MJ, Anderson B. Use of relational agents to improve family communication in type 1 diabetes: methods. JMIR Res Protoc 2016 Jul 28;5(3):e151 [FREE Full text] [doi: 10.2196/resprot.5817] [Medline: 27468762]
61. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. JMIR Mhealth Uhealth 2021 Nov 26;9(11):e28577 [FREE Full text] [doi: 10.2196/28577] [Medline: 34842552]
62. Tongpeth J, Du HY, Clark RA. Development and feasibility testing of an avatar-based education application for patients with acute coronary syndrome. J Clin Nurs 2018 Oct;27(19-20):3561-3571. [doi: 10.1111/jocn.14528] [Medline: 29777554]
63. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
64. Hahn L, Schmidt MD, Rathbun SL, Johnsen K, Annesi JJ, Ahn SJ. Using virtual agents to increase physical activity in young children with the virtual fitness buddy ecosystem: study protocol for a cluster randomized trial. Contemp Clin Trials 2020 Dec;99:106181 [FREE Full text] [doi: 10.1016/j.cct.2020.106181] [Medline: 33096225]
65. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22(1):77-89. [doi: 10.3109/14639239709089835] [Medline: 9183781]

66. Thompson D, Cantu D, Rajendran M, Rajendran M, Bhargava T, Zhang Y, et al. Development of a teen-focused exergame. Games Health J 2016 Oct;5(5):342-356 [FREE Full text] [doi: 10.1089/g4h.2016.0035] [Medline: 27680385]
67. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
68. Aparicio-Ugarriza R, Cuenca-García M, Gonzalez-Gross M, Julián C, Bel-Serrat S, Moreno LA, et al. Relative validation of the adapted Mediterranean diet score for adolescents by comparison with nutritional biomarkers and nutrient and food intakes: the healthy lifestyle in Europe by nutrition in adolescence (HELENA) study. Public Health Nutr 2019 Sep;22(13):2381-2397. [doi: 10.1017/S1368980019001022] [Medline: 31204628]
69. Stavropoulos V, Pontes HM, Gomez R, Schivinski B, Griffiths M. Proteus effect profiles: how do they relate with disordered gaming behaviours? Psychiatr Q 2020 Sep;91(3):615-628. [doi: 10.1007/s11126-020-09727-4] [Medline: 32140972]
70. Dauchot N. Introduction to the proteus effect. Medium. 2018. URL: https://medium.com/uxxr/the-proteus-effect-aeb46d6dfd86 [accessed 2022-09-01]
71. Hariton E, Locascio J. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018 Dec;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
72. Bondemark L, Ruf S. Randomized controlled trial: the gold standard or an unobtainable fallacy? Eur J Orthod 2015 Oct;37(5):457-461. [doi: 10.1093/ejo/cjv046] [Medline: 26136438]
73. Yee N, Bailenson J. The proteus effect: the effect of transformed self-representation on behavior. Human Comm Res 2007 Jul;33(3):271-290 [FREE Full text] [doi: 10.1111/j.1468-2958.2007.00299.x]
74. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2018 May;24(4):290-302. [doi: 10.1177/1357633X17692722] [Medline: 28181859]
75. Yanai H, Tomono Y, Ito K, Furutani N, Yoshida H, Tada N. The underlying mechanisms for development of hypertension in the metabolic syndrome. Nutr J 2008 Apr 17;7:10 [FREE Full text] [doi: 10.1186/1475-2891-7-10] [Medline: 18416854]
76. Franklin SS. Hypertension in the metabolic syndrome. Metab Syndr Relat Disord 2006;4(4):287-298. [doi: 10.1089/met.2006.4.287] [Medline: 18370747]
77. Sparrenberger F, Cichelero FT, Ascoli AM, Fonseca FP, Weiss G, Berwanger O, et al. Does psychosocial stress cause hypertension? A systematic review of observational studies. J Hum Hypertens 2009 Jan;23(1):12-19. [doi: 10.1038/jhh.2008.74] [Medline: 18615099]
78. Nguyen Q, Dominguez J, Nguyen L, Gullapalli N. Hypertension management: an update. Am Health Drug Benefits 2010 Jan;3(1):47-56 [FREE Full text] [Medline: 25126308]
79. Swarup S, Goyal A, Grigorova Y, Zeltser R. Metabolic syndrome. In: StatPearls. Treasure Island, FL, USA: StatPearls Publishing; 2021.
80. Lee M, Han K, Kim MK, Koh ES, Kim ES, Nam GE, et al. Changes in metabolic syndrome and its components and the risk of type 2 diabetes: a nationwide cohort study. Sci Rep 2020 Feb 11;10(1):2313 [FREE Full text] [doi: 10.1038/s41598-020-59203-z] [Medline: 32047219]
81. Seo EH, Kim H, Kwon O. Association between total sugar intake and metabolic syndrome in middle-aged Korean men and women. Nutrients 2019 Sep 01;11(9):2042 [FREE Full text] [doi: 10.3390/nu11092042] [Medline: 31480603]
82. Zając-Gawlak I, Pelclová J, Groffik D, Přidalová M, Nawrat-Szołtysik A, Kroemeke A, et al. Does physical activity lower the risk for metabolic syndrome: a longitudinal study of physically active older women. BMC Geriatr 2021 Jan 06;21(1):11 [FREE Full text] [doi: 10.1186/s12877-020-01952-7] [Medline: 33407205]
83. Bankoski A, Harris TB, McClain JJ, Brychta RJ, Caserotti P, Chen KY, et al. Sedentary activity associated with metabolic syndrome independent of physical activity. Diabetes Care 2011 Feb;34(2):497-503 [FREE Full text] [doi: 10.2337/dc10-0987] [Medline: 21270206]
84. Moyer AE, Rodin J, Grilo CM, Cummings N, Larson LM, Rebuffé-Scrive M. Stress-induced cortisol response and fat distribution in women. Obes Res 1994 May;2(3):255-262 [FREE Full text] [doi: 10.1002/j.1550-8528.1994.tb00055.x] [Medline: 16353426]
85. Kandiah J, Yake M, Jones J, Meyer M. Stress influences appetite and comfort food preferences in college women. Nutr Res 2006 Mar;26(3):118-123 [FREE Full text] [doi: 10.1016/j.nutres.2005.11.010]
86. Torres SJ, Nowson CA. Relationship between stress, eating behavior, and obesity. Nutrition 2007;23(11-12):887-894. [doi: 10.1016/j.nut.2007.08.008] [Medline: 17869482]
87. Wichianson JR, Bughi SA, Unger JB, Spruijt-Metz D, Nguyen-Rodriguez ST. Perceived stress, coping and night-eating in college students. Stress Med 2009 Aug;25(3):235-240 [FREE Full text] [doi: 10.1002/smi.1242]
88. Yau YH, Potenza MN. Stress and eating behaviors. Minerva Endocrinol 2013 Sep;38(3):255-267 [FREE Full text] [Medline: 24126546]
89. Chiesa A, Malinowski P. Mindfulness-based approaches: are they all the same? J Clin Psychol 2011 Apr;67(4):404-424. [doi: 10.1002/jclp.20776] [Medline: 21254062]
90. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]

91. Labee EE. Psychology Moment by Moment: A Guide to Enhancing Your Clinical Practice with Mindfulness and Meditation. Oakland, CA, USA: New Harbinger Publications; 2011.
92. O'Reilly GA, Cook L, Spruijt-Metz D, Black DS. Mindfulness-based interventions for obesity-related eating behaviours: a literature review. Obes Rev 2014 Jun;15(6):453-461 [FREE Full text] [doi: 10.1111/obr.12156] [Medline: 24636206]
93. Olson KL, Emery CF. Mindfulness and weight loss: a systematic review. Psychosom Med 2015 Jan;77(1):59-67. [doi: 10.1097/PSY.0000000000000127] [Medline: 25490697]

Abbreviations

BCT: behavior change technique
CVD: cardiovascular disease
MetS: metabolic syndrome
mHealth: mobile health
PA: physical activity
RCT: randomized controlled trial
RQ: research question

Edited by L Buis, M Dorsch; submitted 16.05.22; peer-reviewed by N Maglaveras, YC Wang; comments to author 12.08.22; revised version received 04.10.22; accepted 23.12.22; published 25.05.23.

Please cite as:
Lyzwinski LN, Elgendi M, Menon C
Conversational Agents and Avatars for Cardiometabolic Risk Factors and Lifestyle-Related Behaviors: Scoping Review
JMIR Mhealth Uhealth 2023;11:e39649
URL: https://mhealth.jmir.org/2023/1/e39649
doi:10.2196/39649
PMID:37227765

Original Paper

Understanding Mobile Health and Youth Mental Health: Scoping Review

Xiaoxu Ding[1,2], MSc; Kelli Wuerth[3], PhD; Brodie Sakakibara[1,2,4], PhD; Julia Schmidt[1,2], PhD; Natalie Parde[5,6], PhD; Liisa Holsti[1,2], PhD; Skye Barbic[1,2,3,7], PhD

[1]Faculty of Medicine, Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, BC, Canada
[2]Faculty of Graduate Studies, Rehabilitation Science, University of British Columbia, Vancouver, BC, Canada
[3]Foundry, Providence Health Care, Vancouver, BC, Canada
[4]Faculty of Medicine, Centre for Chronic Disease Prevention and Management, University of British Columbia, Kelowna, BC, Canada
[5]Department of Computer Science, University of Illinois Chicago, Chicago, IL, United States
[6]Natural Language Processing Laboratory, University of Illinois Chicago, Chicago, IL, United States
[7]Centre for Health Evaluation & Outcome Sciences, Vancouver, BC, Canada

Corresponding Author:
Xiaoxu Ding, MSc
Faculty of Medicine
Department of Occupational Science and Occupational Therapy
University of British Columbia
T325 - 2211 Wesbrook Mall | Musqueam Territory
Vancouver, BC, V6T 2A1
Canada
Phone: 1 236 992 8222
Fax: 1 604 822 7624
Email: xxd51@student.ubc.ca

Abstract

Background: A total of 75% of people with mental health disorders have an onset of illness between the ages of 12 and 24 years. Many in this age group report substantial obstacles to receiving quality youth-centered mental health care services. With the rapid development of technology and the recent COVID-19 pandemic, mobile health (mHealth) has presented new opportunities for youth mental health research, practice, and policy.

Objective: The research objectives were to (1) synthesize the current evidence supporting mHealth interventions for youths who experience mental health challenges and (2) identify current gaps in the mHealth field related to youth's access to mental health services and health outcomes.

Methods: Guided by the methods of Arksey and O'Malley, we conducted a scoping review of peer-reviewed studies that used mHealth tools to improve youth mental health (January 2016-February 2022). We searched MEDLINE, PubMed, PsycINFO, and Embase databases using the following key terms: (1) mHealth; (2) youth and young adults; and (3) mental health. The current gaps were analyzed using content analysis.

Results: The search produced 4270 records, of which 151 met inclusion criteria. Included articles highlight the comprehensive aspects of youth mHealth intervention resource allocation for targeted conditions, mHealth delivery methods, measurement tools, evaluation of mHealth intervention, and youth engagement. The median age for participants in all studies is 17 (IQR 14-21) years. Only 3 (2%) studies involved participants who reported their sex or gender outside of the binary option. Many studies (68/151, 45%) were published after the onset of the COVID-19 outbreak. Study types and designs varied, with 60 (40%) identified as randomized controlled trials. Notably, 143 out of 151 (95%) studies came from developed countries, suggesting an evidence shortfall on the feasibility of implementing mHealth services in lower-resourced settings. Additionally, the results highlight concerns related to inadequate resources devoted to self-harm and substance uses, weak study design, expert engagement, and the variety of outcome measures selected to capture impact or changes over time. There is also a lack of standardized regulations and guidelines for researching mHealth technologies for youths and the use of non–youth-centered approaches to implementing results.

Conclusions: This study may be used to inform future work as well as the development of youth-centered mHealth tools that can be implemented and sustained over time for diverse types of youths. Implementation science research that prioritizes youths' engagement is needed to advance the current understanding of mHealth implementation. Moreover, core outcome sets may support a youth-centered measurement strategy to capture outcomes in a systematic way that prioritizes equity, diversity, inclusion, and robust measurement science. Finally, this study suggests that future practice and policy research are needed to ensure the risk of mHealth is minimized and that this innovative health care service is meeting the emerging needs of youths over time.

(JMIR Mhealth Uhealth 2023;11:e44951) doi:10.2196/44951

KEYWORDS

adolescent; COVID-19; engagement; health outcome; illness; implementation; mental disorder; mental health; mHealth intervention; mHealth tools; mHealth; policy; scoping review; young adult; youth

Introduction

Mental Illness

Mental health disorders are the leading cause of disability worldwide and are considered a global public health challenge [1,2]. Globally, 1 in 4 people are affected by mental health disorders each year; up to 50% of people experience mental health challenges in their lifetime [3]. The global burden of disease (GBD) is compounded by the "mental health treatment gap," referring to those people in need of mental health treatment but who have not received it. People with mental health challenges experience different levels of barriers to accessing specialized care when needed [4]. Such barriers are multifaceted and may include poor mental health literacy, social stigma, trust and confidentiality issues with health professionals, and systemic difficulties such as financial hardship [5].

A total of 75% of people with mental health disorders have an onset of illness between the ages of 12 and 24 years [6]. This is a peak period of development for youths (defined here as ages 12-24 years); it is often the life stage to pursue education or begin a career, to build social relationships, and to explore new interests [7]. Yet, youths experience the worst levels of access to mental health care from poorly designed, grossly underresourced, and typically unfriendly health care services [8]. Current research indicates the need for a full range of interventions for youths [9], including health promotion [10,11], early intervention [12], and long-term supports such as integrated self-management [13], community outreach [14], and hospital care [15]. With existing barriers compounded by the COVID-19 pandemic, in-person mental health care is more challenging than ever to navigate and access for youths [16]. In response, mobile health (mHealth), a term used to describe the collective set of digital mental health interventions, has been proposed as a solution to meet the needs of youths across the world.

mHealth

In 2022, mobile phone users reached 6.5 billion worldwide, accounting for approximately 80% of the global population. In developed countries such as Canada and the United States, 97% of people own a mobile phone and 85% use a smart phone [17,18]. Such mass usage is motivating the rapid growth of medical software and apps for diverse populations, including youths [19]. These mobile device–based apps cover multiple domains of technologies used to assess, capture, or support areas of health, including physical activity and fitness, diet, emotional and mental health, and health services. mHealth technologies have been used to track vital body signs, such as blood pressure, heart rate, exercise, sleep activity, nutritional values in meals, mental health and wellness, anxiety, and mood [20]. A total of 54,603 health care and medical apps were available on the Google Play Store in August 2022, up by almost 4% compared to the previous quarter [21].

While the pace of new mHealth technologies is growing rapidly, the regulations and standards for their use have not yet followed. In a recent study of nearly 300 apps for mental health, less than one-third received input from a mental health expert [22]. There is also little consensus regarding standards for mHealth tools, not to mention the effectiveness of mHealth interventions on diverse populations, including youths experiencing mental health challenges.

In summary, the field of mHealth often presents a dichotomy of opportunity and risk. On the one hand, the lack of standards and regulation for mHealth apps presents global concerns for the safety of youths, notably personal security and the uptake of misinformation [23-25]. On the other hand, mHealth may have an increasingly important role in health promotion, education, and interventions to bridge the gap for those who cannot access in-person services [26]. A greater understanding is needed to learn about how youths use mHealth technologies in their daily lives, with specific emphasis on understanding how they navigate safety risks and use these technologies to improve health and wellness outcomes.

The purpose of this scoping review is to synthesize the current evidence supporting mHealth interventions for youths accessing support for mental health challenges. This will facilitate understanding of what is missing in the field of mHealth and support recommendations for research, practice, and policy. The specific objectives are to (1) synthesize the current evidence supporting mHealth interventions for youths who experience mental health challenges and (2) identify current gaps in the mHealth field, with an overarching goal of improving youths' mental health service access, outcomes, and experiences.

Methods

Overview

We conducted a scoping review to examine the extent, range, and nature of mHealth and to identify gaps in the existing literature on this emerging topic. This scoping study followed the 5 stages of Arksey and O'Malley's scoping study framework [27] to (1) identify the research question; (2) identify relevant studies; (3) select studies; (4) chart the data; and (5) collate, summarize, and report the results.

Step 1: Identify the Research Question

What is known in the existing literature about the feasibility and effectiveness of mHealth intervention for youths, aged 12-24 years, facing mental health challenges?

Step 2: Identify Relevant Studies

Under the guidance of a medical librarian, a comprehensive search of the following electronic databases was conducted: MEDLINE (Ovid), Embase (Ovid), PsycINFO (EBSCO), and PubMed (see Multimedia Appendix 1 for example search). We consulted with mHealth stakeholders and youths to decide on the range of dates to search. Our expert team highlighted rapid changes in the field, notably the influence of TikTok after its launch in 2016. To ensure relevance and reference value, we decided to only review articles published during the past 6 years (January 1, 2016, to February 7, 2022) to manage the scope, breadth, and rapidly changing information available. Key terms derived from the research question were selected and expanded to create a comprehensive list of search terms, including "telemedicine," "telerehabilitation," "mobile applications," "mHealth (mobile health)," "eHealth (electronic/digital health)," and "telehealth," as well as a combination of the following mental health condition-related terms: "mental disorders," "anxiety," "depression," "eating disorder," "schizophrenia," "bipolar," "obsessive compulsive disorder," and "posttraumatic stress disorder," along with a list of key terms to define the age group of this review: "adolescent," "teen," "youth," and "young adult." Combinations of these terms, along with Medical Subject Heading (MeSH) terms, were tested iteratively in each of the databases selected to inform the new combination of different terms leading to relevant literature. All searches included at least one identifier for mHealth (eg, telehealth and eHealth), 1 identifier for mental health condition (eg, depression and anxiety), and 1 identifier for age range (eg, youth and young adult). The lead author reviewed the title and abstract of each study to determine eligibility based on predetermined inclusion and exclusion criteria (described below) after duplicates were removed. After the completion of the initial review, the articles were thoroughly reviewed by the lead author based on the research topic.

Step 3: Select Studies

The following inclusion criteria were considered: (1) published in English; (2) published between January 1, 2016, and February 7, 2022; (3) included human subjects, whose ages fall between 12 and 24 years; (4) included at least one mHealth intervention tool targeting 1 or more mental health conditions for youths; and (5) referenced literature from peer-reviewed journals and book chapters. Exclusion criteria were as follows: (1) editorial comments, commentaries, book reviews, and opinion articles; (2) incomplete studies (eg, description of intervention, protocols, ideas from symposia, and conference summaries); (3) articles without full text available. Relevant systematic reviews were included in the study to serve as background literature but were excluded from the data extraction and analysis process to focus on intervention studies.

Step 4: Chart the Data

Through careful review of the literature, the researchers (XD and SB) identified the key components and issues discussed in all relevant studies. This information was recorded in a data extraction sheet, along with information on each study (author, publication year, population demographics, location, study design, level of evidence, characteristics of the intervention, targeted health condition, and outcomes).

Step 5: Collate, Summarize, and Report Results

The information was synthesized and used to map out the scope and breadth of included literature on the topic of mHealth intervention for youths' mental health challenges.

Results

Overview

As noted in Figure 1, the search identified 4270 citations for initial screening. As noted in Figure 1, a total of 1411 duplicates were removed, resulting in 2859 citations for title and abstract reviews. A further 2413 articles were excluded because their title or abstract did not address mHealth interventions for youths' mental health, leaving 446 citations. After full text review, a total of 151 articles met the inclusion criteria for this study. Table 1 summarizes the countries of origin of the included studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram describing the search process and the number of articles meeting inclusion criteria for this study (N=151).

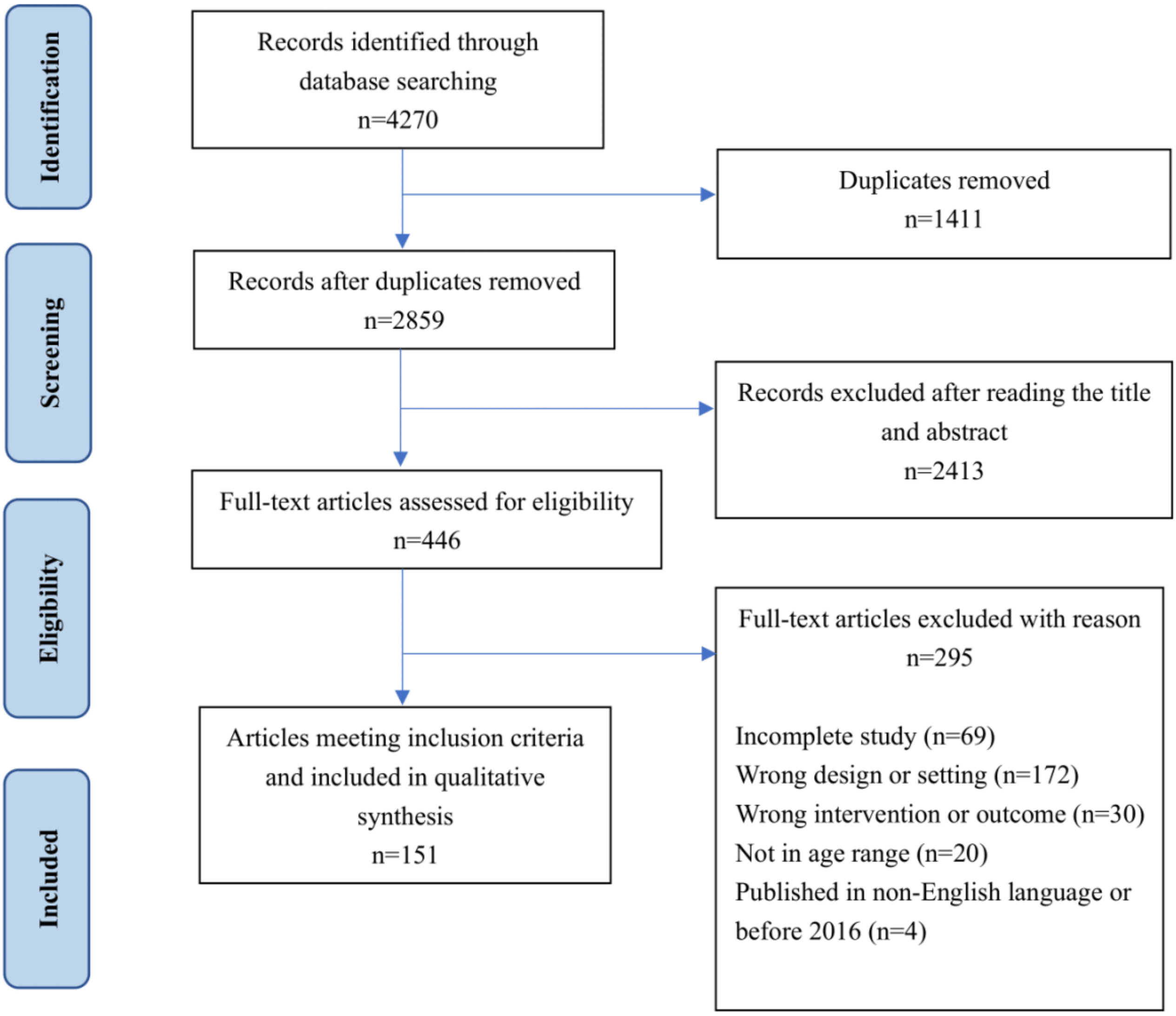


Table 1. Countries of origin of papers included in the scoping review.

Country	Articles (N=151), n %
United States	61 (40)
Australia	28 (19)
Canada	10 (7)
Sweden	10 (7)
United Kingdom	8 (6)
New Zealand	7 (5)
China	4 (3)
Nigeria	3 (2)
Finland	3 (2)
The Netherlands	3 (2)
Japan	2 (1)
Germany	2 (1)
Korea	1 (1)
India	1 (1)
Spain	1 (1)
Iceland	1 (1)
Denmark	1 (1)
Belgium	1 (1)
Italy	1 (1)
Norway	1 (1)
Israel	1 (1)
France	1 (1)

Level of Evidence

As shown in Table 2, four levels of evidence were adopted to show the power of evidence among included youths' mHealth interventions for mental health research [28,29].

Table 2. Design of studies included in the scoping review.

Level of evidence	Study design	Articles, n
I	Randomized controlled trials	60
II	Nonrandomized experimental studies (quasi-experimental, pre-post, and cohort)	34
III	Cross-sectional, longitudinal surveys, and mixed method study with nonexperimental quantitative part	27
IV	Descriptive, case studies, and qualitative studies	30

As shown in Table 2, study types and designs varied. The highest level of evidence included experimental studies identified as randomized controlled trials (60/151, 40%). The second highest level of evidence included studies that were nonrandomized trials, such as quasi-experimental design (usually with 2 groups) or 1-group pre-post design (34/151, 23%). The third level of evidence included studies that only conducted feedback surveys to their participants once after the intervention, longitudinal observational studies, and mixed methods design with these mentioned quantitative parts (27/151, 18%). Last, some of the included studies had lower levels of evidence, including descriptive studies, case studies, and qualitative studies describing the outcome of mHealth interventions (30/151, 20%).

Age

All included studies have a general targeted age group of "young people," "adolescent," or "youth." The median age for participants in all studies was 17 (IQR 14-21) years, with 2 outliers (with participants aged 61 years [30] and 62 years [31]) coming from the studies targeting "university students."

Sex and Gender

A total of 20 (13%) papers did not provide any statistics on sex or gender. Only 3 (2%) studies involved participants who reported their sex or gender beyond binary options. One study focused on eating disorders and had 1 intersex participant [32]. The other 2 studies prioritized mental health in 2-spirited, lesbian, gay, bisexual, transgender, queer/questioning, intersex, and agender (2SLGBTQIA+) groups: 1 included 108 sexual minority young adults [33] and the other included 565 participants with a combination of transgender, genderqueer, gender expansive, intersex, agender, 2-spirited, and third gender identities [34]. The remaining 148 (98%) studies had a total of 6478 (34.2%) men and 12,480 (65.8%) women as participants.

Organizational Affiliation

The included studies took place in various settings. They were coded into the following 4 different categories, inspired by Marshall et al [22]: (1) research developed under clinic or hospital-related setting (97/151, 64%); (2) research developed in a university (28/151, 19%); (3) research developed in other nonuniversity schools and institutions (9/151, 6%); and (4) insufficient information to tell (17/151, 11%).

Modes of Delivery of mHealth Interventions

As noted in Table 3, a variety of modes are used to deliver services. These included web pages (54/151, 36%), smartphone apps (51/151, 34%), phone calls and SMS text messages (10/151, 9%), and innovative virtual reality (VR) tools (3/151, 2%).

Table 3. Delivery modes among studies included in the review.

Delivery modes	Studies, n (%)	Randomized controlled trials, n/N (%)	Positive[a] results, n/N (%)
Web pages	54 (36)	33/54 (61)	48/54 (89)
Smartphone apps	51 (34)	13/51 (26)	40/51 (78)
Video conferencing	20 (13)	3/20 (15)	15/20 (75)
SMS text messages	10 (7)	5/10 (50)	7/10 (70)
Chatbox	4 (3)	2/4 (50)	4/4 (100)
Phone calls	3 (2)	1/3 (33)	2/3 (67)
Virtual reality	3 (2)	0/3 (0)	3/3 (100)
Mixed modes	6 (4)	3/6 (50)	4/6 (67)
Total	151 (100)	60/151 (N/A[b])	123/151 (N/A)

[a]Positive results: researchers stated overall positive effect of intervention or received positive feedback from users to show promising practical usage of the intervention. Negative results: researchers see no changes in mental health symptoms before and after the intervention or receive negative feedback from users.

[b]N/A: Not applicable.

Targeted Conditions

The number and percentage of health conditions targeted using mHealth interventions are summarized in Table 4.

Table 4. Health conditions and measurement tools among included studies.

Health condition	Studies, n (%)	Measurement tools (scales) used, n
General health	26 (15)	66
Anxiety	35 (20)	52
Depression	32 (19)	68
Suicide, self-harm, or violence	11 (6)	15
Substance use	10 (6)	22
Eating disorder	10 (6)	10
Stress or mood related	9 (5)	20
Sleep disorder	6 (4)	9
ASD[a]	5 (3)	4
STD or STI[b]	5 (3)	10
Psychosis	4 (2)	5
ADHD[c]	4 (2)	8
PTSD[d]	4 (2)	3
OCD[e]	3 (2)	1
FASD[f]	2 (1)	N/A[g]
Others	6 (4)	N/A
Total	172[h] (N/A)	N/A

[a]ASD: autism spectrum disorder.
[b]STD or STI: sexually transmitted disease or sexually transmitted infection.
[c]ADHD: attention-deficit/hyperactivity disorder.
[d]PTSD: posttraumatic stress disorder.
[e]OCD: obsessive-compulsive disorder.
[f]FASD: fetal alcohol spectrum disorder.
[g]N/A: not applicable.
[h]The total number of conditions exceeds the total number of studies because there are articles targeting more than 1 condition.

Measurement Approaches Used in the Studies

A wide range of tools were employed to measure outcomes in mHealth intervention research (Table 4). A detailed table listing the outcome measures and the references to the studies using them can be found in Multimedia Appendix 2.

COVID-19

Of the 151 studies included within the given time period, 68 (45%) were published after the COVID-19 outbreak, with most (44/151, 65%) published in 2020. The numbers of studies published each year are as follows: 11 studies in 2016; 21 studies in 2017; 19 studies in 2018; 32 studies in 2019; 44 studies in 2020; 23 studies in 2021; and 1 study in February 2022.

The following 8 studies directly addressed health conditions or service delivery modalities influenced by the pandemic: (1) treatment of eating disorders in adolescents [35]; (2) group-based psychiatric care using telehealth [36]; (3) a web-based art therapy group for learning disabled young adults using WhatsApp [37]; (4) telehealth versus in‐person intensive outpatient program (IOP) for eating disorders during versus before COVID‐19 [32]; (5) a well-being app to support young people living in New Zealand [38]; (6) peer-to-peer live-streaming intervention to promote physical activity and reduce anxiety during homeschooling [39]; (7) a mindfulness-based mHealth intervention among psychologically distressed university students in quarantine [40]; and (8) smartphone application for adolescents with anorexia nervosa [41].

Discussion

Overview

This scoping review provides a comprehensive synthesis of mobile mental health intervention research for youths. The study identified that the number of interventions is proliferating over time, with limited emphasis on study quality, youths' engagement, youth-centered outcome assessment, implementation standards, or consideration for equity, diversity, and inclusion in research. Youths' engagement was rarely mentioned as a part of any research study method, and most studies focused primarily on youths who identified within the gender binary. Few studies discussed the implementation and scale of interventions in diverse settings (Table 3) and even fewer studies showed an impact of mHealth interventions over time for diverse types of youths. With significant growth and investment in the area, this review may inform direction for future research and advance practices within mHealth intervention approaches for youths who experience mental health challenges or illnesses.

Among the studies that met the inclusion criteria, depression (32/151, 19%), anxiety (35/151, 20%), and general mental health concerns (26/151, 15%) accounted for over half of the studies, and the remaining addressed 17 other mental health concern categories. According to a GBD systematic analysis of mental health conditions relevant to this study, the top concerns for all youths are self-harm, depressive disorders, interpersonal violence, anxiety disorders, HIV/AIDS, conduct disorder, and drug use disorder [42]. The point estimate of alcohol and drug use disorders combined is very close to the prevalence of depression, serving as 2 of the most concerning mental health disorders worldwide [43]. The high prevalence of depressive and anxiety disorders corresponds with the allocation of research priorities in studies included in this review, suggesting the current focus for youths' mental health care services. Nevertheless, while self-harm and interpersonal violence ranked high on the GBD mental health conditions list, there were comparatively limited research studies for these categories. Substance use disorders are also underrepresented in existing mHealth studies for youths. More attention and resources should be given to the development of mHealth intervention tools targeting self-harm, interpersonal violence, and substances use among diverse youths.

From a global health perspective, 95% (143/151) of studies came from developed countries. This result is not surprising, as mHealth resources are almost exclusively concentrated in high-income countries, although the prevalence of depression and anxiety in high-income countries is not significantly higher than in the rest of the world [44]. Therefore, more future mHealth research needs to be conducted in low- and middle-income countries, especially when one of the advantages of remote health care access is cost-effectiveness. Previous studies have demonstrated adapting interventions developed in high-income countries for use in low- and middle-income countries, such as India, Sierra Leone, Romania, Malaysia, and South Africa [45,46]; these lessons could be adapted for mHealth interventions for youths' mental health.

Delivery Modes

With respect to delivering mHealth interventions, most studies employed web pages (54/151, 36%) and smartphone apps (51/151, 34%), and these modes also had the largest number of randomized controlled trials conducted. The results show web-based mHealth tools have the strongest evidence for improving mental health conditions in youths, but the effectiveness of other intervention modes cannot be ruled out. Web-based intervention tools require youths to have internet access, and the use of computers may not be convenient in home and school settings, so the development of mHealth interventions has gradually evolved to include smartphones, smart devices, wearables, and newer technologies, including VR and augmented reality [47]. However, with the rising number of options available for mHealth intervention delivery, youths' intention to use has become a more complicated question. Currently, many studies contributing to the adoption research of mHealth for youths are based on the technology acceptance model, an information systems theory that models user's acceptance of technology mainly based on perceived usefulness and perceived ease of use [48-50]. Yet, it is essential to realize that participants in research studies are provided with a predetermined type of intervention by the researchers, often with limited youths' engagement. That is, unlike in a real-world environment where diverse users who need support must seek it themselves, participants in the research settings did not actively choose which kind of mHealth intervention to use. Therefore, future studies need to consider the youth-intended delivery modes during mHealth intervention development from a user perspective, especially when considering innovative technologies, such as VR, artificial intelligence chat, and gaming. Future research should also consider implementation research, not only to understand the efficacy and effectiveness of the interventions but also their capacity to be sustained over time in diverse settings.

Measurement

The measurement tools used in the studies (Multimedia Appendix 2) can be broadly defined as either (1) measuring the outcome of a patient's condition or (2) measuring the subject usability of a product or service. As displayed in Table 4, depression, anxiety, and general mental health categories each adopted more than 50 measurement tools. The review also identified numerous measurement scales used for each health condition. For example, 7 sleep-relevant measurements were used in 6 sleep-related studies [51-56], but the researchers did not provide a rationale as to why they chose the scale used among all available sleep-related measurements, nor was evidence provided about the fitness of the tools for youths in the varying contexts. Similar complications were presented in other included studies with varying health conditions. Future research should focus on developing a guideline for researchers to follow when selecting the most appropriate measurement scales in both research and clinical settings and on validating measurement scales designed for use with youths with consideration for equity, diversity, inclusion, and psychometric rigor.

Evaluation of mHealth Products

The percentage of positive results is considerably high among all included studies (123/151, 81%). Study aims included "feasibility," "effect," "acceptability," "efficacy," "fidelity," "effectiveness," and "cost-effectiveness," with research designs including experiments, surveys, and interviews. Questions remain as to whether positive results translate to real-world applications or which types of outcomes possess high level of evidence in the knowledge translation process for health care services. In a Canadian study where participants were asked to rate mHealth apps objectively, results were highly variable, and 28% of reviewers were not even sure about the overall quality of the health product [57]. The same characteristics in different studies also vary within a wide range. For instance, the intervention durations ranged from as short as 7 days [58] or 2 sessions [59] to as long as 24 weeks [60] or 24 months [61]. The number of participants among all studies ranges from 2 [62] to 2532 [63]. Intervention types also varied significantly. Such variations raise concerns about the lack of a standardized evaluation strategy. To address the complex uncertainty in evaluating mHealth tools, a multifaceted evaluation framework needs to be adopted to assess the different perspectives, elements, and features of an intervention that leads to a final mHealth product. Using non–youth-centered frameworks to evaluate mHealth products can consequently produce ambiguous or incorrect information on their effectiveness, leading to misuse, misdiagnosis, wasted time, and, worst of all, negative health impacts and experiences [64].

Remote Care Transition

Several included studies discuss how to support youths during the COVID-19 pandemic [65,66]. With many people transitioning to work-from-home or hybrid arrangements during and after the pandemic, there arose an imminent need for a mental health technology revolution, and web-based health service delivery has emerged as a preferred tool [67].

Childs and colleagues [36] illustrated the feasibility of a rapid transition to telemedicine services during the pandemic. Yale New Haven Psychiatric Hospital decided to discontinue in-person IOP services within 3 business days of the World Health Organization's pandemic declaration [36]. The first mobile service was available within a week, and subsequent treatment plans and adolescent ambulatory services were developed to reach IOP level. The study demonstrated that it took a comprehensive program 2 months to transition from 100% in-person service prepandemic (March) to 100% telehealth service after the start of the pandemic (May 2020), showing the feasibility of the deployment of mHealth tools in clinical settings and the smooth transition from physical to virtual health care access. One notable limitation of this study is that the clinicians focused on the transition process rather than the effectiveness of the intervention tool.

Another youths' mHealth intervention study showed the transition to virtual services was not always desired by clients. The authors presented a case where a participant refused to cooperate, and the telemedicine service increased the tension between the participant and family members [35]. When developing mHealth interventions targeting youths' mental health, it is important to consider how to achieve optimal patient engagement when physical contact with a service provider is not an option. Last, it is still unknown whether these transitioned services will continue to be provided virtually on an ongoing basis. Future research is needed to investigate the long-term influence of such transitions and determine the feasibility of normalizing mHealth services.

Youths' and Stakeholder Engagement

Engagement with mHealth interventions is thought to be important for intervention effectiveness by increasing acceptability, satisfaction, intervention adherence, and levels of attention and enjoyment [68]. This can be extended to engaging youths in mHealth research to provide comprehensive, ongoing, tailored, and interactive support to improve health [69].

Current youth mHealth research often engages youths by asking for feedback in a survey or interview to test usability, feasibility, and acceptability [31,70,71], or by including young users as participants in experiments to evaluate the effectiveness and efficiency of an intervention. However, few studies mentioned how they engaged youths in the development phase and followed design thinking with the priority population. Youths and other stakeholders (eg, family and caregivers, service providers, and graphic designers) can contribute more than just feedback on the provided services; they can be offered opportunities to participate in the product and service design stages to make sure the end product is tailored to their needs and preferences. A previous conceptual model indicated that, during the optimization phase of an intervention [72], participants need to understand how the provided materials can inspire them and facilitate their thoughts to improve self-efficacy, which increases capability for self-monitoring and self-regulation and can lead to improved health outcomes and behavioral changes [72,73]. Researchers proposed a supportive accountability model that emphasized the importance of human support in mHealth interventions to increase adherence to trustworthy, benevolent, and professional information [74]. To summarize, it is crucial to apply such theoretical models to interventions targeting youths' mental health as well. mHealth researchers and developers ought to involve youths in every stage of design, development, and implementation. Our team is currently studying youths' engagement in the mHealth development phase to understand youths' information preferences and make sure mHealth interventions are designed to convey the benefits of human support, similar to in-person services.

Policy

More than two-thirds of the studies took place in a clinical setting, yet none of the studies reviewed provided systematic frameworks or models to help translate, scale, and sustain available mHealth tools to clinical practice. If health care stakeholders and policy makers aim to scale up and normalize mHealth services in the near future, it is essential to understand the feasibility and impact of implementing new mHealth tools in current models of care (eg, health care and schools). Guidelines and standards may be critical to ensuring that mHealth interventions are trustworthy and can be value added to health services that are delivered to youths and their families.

Strengths and Limitations

This review has a broad scope of attempting to draw a picture of existing mHealth intervention tools specifically designed for younger populations and how their effectiveness is being assessed. This scoping review addressed a broad term list and a large number of parameters. Inclusion and exclusion criteria were strictly set from the beginning and determined by experts in the field and a medical librarian, and diversity is presented for all included studies. There are unlimited possibilities for future work, particularly with the uncertainty of the COVID-19 pandemic and the ongoing response of the health care system to remote health access. Regarding weaknesses, the lack of critical appraisal is a widely recognized limitation for scoping reviews [75]. The scope of this review may be broad, but the depth and the quality of all included papers were not systematically critiqued. We also acknowledge that our search strategy may have missed key terms (eg, internet-based interventions) and intervention descriptions (eg, asynchronous vs synchronous) that may have limited our ability to completely summarize all relevant articles. In addition, the COVID-19 pandemic has not come to an end, and there is still ongoing research about long-term COVID-19 symptoms. Thus, the results relevant to remote care transition and COVID-19 should be interpreted with some caution.

Conclusions

As the need for mental health services continues to accelerate [76], mHealth technologies provide a solution to support diverse youths who may not be able to access in-person services. The impact of mHealth interventions on youths' mental health has been increasingly recognized by researchers, service providers, and policy makers. Results of our scoping review demonstrate a range of studies that capture the exponential growth of mHealth interventions for youths, with significant potential to be value-added for youths who are seeking support for mental health challenges. However, the review also highlighted notable gaps in research that include youths' voice throughout the research process, notably diverse youths in both developing and developed countries. Future research is needed that adopts an equity, diversity, and inclusion lens, prioritizes understanding how current mHealth technologies can be adopted into existing models of care, and develops guidelines, standards, and evaluation frameworks to support future mHealth development and implementation. As the field continues to expand rapidly, more global resources are needed to monitor technological advancements to provide quality mHealth services to every youth where and when they need them.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample Literature Search Sheet.

[DOCX File , 13 KB - mhealth_v11i1e44951_app1.docx]

Multimedia Appendix 2

List of measurement scales and studies using them.

[DOCX File , 41 KB - mhealth_v11i1e44951_app2.docx]

References

1. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013;382(9904):1575-1586. [doi: 10.1016/s0140-6736(13)61611-6] [Medline: 23993280]
2. Wainberg ML, Scorza P, Shultz JM, Helpman L, Mootz JJ, Johnson KA, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. Curr Psychiatry Rep 2017;19(5):28 [FREE Full text] [doi: 10.1007/s11920-017-0780-z] [Medline: 28425023]
3. WHO. The World Health Report 2001. Geneva, Switzerland: World Health Organization; 2001.
4. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
5. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
6. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602 [FREE Full text] [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
7. Magnusson D. The lifespan development of individuals: behavioral, neurobiological, and psychosocial perspectives: a synthesis. Cambridge England; New York: Cambridge University Press; 1996.
8. McGorry PD, Mei C. Early intervention in youth mental health: progress and future directions. Evid Based Ment Health 2018;21(4):182-184 [FREE Full text] [doi: 10.1136/ebmental-2018-300060] [Medline: 30352884]
9. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, DCP MNS Author Group. Addressing the burden of mental, neurological, and substance use disorders: key messages from disease control priorities, 3rd edition. Lancet 2016;387(10028):1672-1685. [doi: 10.1016/S0140-6736(15)00390-6] [Medline: 26454360]
10. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
11. O'Connor CA, Dyson J, Cowdell F, Watson R. Do universal school-based mental health promotion programmes improve the mental health and emotional wellbeing of young people? A literature review. J Clin Nurs 2018;27(3-4):e412-e426. [doi: 10.1111/jocn.14078] [Medline: 28926147]

12. Ospina-Pinillos L, Davenport T, Iorfino F, Tickell A, Cross S, Scott EM, et al. Using new and innovative technologies to assess clinical stage in early intervention youth mental health services: evaluation study. J Med Internet Res 2018;20(9):e259 [FREE Full text] [doi: 10.2196/jmir.9966] [Medline: 30201602]
13. Lean M, Fornells-Ambrojo M, Milton A, Lloyd-Evans B, Harrison-Stewart B, Yesufu-Udechuku A, et al. Self-management interventions for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2019;214(5):260-268 [FREE Full text] [doi: 10.1192/bjp.2019.54] [Medline: 30898177]
14. Saw A, Steltenpohl CN, Bankston-Lee K, Tong EK. A community-based "street team" tobacco cessation intervention by and for youth and young adults. J Community Health 2018;43(2):383-390. [doi: 10.1007/s10900-017-0435-3] [Medline: 29071569]
15. Brown WN, D'Errico E, Morrell HER. The youth alternative solutions program: evaluating a hospital-based intervention for adolescent substance use. J Addict Nurs 2015;26(1):32-40. [doi: 10.1097/JAN.0000000000000062] [Medline: 25761161]
16. McGorry PD, Mei C, Chanen A, Hodges C, Alvarez-Jimenez M, Killackey E. Designing and scaling up integrated youth mental health care. World Psychiatry 2022;21(1):61-76 [FREE Full text] [doi: 10.1002/wps.20938] [Medline: 35015367]
17. Pew Research Center. Mobile Fact Sheet. Washington, D.C; 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-05-25]
18. Outlook SDM. Number of smartphone users in Canada from 2018 to 2024. Statista. 2019. URL: https://www.statista.com/statistics/467190/forecast-of-smartphone-users-in-canada/ [accessed 2023-05-25]
19. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014;39(5):356-364 [FREE Full text] [Medline: 24883008]
20. Oyebode O, Ndulue C, Alhasani M, Orji R. Persuasive mobile apps for health and wellness: a comparative systematic review. In: Gram-Hansen SB, Jonasen TS, Midden C, editors. Persuasive Technology Designing for Future Change. Cham: Springer International Publishing; 2020:163-181.
21. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 3nd quarter 2022. Statista. 2022. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2023-05-25]
22. Marshall JM, Dunstan DA, Bartik W. The digital psychiatrist: in search of evidence-based apps for anxiety and depression. Front Psychiatry 2019;10:831 [FREE Full text] [doi: 10.3389/fpsyt.2019.00831] [Medline: 31803083]
23. Nesi J. The impact of social media on youth mental health: challenges and opportunities. N C Med J 2020;81(2):116-121 [FREE Full text] [doi: 10.18043/ncm.81.2.116] [Medline: 32132255]
24. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. Int J Adolesc Youth 2019;25(1):79-93 [FREE Full text] [doi: 10.1080/02673843.2019.1590851]
25. Kelly Y, Zilanawala A, Booker C, Sacker A. Social media use and adolescent mental health: findings from the UK millennium cohort study. EClinicalMedicine 2018;6:59-68 [FREE Full text] [doi: 10.1016/j.eclinm.2018.12.005] [Medline: 31193561]
26. Zenone M, Ow N, Barbic S. TikTok and public health: a proposed research agenda. BMJ Glob Health 2021;6(11):e007648 [FREE Full text] [doi: 10.1136/bmjgh-2021-007648] [Medline: 34819326]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Sackett D. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 1989 Feb 01;95(2):2S-24. [doi: 10.1378/chest.95.2.2s] [Medline: 2914516]
29. No authors listed. The periodic health examination. Canadian Task Force on the Periodic Health Examination. Can Med Assoc J 1979 Nov 03;121(9):1193-1254 [FREE Full text] [Medline: 115569]
30. Dear BF, Johnson B, Singh A, Wilkes B, Brkic T, Gupta R, et al. Examining an internet-delivered intervention for anxiety and depression when delivered as a part of routine care for university students: a phase IV trial. J Affect Disord 2019;256:567-577 [FREE Full text] [doi: 10.1016/j.jad.2019.06.044] [Medline: 31280082]
31. Viskovich S, Pakenham KI. Pilot evaluation of a web-based acceptance and commitment therapy program to promote mental health skills in university students. J Clin Psychol 2018;74(12):2047-2069. [doi: 10.1002/jclp.22656] [Medline: 29962090]
32. Levinson CA, Spoor SP, Keshishian AC, Pruitt A. Pilot outcomes from a multidisciplinary telehealth versus in-person intensive outpatient program for eating disorders during versus before the Covid-19 pandemic. Int J Eat Disord 2021 Sep;54(9):1672-1679. [doi: 10.1002/eat.23579] [Medline: 34245028]
33. Pachankis JE, Williams SL, Behari K, Job S, McConocha EM, Chaudoir SR. Brief online interventions for LGBTQ young adult mental and behavioral health: a randomized controlled trial in a high-stigma, low-resource context. J Consult Clin Psychol 2020;88(5):429-444 [FREE Full text] [doi: 10.1037/ccp0000497] [Medline: 32271053]
34. Dobias ML, Schleider JL, Jans L, Fox KR. An online, single-session intervention for adolescent self-injurious thoughts and behaviors: results from a randomized trial. Behav Res Ther 2021;147:103983. [doi: 10.1016/j.brat.2021.103983] [Medline: 34688102]
35. Yaffa S, Adi EL, Itai P, Marit JM, Doron G, Daniel S. Treatment of eating disorders in adolescents during the COVID-19 pandemic: a case series. J Eat Disord 2021;9(1):17 [FREE Full text] [doi: 10.1186/s40337-021-00374-z] [Medline: 33568207]

36. Childs AW, Unger A, Li L. Rapid design and deployment of intensive outpatient, group-based psychiatric care using telehealth during coronavirus disease 2019 (COVID-19). J Am Med Inform Assoc 2020;27(9):1420-1424 [FREE Full text] [doi: 10.1093/jamia/ocaa138] [Medline: 32687151]
37. Datlen GW, Pandolfi C. Developing an online art therapy group for learning disabled young adults using WhatsApp. Int J Art Ther 2020;25(4):192-201. [doi: 10.1080/17454832.2020.1845758]
38. Serlachius A, Boggiss A, Lim D, Schache K, Wallace-Boyd K, Brenton-Peters J, et al. Pilot study of a well-being app to support New Zealand young people during the COVID-19 pandemic. Internet Interv 2021;26:100464 [FREE Full text] [doi: 10.1016/j.invent.2021.100464] [Medline: 34631433]
39. Zheng Y, Wang W, Zhong Y, Wu F, Zhu Z, Tham YC, et al. A peer-to-peer live-streaming intervention for children during COVID-19 homeschooling to promote physical activity and reduce anxiety and eye strain: cluster randomized controlled trial. J Med Internet Res 2021;23(4):e24316 [FREE Full text] [doi: 10.2196/24316] [Medline: 33882021]
40. Sun S, Lin D, Goldberg S, Shen Z, Chen P, Qiao S, et al. A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: a randomized controlled trial. J Couns Psychol 2022;69(2):157-171 [FREE Full text] [doi: 10.1037/cou0000568] [Medline: 34264696]
41. Naccache B, Mesquida L, Raynaud JP, Revet A. Smartphone application for adolescents with anorexia nervosa: an initial acceptability and user experience evaluation. BMC Psychiatry 2021;21(1):467 [FREE Full text] [doi: 10.1186/s12888-021-03478-7] [Medline: 34563166]
42. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet 2020;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
43. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. Curr Psychiatry Rep 2019;21(2):10. [doi: 10.1007/s11920-019-0997-0] [Medline: 30729322]
44. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/s0140-6736(21)02143-7]
45. Fendt-Newlin M, Jagannathan A, Webber M. Cultural adaptation framework of social interventions in mental health: evidence-based case studies from low- and middle-income countries. Int J Soc Psychiatry 2020;66(1):41-48 [FREE Full text] [doi: 10.1177/0020764019879943] [Medline: 31580173]
46. Sivaraman B, Nye E, Bowes L. School-based anti-bullying interventions for adolescents in low- and middle-income countries: a systematic review. Aggress Violent Behav 2019;45:154-162. [doi: 10.1016/j.avb.2018.07.007]
47. Giansanti D. Towards the evolution of the mHealth in mental health with youth: the cyber-space used in psychological rehabilitation is becoming wearable into a pocket. Mhealth 2020;6:18 [FREE Full text] [doi: 10.21037/mhealth.2019.11.02] [Medline: 32270010]
48. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319. [doi: 10.2307/249008]
49. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
50. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
51. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
52. Chen RY, Feltes JR, Tzeng WS, Lu ZY, Pan M, Zhao N, et al. Phone-based interventions in adolescent psychiatry: a perspective and proof of concept pilot study with a focus on depression and autism. JMIR Res Protoc 2017;6(6):e114 [FREE Full text] [doi: 10.2196/resprot.7245] [Medline: 28623183]
53. Dewald JF, Short MA, Gradisar M, Oort FJ, Meijer AM. The chronic sleep reduction questionnaire (CSRQ): a cross-cultural comparison and validation in Dutch and Australian adolescents. J Sleep Res 2012;21(5):584-594 [FREE Full text] [doi: 10.1111/j.1365-2869.2012.00999.x] [Medline: 22329363]
54. Kerkhof GA, Geuke MEH, Brouwer A, Rijsman RM, Schimsheimer RJ, Van Kasteel V. Holland sleep disorders questionnaire: a new sleep disorders questionnaire based on the international classification of sleep disorders-2. J Sleep Res 2013;22(1):104-107 [FREE Full text] [doi: 10.1111/j.1365-2869.2012.01041.x] [Medline: 22924964]
55. Lebrun C, Gély-Nargeot MC, Maudarbocus KH, Bayard S. Assessing sleep-related safety behaviors: adaptation and validation of a french version of the sleep-related behaviors questionnaire in a nonclinical sample. Behav Sleep Med 2020;18(1):107-119. [doi: 10.1080/15402002.2018.1546178] [Medline: 30462561]
56. Morin CM, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]

57. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-Mental health apps in Canada: results of a modified delphi process. JMIR Mhealth Uhealth 2018;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
58. Smelror RE, Bless JJ, Hugdahl K, Agartz I. Feasibility and acceptability of using a mobile phone app for characterizing auditory verbal hallucinations in adolescents with early-onset psychosis: exploratory study. JMIR Form Res 2019;3(2):e13882 [FREE Full text] [doi: 10.2196/13882] [Medline: 31094321]
59. King SC, Richner KA, Tuliao AP, Kennedy JL, McChargue DE. A comparison between telehealth and face-to-face delivery of a brief alcohol intervention for college students. Subst Abus 2020;41(4):501-509 [FREE Full text] [doi: 10.1080/08897077.2019.1675116] [Medline: 31644389]
60. Ritvo P, Knyahnytska Y, Pirbaglou M, Wang W, Tomlinson G, Zhao H, et al. Online mindfulness-based cognitive behavioral therapy intervention for youth with major depressive disorders: randomized controlled trial. J Med Internet Res 2021;23(3):e24380 [FREE Full text] [doi: 10.2196/24380] [Medline: 33688840]
61. Van Voorhees B, Gladstone TRG, Sobowale K, Brown CH, Aaby DA, Terrizzi DA, et al. 24-month outcomes of primary care web-based depression prevention intervention in adolescents: randomized clinical trial. J Med Internet Res 2020;22(10):e16802 [FREE Full text] [doi: 10.2196/16802] [Medline: 33112254]
62. Stewart RW, Orengo-Aguayo R, Young J, Wallace MM, Cohen JA, Mannarino AP, et al. Feasibility and effectiveness of a telehealth service delivery model for treating childhood posttraumatic stress: A community-based, open pilot trial of trauma-focused cognitive–behavioral therapy. Journal of Psychotherapy Integration 2020 Jun;30(2):274-289 [FREE Full text] [doi: 10.1037/int0000225]
63. Fendrich M, Ives M, Kurz B, Becker J, Vanderploeg J, Bory C, et al. Impact of mobile crisis services on emergency department use among youths with behavioral health service needs. Psychiatr Serv 2019 Oct 01;70(10):881-887 [FREE Full text] [doi: 10.1176/appi.ps.201800450] [Medline: 31215355]
64. Quintana Y, Torous J. A Framework for Evaluation of Mobile Apps for Youth Mental Health. Guelph, Ontario, Canada: Homewood Research Institute; 2020. URL: https://bp-net.ca/wp-content/uploads/2020/11/A-Framework-for-Evaluation-of-Mobile-Apps-for-Youth-Mental-Health_May-20....pdf [accessed 2023-05-25]
65. Phillipou A, Meyer D, Neill E, Tan EJ, Toh WL, Van Rheenen TE, et al. Eating and exercise behaviors in eating disorders and the general population during the COVID-19 pandemic in Australia: Initial results from the COLLATE project. Int J Eat Disord 2020 Jul;53(7):1158-1165 [FREE Full text] [doi: 10.1002/eat.23317] [Medline: 32476163]
66. Schlegl S, Maier J, Meule A, Voderholzer U. Eating disorders in times of the COVID-19 pandemic-Results from an online survey of patients with anorexia nervosa. Int J Eat Disord 2020 Nov;53(11):1791-1800 [FREE Full text] [doi: 10.1002/eat.23374] [Medline: 32841413]
67. Figueroa CA, Aguilera A. The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. Front Psychiatry 2020;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
68. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
69. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
70. Saberi P, McCuistian C, Agnew E, Wootton AR, Legnitto Packard DA, Dawson-Rose C, et al. Video-counseling intervention to address HIV care engagement, mental health, and substance use challenges: a pilot randomized clinical trial for youth and young adults living with HIV. Telemed Rep 2021;2(1):14-25 [FREE Full text] [doi: 10.1089/tmr.2020.0014] [Medline: 33575683]
71. Kajitani K, Higashijima I, Kaneko K, Matsushita T, Fukumori H, Kim D. Short-term effect of a smartphone application on the mental health of university students: A pilot study using a user-centered design self-monitoring application for mental health. PLoS One 2020;15(9):e0239592 [FREE Full text] [doi: 10.1371/journal.pone.0239592] [Medline: 32976515]
72. Pfammatter AF, Marchese SH, Pellegrini C, Daly E, Davidson M, Spring B. Using the preparation phase of the multiphase optimization strategy to develop a messaging component for weight loss: formative and pilot research. JMIR Form Res 2020 May 13;4(5):e16297 [FREE Full text] [doi: 10.2196/16297] [Medline: 32347804]
73. Bandura A. Health promotion from the perspective of social cognitive theory. Psychology & Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
74. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
75. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]

76. WHO. World Health Organization. WHO highlights urgent need to transform mental health and mental health care. Geneva, Switzerland: World Health Organization; 2022. URL: https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care [accessed 2022-06-13]

Abbreviations

2SLGBTQIA+: 2-spirited, lesbian, gay, bisexual, transgender, queer/questioning, intersex, and agender
GBD: global burden of disease
IOP: intensive outpatient program
MeSH: Medical Subject Heading
mHealth: mobile health
VR: virtual reality

Edited by L Buis; submitted 09.12.22; peer-reviewed by L Harber, T Gladstone; comments to author 08.02.23; revised version received 30.03.23; accepted 16.05.23; published 16.06.23.

Please cite as:
Ding X, Wuerth K, Sakakibara B, Schmidt J, Parde N, Holsti L, Barbic S
Understanding Mobile Health and Youth Mental Health: Scoping Review
JMIR Mhealth Uhealth 2023;11:e44951
URL: https://mhealth.jmir.org/2023/1/e44951
doi:10.2196/44951
PMID:37220197

Review

Influencing Factors to mHealth Uptake With Indigenous Populations: Qualitative Systematic Review

Andrew Goodman[1,2*], AdvDip; Ray Mahoney[3*], PhD; Geoffrey Spurling[4*], MBBS, PhD; Sheleigh Lawler[1*], PhD

[1]School of Public Health, The University of Queensland, Turrbal, Jagera Country, Herston, Australia

[2]Australian eHealth Research Centre (AEHRC), Commonwealth Scientific and Industrial Research Organisation (CSIRO), Turrbal, Jagera Country, Herston, Australia

[3]College of Medicine and Public Health, Flinders University, Kaurna Country, Adelaide, Australia

[4]General Practice Clinical Unit, The University of Queensland, Turrbal, Jagera Country, Herston, Australia

[*]all authors contributed equally

Corresponding Author:
Andrew Goodman, AdvDip
School of Public Health
The University of Queensland
266 Herston Road
Turrbal, Jagera Country, Herston, 4006
Australia
Phone: 61 733655393
Fax: 61 733655442
Email: Andrew.Goodman@csiro.au

Abstract

Background: The advancements and abundance of mobile phones and portable health devices have created an opportunity to use mobile health (mHealth) for population health systems. There is increasing evidence for the feasibility and acceptance of mHealth with Indigenous populations. Providing a synthesis of qualitative findings of mHealth with Indigenous populations will gain insights into the strengths and challenges to mHealth use in Indigenous populations.

Objective: This review aimed to identify and synthesize qualitative data pertaining to the experiences and perceptions of mHealth from the perspectives of end users (patients and service providers) living in the colonial settler democracies of Canada, Australia, New Zealand, the United States, the Pacific Islands, and the Sápmi region of northern Europe.

Methods: In May 2021, systematic searches of peer-reviewed, scientific papers were conducted across the 5 databases of PubMed, CINAHL, Embase, PsycINFO, and Web of Science. Qualitative or mixed method studies were included where a mHealth intervention was the primary focus for responding to health challenges with Indigenous populations. Two authors independently screened papers for eligibility and assessed the risk of bias using a modified version of the Critical Appraisal Skills Programme. A meta-aggregative approach was used to analyze the findings of included studies.

Results: Seventeen papers met the eligibility criteria, 8 studies with patients, 7 studies with service providers, and 2 studies that included both patients and service providers. Studies were conducted in Australia (n=10), Canada (n=2), New Zealand (n=2), Papua New Guinea (n=1), the United States (n=1), and Samoa (n=1). Our interpretation of these qualitative findings shows commonalities between Indigenous patients' and service providers' perceptions of mHealth. We summarize our findings in six themes: (1) mHealth literacy, (2) mHealth as a facilitator for connection and support, (3) mHealth content needed to be culturally relevant, (4) mHealth security and confidentiality, (5) mHealth supporting rather than replacing service providers, and (6) workplace and organizational capacity.

Conclusions: This research suggests that mHealth can meet the needs of both patients and service providers when the mHealth intervention is culturally relevant, accounts for digital and health literacy, incorporates interactive components, is supported by workplaces, fits into health provider workflows, and meets security and confidentiality standards. Future mHealth research with Indigenous populations should partner with key representatives (eg, patients, service providers, and executive leaders) in the mHealth design appropriate to the purpose, people, setting, and delivery.

(JMIR Mhealth Uhealth 2023;11:e45162) doi:10.2196/45162

KEYWORDS

mHealth; Indigenous; Canada; Australia; New Zealand; United States; Papua New Guinea; Samoa; qualitative; systematic review; feasibility; acceptability; users; design; workflow

Introduction

The technological advancements and abundance of mobile phones and portable health devices have created a plethora of mobile health (mHealth) tools. mHealth is defined as "the use of mobile devices—such as mobile phones, patient monitoring devices, personal digital assistants and wireless devices—for medical and public health practice" [1]. These include mobile phone apps, text messages, portable monitoring devices and electronic patient information.

Systematic reviews globally have suggested mHealth is a broadly feasible and effective resource for a range of health conditions including; behavior change [2,3], noncommunicable disease management [4-9], perinatal care [10,11] medication adherence [12], and mental health well-being [13,14]. Likewise, health care workers suggest mHealth improves patient health outcomes and increases peer communication and care coordination [15,16].

There is a growing number of qualitative studies exploring the views and perceptions of mHealth from 2005 onward, resulting in a number of qualitative systematic reviews [16-21]. Findings from these reviews provide a collective insight into user perceptions and experience of mHealth to influence future research and implementation. These systematic reviews predominantly focus on non-Indigenous populations and fail to explore the user experiences of Indigenous people and their service providers. We need to ensure a space is kept privileging Indigenous worldviews as it pertains to mHealth. mHealth interventions are being explored with Indigenous populations with increasing interest [22-24]. Reviews examining the applicability of mHealth for Indigenous populations exist, and these indicate it is an acceptable health resource [23,24]. Yet, these reviews include qualitative data as only a peripheral focus and are inconsistent with the intervention type [23], and outcomes [24].

Providing a synthesis of qualitative findings of mHealth with Indigenous populations will gain insights to the strengths and challenges to mHealth use in Indigenous populations. This review aimed to identify and synthesize qualitative data pertaining to the experiences and perceptions of mHealth with Indigenous populations and the service providers that work with Indigenous populations.

Methods

Overview

A systematic search was conducted of peer-reviewed literature for this qualitative synthesis. A protocol of this qualitative synthesis was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42021251861). We extracted qualitative data pertaining to the experiences and perceptions of both patients (Indigenous peoples) and service providers (either Indigenous or non-Indigenous health policy makers, health care professionals, and researchers) who work with Indigenous peoples from Canada, Australia, New Zealand, the United States, the Pacific Islands, and the Sápmi region of northern Europe. We define Indigenous Peoples as "distinct social and cultural groups that share collective ancestral ties to the lands and natural resources where they live, occupy or from which they have been displaced" [25].

Search Strategy and Selection Criteria

A comprehensive list of search terms and strings were developed with the assistance of a librarian with expertise in systematic reviews. Systematic searches of peer-reviewed, scientific papers in English were conducted across 5 databases in May 2021: PubMed, CINAHL, Embase, PsycINFO, and Web of Science. Qualitative or mixed method studies were included where a mHealth intervention was the primary focus for responding to health challenges with Indigenous populations. As such, experimental and quasi-experimental studies were considered, as long as they met the following inclusion criteria:

- Participants: Indigenous people of all ages from Canada, Australia, New Zealand, United States, the Pacific Islands, the Sápmi region of northern Europe; OR are service providers (either Indigenous or non-Indigenous) who work with Indigenous persons from Canada, Australia, New Zealand, the United States, the Pacific Islands, the Sápmi region of northern Europe; OR where participants are multicultural, outcomes for Indigenous persons are reported specifically.
- Interventions: primary focus was a mHealth intervention delivered using a wireless device (eg, mobile or tablet app, website designed for mobile, messaging [SMS, voice, multimedia messaging system, etc]). The mHealth intervention aims to address a health challenge (eg, diagnosis of disease, substance use, health behaviors, quality of life, health knowledge, self-efficacy, caregiver support, etc).
- Outcomes: studies reported on one or more outcomes including user; experiences, perceptions, barriers, and enablers via qualitative research methods (eg, interviews and focus groups).

A sample of the search strings using text words and subject heading keywords for PubMed can be found in Multimedia Appendix 1. The use of proximity operators, truncation, and phrase searching was used to widen the search to capture all iterations of both the mHealth and Indigenous themes. The 2 search strings were then combined to narrow the results—enabling discovery of all possible scientific papers, which capture mHealth interventions with Indigenous populations from Canada, Australia, New Zealand, the United States, the Pacific Islands, and the Sápmi region of northern Europe. The qualitative papers were then identified via screening by 2 researchers (AG and SL).

Data Extraction and Quality Appraisal

Initial database searches and duplicate removal were conducted by 1 author (AG). Screening, review, and extraction were assisted by the web-based systematic review program Covidence (Veritas Health Innovation) [26]. Two authors (AG and SL) independently screened titles and abstracts against the inclusion criteria, and papers clearly not meeting the inclusion criteria were excluded.

Subsequently, 2 authors (AG and SL) screened the full-text papers independently and then discussed for comparison. Any differing views were resolved through discussion. Manual searches of reference lists were conducted on full-text papers included in the review. A final list of full-text papers and their citations which met inclusion criteria were downloaded and saved using Covidence software.

The quality of the included studies was appraised using a modified version of the Critical Appraisal Skills Programme (CASP) qualitative checklist [27]. An additional question from the Joanna Briggs Institute (JBI) was added that related to locating the researchers cultural or theoretical standpoint [28], improving the cultural rigor of this critical appraisal tool.

Data Analysis and Synthesis

The data included in the analysis were all text included in the "Results" or "Findings" sections of the papers (excluding purely quantitative findings) and was extracted from the papers into NVivo 12 Plus software (QSR International) [29]. Characteristics of each study to be extracted for descriptive purposes included: Indigenous identification, study location (country), year, sample size, participant demographic characteristics (age, gender), data collection, and analysis methods.

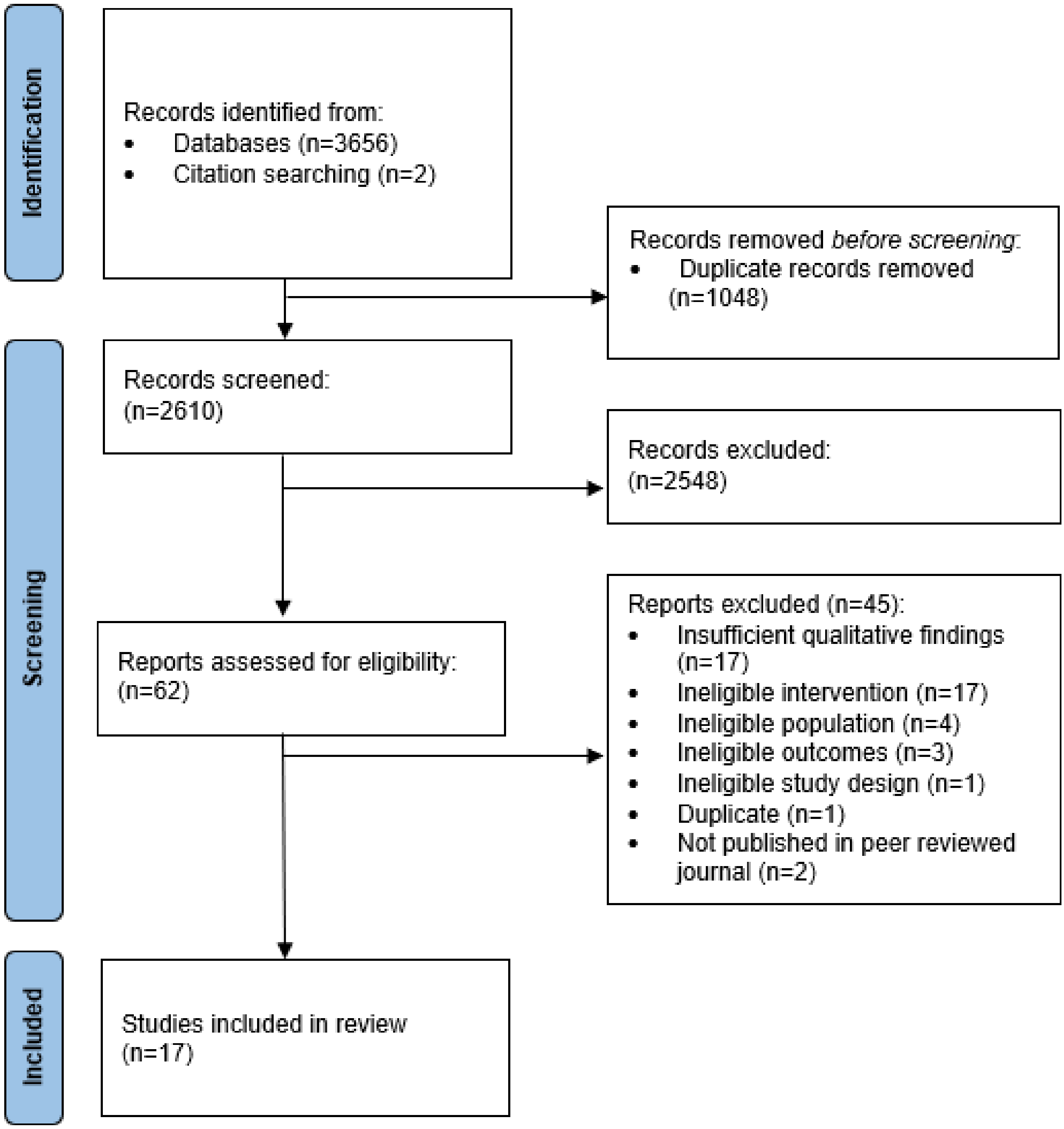
A meta-aggregative approach was used to analyze the findings [30]. This analysis approach privileges the findings, presented as "themes" or "constructs" in qualitative research, as identified by the researchers (not the reviewer). This method helps ensure the expanse and breadth of cultural learnings identified by researchers conducting the original studies are not lost by the reviewers.

Results

Overview

From database searches, 2608 unique papers were identified; 2 additional papers were located by manual searches. In total, 2610 titles and abstracts were reviewed against the inclusion criteria, of which 2548 were excluded, leaving 62 papers for full-text review. Following the full-text review, 45 papers were excluded, leaving 17 studies included in this qualitative systematic review (Figure 1).

Figure 1. Preferred reporting items for systematic reviews and meta-analyses flow diagram of study selection.



Description of Included Studies

All 17 studies included in this review were published between 2013 and 2021. Eight were studies specifically with Indigenous patients [31-38]. Seven studies were with service providers (either Indigenous or non-Indigenous) who work with Indigenous peoples [39-45]. Two studies involved both patients and service providers in data collection [46,47], so findings were considered for both.

Characteristics of the 17 studies are shown in Table 1. Ten papers involved Aboriginal and Torres Strait Islander peoples of Australia [31,36-38,40,42,44-47], 2 with the First Nations, Inuit, or Métis peoples of Canada [34,39], 2 with the Māori peoples of Aotearoa, New Zealand [33,35], 1 paper with the Indigenous peoples of Papua New Guinea [43], 1 paper with the Native Hawaiian and Pacific Islander peoples of Hawaii, the United States [41], and 1 paper with the Indigenous peoples of Samoa [48]. We were unable to identify any papers with Indigenous people of the Sápmi region of northern Europe that met the review criteria.

Table 1. Characteristics of included studies.

Studies	Indigenous peoples (country)	Focus area of intervention	Type of mHealth[a] delivery	Participants (roles)	Method	CAT[b] [27,28]
Indigenous patients						
Kennedy et al [38]	Aboriginal and Torres Strait Islander (Australia)	Perinatal health care	App[c]	8 Indigenous patients	Interviews[d]	10
Tighe et al [37]	Aboriginal and Torres Strait Islander (Australia)	Suicide prevention	App[e]	13 Indigenous patients	Interviews[d,f]	8
Jongbloed et al [34]	First Nations, Inuit, and Métis (Canada)	Illicit drug use	mHealth broad concept	130 Indigenous patients	Questionnaire[g]	6
Peiris et al [36]	Aboriginal and Torres Strait Islander (Australia)	Smoking cessation	App[e]	15 Indigenous patients	Interviews[h]	8
Gasteiger et al [35]	Māori (New Zealand)	Pregnancy or perinatal health care	mHealth broad concept	Nine Indigenous patients	Interviews[d,f,i]	11
Te Morenga et al [33]	Māori (New Zealand)	Healthy lifestyle	mHealth broad concept	21 Indigenous patients	Focus group and "bus stop activity"[d,i]	9
McCool et al [48]	Samoa	Smoking cessation	Text message	36 Indigenous patients	Focus group[d,f]	9
Povey et al [31]	Aboriginal and Torres Strait Islander (Australia)	Mental health well-being and suicide prevention	App[e]	9 Indigenous patients	Focus group[h]	10
Service providers						
Akearok et al [39]	Inuit (Canada)	Social determinants view of health	App[e]	5 (health service recruitment and education staff)	Interviews and survey[j]	3
Raphiphatthana et al [42]	Aboriginal and Torres Strait Islander (Australia)	Mental health well-being	App[e]	57 (nurses, support workers, Indigenous health workers, psychologists, and alcohol and other drug workers)	Interviews[d,h]	10
Macniven et al [45]	Aboriginal and Torres Strait Islander (Australia)	Cardiovascular	ECG attached to a mobile phone (iECG)	18 (Indigenous health workers, registered nurses)	Interviews[d,f]	10
Bennett-Levy et al [44]	Aboriginal and Torres Strait Islander (Australia)	Mental health well-being	App[e]	28 (consultant trainers, youth workers, Indigenous service workers, drug and alcohol worker, family development worker, well-being coordinator, Aboriginal health education officer, mental health support worker, and healthy lifestyle worker)	Interviews and field notes[d]	9
Yazdanshenas et al [41]	Native Hawaiian and Pacific Islander (United States)	Hypertension	Text message	20 (executive leader, church leader, community advocate, and health care providers)	Interviews[h]	7
Dingwall et al [40]	Aboriginal and Torres Strait Islander (Australia)	Mental health well-being	App[e]	15 (health professionals, managers, program coordinators, and an Aboriginal elder)	Interviews[d]	10
Kurumop et al [43]	Papua New Guinea	Malaria	SMS Text message	17 (health workers)	Focus group[h]	9
Both Indigenous patients and service providers						

Studies	Indigenous peoples (country)	Focus area of intervention	Type of mHealth[a] delivery	Participants (roles)	Method	CAT[b] [27,28]
Brown et al [46]	Aboriginal and Torres Strait Islander (Australia)	Mental health well-being	mHealth broad concept	12 (8 Indigenous health workers, 4 Indigenous patients).	Focus group[h,k]	11
Houston et al [47]	Aboriginal and Torres Strait Islander (Australia)	Perinatal health care or parenting	App[e] and website	31 (21 administration staff, pediatricians, child health nurses, general practitioners, and Indigenous health workers; 10 Indigenous patients)	Interviews and focus group[d]	8

[a]mHealth: mobile health.

[b]CAT: critical appraisal tool, maximum score is 11.

[c]mHealth interactive: integrated app used for access to health information or personal monitoring of health determinates that allows for information exchange (eg, peers and service providers).

[d]Thematic analysis.

[e]mHealth personal: autonomous app used for access to health information or personal monitoring of health determinates with no interactive capabilities.

[f]Inductive analysis.

[g]Rapid qualitative analysis.

[h]Hybrid approach to qualitative analysis.

[i]Kaupapa Maori approach.

[j]Narrative approach to qualitative analysis.

[k]Yarning approach.

Thematic Synthesis

Overview

Our interpretation of these qualitative findings shows commonalities between Indigenous patients and service providers perceptions of mHealth. We have collectively termed both as “end users” hereafter unless explicitly stated otherwise. Common themes across end users were: the importance of mHealth or digital literacy, mHealth as a facilitator for connection and support, mHealth content that needed to be culturally relevant, and data security and confidentiality. Two themes emerged that were unique to service providers including the importance of mHealth supporting rather than replacing service providers, and the role of workplace champions and organizational capacity for influencing uptake and sustainability of mHealth.

mHealth Literacy

Access to the required hardware (mobile or smartphones and touch screen tablets) and systems (network coverage and IT) was identified as an important influence to mHealth uptake by end users. Service providers noted barriers to accessing the mHealth hardware and systems with reasons including regionality and workplace restrictions [40,42,43,45,46]. Service providers held a perception that mobile phones were not prevalent or accessible to patients due to cost [46] and remote location [42]. However, Indigenous patients saw themselves as competent and confident users of technology and mobile phones for everyday life [31,33-38,46,47]. Yet, technology difficulties and lack of device access were still raised in several studies [31,34-36,46,47]. Some studies noted concerns end users had relating to the digital literacy required for mHealth [40,42,44].

> *Low levels of IT literacy pose a challenge to electronic mental health adoption. Unfamiliarity with different ways of using technologies impedes the utilization of the approach by both service providers and community members. Poor IT literacy within communities was attributed to limited access to technology…* [42]

Limited confidence in using new technology such as mHealth initiatives was identified as a barrier to uptake for service providers [40,44]. The investment of time and effort into appropriate mHealth training and ongoing support was suggested as a mitigation strategy for technical difficulties for end users [36,40].

Age and generational implications were raised as influential factors to the uptake of mHealth. Whether implicitly or explicitly, end users perceived mHealth would be more applicable and accepted by younger people [31,33,40-42,44,46]. Service providers perceived that older people have limited or no access to mobile phones and thus would have lower digital literacy [40,41,46]. Interestingly, these age-related barriers were not reflected by Indigenous patients in this Australian study.

> *Most older interviewees did not appear to have any major issues with knowledge on how to access phone features.* [36]

End users noted the importance of mHealth resources easy to understand and use for a confident user experience. A notable motivation for service providers to use a mHealth resource was for it to be uncomplicated and easy to use [39,40,45-47]. Likewise, Indigenous patients advised that if mHealth platforms were complex, slow, or used too much data, uptake and sustained use was less likely [31,33,35,38,46]. The importance of clear and concise language and the avoidance of jargon in mHealth messaging was noted to encourage comprehension for end users [31,37,38,41,43,47,48]. Service providers stated the importance of mHealth content being appropriate to the learning styles, health knowledge, and communication styles of their patients [39-41,43,44,46,47]. Indigenous patients were enthusiastic about the potential benefits mHealth provided in accessing relevant information for their health journey [33-35,37,38,46-48].

> *Participants spoke of parents being technologically savvy, and parents referred to accessing apps, YouTube clips, social media and the internet from their mobile phones for infant feeding information and support prior to the Growing healthy program.* [47]

The incorporation of visual and audio capabilities was suggested by end users to create a better understanding of the content [31,35-41,44,46-48]. mHealth may provide an appropriate tool to bridge health knowledge gaps [42] and enable education and empowerment for health care.

> *Several interviewees described how the iECG device provided unique opportunities to engage patients in education around AF and their heart, and to empower patients to find out more about their heart health.* [45]

mHealth as a Facilitator for Connection and Support

End users found mHealth an appropriate resource to facilitate engagement, connection, and support within health care systems. Indigenous patients appreciate that mHealth facilitated connection to support people, along with health care providers [31,33-38,47,48]. Likewise, service providers viewed mHealth as important for connection to patients with the added facility to connect with professional colleagues [40-42,46].

mHealth was found to provide a sense of reassurance and encouragement across a range of health journeys for Indigenous patients, including perinatal health [35,47], patients living with mental health challenges [34,37], and people on a smoking cessation journey [36,48]. Indigenous patients suggested mHealth could enable a web-based community to connect with others on similar health journeys [31,33-38,46]. Moreover, Indigenous patients appreciated the capability of mHealth to share health knowledge with family and support people in their lives [33,35,47,48].

> *Participants valued sharing advice and experience-based information with their families, partners or wider virtual communities, such as Facebook groups.* [35]

Service providers found that mHealth encouraged trust with patients while creating a collaborative environment with other health staff [40-42,46]. mHealth was found to provide professional peer support [40,46] while streamlining clinical communication and encouraging service provider collaborations [42].

> *Communication across services working with the same client may help to ensure nonoverlapping of interventions and resources.* [42]

Australian service providers noted mHealth broke down the barriers of patient engagement, "equalising the power imbalance often present in their relationships with clients" [40]. Service providers attributed this to the app acting as an impartial entity, encouraging person-centered care [40]. First Nations, Inuit, and Métis youth in Canada explained that having a mobile phone would enable them to connect with health professionals as well as on behalf of peers in emergency situations [34]. Youth in Australia found that mHealth provided connection to service providers while adding anonymity and privacy to the navigation of their mental health journey.

> *Some may have felt known in a small community or simply hesitant to engage a service because they felt uncomfortable. The app allowed them a choice in health care that was previously unavailable.* [37]

mHealth Content Needed to Be Culturally Relevant

End users stated the importance of mHealth including culturally relevant imagery and language to enable engagement, trust, and relatable connection. The inclusion of culturally relevant language and imagery was important for Indigenous patients to encourage engagement and build trust in mHealth content [31-33,37,38,46,47]. Likewise, service providers suggested the need for culturally applicable imagery and language in mHealth content in several studies [39-41,43,44,46,47]. End users suggested the translation of mHealth content to traditional language would enable comprehension of content as well as increase uptake [31,33,40,43,46].

> *…participants were keen to engage with apps that included Māori language, tikanga and knowledge.* [33]

> *Culturally relevant graphics, voices, animation, and optional short video clips may assist in engagement with the content, improve understanding, and overcome literacy issues.* [31]

> *Recommended features of a technology resource included a look and feel that was user-friendly, aesthetically pleasing (e.g., more visuals, Indigenous artwork and potentially Indigenous language for more remote communities), easy to read, quick to navigate, and interactive (e.g., notifications, touch screen, user online status shown).* [46]

Yet, the acknowledgment of diversity in cultural relevance was an important implication noted by Indigenous patients [31,33,38], namely, the tailoring of dialect [31,38] and that content be appropriate to the local cultural peoples [33], to ensure mHealth is not dismissive of cultural diversity.

Findings suggest mHealth can assist in developing cultural competence through gaining a better understanding of cultural diversity, histories, and traditional languages. In Australia, Indigenous patients advised the importance of including cultural determinants such as colonization, intergenerational trauma, and identity within mHealth content [31,38]. In Aotearoa, New Zealand, Māori patients chose to use traditional terminology in the thematic findings of mHealth exploration, acknowledging the importance of the cultural determinants of health [33,35]. mHealth was found to be an important resource to support culturally competent health care delivery for locum service providers in Canada.

> *Respondents expressed gratitude that the app now exists as an important tool for use in training and orienting new hires to Nunavut's cultural and language context.* [39]

mHealth Security and Confidentiality

Security and privacy consistently emerged with Indigenous patients across several studies with differing views and implications [31,34,35,37,46]. Povey et al [31] found Indigenous patients were largely dismissive of privacy issues with regard to mHealth, noting that personal information held on phones such as photos, or emails being seen would worry them more. There were, however, concerns raised about the privacy and confidentiality of information being shared during group discussions embedded in mHealth [46]. In addition, Māori women felt a sense of intrusion when using their mobile phone to seek health advice [35]. This intrusion was due to third-party systems, not necessarily a mHealth resource.

> *…emphasised privacy concerns whereby they encountered personalised advertising on Google and Facebook that was based on previous searches done on the device.* [35]

Importantly, mHealth offered the opportunity of anonymous support for patients wishing not to engage with health services face to face [31,34,37]. Access to their own phone provided a sense of privacy and a safety net for Indigenous patients in Canada [34]. Likewise, Indigenous patients in Australia appreciated the facility of remote support seeking with the avoidance of unwanted in-person contact.

> *The ability to interact with the app privately, without anyone else needing to be present, meant that youth who may have been reluctant or afraid to speak to family members or health care professionals in a face-to-face setting could still access support.* [37]

mHealth Supporting Rather Than Replacing Service Providers

Service providers stated the importance of mHealth needing to support established workloads and practices rather than being an onerous addition to established workloads. Service providers raised uncertainties about the sustainability of mHealth, and how their roles and responsibilities may change with the implementation of mHealth [40-42,44,46]. The perceived "lack of fit" with established work practices was a professional barrier identified [40,42,44]. Service providers suggested that mHealth should be considered as a complementary resource in addition to "in person" and physical resources [40,41,46]. Service providers in Australia found mHealth may be more useful for staff lacking experience and confidence in health practice.

> *Gatekeepers less experienced in suicide prevention may find a resource more useful than more experienced or confident gatekeepers.* [46]

Service providers saw the benefit of mHealth as an educational tool to develop skills and knowledge. Service providers in Australia liked the health promotion opportunity a smartphone-enabled electrocardiogram (ECG) provided [45]. Service providers in Papua New Guinea valued the guidance capabilities mHealth provided them for clinical malaria treatment procedures [43]. Service providers in Australia identified mHealth as an appropriate resource to gain professional skills and knowledge in interviewing and counseling [40,44]. A smartphone-enabled ECG (ie, iECG) was found to have an indirect educational effect on service providers in Australia.

> *Some staff also spoke of how using the device for screening led them to want to learn more about AF and cardiovascular disease themselves in their professional role.* [45]

Workplace and Organizational Capacity

Workplace leadership, capacity, and strategic direction emerged as influencing factors to the uptake and sustainability of mHealth for service providers working with Indigenous populations [39,40,42,44,45].

Workplaces that have leaders and champions to drive and support mHealth were a central factor in enabling mHealth uptake. The presence of enthusiastic managers and eager IT champions had a positive effect on the workforce's interest in mHealth resources with service providers in Australia [42,44]. Workplace leaders that did not perceive the need for or effectiveness of mHealth were often a barrier to the uptake by service providers [42,44]. The advocacy of mHealth from leadership was an influencing factor to acceptance:

> *Having leaders within the organization showing interest and providing direct support was perceived to facilitate uptake. It created incentives and provided opportunities for service providers to reflect and evaluate the utility of the electronic mental health approach.* [42]

Workplace staff capacity and retention contributed to the opportunities service providers had to commit to mHealth implementation [40,42,44]. High turnover of staff contributed to a lack of sustained mHealth knowledge and skill within the workplace [40,42,44]. The significance of investment into continued staff training and development was seen as important for mHealth success [40,42,45]. Limited workload capacity due to underresourcing impeded mHealth delivery [40,42,44] and restricted service providers' capacity to engage in mHealth.

> *…in many services, demanding workloads left the workers with little or no opportunity to incorporate new skills into their existing work practices…* [44]

A workplace culture that supports and drives the use of health innovations was shown to positively impact service providers' perception of mHealth. The absence of health innovation priorities in workplace strategies caused a sense of ambivalence and ineptness toward the need for mHealth among service providers [39,42,44]. Workplaces that invested in systems, valued innovation, and had supportive leadership, positively influenced service providers' perception and engagement with mHealth tools [40,42,44]. Alignment of the health innovation with organizational principles was found to influence uptake.

Uptake of electronic mental health approaches was dependent upon the perceived fit of the innovation to the organization's priorities. [42]

Discussion

Principal Results

This review found that both Indigenous patients and service providers are enthusiastic about the role that mHealth can play in health service delivery.

Common themes across end users were: importance of mHealth or digital literacy, mHealth as a facilitator for connection and support, mHealth content needed to be culturally relevant, and data security and confidentiality are a priority. Two themes emerged that were unique to service providers: the importance of mHealth supporting rather than replacing service providers and the role of workplace champions and organizational capacity for influencing the uptake and sustainability of mHealth.

In this review, most included studies stated the importance of relevant cultural imagery and language, which enabled greater comprehension of mHealth messaging and increased engagement by end users [31-33,37-41,43,44,46,47]. Cultural content needs to account for the heterogeneity of Indigenous peoples, appropriate to location, language, people, and knowledge systems. This creates a challenge for mHealth developers and researchers alike in having 1 product with the capability to be distributed to a culturally diverse audience. Regarding language, Varnfield et al [49] increased their scope of patient engagement with their mHealth app being "available in several different selected languages." This demonstrates that mHealth has the potential to be adaptive with its content.

Similar to the included study findings of this review, mHealth has been shown to enable patients to engage with their health care providers more effectively as well as connect with peers on similar health care journeys [21]. Moreover, our findings support other reviews reporting health care providers who found mHealth improved communication between their patients and colleagues [15,16].

Our findings showed the importance of workplaces and their leadership in influencing the uptake of mHealth [39,40,42,44,45]. Likewise, Palacholla et al [50] found leadership and organizations that were supportive and facilitated digital health adoption in clinical settings. An important factor when implementing health service innovation is localized agenda setting being led by need, want, and appropriateness [51]. Within a mHealth context, Gagnon et al [52] found health professionals considered their workplace environment as one of the top contributing factors to adoption. Engaging health care organizations as a partner to support mHealth may offer the greatest opportunity for sustained uptake.

Other systematic reviews conducted to understand the influencing factors to mHealth uptake show a strong correlation with the findings presented here. Namely, the principal influencers for adoption are the mHealth design, personal perceptions of mHealth, and the workplace environment [16,21,52], which suggest that co-design may offer an effective methodology for sustained mHealth uptake with Indigenous populations and service providers that work with Indigenous populations.

Early engagement with the Indigenous community within eHealth research and implementation has shown to offer the greatest opportunity for acceptability, and local advocacy [22,23,53,54]. Moreover, this model of prioritizing community partnership and co-design is recommended by governing ethical guidelines on research with Indigenous peoples internationally to achieve beneficial research outcomes [55-58]. Despite this, Eyles et al [59] found a lack of co-design methods for minority and Indigenous groups internationally in the development of mHealth interventions. With the novelty of mHealth along with the cultural considerations involved in the study population, it would be practical to enter a colearning and cocreation relationship to achieve mutually beneficial outcomes.

In conclusion, there has been considerable growth in qualitative research exploring contextual factors in relation to mHealth uptake in non-Indigenous populations, yet less so for Indigenous populations. To our knowledge, this is the first review of qualitative studies that provides an understanding of the influential factors for both patients and service providers for Indigenous populations in relation to mHealth.

Strengths, Limitations, and Future Directions

Having 2 reviewers from diverse cultural backgrounds and gender orientations independently screening improved the quality of this meta-synthesis. The authors are a multidisciplinary team with a breadth of expertise in this review focus (psychology, digital health, qualitative research, and Indigenous health). Using a meta-aggregative approach to analyze the findings ensured cultural learnings identified by researchers' conducting the original studies were not lost by the reviewers. The quality appraisal tool used a modified version of the CASP qualitative checklist, with the additional question locating the researchers' cultural or theoretical standpoint, improving the cultural rigor of this critical appraisal tool. Most studies were of medium to high quality, and the quality appraisal tool can be found in Multimedia Appendix 2.

Our review has some limitations; first, the searches were restricted to peer-reviewed literature published in 5 databases (PubMed, CINAHL, Embase, PsycINFO, and Web of Science). Second, publication bias may have occurred due to the subjective quantifying of studies reporting on one or more outcomes via qualitative research methods. Finally, the results of this study are based on the meta-synthesis of qualitative data, which is inherently subjective. There are studies included from all countries (except the Sápmi region), but there are still only a few studies in each country, and so more work is needed. Papers need to report not only on patients' perspective but other end users to gain a full understanding of the perceptions of mHealth in supporting health care with Indigenous populations.

Conclusions

This review used meta-aggregation to summarize the findings of 17 qualitative studies on the experiences and perceptions of mHealth with Indigenous populations and the service providers that work with Indigenous populations. mHealth end users are enthusiastic about the role that mHealth can play in Indigenous health service delivery. There is a need for mHealth design to center end users within a co-designed approach with Indigenous people. There is recent work driving this agenda in an Australian context [60]. Allowing end users to suggest localized agenda setting through co-design may provide an opportunity for ownership, championship, and mitigation of barriers in mHealth implementation. Future research should partner with key representatives (eg, patients, health care professionals, and executive leaders) in the mHealth design appropriate to the purpose, people, setting, and delivery.

Acknowledgments

To ensure a culturally inclusive lens, our authorship reflects a diversity of the background, career stage, gender, and race. Specific to the focus of the manuscript AG and RM are Indigenous, and GS and SL are non-Indigenous. AG is an Aboriginal PhD candidate from Iningai country in Central West Queensland, Australia. AG has spent more than 13 years as an Indigenous health worker in Queensland alongside rural and remote Aboriginal and Torres Strait Islander people in the discipline of cardiac and health care services. RM is Aboriginal, a descendant of the Bidjara people of Central Western Queensland, Australia. RM is an Aboriginal health leader and researcher and has worked extensively to implement best practice cardiovascular care, particularly for Aboriginal and Torres Strait Islander peoples. GS is a General Practitioner at the Inala Indigenous Health Service in Brisbane, Queensland, and a General Practice academic at The University of Queensland. SL is a public health academic, with research interests in broad reach interventions to improve health outcomes in priority populations. We thank The University of Queensland for the support of this review via an Indigenous higher degree by research Development Grant. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to the full data in the study and accept responsibility to submit for publication.

Data Availability

The search strategy is available in Multimedia Appendix 1. Any additional data are available upon request from the corresponding author.

Authors' Contributions

AG and SL conceptualized the study design, retrieved, analyzed, and interpreted the research data on which the scholarly work is based. AG and SL prepared the manuscript with significant input and critical review from RM and GS. All authors have read and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative systematic review search string example.
[PNG File , 250 KB - mhealth_v11i1e45162_app1.png]

Multimedia Appendix 2
CASP tool with additional JBI question.
[PDF File (Adobe PDF File), 98 KB - mhealth_v11i1e45162_app2.pdf]

References

1. Global difusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. WHO. Geneva: World Health Organization; 2016. URL: https://www.who.int/publications/i/item/9789241511780 [accessed 2023-05-30]
2. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review (m-health technology interventions for lay people). PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362]
3. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
4. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

5. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
6. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
7. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
8. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
9. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
10. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
11. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018;20(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
12. Palmer MJ, Machiyama K, Woodd S, Gubijev A, Barnard S, Russell S, et al. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018:1-54 [FREE Full text] [doi: 10.1002/14651858.cd012675.pub2]
13. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
14. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
15. White A, Thomas DSK, Ezeanochie N, Bull S. Health worker mHealth utilization: a systematic review. Comput Inform Nurs 2016;34(5):206-213 [FREE Full text] [doi: 10.1097/CIN.0000000000000231] [Medline: 26955009]
16. Odendaal WA, Watkins JA, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
17. Carter DD, Robinson K, Forbes J, Hayes S. Experiences of mobile health in promoting physical activity: a qualitative systematic review and meta-ethnography. PLoS One 2018;13(12):e0208759 [FREE Full text] [doi: 10.1371/journal.pone.0208759] [Medline: 30557396]
18. Hu R, van Velthoven MH, Meinert E. Perspectives of people who are overweight and obese on using wearable technology for weight management: systematic review. JMIR Mhealth Uhealth 2020;8(1):e12651 [FREE Full text] [doi: 10.2196/12651] [Medline: 31929104]
19. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
20. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
21. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
22. Reilly R, Stephens J, Micklem J, Tufanaru C, Harfield S, Fisher I, et al. Use and uptake of web-based therapeutic interventions amongst indigenous populations in Australia, New Zealand, the United States of America and Canada: a scoping review. Syst Rev 2020;9(1):123 [FREE Full text] [doi: 10.1186/s13643-020-01374-x] [Medline: 32475342]
23. Jones L, Jacklin K, O'Connell ME. Development and use of health-related technologies in Indigenous communities: critical review. J Med Internet Res 2017;19(7):e256 [FREE Full text] [doi: 10.2196/jmir.7520] [Medline: 28729237]
24. Hobson GR, Caffery LJ, Neuhaus M, Langbecker DH. Mobile health for first nations populations: systematic review. JMIR Mhealth Uhealth 2019;7(10):e14877 [FREE Full text] [doi: 10.2196/14877] [Medline: 31593537]
25. The World Bank. Indigenous Peoples. The World Bank. 2023. URL: https://www.worldbank.org/en/topic/indigenouspeoples [accessed 2023-07-03]
26. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2023-05-25]

27. Critical Appraisal Skills Programme (CASP). Critical appraisal skills programme qualitative checklist. Qual Res checklist 2017;31(13):449. [doi: 10.7717/peerj.12951/table-2]
28. The Joanna Briggs Institute. The joanna briggs institute critical appraisal tools for use in JBI systematic reviews. Checklist Qual Res 2017;9(Supplement):e025789 [FREE Full text] [doi: 10.11124/01938924-201109481-00003]
29. NVivo (Version 12). QSR International. 2018. URL: https://qsr-nvivo.software.informer.com/12.2/ [accessed 2023-05-25]
30. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. JBI Reports 2004;2(2):45-64. [doi: 10.1111/j.1479-6988.2004.00008.x]
31. Povey J, Mills PPJR, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for aboriginal and torres strait islander australians: a qualitative study. J Med Internet Res 2016;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
32. McCool J, Dobson R, Whittaker R. Moving beyond the individual: mHealth tools for social change in low-resource settings. BMJ Glob Health 2018;3(6):e001098 [FREE Full text] [doi: 10.1136/bmjgh-2018-001098] [Medline: 30498591]
33. Te Morenga L, Pekepo C, Corrigan C, Matoe L, Mules R, Goodwin D, et al. Co-designing an mHealth tool in the New Zealand Māori community with a “kaupapa māori” approach. Altern Int J Indig Peoples 2018;14(1):90-99. [doi: 10.1177/1177180117753169]
34. Jongbloed K, Pearce ME, Thomas V, Sharma R, Pooyak S, Demerais L, Cedar Project Partnership. The cedar project - mobile phone use and acceptability of mobile health among young indigenous people who have used drugs in British Columbia, Canada: mixed methods exploratory study. JMIR Mhealth Uhealth 2020;8(7):e16783 [FREE Full text] [doi: 10.2196/16783] [Medline: 32716311]
35. Gasteiger N, Anderson A, Day K. Rethinking engagement: exploring women's technology use during the perinatal period through a Kaupapa Māori consistent approach. J (New Zealand Coll Midwives) 2019;55(55):20-26 [FREE Full text]
36. Peiris D, Wright L, News M, Rogers K, Redfern J, Chow C, et al. A smartphone app to assist smoking cessation among aboriginal Australians: findings from a pilot randomized controlled trial. JMIR Mhealth Uhealth 2019;7(4):e12745 [FREE Full text] [doi: 10.2196/12745] [Medline: 30938691]
37. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in aboriginal and torres strait islander youth. JMIR Ment Health 2020;7(12):e14296 [FREE Full text] [doi: 10.2196/14296] [Medline: 33258782]
38. Kennedy M, Kumar R, Ryan NM, Bennett J, Fuentes GLH, Gould GS. Codeveloping a multibehavioural mobile phone app to enhance social and emotional well-being and reduce health risks among aboriginal and torres strait islander women during preconception and pregnancy: a three-phased mixed-methods study. BMJ Open 2021;11(11):e052545 [FREE Full text] [doi: 10.1136/bmjopen-2021-052545] [Medline: 34819285]
39. Akearok GH, Tabish T, Cherba M. Cultural orientation and safety app for new and short-term health care providers in Nunavut. Can J Public Health 2020;111(5):694-700 [FREE Full text] [doi: 10.17269/s41997-020-00311-8] [Medline: 32219728]
40. Dingwall KM, Puszka S, Sweet M, Nagel T. “Like drawing into sand”: acceptability, feasibility, and appropriateness of a new e-mental health resource for service providers working with aboriginal and torres strait islander people. Aust Psychol 2020;50(1):60-69. [doi: 10.1111/ap.12100]
41. Yazdanshenas H, Bazargan M, Jones L, Vawer M, Seto TB, Farooq S, et al. Engaging gatekeeper-stakeholders in development of a mobile health intervention to improve medication adherence among African American and pacific islander elderly patients with hypertension. JMIR Mhealth Uhealth 2016;4(4):e116 [FREE Full text] [doi: 10.2196/mhealth.5905] [Medline: 27784651]
42. Raphiphatthana B, Sweet M, Puszka S, Whitty M, Dingwall K, Nagel T. Evaluation of electronic mental health implementation in northern territory services using the integrated "promoting action on research implementation in health services" framework: qualitative study. JMIR Ment Health 2020;7(5):e14835 [FREE Full text] [doi: 10.2196/14835] [Medline: 32452812]
43. Kurumop SF, Bullen C, Whittaker R, Betuela I, Hetzel MW, Pulford J. Improving health worker adherence to malaria treatment guidelines in papua New Guinea: feasibility and acceptability of a text message reminder service. PLoS One 2013;8(10):e76578 [FREE Full text] [doi: 10.1371/journal.pone.0076578] [Medline: 24116122]
44. Bennett-Levy J, Singer J, DuBois S, Hyde K. Translating E-Mental health into practice: what are the barriers and enablers to E-Mental health implementation by aboriginal and torres strait islander health professionals? J Med Internet Res 2017;19(1):e1 [FREE Full text] [doi: 10.2196/jmir.6269] [Medline: 28077347]
45. Macniven R, Gwynn J, Fujimoto H, Hamilton S, Thompson SC, Taylor K, et al. Feasibility and acceptability of opportunistic screening to detect atrial fibrillation in aboriginal adults. Aust N Z J Public Health 2019;43(4):313-318 [FREE Full text] [doi: 10.1111/1753-6405.12905] [Medline: 31141280]
46. Brown K, Toombs M, Nasir B, Kisely S, Ranmuthugala G, Brennan-Olsen SL, et al. How can mobile applications support suicide prevention gatekeepers in Australian indigenous communities? Soc Sci Med 2020;258:113015 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113015] [Medline: 32559573]

47. Houston A, Laws R, Askew D, Saldanha T, Denney-Wilson E. Exploring the cultural appropriateness and usefulness of a mHealth promotion program for infant feeding in an urban aboriginal health service: a qualitative study. Aust Indig Health Bull 2017;17(3):1-10.
48. McCool J, Tanielu H, Umali E, Whittaker R. Assessing the cross-cultural adaptation and translation of a text-based mobile smoking cessation program in samoa (TXTTaofiTapaa): pilot study. JMIR Mhealth Uhealth 2018;6(8):e173 [FREE Full text] [doi: 10.2196/mhealth.9033] [Medline: 30170994]
49. Varnfield M, Redd C, Stoney RM, Higgins L, Scolari N, Warwick R, et al. M♡THer, an mHealth system to support women with gestational diabetes mellitus: feasibility and acceptability study. Diabetes Technol Ther 2021;23(5):358-366 [FREE Full text] [doi: 10.1089/dia.2020.0509] [Medline: 33210954]
50. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
51. Fox A, Gardner G, Osborne S. A theoretical framework to support research of health service innovation. Aust Health Rev 2015;39(1):70-75 [FREE Full text] [doi: 10.1071/AH14031] [Medline: 25513878]
52. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
53. Verbiest MEA, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: a case study in new zealand. Transl Behav Med 2019;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]
54. Kip H, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multidisciplinary Approach. 1 ed. Milton Park, Abingdon, Oxfordshire: Routledge; 2018.
55. Keeping research on track II: A companion document to Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: guidelines for researchers and stakeholders. National Health and Medical Research Council. Canberra: Commonwealth of Australia; 2018. URL: https://www.arts.unsw.edu.au/sites/default/files/documents/keepingresearchontrack.pdf [accessed 2023-05-25]
56. Code of ethics for Aboriginal and Torres Strait Islander research. Australian Institute of Aboriginal and Torres Strait Islander Studies. 2020. URL: https://aiatsis.gov.au/sites/default/files/2020-10/aiatsis-code-ethics.pdf [accessed 2023-05-25]
57. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, Canadian Institutes of Health Research Natural Sciences and Engineering Research Council of Canada Social Sciences and Humanities Research Council, Editor. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council. 2018. URL: https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html [accessed 2023-05-25]
58. Pūtaiora Writing Group. Te Ara Tika: Guidelines for Māori Research Ethics: A Framework for Researchers and Ethics Committee Members. Auckland: Health Research Council of New Zealand; 2010.
59. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
60. Chelberg GR, Butten K, Mahoney R, eHRCATSIH Group. Culturally safe eHealth interventions with aboriginal and torres strait Islander people: protocol for a best practice framework. JMIR Res Protoc 2022;11(6):e34904 [FREE Full text] [doi: 10.2196/34904] [Medline: 35687420]

Abbreviations

CASP: Critical Appraisal Skills Programme
ECG: electrocardiogramJBI: Joanna Briggs Institute
mHealth: mobile health
PROSPERO: International Prospective Register of Systematic Reviews

Edited by L Buis; submitted 18.12.22; peer-reviewed by U Nair, P Camp, A Al; comments to author 08.02.23; revised version received 07.03.23; accepted 21.04.23; published 23.06.23.

Please cite as:
Goodman A, Mahoney R, Spurling G, Lawler S
Influencing Factors to mHealth Uptake With Indigenous Populations: Qualitative Systematic Review
JMIR Mhealth Uhealth 2023;11:e45162
URL: https://mhealth.jmir.org/2023/1/e45162
doi:10.2196/45162
PMID:37351947

Review

Knowledge Discovery in Ubiquitous and Personal Sleep Tracking: Scoping Review

Nhung Huyen Hoang[1*], BA; Zilu Liang[1*], PhD

Graduate School of Engineering, Kyoto University of Advanced Science, Kyoto, Japan

[*]all authors contributed equally

Corresponding Author:
Nhung Huyen Hoang, BA
Graduate School of Engineering
Kyoto University of Advanced Science
Ukyo Ward, Yamanouchi Gotandacho, 18
Kyoto, 615-0096
Japan
Phone: 81 0754067000
Email: 2021mm12@kuas.ac.jp

Abstract

Background: Over the past few decades, there has been a rapid increase in the number of wearable sleep trackers and mobile apps in the consumer market. Consumer sleep tracking technologies allow users to track sleep quality in naturalistic environments. In addition to tracking sleep per se, some sleep tracking technologies also support users in collecting information on their daily habits and sleep environments and reflecting on how those factors may contribute to sleep quality. However, the relationship between sleep and contextual factors may be too complex to be identified through visual inspection and reflection. Advanced analytical methods are needed to discover new insights into the rapidly growing volume of personal sleep tracking data.

Objective: This review aimed to summarize and analyze the existing literature that applies formal analytical methods to discover insights in the context of personal informatics. Guided by the problem-constraints-system framework for literature review in computer science, we framed 4 main questions regarding general research trends, sleep quality metrics, contextual factors considered, knowledge discovery methods, significant findings, challenges, and opportunities of the interested topic.

Methods: Web of Science, Scopus, ACM Digital Library, IEEE Xplore, ScienceDirect, Springer, Fitbit Research Library, and Fitabase were searched to identify publications that met the inclusion criteria. After full-text screening, 14 publications were included.

Results: The research on knowledge discovery in sleep tracking is limited. More than half of the studies (8/14, 57%) were conducted in the United States, followed by Japan (3/14, 21%). Only a few of the publications (5/14, 36%) were journal articles, whereas the remaining were conference proceeding papers. The most used sleep metrics were subjective sleep quality (4/14, 29%), sleep efficiency (4/14, 29%), sleep onset latency (4/14, 29%), and time at lights off (3/14, 21%). Ratio parameters such as deep sleep ratio and rapid eye movement ratio were not used in any of the reviewed studies. A dominant number of the studies applied simple correlation analysis (3/14, 21%), regression analysis (3/14, 21%), and statistical tests or inferences (3/14, 21%) to discover the links between sleep and other aspects of life. Only a few studies used machine learning and data mining for sleep quality prediction (1/14, 7%) or anomaly detection (2/14, 14%). Exercise, digital device use, caffeine and alcohol consumption, places visited before sleep, and sleep environments were important contextual factors *substantially* correlated to various dimensions of sleep quality.

Conclusions: This scoping review shows that knowledge discovery methods have great potential for extracting hidden insights from a flux of self-tracking data and are considered more effective than simple visual inspection. Future research should address the challenges related to collecting high-quality data, extracting hidden knowledge from data while accommodating within-individual and between-individual variations, and translating the discovered knowledge into actionable insights.

(JMIR Mhealth Uhealth 2023;11:e42750) doi:10.2196/42750

KEYWORDS

sleep tracking; knowledge discovery; data mining; personal informatics; self-experimentation; sleep health; scoping review; mobile phone

Introduction

In tandem with the advent of consumer wearable technologies, there has been a growing interest in using consumer sleep tracking technologies for personal health management. Being aware of the importance of having a good night's sleep, many individual users are routinely monitoring their sleep [1-3], and sleep tracking has been a popular topic, especially in the quantified-self community. Consumer sleep tracking technologies are largely divided into 2 types: smartphone dependent and smartphone independent. Smartphone-dependent sleep tracking technologies leverage the integrated sensors of a smartphone (eg, accelerometer, gyroscope, and microphone) to measure body movements and ambient sound, based on which a user's sleep states can be estimated. Smartphone-independent sleep tracking technologies use independent hardware with multiple sensing modalities, such as accelerometers and photoplethysmography. These devices often come in the form of a wristband (eg, Fitbit [Fitbit Inc] and Apple Watch [Apple]), headband (eg, SleepSheperd [Sleep Shepherd] and Neuroon [Vandrico Inc]), or finger ring (eg, Oura [Oura Health Oy]), and they rely on proprietary sleep staging algorithms to calculate sleep metrics based on measurable physiological signals [4]. The accuracy of these consumer technologies has been significantly improved over the years. Recent models have proven to be reasonably accurate, especially in measuring the time of sleep onset and offset, total sleep duration, and sleep efficiency (SE) [5-7]. A recent study comparing 7 consumer sleep tracking devices with polysomnography (the gold standard of sleep measurement) demonstrated that their validity could outperform medical-grade actigraphy [6]. In the past decade, sleep tracking has become one of the most studied topics in the research field of personal informatics. At the intersection of ubiquitous computing, human-computer interaction, and sleep science, researchers from multiple disciplines have made joint efforts to investigate the validity of existing consumer sleep tracking devices [5-10], develop accurate sleep staging algorithms tailored to consumer sleep tracking devices [11-15], develop smartphone apps for visualizing personal sleep data [16-19], develop artificial intelligence–based sleep coaching systems that help people improve sleep hygiene [20], and understand the challenges for sleep tracking technologies to eventually improve sleep health [3,21-23]. At a higher level, sleep tracking studies have mostly been guided by general self-tracking frameworks, such as the lived informatics model [24] and the Prevetiver Health care on Individual Level framework [25]. Both frameworks emphasize the iterative exploration and analysis of the self-tracking data to gain insight and drive behavioral changes.

One of the known challenges in sleep tracking is how to empower layperson users to make sense of their sleep data and to identify lifestyle or environmental factors that they can modify for better sleep [22]. In the field of health informatics, health data analytics could be divided into multiple levels according to their analytical capabilities [26]. Depending on its outcomes, health analytics can be descriptive, diagnostic, predictive, or prescriptive in nature [27]. At the lowest level lies *descriptive analytics*, which answers the question, *what has happened?* This level of analytics describes data *as is* without applying complex calculations and exploration. Common techniques at this level, such as standard reports and alerts, focus on categorizing, characterizing, aggregating, and classifying data to understand the past and current states. Existing sleep tracking analytics is mostly centered on this level, which aims to help users gain a *nice-to-know* validation of their subjective perception of sleep. At the second level is *diagnostic analytics*, which focuses on possible antecedents and answers the question, *why did it happen?* This level of analytics requires extensive exploration and directed analysis and inference based on existing data to identify the potential problems and their probable causes. At the third level lies the *predictive analytics*, which focuses on the possible consequences and answers the question, *what is likely to happen next?* Sleep tracking technologies at this level should be able to predict users' sleep quality in the near or far future by examining their historical self-tracking data, detecting patterns, and then leveraging the patterns to forecast. The highest level of analytics is *prescriptive analytics*, which answers the question, *what should be done about it?* It uses domain knowledge in medicine and health science in addition to data to generate recommendations for health interventions (eg, recommendations for better sleep).

Table 1 provides a mapping of the 4 levels of the analytics framework by Burke [26] for sleep tracking. The descriptive analytics would focus on answering questions such as "How many hours did I sleep last night?" "How many awakenings did I have last night?" and "What is the average deep sleep ratio during the past one month?" So far, sleep tracking has predominantly centered on such descriptive analytics, typically by visualizing data with charts and tables on a dashboard. This type of application could be meaningful in understanding users' current sleep patterns. However, with a flux of multiple models of sensor data, simple data visualization may miss important patterns that are not easily observable through visual inspection. As the complexity of sleep tracking data increases, it becomes necessary to examine the data in a more structured manner using advanced analytics. For example, diagnostic analytics could help answer questions such as "Was my sleep normal?" or "Why did I have so many awakenings last night?" Predictive and prescriptive analytics could answer questions such as "How will my sleep quality be in five years if I keep going to bed at 2:00 am?" or "Will I sleep better tonight if I work out 10 minutes longer in the morning?" To achieve advanced analytics, it is necessary to combine different streams of contextual information into a sleep analysis. Although many consumers' sleep tracking systems support the simultaneous collection of multiple streams of contextual information, these data are often visualized separately and rarely integrated with sleep analysis. Currently, there is a paucity of advanced analytics in sleep tracking research.

Knowledge discovery is the process of finding meaningful patterns from data, and data mining is a central step within a knowledge discovery process. Popular data mining techniques, such as association rules mining and anomaly detection, are widely used in many application domains to detect hidden patterns in large data sets [26]. In this paper, we present a scoping review on the application of knowledge discovery methods in sleep tracking. Such a review is useful for technical researchers interested in applying a wide range of machine learning and data mining techniques to the personal health domain as well as for sleep scientists who want to leverage the latest wearable technology combined with a data-driven approach for personalized and nonpharmaceutical interventions. Previous reviews on consumer sleep tracking technologies have dominantly focused on the utility and validity of these devices, especially in terms of their strengths and limitations relative to more widely accepted devices [4,28,29]. To the best of our knowledge, this scoping review is the first to focus on the advanced data analytics related to sleep tracking. An advanced data-driven approach has the potential to discover meaningful patterns or hidden correlations that could be used to guide behavioral change for better sleep. On the basis of the scoping review, we highlight the research opportunities for data-driven sleep computing.

Table 1. Level of analytics and its mapping to sleep tracking.

Analytics level	Questions answered	Mapping to sleep tracking
Descriptive	What has happened?	"How many hours did I sleep last night?"
Diagnostic	Why did it happen?	"Why did I have so many awakenings last night?"
Predictive	What is likely to happen next?	"How will my sleep quality be in five years if I keep going to bed at 2:00 am?"
Prescriptive	What should be done about it?	"Will I sleep better tonight if I work out 10 minutes longer in the morning?"

Methods

Research Questions

This scoping review was guided by the problem-constraints-system framework, which is widely used for conducting a literature review in computer science. The problem-constraints-system framework focuses on 3 different aspects of a research topic in computing research: a specific problem of interest (P); systems, applications, or algorithms (S) for tackling the problem; and constraints (C). After iterative brainstorming, we proposed the following 4 research questions (RQs) to anchor the entire review process:

- RQ1: What is the general research trend of knowledge discovery in sleep tracking?
- RQ2: What sleep quality metrics and contextual factors were considered, and how were they measured?
- RQ3: What knowledge discovery methods or algorithms were applied? What knowledge was discovered?
- RQ4: What challenges exist? What are the opportunities for future research?

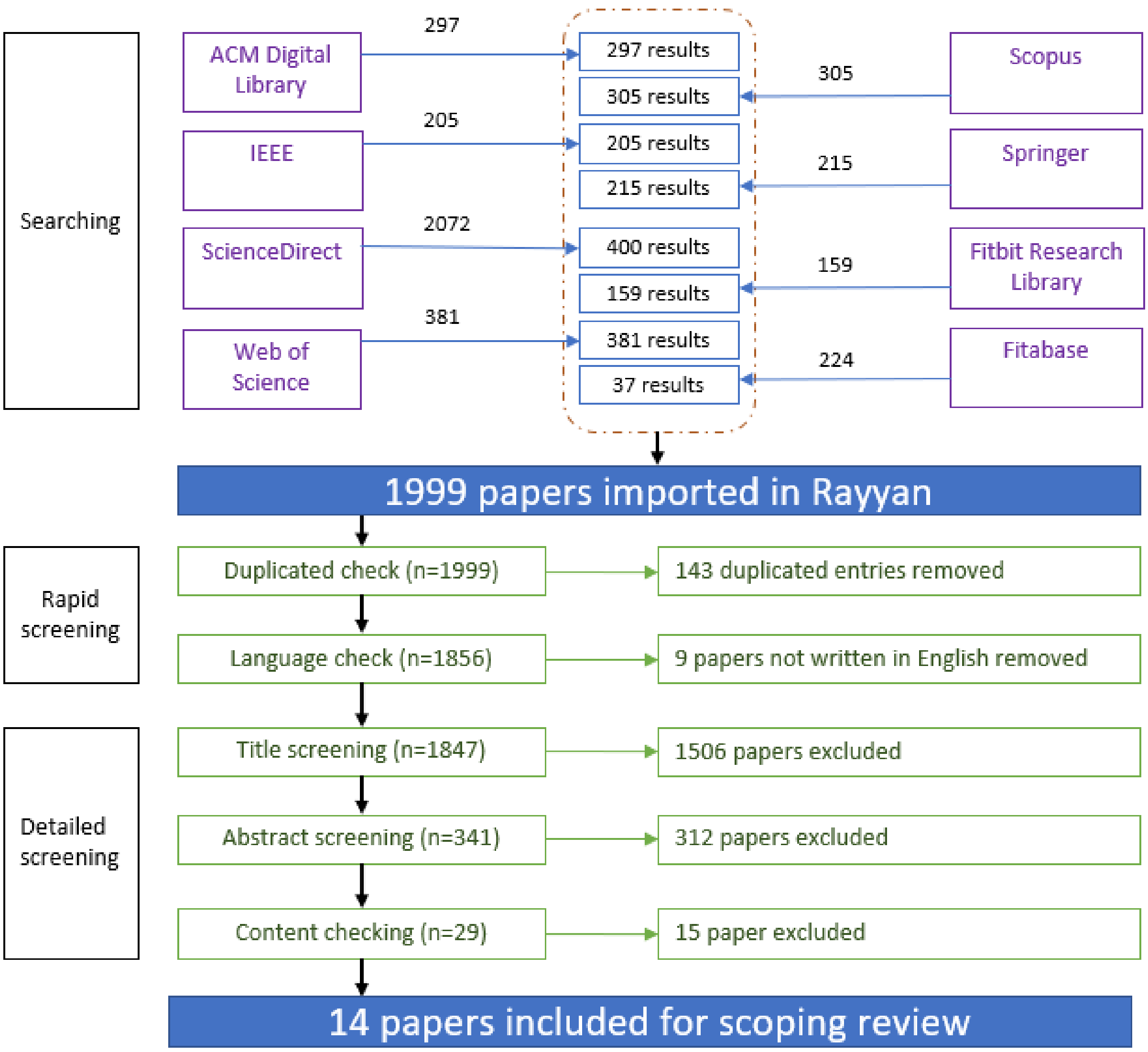
Search Strategy and Query String

In this review, we focused on the application of data mining to identify the relationships between sleep and contextual factors with consumer wearable devices. Therefore, search results must contain all 4 pertinent aspects: sleep metrics, contextual factors, wearable devices, and knowledge discovery. Automated searches were conducted in a number of databases, including the Web of Science, Scopus, ACM Digital Library, IEEE Xplore, ScienceDirect, Springer, Fitbit Research Library, and Fitabase. As our review was centered on the computing aspect rather than the clinical aspect of sleep tracking, we used databases that focus more on engineering and computer science publications, including the ACM Digital Library, IEEE Xplore, ScienceDirect, and Springer. PubMed was not used because we were not interested in cohort studies or medical assessments. We also included gray literature such as the Fitbit Research Library and Fitabase because of their relevance to our topics of interest. The query strings were slightly different for each database but always contained 4 main keyword combinations: "sleep," "lifestyle" AND "contextual," "data mining" OR "knowledge discovery," and "wearable device." Synonyms words and words of related concepts (eg, "self-tracking," "sensors," "machine learning," "correlation," and "statistics") were also listed to avoid missing out on related publications. The search strategy varied for each search engine, as each of them had different rules and options. The query string was truncated in some databases (eg, Web of Science, ScienceDirect, and IEEE Xplore) either because long query strings resulted in many irrelevant entries or because of word limit. We included several types of publications, including journal articles, conference proceeding papers, workshop presentations, patents, and gray literature, to gain a broad scope of the research topic. Editorial articles, theses, and dissertations were also excluded.

Study Selection

The study selection process is shown in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart in Figure 1. We applied the following exclusion criteria to filter out irrelevant papers retrieved from the databases: (1) studies not related to human sleep (eg, animal studies and sleep mode of sensor system), (2) clinical studies aiming at treating sleep disorders, (3) hypothesis-driven laboratory-based studies using invasive sleep tests (eg, polysomnography), (4) validation studies focusing on comparing wearable devices with medical devices, (5) studies focusing on estimating sleep architecture based on concurrent physiological signals, and (6) papers not written in English. We retrieved the returned entries from each database and imported them into the web-based review management software Rayyan (Rayyan). All databases provided tools to export the returned entries in a single file, except for the Fitbit Research Library and Fitabase. We exported only 400 of the 2072 papers from the ScienceDirect because the rest of the papers were not relevant to our review topic. For the ScienceDirect database, the returned items were first listed using the "relevant order" tool provided by the database. We used an a prior condition to include the first 400 items. For the remaining items, we separated them into groups of 200 items and randomly checked 50 items in each group. We found that from the 400th item onward, the studies were off the topic according to our exclusion criteria. Thus, we eliminated these items because they did not help address our central RQs. For example, the keyword "contextual" led to the retrieval of papers in civil engineering on checking bedroom quality with air condition, ambient light, and noise exposure. Those were excluded based on the exclusion criterion 1. The keyword "wearable device" led to the retrieval of papers in electrical and mechanical engineering on hardware design for sensors, batteries, and ergonomic design. Those were excluded based on the exclusion criterion 4. For the Fitbit Research Library, 2 papers were missing because they were either retracted or no longer available. Entries in Fitabase were screened for duplicity before manually adding to Rayyan because Fitabase does not support automatic export. As many of the Fitabase entries were duplicates of publications from other databases, only 37 papers were added to Rayyan.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of article screening process.



In total, 1999 papers were imported to Rayyan. Screening for duplication and written language was performed automatically in Rayyan. After the rapid screening step, the first author performed title and abstract screening and suggested the inclusion of 341 papers. The second author repeated the title and abstract screening of the 341 papers while paying special attention to the exclusion criteria 3 and 5. Any conflicts were resolved through discussion with the first author. After the screening and eligibility checks, 29 papers remained. Both authors read the full text of the 29 papers in detail and finally selected 14 papers for the scoping review.

Results

General Research Trend (RQ1)

Ubiquitous and personal sleep tracking has become an active research area since approximately 2011, when a few pioneer studies were published in the human-computer interaction community [17,18,30]. Since then, a large number of studies on sleep tracking have been published, but most of them are dominantly centered on the validation of existing sleep tracking devices or systems [5-10], on the development of new sleep staging algorithms [11-15], or on the investigation into users' experience with the technologies [3,21-23]. In contrast, studies that focus on knowledge discovery in sleep tracking data are scarce and ad hoc. The screening process identified only 14 relevant studies, of which more than half (8/14, 57%) were conducted in the United States, followed by Japan (3/14, 21%). The other publications were from China, Korea, Finland, and Australia. Only 5 (36%) of the 14 publications were journal articles, and the rest were conference proceeding papers. Chronologically, studies by Jayarajah et al [31] and Gelman and Hill [32] were one of the earliest studies in this field. The authors developed a binary tree model to predict good and poor sleep based on app use activity and social time during the day. Since 2015, the topic of knowledge discovery in sleep tracking has begun to attract more attention along with the advances in consumer sleep tracking technologies. As a result, the number of publications has increased slightly in the subsequent years, but the total amount is still limited.

A common objective of these studies was to help users gain insights into how their sleep quality was associated with other aspects of their daily lives. Some of the specific motivations are as follows:

- Identify aspects of daily life that demonstrate significant associations with personal sleep quality from self-tracking data [16,20,33]
- Highlight the potential for sleep metrics from wearable devices to provide novel insights into data generated from a large cohort [34-36]
- Detect aberrant sleep patterns or typical events during sleep by considering individuals' sleep baselines [37,38]
- Guide users in designing self-experiments to identify personal modifiable lifestyle factors for better sleep health [20]
- Develop a recommender system that provides both general and personalized recommendations for better sleep health [20]

The studies reviewed in this paper demonstrated a tendency to analyze sleep tracking data along with a flux of contextual factors. These factors were used as independent variables for predicting sleep quality and, to a lesser degree, for identifying antecedent events that affect sleep quality. In the scheme of traditional sleep science studies, only a limited number of independent variables were considered, and the confounding effect of noninterested factors needed to be controlled through a rigid experiment design. In comparison, data collection experiments in ubiquitous sleep tracking studies are often conducted in a naturalistic environment, making it challenging to control for confounding factors. Therefore, advanced data analysis techniques are required to control the effects of confounding factors during the analysis. Another line of research effort is the personalized detection of aberrant sleep. Although sleep quality assessment may sound straightforward using clinical standards [39], it remains challenging if personal differences in sleep needs should be taken into consideration. Studies in this direction are limited, and we identified only 2 relevant studies [37,40].

Quantification and Measurement of Sleep and Contextual Factors (RQ2)

Human sleep can be quantified along multiple dimensions, such as sleep duration, sleep continuity, sleep timing, and subjective perception of the sleep event [41]. We found that not all studies used the same set of metrics to characterize sleep quality. Even the same sleep metric may be capsulated in different terminologies across studies. To facilitate cross-study comparisons, we mapped the sleep metrics in each study to standard clinical terms, whenever possible. Original sleep metrics that have no corresponding clinical terms were simply left as they were. As presented in Table 2, the most used sleep metrics among the reviewed studies were subjective sleep quality (4 studies), sleep efficiency (SE, 4 studies), sleep onset latency (SOL; 4 studies), and time at lights off (3 studies). Ratio parameters such as deep sleep ratio and rapid eye movement ratio were not used in any of the studies. Moreover, the cutoff between good and poor sleep varied from study to study. A few studies have adopted clinical cutoffs, particularly for the Pittsburgh Sleep Quality Index (5) [31], SE (85%) [42], wake after sleep onset (30 minutes) [16], and total sleep time (7-9 hours) [16]. The rest chose to use heuristic cutoffs that did not comply with the clinical guidelines.

Table 2. Sleep quality metrics, measurement methods, and cutoff of good or poor sleep.

Clinical term, original term, and measurement method	Cutoff
Subjective sleep quality	
Pittsburgh Sleep Quality Index [31,38]	Good: ≤5 and poor: >5 [31]; good: ≤7 and poor: >7 [38]
Sleep rating (SleepAsAndroid app) [20]	N/A[a]
Leeds Sleep Evaluation Questionnaire [19]	N/A
Sleep rating (SleepApp) [19]	N/A
SE[b]	
SE (Fitbit) [33]	Good: ≥95% and poor: <95%
Efficiency (MS Band) [43]	N/A
SE (Polar) [40]	Good: >mean (SD) and poor: <mean (SD)
SE (Garmin) [42]	Good: ≥85% and poor: <85%
SOL[c] (minutes)	
Minutes to fall asleep (Fitbit) [16]	N/A
SOL (SleepAsAndroid app) [20]	N/A
Sleep latency (Garmin) [42]	Good: ≤15, average: 15-30, and poor: >30
Time to fall asleep (MS Band) [43]	N/A
Time at lights off	
Bedtime (Fitbit) [35]	Normal bedtime: median of a participant's bedtimes; deviation categories were 1-30 minutes, 30-60 minutes, 1-2 hours, 2-3 hours, and ≥3 hours
Bedtime (Fitbit) [37]	N/A
Bedtime as estimated by the time of the last network signal [36]	N/A
Wake after sleep onset (minutes)	
Awake minutes (Garmin) [42]	Good: ≤20 and poor: >20
Minutes awake (Fitbit) [16]	Good: ≤30 and poor: >30
Number of awakenings >5 minutes	
Awakenings >5 minutes (Garmin) [42]	Good: ≤1 and poor: >1
Number of wakeups (MS Band) [43]	N/A
Time in bed (minutes)	
Time in bed (MS Band) [34]	N/A
Total sleep time (minutes)	
Minutes asleep (Fitbit) [16]	Good: 420-540 and poor: <420 or >540
Original metrics (no corresponding medical term)	
Sleep ratio=the ratio of the sleep minutes with normal heart rate versus total sleep time (Fitbit) [44]	Good: >90%, normal: 60%-90%, and bad: <60%
Number of awakenings per hour [20]	N/A
Awakening count including restlessness (Fitbit) [16]	N/A
Permutation entropy of Fitbit measured sleep state time series [37]	N/A

[a]N/A: not applicable.

[b]SE: sleep efficiency.

[c]SOL: sleep onset latency.

Sleep quality metrics were measured using 3 methods: questionnaire based, app based, and wearable based. The Pittsburgh Sleep Quality Index was the most widely used questionnaire to measure the subjective perception of sleep duration, continuity, efficiency, and satisfaction [31,38]. The SleepAsAndroid app was used to collect sleep data based on user movement patterns in the study by Daskalova et al [20], and the SleepApp was used to collect users' subjective sleep quality ratings in the study by Ravichandran et al [19]. Although sleep tracking apps such as SleepAsAndroid were widely downloaded and used by millions of users, their ability to distinguish between quiet awakenings, deep sleep, or empty beds was still limited [20]. In a related vein, studies by Jayarajah et al [31] and Faust et al [36] have defined sleep time as the longest period during which there was no activity on users' smartphones. However, the authors acknowledged that this method was not reliable because not everyone had the habit of using smartphones until they fell asleep. Most studies (9/14, 64%) used commercial wearable devices such as Fitbit [16,33,35,37,44], Microsoft Band (MS Band) [34,43], Garmin [42], and Polar [40]. Although all 3 methods are noninvasive, easy to use, and allow longitudinal collection of sleep data, each method has some limitations. The questionnaire-based method is subject to memory recall bias. Sleep data collected by sleep tracking apps and wearable sleep trackers provide an objective description of sleep quality and sleep structure. However, they may also be prone to measurement errors because of hardware and software limitations [7,28]. They also require users to place the smartphone nearby or to wear the device continuously, which may cause discomfort during long-term use. Moreover, despite being small and convenient, consumer wearable trackers cannot provide hypnogram information that is as detailed as medical devices.

Along with sleep quality metrics, researchers have considered a wide range of contextual factors in their studies. As presented in Table 3, the contextual factors of interest include the sleep environment [16,20,42], daily activities [16,19,20,31,36,37,40,42,43], physiological states [16,35,36,43,44], and mental states [16,19]. It was a common practice to collect demographic information (eg, age, gender, and BMI) and medical history using questionnaires at the beginning of a sleep tracking study [16,35,36,43,44]. Other contextual factors were recorded during the data collection experiments. Researchers have been curious about how web activities of university students could be coupled with their sleep quality [31,34,36]. These studies developed their own tools to track users' web-based behavior and linked them to sleep quality metrics. Using consumer wearable trackers such as Fitbit, MS Band, and Garmin, researchers were able to expand their list of factors to include, for example, bedtime, steps, distance, hours of exercise, and calories burned [16,19,20,31,36,37,40,42,43]. Biosignals such as heart rate during sleep and daytime activities were included in the studies by Faust et al [35], Farajtabar et al [43], and Choi et al [44]. Several studies have also manually collected input features such as coffee, alcohol, mood, and stress [16,19,20]. These factors may have a significant effect on the circadian cycle of hormone secretion and thus may provide useful information for sleep quality prediction. However, collecting these data are nontrivial, as users tend to forget to log the data on a daily basis. How to collect these data more efficiently and how to reduce the risk of missing data remain challenging.

Table 3. Contextual factors and measurement methods.

Category and contextual factor	Data collection method
Physiological factors	
Age	Self-report [43,44]
Sex	Self-report [36,43]
Body weight	Fitbit [16] and self-report [43]
BMI	Self-report [44]
Body temperature	Diary [16]
Heart rate	Fitbit [35,44] and MS Band [43]
Menstrual cycle	Diary [16]
Calorie in and out	Fitbit [16] and MS Band [43]
Activity calorie	Fitbit [16]
Psychological factors	
Stress	Diary [16]
Mood	Diary [16] and SleepApp [19]
Tiredness	Diary [16]
Dream	Diary [16]
Sleep quality the previous night	SleepAsAndroid [20] and MS Band [43]
Cognitive performance	Keystroke time [34] and click time [34]
Behavioral factors	
Steps	Fitbit [16,37] and MS Band [43]
Distance walked	Fitbit [37]
Active time	Fitbit [16,37]
Exercise	SleepAsAndroid [20], MS Band [43], Polar [40], Garmin [42], and SleepApp [19]
Coffee	Diary [16], SleepAsAndroid [20], and SleepApp [19]
Alcohol	Diary [16], SleepAsAndroid [20], and SleepApp [19]
Tobacco	Self-report [44] and SleepApp [19]
Electronic device use	App use time (total and different app categories) [31], diary [16], Bing search logs [43], SleepApp [19], and campus network [36]
Nap	Diary [16], SleepAsAndroid [20], and SleepApp [19]
Location	Campus Wi-Fi [31] and Cortana [43]
Social activity	GruMon (location estimation based on Wi-Fi signals) [31], diary [16], and Twitter [43]
Mealtime	Diary [16], smartphone camera [42], SleepApp [19], and campus smart card [36]
Waketime	SleepAsAndroid [20]
Bedtime	Fitbit [37], IoT[a] sensor [42], and SleepApp [19]
Environmental factors	
Ambient temperature	Diary [16] and IoT sensor [42]
Ambient humidity	Diary [16] and IoT sensor [42]
Ambient light	Diary [16] and SleepAsAndroid [20]
Ambient noise	SleepAsAndroid [20]
Day of week	MS Band [43]

[a]IoT: internet of things.

We found that 13 (93%) of the 14 reviewed studies conducted their own data collection experiments [16,31,33,37,38,43,44] in free-living conditions. In these studies, sleep data were recorded in participants' usual sleep environments (eg, homes and caregiving facilities), whereas contextual factors were recorded while participants were at schools, universities, workplaces, or sports centers. In contrast, only 1 study used an existing data set [35], which can be accessed over the web [45]. Data sharing is not yet a common practice in the field, and the number of public data sets is limited.

Knowledge Discovery in Sleep Tracking (RQ3)

Data Preprocessing

All the reviewed studies used data sets that were collected in free-living environments. Data collection "in the wild" increases the ecological validity of the studies but at the sacrifice of data quality. The first step in the knowledge discovery process is to deal with missing, wrong, and duplicate data. Although there are many methods for data cleaning in the literature, the methods adopted in the reviewed publications are extremely simple and straightforward. Most of the studies (5/14, 36%) simply excluded records that contain missing values (eg, sleep logs with missing fields or sleep duration of 0 minutes) [16,19,37] or excluded users who do not contribute sufficient data (eg, fewer than 30 sleep records) [35,36]. One study excluded a certain data source (eg, search engine interactions originating from mobile devices) to avoid causing distortion to the data distribution [34].

Similarly, data out of logical ranges were removed. Users aged <10 years or >100 years, with weight <22.7 kg or >112.5 kg, or with height <127 cm or >457.2 cm were excluded in the study by Farajtabar et al [43]. Extremely short (<0.5 or 4 hours) and long (>12 hours) sleep records were removed in the studies by Althoff et al [34] and Farajtabar et al [43]. Sleep entries with bedtime between 7:00 AM and 7:00 PM were removed in the study by Liang et al [33]. Exercise time is another criterion for filtering out potentially erroneous data records. For example, exercise events <5 or >180 minutes were removed in the study by Farajtabar et al [43]. Exercises with calorie consumption per hour <50 or >2000 calories or with a duration of <10 minutes were excluded in the study by Liu et al [40]. In addition, data records with steps <1000 and those with a sedentary time of 0 minute were removed in the study by Liang et al [16].

Time stamps are another focus in data preprocessing. The data collection in the study by Ravichandran et al [19] primarily relied on users' manual input in SleepApp and thus had a higher risk of human errors. Consequently, all logs with aberrant timestamps were removed. In addition, 12 hours were added to or subtracted from the recorded bedtime or wake-up time where the users might have forgotten to toggle the AM and PM switch on the app. Dealing with timestamps involves not only data cleaning but also temporal matching among multiple data sources [42] as well as data type conversion (eg, 18:30 to 1830) [16,37].

Other types of data preprocessing included selecting users with larger variations in sleep and exercise [38], removing redundant entries [19], and resampling the raw data (eg, the photoplethysmography-derived heart rate time series was aggregated every 3 minutes to achieve a constant sampling rate [38]).

Some knowledge discovery processes that rely on machine learning or data mining techniques may require a feature engineering process instead of directly using the cleaned data as input. For example, several studies have involved the construction of secondary features from the cleaned data [37,38,40]. Features were normalized to have a mean and SD equal to 0 and 1 [42] or normalized over another feature (eg, exercise intensity features were normalized by dividing the basal metabolic rate [38]). Dimension reduction (eg, principal component analysis) was applied to reduce the number of input features to avoid the adverse effect of the "curse of dimension" [31].

Data Mining

The selection of the data mining method depends on the purpose of the studies and, to a lesser degree, on the size of the available data set. Table 4 provides a summary of the data mining methods and the specific techniques or algorithms used in the reviewed studies. We also listed the independent variables (or input) and dependent variables (or output) of the constructed models. Correlation analysis, regression analysis, and rule induction are the most used methods for finding meaningful associations between contextual factors and sleep quality metrics. In total, 3 correlation analysis techniques, Pearson correlation, Spearman correlation, and repeated measure correlation, were applied to examine the strength of the pairwise linear relationships between sleep and contextual factors [16,19,20]. Similarly, various regression analysis methods have been used, ranging from simple linear regression to linear mixed effects regression to piecewise fixed effects regression [34,35,43]. Least square estimation was the most popular technique for parameter estimation in regression analysis and was used in the studies by Althoff et al [34] and Farajtabar et al [43]. The study by Faust et al [35] provided no information but is highly likely to use the same technique. It is worth noting that the Pearson correlation coefficient is equivalent to the standardized slope of a simple linear regression line.

Table 4. Summary of the data mining methods used in the reviewed studies.

Data mining method and techniques or algorithms	Data size	Independent variable	Dependent variable
Correlation analysis			
Pearson correlation [20]	Preliminary study: 24 users over 20 days; final study: 19 users over 21 days	TST[a] and contextual factors	SOL[b], NAWK[c], and sleep rating
Spearman correlation [16]	12 users over 2 weeks	Contextual factors	TST, WASO[d], NAWK, SOL, and SE[e]
Repeated measure correlation [19]	10 users over 2 weeks	Bedtime, TIB[f], and contextual factors	SE, SOL, NAWK, restlessness, TIB, and LSEQ[g]
Regression analysis			
Piecewise fixed effects regression [34]	31,793 users over 18 months; all American users	Time of day, time after waking up, and sleep duration	Cognitive performance
Simple linear regression [43]	Approximately 20,000 users over 4 months	Contextual factors	SOL, NAWK, and SE
Linear mixed effects model [35]	557 users over 1 year	Bedtime regularity	Resting heart rate
Rule induction			
A priori algorithm [33]	1 user over 180 days; 4 users over 2 weeks	Contextual factors	SE
Learn from Examples using Rough Sets [44]	280 users over 1 month; only the data of males were used	Contextual factors	Sleep ratio
Event mining (+causal inference) [42]	1 user over 800 days	Contextual factors	SOL, WASO, NAWK, and SE
Causal inference			
Stratified propensity score analysis [43]	Approximately 20,000 users over 4 months	Contextual factors	SOL, NAWK, and SE
Bayesian network analysis [36]	5200 users over 6 months	Contextual factors and bedtime	Contextual factors and bedtime
Time series analysis			
Anomaly detection [37]	1 user over 35 days	Fitbit measured intraday time series, TST, WASO, NAWK, and bedtime	Permutation entropy of sleep time series
SAX[h]-based motif matching and principle optimization [38]	100 users over 10 weeks	Heart rate time series data	PSQI[i]
Statistical test			
Unpaired 2-samples Wilcoxon test [40]	271 users over 8 months	Contextual factors	Statistical differences between good and poor sleep
Decision tree			
J4.8 Classifier [31]	400 users over 15 months	Contextual factors	PSQI

[a]TST: total sleep time.

[b]SOL: sleep onset latency.

[c]NAWK: number of awakenings.

[d]WASO: wake after sleep onset.

[e]SE: sleep efficiency.

[f]TIB: time in bed.

[g]LSEQ: Leeds Sleep Evaluation Questionnaire.

[h]SAX: Symbolic Aggregate Approximation.

[i]PSQI: Pittsburgh Sleep Quality Index.

Although correlation analysis and regression analysis (except piecewise constant approximation) capture the relationships between sleep and contextual factors in the entire sampling range, the piecewise regression in the study by Althoff et al [34] shared some resemblance with rule induction methods that capture the relationship between sleep and contextual factors within a constrained range. However, in contrast to piecewise regression, which quantifies the covariance between 2 variables in each partitioned segment, rule induction methods focus on extracting frequent patterns in the data sets that characterize the co-occurrence of 2 variables when their values fall into the corresponding ranges specified in a rule. Moreover, rule induction methods are usually robust to missing data. As the rule induction methods were originally developed to analyze categorical data, numerical data need to be converted to categorical data through a discretization step before rule induction methods can be applied. There are 4 discretization methods used in the reviewed studies: equal size discretization [33,44], equal frequency discretization [33], k-means clustering discretization [33], and discretization with heuristically defined cutoffs [42]. Association rules mining is a popular rule induction method that has been widely used in traditional medical and health informatics applications; however, it has only been used in 1 study among all the studies we reviewed [33]. In that study, the a priori algorithm was applied for rule induction. The quality of the induced association rules was validated by higher local correlation coefficients (ie, the Pearson correlation coefficient when the variables fall into the ranges specified in a rule) than the global correlation coefficients (ie, the Pearson correlation coefficient between 2 variables within the entire sampling range). To better handle the potential inconsistency (eg, conflicting records) in the data set, Rock-Hyun et al [44] applied another rule induction algorithm named learning from examples using rough sets. The global covering algorithm computes the lower and upper approximations of all the target sleep quality metrics (eg, sleep ratio="good") if the input data set contains conflicting records. The quality of the induced rules was assessed based on the predictive accuracy of the target sleep metrics. Although association rules mining and learning from examples using rough sets capture only the parallel co-occurrence of 2 items (ie, when the values of 2 variables fall into the corresponding ranges specified by a rule), event mining can also capture the co-occurrence of 2 items with a time lag (ie, the temporal sequence when the 2 items occur) [46]. To a certain degree, event mining resembles sequential pattern mining [47]; however, this characteristic was not used in the study by Upadhyay et al [42].

Causal inference is a powerful approach to reduce potential bias in the identified relationships between sleep and contextual factors because of observed confounding factors. This method is likely to outperform simple correlation analysis or rule induction–based methods. In the study by Farajtabar et al [43], stratified propensity score analysis was performed to isolate the effects of potential confounding factors. A similar technique was used in the study by Upadhyay [42] to enhance the quality of the induced rules by accommodating confounding factors. In addition, the Bayesian network was applied to explore the relationship between sleep schedules and behavioral factors [36].

Statistical tests and decision trees were also used in the existing literature, but only in 1 study each. The unpaired 2-samples Wilcoxon test was applied to identify significant differences in a set of selected contextual factors between good and poor sleepers [40]. Despite its simplicity, this method does not generate quantitative relationships between the contextual factors and sleep quality. In contrast, decision trees were used in the study by Jayarajah et al [31] to predict sleep quality using contextual factors as input features.

In addition to the abovementioned methods, time series analysis was also used but only in 2 studies [37,38]. Dimension reduction and anomaly detection were combined to identify aberrant sleep recordings while counting in personal sleep baseline in the study by Liang et al [37]. Feng and Narayanan [38] introduced a method to discover motifs in heart rate time series, which were signal patterns that appeared most frequently during sleep [38]. They then used these motifs as features to predict sleep quality. In contrast, although the study by Upadhyay et al [42] adopted a streaming data perspective, it did not incorporate any formal time series analysis technique [39].

Knowledge Discovered

The knowledge discovery process in the reviewed studies identified interesting associations both at the cohort level and the individual level. First, significant associations were found among the sleep quality metrics. Late bedtime was associated with a higher permutation entropy of the Fitbit measured sleep time series (indicating a higher chance of aberrancy) [37]. Bedtime deviation was correlated to longer SOL [19]. In addition, sleep duration was positively associated with subjective sleep satisfaction [19,20].

Regarding the relationship between sleep quality and contextual factors, exercise was the most identified association factor [33,40], but the relationship between exercise and sleep was complex [40,43]. First, not all exercise features have a predictive power of sleep quality. For example, Liu et al [40] found that exercise duration, relative calories consumption, and exercise timing could be used as predictors of sleep quality, but exercise intensity was not significantly associated with sleep quality. Second, exercise may be positively associated with some sleep quality metrics but negatively associated with others. For example, exercise before bed may be linked to shorter SOL and higher SE [43], and exercise seems to improve SOL the most among all the sleep quality metrics [42]. However, the results diverged as different types of exercises were considered. Taking more steps meant fewer awakenings, whereas running and burning more calories were correlated with more awakenings [43]. Moreover, longer exercise duration (eg, >100 minutes) may be associated with good sleep for some users but poor sleep for others [40]. Furthermore, confounding factors may modulate the relationship between sleep and exercise. With causal inference, it was found that pleasant ambient temperature at bedtime significantly strengthened the relationship between exercise and sleep, whereas having a poor sleep the previous night detracted from the beneficial effects of exercise [42].

Digital device use is another important factor that correlates with sleep quality and sleep schedule. No web searches before bed correlated with shorter SOL and higher SE, whereas web searches before bed correlated with more awakenings [43]. The association between app use and sleep quality was modulated by the frequency and timing of use [31]. In particular, low social app use was associated with good sleep, and more app use correlated with good sleep if used for >4 hours before sleep. Reading and gaming app use within 1 hour before bedtime correlated with poor sleep. A strong connection between internet surfing habits and bedtime was identified by Guo et al [36]. Video lovers tended to go to bed later than game fans. Going to bed late, in turn, has negative consequences. Deviations from usual bedtime may result in a higher resting heart rate during sleep [35]. Students who went to bed late were more likely to have a poor academic performance [36]. Late bedtime (in relation to one's circadian cycle) reduced cognitive performance the next day, whereas early bedtime did not have the same negative effect [34]. In contrast, having a sufficient sleep was important for maintaining normal cognitive performance [34].

As expected, caffeine and alcohol consumption were significant association factors. Consuming caffeine late in the day was a universal negative factor in subjective sleep ratings [20]. Alcohol consumption was positively correlated with SE and wake-up freshness and was negatively correlated with wake after sleep onset and number of awakenings for some users but not all users [16].

Not just personal activities or substance consumption associated with sleep, but places visited before bedtime and social life also play a role. One study found that students who spent more time with friends had better sleep quality than those who stayed alone on campus most of the time [31]. In addition, if students spent most of their time outside campus, good quality sleep was found for those who spent <15% of their time being alone. Another study tracked users' location during the day and stated that users took longer to fall asleep if they visited food-related or bank-related locations close to bedtime [43]. In addition, sleep quality may vary depending on the environment in which sleep takes place. The most common relationship identified for many users by Daskalova et al [20] was the pair of noisiness and number of awakenings [16]. Temperature was positively associated with all sleep metrics except for SOL [42].

Challenges and Opportunities (RQ4)

Knowledge discovery in sleep tracking is the process of extracting nonobvious hidden knowledge from self-tracking sleep data and other available contextual information. Preparing a data set of a sufficient size is the first step in this process. Almost all the reviewed studies (13/14, 93%) conducted original data collection experiments using noninvasive wearable and mobile sensors. The existing literature highlighted several challenges of data collection in sleep tracking. First, the absence of an objective, quantifiable, and universal definition of good sleep places a big challenge in annotating the collected sleep data [16,19]. Without a well-annotated data set, it is not feasible to apply supervised data mining techniques, and the absence of ground truth impedes the unbiased evaluation of the knowledge discovery process. Second, some contextual factors are considered difficult to quantify. These factors include digital device use, caffeine and alcohol consumption, and social interaction [16], to name but a few. The timing of data logging may also influence the results [20]. For example, users were advised to avoid using digital devices 2 hours before bedtime but had to log on to their smartphones to submit daily data at the end of the day. Third, existing passive sensing methods may have strong limitations [34]. For example, some studies (2/14, 14%) assume that users check their smartphone right before bedtime and immediately after waking up [31,36] and thus may miss out on users who have no such habits. Consumer wearables may have limited accuracy in measuring sleep stages and other factors [7], but the issue of data quality was not considered in the reviewed studies [35].

Moreover, interpreting the knowledge discovery outcome is not always straightforward. Correlation analysis essentially captures the covariance of 2 variables. Users with regular sleep and daily life routines may end up with no significant correlations found because of the lack of variability in their data. However, users may misinterpret this as having no relationship [16]. Rule induction methods usually generate a large number of rules, but not all of them are useful. Long rules with too many factors in the antecedent, despite of being explainable, provide no actionable insights because of their complexity (eg, "IF 17.85 < BMI < 25.21 AND Smoking is Yes AND 61.81 < Normal_Avg_HR < 79.0 AND 0 < Normal_Awake <19.0 AND 1.5 < Normal_Really_Awake < 24.0 AND 1.5 < High_Asleep < 606.5 AND 1.5 < High_Awake < 155.0 AND 0.5 < High_Really_Awake < 175.5 THEN Sleep Quality Status is 'Bad' with support 8") [44]. In contrast, short association rules may be more comprehensible (eg, "minutes very active={33; 38}=> good sleep or steps={18,658; 20,263}=> good sleep" [33]). However, heuristic discretization without a semantic meaning may impede understandability [20].

Despite the challenges, the reviewed studies highlighted several opportunities for future research. In total, 3 studies suggested considering more contextual factors in addition to the ones already studied, such as emotion, diet, productivity, and chronotypes [31,34,36,42]. Acknowledging that the correlations at a cohort level may be weak [31], argues for an individual-centric approach to identifying the most important contextual factors for each user. Along the same line [43], it was pointed out that building predictive models within similar user groups is more practical. They proposed a hierarchical modeling scheme with a top layer containing population parameters and lower layers personalized to individual users. Similar user profiling and segmented modeling proposals were presented in the studies by Liang et al [33] and Farajtabar et al [43].

Discussion

Principal Findings

Sleep tracking using consumer wearable devices and mobile apps has attracted remarkable attention from the research community. However, sleep tracking studies have focused on developing sleep tracking technologies for accurately measuring sleep per se, and little attention has been directed to the extraction of patterns and insights from these data. To the best of our knowledge, this scoping review is the first to map the existing literature from a knowledge discovery perspective in sleep tracking.

Our analysis results showed that the number of publications on the topic of interest has slightly increased over the years but is still low, probably because the data-driven scheme has not been fully embraced in personal informatics. Nonetheless, we found that the existing literature covered all 4 levels of analytics, as presented in Table 1. Most of the 14 studies that we reviewed applied simple correlation analysis, regression analysis, and rule induction methods to discover the associations between sleep and other aspects of life. Although most consumer sleep tracking technologies allow users to visually inspect their sleep data (which is descriptive in nature), the reviewed studies demonstrated the feasibility of diagnostic analysis with a flux of sleep and contextual data. Although correlation does not necessarily indicate causality, a combination of association analysis and causal inference—as was done in the study by Upadhyay et al [42]—may help users narrow down the scope of possible modifiable factors that are likely to affect their sleep quality. Machine learning and data mining techniques were only used in a few studies for anomaly detection (which is diagnostic) [37] or sleep quality prediction (which is predictive) [31,43]. In total, 2 studies developed computational models to generate personalized recommendations for better sleep and showed promise in prescriptive analysis of sleep tracking [20,42]. The most used sleep metrics among the reviewed studies were subjective sleep quality, SE, SOL, and time at lights off. Exercise, digital device use, places visited during the day and before bedtime, and sleep environment are the major factors that significantly correlate with various dimensions of sleep quality.

Taken together, there are a few key challenges that are relevant to the findings. On the one hand, it is nontrivial to collect high-quality data in naturalistic settings. Challenges include how to motivate users to overcome tracking fatigue, how to enhance the reliability of consumer wearables and apps, and how to quantify and automate the collection of contextual information to represent the current challenges surrounding the collection of sleep tracking data sets. On the other hand, how to extract hidden knowledge from data, how to accommodate commonness and individuality, and how to interpret data mining results are topics for future studies.

Nuance in Handling Within-Individual Variation

The selection of appropriate data mining methods relies on a correct understanding of the nature of the sleep tracking data set. Researchers often conduct longitudinal data collection experiments that involve the collection of multiple measurements of the same variables (eg, sleep quality, exercise, and ambient light) from each individual user. The data form a hierarchical or clustered structure when aggregated at the cohort level. Caution must be exercised when applying traditional analytic and modeling techniques developed for single-level data, as hierarchical data are likely to violate the assumption of independent errors of those techniques. In particular, although a hierarchical data set offers the benefit of a larger amount of data, the within-individual variation at the individual level needs to be addressed carefully through multilevel analysis and modeling.

In a multilevel analysis framework, the repeated measures are clustered within the level of an individual, and each individual is treated as a cluster unit. Depending on whether the analysis of one cluster involves pooling the data of other clusters, there are 3 approaches to analyzing a hierarchical data set: complete pooling, no pooling, and partial pooling. Complete pooling completely ignores the variation between individual users and treats all samples as being drawn from the same population. A dominant portion of the studies reviewed in this work [16,31,33,34,36,38,40,43,44] adopted this approach. Nonetheless, this approach is undesirable, as it violates the assumption of independence. The results could have been distorted when the between-individual variation was high. At the other end of the spectrum lies the no pooling approach, where the analysis of the relationship between sleep and contextual factors was performed only on the data of each individual user without considering data from other users. Daskalova et al [20] Upadhyay et al [42] embraced an N-of-1 design and correspondingly adopted the no pooling approach to analyze the collected data. At the surface, this approach is plausible for fully handling the within-individual variation. However, it bears the risk of overstating the variation between individual users because of potential overfitting when the number of samples from individual users is small. Partial pooling or multilevel modeling compromises between pooled and unpooled estimates, with the relative weights of pooling determined by the sample size of each individual user and the variation within and between individuals. Multilevel modeling automatically adjusts the degree of pooling with a "soft constraint," which ensures strong pooling for users with fewer records and weak pooling for users with abundant records in the data set [32]. We found that only Ravichandran et al [19] and Faust et al [35] used the multilevel modeling approach and explicitly considered the within-individual variation.

Most reviewed studies (8/14, 57%) seem to have relied on the undue assumption of an independent and identically distributed data set. As a result, some studies (2/14, 14%) found mixed or even conflicting results on the relationships between sleep and contextual factors at the cohort level [40,44] and, consequently, generated no valuable insights. Studies using an N-of-1 design are interesting exceptions. In these studies, analysis was conducted on each user's data, thus eliminating the effect at the cohort level. However, the robustness and generalizability of the findings are questionable. Even in studies with an N-of-1 design, it may still be helpful to partially pool some samples from the population to increase the reliability of the model parameter estimates. As such, Gelman et al [48] suggested always using multilevel modeling (ie, "random effects") as a rule of thumb, for example, linear mixed effect model over simple linear regression model and generalized linear mixed model trees [49] over the J4.8 classifier.

Limitations of the Study

There is room for improvement in several aspects of this study. First, because of the limitations inherent in scoping reviews, this study is exploratory and primarily qualitative in nature.

Limited by the review methodology, we were unable to generate a quantitative "summary of findings, " as required for systematic reviews or meta-analyses. Second, our method is nonstandard in a sense that we performed prescreening on the items identified in the ScienceDirect database before importing all entries into Rayyan. Although we made an effort to ensure that the removed items were not relevant, we cannot rule out the possibility of missing publications that should have been included. Third, we did not conduct critical appraisal on the quality of the selected papers or perform a risk of bias assessment, which may have led to potential bias in the selection and interpretation of the papers. Despite these limitations, this review provides a well-scoped summary of existing research and could lay the groundwork for future systematic reviews. The research gaps that we identified can be used to inform future research agendas.

Conclusions

This scoping review built an understanding of the scope and nature of existing literature on knowledge discovery in ubiquitous and personal sleep tracking. To the best of our knowledge, this is the first review that exclusively focused on the knowledge discovery aspect of self-tracking in the realm of sleep health. In total, 14 studies were included in the review based on the exclusion criteria. We found that the existing literature covered all 4 levels of the analytics framework in health informatics. However, half (7/14, 50%) of the studies have only applied simple correlation analysis and regression analysis, aiming to discover significant associations between sleep and available contextual information. Machine learning and data mining techniques have not yet been widely used, probably because of the lack of large and quality data sets. Exercise, digital device use, places visited during the day and before bedtime, and sleep environment were the most identified factors associated with sleep quality. We identified key challenges surrounding the collection of high-quality sleep tracking data sets with consumer-grade sensors and in naturalistic settings as well as the extraction of hidden knowledge that could be translated into actionable insights and personalized behavior interventions. We highlight that future research should develop data analytics techniques and prediction models that properly handle the within-individual variation and between-individual variation in sleep tracking data sets. We hope that this scoping review could lay the groundwork for future research on ubiquitous and personal sleep tracking.

Acknowledgments

This study was sponsored by a Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid for Early-Career Scientists (grant 21K17670). The funder had no role in the study design, data collection and analysis, or manuscript preparation.

Authors' Contributions

ZL contributed to study conception. ZL and NHH contributed to study design, analysis and interpretation of results, and manuscript drafting and revision. NHH collected the data. All authors have reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Grandner MA, Lujan MR, Ghani SB. Sleep-tracking technology in scientific research: looking to the future. Sleep 2021 May 14;44(5):zsab071 [FREE Full text] [doi: 10.1093/sleep/zsab071] [Medline: 33782687]
2. Choi YK, Demiris G, Lin SY, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
3. Liu W, Ploderer B, Hoang T. In bed with technology: challenges and opportunities for sleep tracking. 2015 Presented at: OzCHI '15: The Annual Meeting of the Australian Special Interest Group for Computer Human Interaction; Dec 7-10, 2015; Parkville, VIC, Australia. [doi: 10.1145/2838739.2838742]
4. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Rev Med Devices 2016 May;13(5):497-506. [doi: 10.1586/17434440.2016.1171708] [Medline: 27043070]
5. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 2021 Feb 12;44(2):zsaa170 [FREE Full text] [doi: 10.1093/sleep/zsaa170] [Medline: 32882005]
6. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291 [FREE Full text] [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
7. Liang Z, Chapa Martell MA. Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. J Healthc Inform Res 2018 Jun 20;2(1-2):152-178 [FREE Full text] [doi: 10.1007/s41666-018-0013-1] [Medline: 35415400]

8. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e13384 [FREE Full text] [doi: 10.2196/13384] [Medline: 31172956]
9. Liang Z, Nishimura T. Are wearable EEG devices more accurate than fitness wristbands for home sleep tracking? Comparison of consumer sleep trackers with clinical devices. 2017 Presented at: IEEE 6th Global Conference on Consumer Electronics (GCCE); Oct 24-27, 2017; Nagoya, Japan. [doi: 10.1109/gcce.2017.8229188]
10. Liang Z, Chapa-Martell MA. Combining numerical and visual approaches in validating sleep data quality of consumer wearable wristbands. 2019 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); Mar 11-15, 2019; Kyoto, Japan. [doi: 10.1109/percomw.2019.8730805]
11. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. Sleep 2020 Jul 13;43(7):zsaa045 [FREE Full text] [doi: 10.1093/sleep/zsaa045] [Medline: 32215550]
12. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep 2019 Dec 24;42(12):zsz180 [FREE Full text] [doi: 10.1093/sleep/zsz180] [Medline: 31579900]
13. Liang Z, Chapa-Martell MA. A multi-level classification approach for sleep stage prediction with processed data derived from consumer wearable activity trackers. Front Digit Health 2021 May 28;3:665946 [FREE Full text] [doi: 10.3389/fdgth.2021.665946] [Medline: 34713139]
14. Liang Z, Chapa-Martell MA. Achieving accurate ubiquitous sleep sensing with consumer wearable activity wristbands using multi-class imbalanced classification. 2019 Presented at: IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech); Aug 5-8, 2019; Fukuoka, Japan. [doi: 10.1109/dasc/picom/cbdcom/cyberscitech.2019.00143]
15. Sirithummarak P, Liang Z. Investigating the effect of feature distribution shift on the performance of sleep stage classification with consumer sleep trackers. 2021 Presented at: IEEE 10th Global Conference on Consumer Electronics (GCCE); Oct 12-15, 2021; Kyoto, Japan. [doi: 10.1109/gcce53005.2021.9621361]
16. Liang Z, Ploderer B, Liu W, Nagata Y, Bailey J, Kulik L, et al. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. Pers Ubiquitous Comput 2016 Sep 16;20(6):985-1000. [doi: 10.1007/s00779-016-0960-6]
17. Bauer JS, Consolvo S, Greenstein B, Schooler J, Wu E, Watson NF, et al. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX, USA. [doi: 10.1145/2207676.2208600]
18. Kay M, Choe EK, Shepherd J, Greenstein B, Watson N, Consolvo S, et al. Lullaby: a capture and access system for understanding the sleep environment. 2012 Presented at: Ubicomp '12: The 2012 ACM Conference on Ubiquitous Computing; Sep 5-8, 2012; Pittsburgh, PE, USA. [doi: 10.1145/2370216.2370253]
19. Ravichandran R, Patel SN, Kientz JA. SleepApp: providing contextualized and actionable sleep feedback. 2020 Presented at: PervasiveHealth '20: 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 18-20, 2020; Atlanta, GA, USA. [doi: 10.1145/3421937.3421996]
20. Daskalova N, Metaxa-Kakavouli D, Tran A, Nugent N, Boergers J, McGeary J, et al. SleepCoacher: a personalized automated self-experimentation system for sleep recommendations. 2016 Presented at: UIST '16: The 29th Annual ACM Symposium on User Interface Software and Technology; Oct 16-19, 2016; Tokyo, Japan. [doi: 10.1145/2984511.2984534]
21. Ploderer B, Rodgers S, Liang Z. What's keeping teens up at night? Reflecting on sleep and technology habits with teens. Pers Ubiquitous Comput 2022 Jan 28;27(2):249-270. [doi: 10.1007/s00779-021-01661-x]
22. Liang Z, Ploderer B. How does Fitbit measure brainwaves: a qualitative study into the credibility of sleep-tracking technologies. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-29 [FREE Full text] [doi: 10.1145/3380994]
23. Liang Z, Ploderer B. Sleep tracking in the real world: a qualitative study into barriers for improving sleep. 2016 Presented at: OzCHI '16: The 28th Australian Conference on Human-Computer Interaction; Nov 29-Dec 2, 2016; Launceston, Tasmania, Australia. [doi: 10.1145/3010915.3010988]
24. Epstein DA, An P, Fogarty J, Munson SA. A lived informatics model of personal informatics. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2804250]
25. Liang Z, Chapa-Martell M. Framing self-quantification for individual-level preventive health care. 2015 Presented at: HEALTHINF '15: 8th International Conference on Health Informatics; January 12-15, 2015; Lisbon, Portugal p. 336-343 URL: https://www.scitepress.org/Papers/2015/52025/52025.pdf [doi: 10.5220/0005202503360343]
26. Burke J. The world of health analytics. In: Kudyba SP, editor. Healthcare Informatics: Improving Efficiency through Technology, Analytics, and Management. Boca Raton, FL, USA: CRC Press; 2016:161-180.
27. Banerjee A, Bandyopadhyay T, Acharya P. Data analytics: hyped up aspirations or true potential? Vikalpa 2013 Oct 01;38(4):1-12. [doi: 10.1177/0256090920130401]

28. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557 [FREE Full text] [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
29. Van de Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200. [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
30. Choe EK, Consolvo S, Watson NF, Kientz JA. Opportunities for computing technologies to support healthy sleep behaviors. 2011 Presented at: CHI '11: the SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, Canada p. 3053-3062 URL: https://dl.acm.org/doi/10.1145/1978942.1979395 [doi: 10.1145/1978942.1979395]
31. Jayarajah K, Radhakrishnan M, Hoi SC, Misra A. Candy crushing your sleep. 2015 Presented at: UbiComp/ISWC '15: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2015 ACM International Symposium on Wearable Computers; September 7-11, 2015; Osaka, Japan p. 753-762 URL: https://dl.acm.org/doi/10.1145/2800835.2804393 [doi: 10.1145/2800835.2804393]
32. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge, UK: Cambridge University Press; 2006.
33. Liang Z, Chapa Martell MA, Nishimura T. Mining hidden correlations between sleep and lifestyle factors from quantified-self data. 2016 Sep Presented at: UbiComp '16: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12-16, 2016; Heidelberg, Germany p. 547-552 URL: https://dl.acm.org/doi/10.1145/2968219.2968319 [doi: 10.1145/2968219.2968319]
34. Althoff T, Horvitz EJ, White RW, Zeitzer J. Harnessing the web for population-scale physiological sensing: a case study of sleep and performance. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW '17; Apr 3-7, 2017; Perth, Australia p. 113-122 URL: https://dl.acm.org/doi/10.1145/3038912.3052637 [doi: 10.1145/3038912.3052637]
35. Faust L, Feldman K, Mattingly SM, Hachen D, Chawla NV. Deviations from normal bedtimes are associated with short-term increases in resting heart rate. NPJ Digit Med 2020 Mar 23;3:39 [FREE Full text] [doi: 10.1038/s41746-020-0250-6] [Medline: 32219180]
36. Guo T, Li L, Zhang D, Xia F. Surf or sleep?: understanding the influence of bedtime patterns on campus. SIGKDD Explor Newsl 2021 Dec 30;23(2):3-12 [FREE Full text] [doi: 10.1145/3510374.3510378]
37. Liang Z, Chapa-Martell MA, Nishimura T. A personalized approach for detecting unusual sleep from time series sleep-tracking data. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics. 2016 Presented at: ICHI '16; Oct 4-7, 2016; Chicago, IL, USA p. 18-23 URL: https://ieeexplore.ieee.org/document/7776322 [doi: 10.1109/ichi.2016.99]
38. Feng T, Narayanan SS. Discovering optimal variable-length time series motifs in large-scale wearable recordings of human bio-behavioral signals. 2019 Presented at: ICASSP '19: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing; May 12-17, 2019; Brighton, UK p. 7615-7619 URL: https://ieeexplore.ieee.org/abstract/document/8682427 [doi: 10.1109/icassp.2019.8682427]
39. Ohayon M, Wickwire EM, Hirshkowitz M, Albert SM, Avidan A, Daly FJ, et al. National Sleep Foundation's sleep quality recommendations: first report. Sleep Health 2017 Feb;3(1):6-19. [doi: 10.1016/j.sleh.2016.11.006] [Medline: 28346153]
40. Liu X, Tamminen S, Korhonen T, Röning J. How physical exercise level affects sleep quality? Analyzing big data collected from wearables. Procedia Comput Sci 2019;155:242-249 [FREE Full text] [doi: 10.1016/j.procs.2019.08.035]
41. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]
42. Upadhyay D, Pandey V, Nag N, Jain R. Personalized user modelling for sleep insight. 2020 Presented at: HuMA'20: 1st International Workshop on Human-centric Multimedia Analysis; Oct 12, 2020; Seattle, WA, USA p. 13-20 URL: https://dl.acm.org/doi/10.1145/3422852.3423478 [doi: 10.1145/3422852.3423478]
43. Farajtabar M, Kıcıman E, Nathan G, White RW. Modeling behaviors and lifestyle with online and social data for predicting and analyzing sleep and exercise quality. Int J Data Sci Anal 2019;8(4):367-383 [FREE Full text] [doi: 10.1007/s41060-018-0136-8]
44. Kang WS, Choi RS, Son CS. Explainable sleep quality evaluation model using machine learning approach. 2017 Presented at: MFI '17: 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems; Nov 16-18, 2017; Daegu, South Korea p. 542-546 URL: https://ieeexplore.ieee.org/document/8170377 [doi: 10.1109/mfi.2017.8170377]
45. Nethealth Data. University of Notre Dame. URL: http://sites.nd.edu/nethealth/data-2/ [accessed 2023-06-19]
46. Jalali L. Interactive Event-driven Knowledge Discovery from Data Streams. Irvine, CA, USA: University of California; 2016.
47. Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. J Biomed Inform 2015 Feb;53:73-80 [FREE Full text] [doi: 10.1016/j.jbi.2014.09.003] [Medline: 25236952]
48. Bloom BS. Bloom's Human Characteristics and School Learning. New York, NY, USA: McGraw-Hill; 1976.

49. Fokkema M, Edbrooke-Childs J, Wolpert M. Generalized linear mixed-model (GLMM) trees: a flexible decision-tree method for multilevel and longitudinal data. Psychother Res 2021 Mar;31(3):313-325. [doi: 10.1080/10503307.2020.1785037] [Medline: 32602811]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
SE: sleep efficiency
SOL: sleep onset latency

Edited by L Buis; submitted 16.09.22; peer-reviewed by SM Ayyoubzadeh, M Ribeiro; comments to author 20.12.22; revised version received 03.02.23; accepted 05.06.23; published 28.06.23.

Please cite as:
Hoang NH, Liang Z
Knowledge Discovery in Ubiquitous and Personal Sleep Tracking: Scoping Review
JMIR Mhealth Uhealth 2023;11:e42750
URL: https://mhealth.jmir.org/2023/1/e42750
doi:10.2196/42750
PMID:37379057

Review

Digital Technologies for Women's Pelvic Floor Muscle Training to Manage Urinary Incontinence Across Their Life Course: Scoping Review

Stephanie J Woodley[1], BPhty, MSc, PhD; Brittany Moller[1], BSc (Hons); Alys R Clark[2], BA, MMathSc, PhD; Melanie D Bussey[3], BPE, MSc, PhD; Bahram Sangelaji[1,4], PT, PhD; Meredith Perry[5], BPhty, MManipTh, PhD; Jennifer Kruger[2], BSc, MSc, PhD

[1]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand

[2]Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

[3]School of Physical Education, Sport and Exercise Sciences, University of Otago, Dunedin, New Zealand

[4]Southern Queensland Rural Health, Brisbane, Australia

[5]School of Physiotherapy, University of Otago, Wellington, New Zealand

Corresponding Author:
Stephanie J Woodley, BPhty, MSc, PhD
Department of Anatomy
School of Biomedical Sciences
University of Otago
270 Great King Street
Dunedin, 9016
New Zealand
Phone: 64 276680828
Email: stephanie.woodley@otago.ac.nz

Abstract

Background: Women with urinary incontinence (UI) may consider using digital technologies (DTs) to guide pelvic floor muscle training (PFMT) to help manage their symptoms. DTs that deliver PFMT programs are readily available, yet uncertainty exists regarding whether they are scientifically valid, appropriate, and culturally relevant and meet the needs of women at specific life stages.

Objective: This scoping review aims to provide a narrative synthesis of DTs used for PFMT to manage UI in women across their life course.

Methods: This scoping review was conducted in accordance with the Joanna Briggs Institute methodological framework. A systematic search of 7 electronic databases was conducted, and primary quantitative and qualitative research and gray literature publications were considered. Studies were eligible if they focused on women with or without UI who had engaged with DTs for PFMT, reported on outcomes related to the use of PFMT DTs for managing UI, or explored users' experiences of DTs for PFMT. The identified studies were screened for eligibility. Data on the evidence base for and features of PFMT DTs using the Consensus on Exercise Reporting Template for PFMT, PFMT DT outcomes (eg, UI symptoms, quality of life, adherence, and satisfaction), life stage and culture, and the experiences of women and health care providers (facilitators and barriers) were extracted and synthesized by ≥2 independent reviewers.

Results: In total, 89 papers were included (n=45, 51% primary and n=44, 49% supplementary) involving studies from 14 countries. A total of 28 types of DTs were used in 41 primary studies, including mobile apps with or without a portable vaginal biofeedback or accelerometer-based device, a smartphone messaging system, internet-based programs, and videoconferencing. Approximately half (22/41, 54%) of the studies provided evidence for or testing of the DTs, and a similar proportion of PFMT programs were drawn from or adapted from a known evidence base. Although PFMT parameters and program compliance varied, most studies that reported on UI symptoms showed improved outcomes, and women were generally satisfied with this treatment approach. With respect to life stage, pregnancy and the postpartum period were the most common focus, with more evidence needed for women of various age ranges (eg, adolescent and older women), including their cultural context, which is a factor that is rarely considered. Women's perceptions and experiences are often considered in the development of DTs, with qualitative data highlighting factors that are usually both facilitators and barriers.

Conclusions: DTs are a growing mechanism for delivering PFMT, as evidenced by the recent increase in publications. This review highlighted the heterogeneity in types of DTs, PFMT protocols, the lack of cultural adaptations of most of the DTs reviewed, and a paucity in the consideration of the changing needs of women across their life course.

(JMIR Mhealth Uhealth 2023;11:e44929) doi:10.2196/44929

KEYWORDS

apps; culture; life course; mobile health; mHealth; pelvic floor muscle training; urinary incontinence; women's health; mobile phone

Introduction

Background

Pelvic floor muscle (PFM) dysfunction, which most commonly manifests as urinary incontinence (UI), pelvic organ prolapse, and pain, is a major and often unreported problem for women. UI, defined as "any involuntary leakage of urine" [1], affects between 25% and 45% of women worldwide, yet the true prevalence is likely to be higher, with women underreporting UI because of the associated shame and embarrassment [1]. UI substantially affects quality of life (QoL) in relation to both women's physical and mental health and well-being and also represents a major economic burden (eg, costs associated with routine care and treatment) [2].

PFM training (PFMT), which includes exercises to increase PFM strength and endurance, is recommended as the first choice for managing UI, especially stress UI [1]. PFMT can be undertaken by women to maintain pelvic health by preventing the onset of UI or can cure or improve symptoms and enhance QoL in adult and older women, including during pregnancy and the postpartum period [3,4]. However, despite its effectiveness, approaches to PFMT vary across communities and countries, and maintaining exercise programs, which are often undertaken at home, is difficult [5]. In addition, many women avoid seeking treatment for UI based on the belief that UI is an inevitable consequence of aging or childbirth, or the perception that little can be done to improve symptoms or QoL, or because of limited access to health services [6].

Digital technologies (DTs; such as the World Wide Web; eHealth; and mobile health [mHealth], including SMS text messaging and apps) provide an avenue for women with UI to seek guidance with PFMT and potentially improve their symptoms and QoL [5-8]. To elicit health benefits, women require access to DTs based on the best scientific evidence. However, although the market appears to be flooded with PFMT apps, few have been scientifically validated in terms of content, quality, or appropriateness [9,10]. In addition, knowing whether PFMT delivered via DTs is sound from a clinical perspective is equally important, but there is a lack of information as to whether PFMT in this context is based on contemporary evidence [11]. This is potentially compounded by the notion that few mHealth apps have been developed in collaboration with key stakeholders, such as women experiencing UI or health care professionals [12].

Factors such as age and culture may influence how women engage with PFMT DTs. For example, there is evidence that women are more vulnerable to developing UI at certain stages in life, including (1) young athletic women, particularly those participating in high-impact sports [13]; (2) during and after pregnancy, when one-third of women giving birth for the first time have UI, which may persist for at least 3 months post partum; (3) menopause, where a peak in UI occurs; and (4) older women (UI prevalence ranges from 43% to 77%), particularly those in residential care, where UI is a substantial risk factor for falls [1]. On the basis of this evidence, age-appropriate and specific PFMT programs seem imperative to best cater to women, yet there appears to be a distinct lack of information related to the uptake of DTs to manage UI at different stages in life [11]. Culture, which encompasses particular spiritual, intellectual, and emotional features, including lifestyle, value systems, traditions, and beliefs [14], not only affects how women interact with DTs [15,16] but also shapes their experiences and attitudes toward UI [15-17]. Although it is essential to understand how culture may affect the use of and engagement with PFMT DTs, with reference to UI, it is unclear whether the experiences and needs of women from different cultures or ethnic groups are considered when developing these types of DTs.

Objectives

Several systematic reviews have recently been published in this field, mostly focusing on the effectiveness of PFMT DTs in terms of improving symptoms of UI and QoL along with adherence to the prescribed PFMT program [7,8,18-21]. In this context, knowledge of the quality and content of PFMT DTs is also important, as is an understanding of whether such DTs are designed for women across their life course and take into account the cultural contexts and experiences of women and other relevant stakeholders. The main aim of this scoping review was to provide a narrative synthesis of digital health technologies used for women's PFMT to manage UI. The key objectives of this review were to (1) explore whether PFMT DTs follow best-practice guidelines and describe outcomes related to their use, (2) establish whether DTs have been designed for PFMT at specific stages in life or consider culture, and (3) describe users' experiences of DTs for PFMT.

Methods

This scoping review was conducted in accordance with the Joanna Briggs Institute methodological framework [22] and the

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [23]. A protocol was prospectively registered with the Open Science Framework [24].

Search Strategy

Following an initial search in PubMed, a systematic search of 7 electronic databases (AMED, CINAHL, Embase, MEDLINE, SPORTDiscus, Scopus, and PsycINFO) was conducted to identify relevant literature (from inception to December 2021). Key search concepts related to (1) DTs (eg, smartphones, cell phones, apps, telemedicine, and mHealth), (2) UI, (3) PFMT or exercise, and (4) key life stages (eg, pregnancy and menopause). Hand searching of reference lists of included articles, as well as citation tracking (eg, Web of Science for the last 5 years), was used to identify additional articles that may have been eligible for screening and inclusion (Multimedia Appendix 1).

Study Selection Criteria

Population

Studies were included if they (1) focused on women aged ≥14 years with or without UI who were using or had used some form of DT to engage in PFMT, (2) evaluated outcomes related to the use of PFMT DTs for managing UI, or (3) explored users' experiences of DTs for PFMT. Studies were excluded if the research focus was on women with overactive bladder or enuresis or neurological conditions, they reported collective data for men and women unless the data specific to women could be extracted separately, they used biofeedback devices that were not connected to an app or any other form of eHealth, and they were published in a language other than English for which a translation could not be acquired (eg, through Google Translate).

Concept

This review considered studies that explored PFMT delivered via DTs for the management of UI—with a focus on the evidence base for DTs and PFMT—the life course, culture, and users' experiences. The use of DTs for health (PFMT) is defined as "a broad umbrella term encompassing eHealth (which includes mHealth), as well as emerging areas, such as the use of advanced computing sciences in 'big data,' genomics, and artificial intelligence" [25]. In addition to eHealth, other applications such as wearable devices with a digital component (eg, a vaginal biofeedback probe connected to a mobile app), telehealth, and personalized medicine are encompassed within the scope of DTs.

Context

Studies that met the previously defined criteria were included to establish the widest coverage of information related to PFMT delivered via DTs for managing UI. This encompassed a large and heterogeneous group of women with or without UI, health care providers (HCPs) or researchers, and other disciplines (eg, IT experts). Any type of health care setting (eg, primary care and community) or discipline (eg, physiotherapy and general practitioners) was considered.

Study Design

Research involving quantitative and qualitative study designs and other forms of gray publications, such as opinion pieces, editorials, conference abstracts, theses, and case studies or series, were considered [26].

Study Selection

Titles and abstracts were independently screened by 5 authors (AC, BS, JK, MB, and SW) using web-based software (Covidence systematic review software; Veritas Health Innovation). The full texts were read and assessed by 2 of the 3 authors (BM, BS, and SW), with discrepancies resolved through consensus or discussion with another member of the team.

Data Extraction and Verification

A customized template was developed in Microsoft Word (Microsoft Corp) and piloted on 5 of the included studies. Data were independently extracted into the template by 5 authors (MB, AC, JK, BM, BS, and SW), transferred to a Microsoft Excel (Microsoft Corp) spreadsheet, and cross-checked by 2 authors (BS and SW). Any disagreements were resolved through consensus or consultation with a third reviewer when necessary.

The data extracted related to general study and participant characteristics included authors; year of publication; country; study aims; sample size; inclusion and exclusion criteria (intervention and comparator groups if relevant); age, gender, and level of education of participants; type of UI; and duration of symptoms.

To address the key objectives of this review, the data outlined in Textbox 1 were extracted.

Textbox 1. Data extracted from the included studies.

- Data related to the digital technologies (DTs)
  - Evidence base for or validation through previous testing
  - Type and features of DTs—capacity to extract, educational features, gamification, reminders and reinforcements, social media and self-monitoring, and technical support [7,27]
- Data specific to pelvic floor muscle training (PFMT)
  - Evidence base for the PFMT program
  - Descriptions of PFMT, which were charted according to the 16 key elements in the PFMT variation of the Consensus on Exercise Reporting Template (CERT-PFMT) [28]. In the context of this review, there was overlap between some CERT-PFMT items and DT features. Item 1 (exercise equipment) in the CERT-PFMT refers to the DT of interest, which incorporates descriptions of the device and related features (eg, biofeedback and mobility requirements) [7]. Item 5 (adherence) is covered by "self-monitoring," item 6 (motivation) relates to "reminders and reinforcements," and item 10 (nonexercise components) equates to "educational features." In the case of overlap, the data were extracted and synthesized under the umbrella of DT.
- Outcomes related to the use of PFMT DTs for managing urinary incontinence (UI), including UI symptoms, quality of life, and adherence to and satisfaction with the program
- Information related to key life stages and the culture of the women engaging with the PFMT DTs
- Experiences (facilitators and barriers) of women and health care providers with PFMT DTs for managing UI

Data Synthesis

The included studies that shared common author teams or apps were grouped accordingly. Descriptive statistics were used to summarize the data (BM, BS, and SW). With the exception of the study protocols, the methodological quality of the included studies was independently appraised using the relevant Joanna Briggs Institute critical appraisal tools [29] by pairs of reviewers, with a third reviewer consulted to reach a consensus if required.

For qualitative studies or the qualitative components of mixed methods studies, thematic synthesis, with the development of analytical themes driven by our review questions (ie, deductive analysis), was used for data synthesis [30] (MP and SW). The analysis occurred over 3 steps, with the last step designed to present clear implications for HCPs and policy makers. First, coding of text segments from the results and discussion specific to the review objectives was performed from sections of the included articles. Next, the raw codes were grouped and named in an iterative manner to form descriptive themes (grouped by the study's reported main themes and women's or clinician's perceptions of facilitators of and barriers to the use of DTs). Finally, analytical themes were generated from descriptive themes, and these analytical themes extended the synthesis beyond the conclusions of the included articles. Data were grouped for both barriers and facilitators under the headings of interactions between users and eHealth, interactions between users and PFMT exercises, and interactions between PFMT exercises and eHealth [31,32]. Although other tangential themes were generated, we presented the themes that were most coherently related to the study objectives.

Deviations From the Protocol

Owing to the large number of papers retrieved, a decision was made to exclude systematic reviews, meta-analyses, and scoping reviews from the analysis, which represents a deviation from the study protocol. Similarly, because of the number of DTs included in this review, we did not classify the types of DTs using the World Health Organization (WHO) classification [33] or rate the apps using the Mobile App Rating Scale [34].

Results

Search Results and Characteristics of the Included Studies

From the 7444 records screened for titles and abstracts, and after the removal of duplicates, 288 (3.87%) full-text reports were reviewed (Figure 1). A total of 89 papers met the inclusion criteria, of which 45 (51%) were classified as primary papers, with the other 44 (49%) considered supplementary papers (Table S1 in Multimedia Appendix 2 [5,6,11,12,31,35-118]; Table S2 in Multimedia Appendix 2 presents the inclusion and exclusion criteria for the included studies [5,6,11,12,31,35-118]). Of the 45 primary studies, many were randomized controlled trials (RCTs; n=13, 29%), with various other designs including cross-sectional studies (n=7, 16%); qualitative studies (n=6, 13%); mixed methods studies combining either RCTs or quasi-experimental trials with qualitative research (n=4, 9%); quasi-experimental studies (n=4, 9%); cohort studies (n=4, 9%); case series (n=4, 9%); and a case report, case-control, and validation study; of these 45 studies, 6 (13%) were study protocols and 2 (4%) were published in a language other than English (Dutch [35] and Portuguese [36]). The supplementary articles consisted of follow-up studies, secondary analyses, associated abstracts reporting a subset of data from the primary article, and author comments and letters to the editor (eg, [89,90,92-99,102-106,108-118]). Publications in this area have increased rapidly since the 2010s, with most protocols registered since 2019 (Figure 2). The methodological quality was rated for 84% (38/45) of the primary studies and was predominantly high (15/38, 39%) or fair (14/38, 37%), with 24% (9/38) considered poor (Multimedia Appendix 3 [5,6,36-41,43-46,48-51,53,57-59,61-77,80]). Key methodological areas for consideration included blinding of participants, therapists, and outcome assessors; measuring outcomes in a valid and reliable way and identifying confounding factors (cross-sectional studies); and consecutive recruitment of participants (case series), although aspects such as double-blinding are recognized as problematic in pragmatic and clinical trials.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flow diagram—search for papers related to digital technologies (DTs) and pelvic floor muscle training (PFMT) for women.

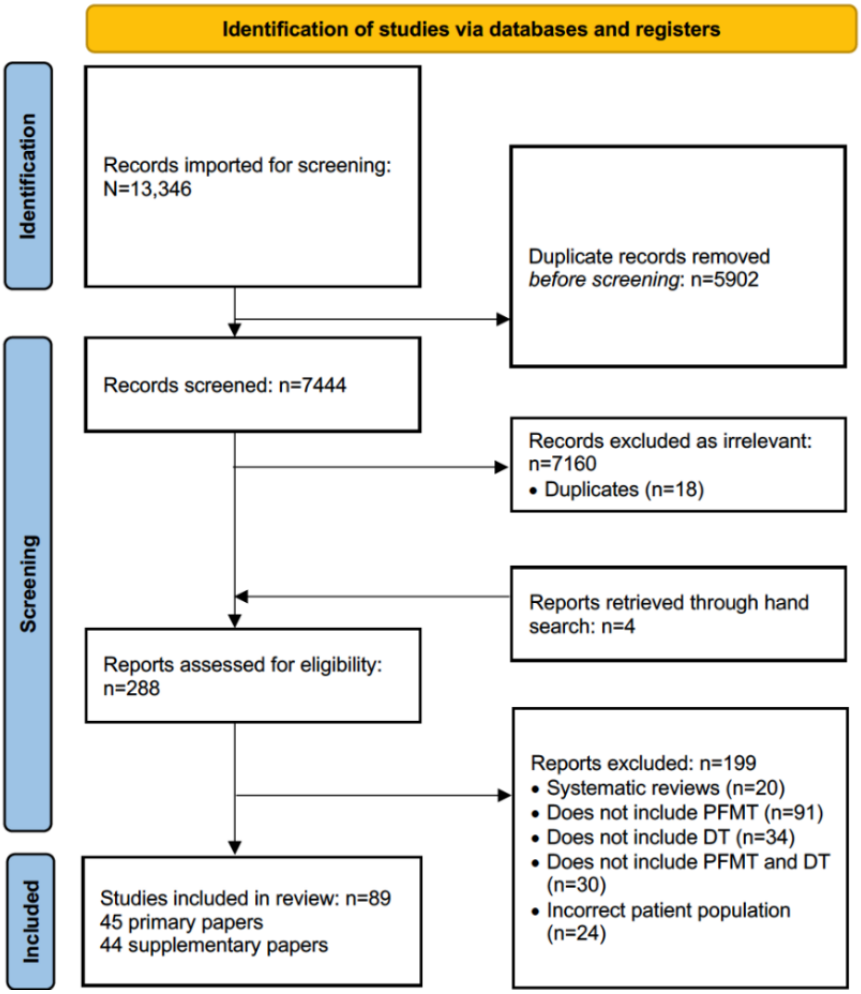
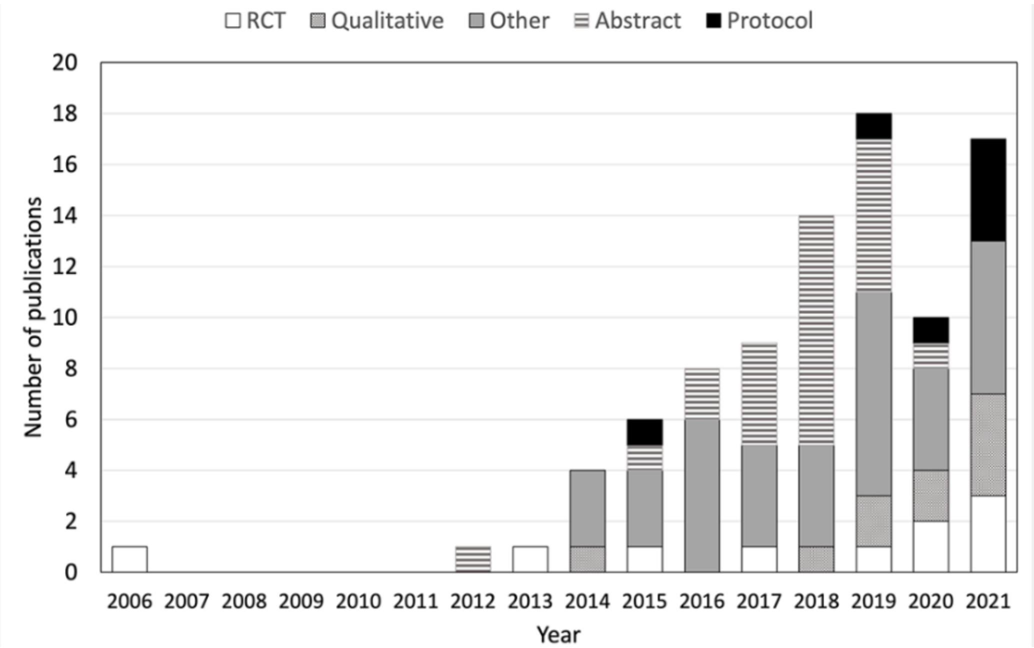


Figure 2. Publications included in this review presented by year of publication. RCT: randomized controlled trial.



The studies originated in 14 countries, primarily from Europe (9/45, 20% from Sweden; 7/45, 16% from the Netherlands; 3/45, 7% from Spain; and 1/45, 2% from Germany), as well as the United States (6/45, 13%), the United Kingdom (4/45, 9%), Brazil (4/45, 9%), China (4/45, 9%, including Hong Kong), Australasia (4/45, 9%), Canada (1/45, 2%), Japan (1/45, 2%), and Malaysia (1/45, 2%). A total of 47% (21/45) of the studies had a total sample size of <50, 9% (4/45) recruited between 50 and 100 participants, and 44% (20/45) of the studies sampled >100 participants (2/20, 10% of which analyzed data from a sample of >10,000 women; Multimedia Appendix 3).

Participant Characteristics

Age

The age of the participants ranged from 18 to 98 years, with approximately 35% of studies (16/45, 36%) including women in their 40s or 50s as the average age. The level of education of the recruited participants was stated for just over half (24/45, 53%) of the primary studies. Among these, most women in each study were shown to be educated at the university level.

UI in the Included Studies

Studies mostly recruited women with stress UI only (17/45, 38%) [5,6,37-51], whereas 7% (3/45) of the studies included those with stress and mixed UI if stress symptoms were predominant [52-54], and 24% (11/45) included a mixture of stress, urge, and mixed UI types [55-65]. In total, 11% (5/45) of the studies included healthy (continent) and incontinent participants [66-70], and 11% (5/45) did not clearly specify the UI type [71-75], and this was irrelevant in 7% (3/45) [36,70,80]. Only 2% (1/45) of the studies [76], a case study, looked at urge UI only. A total of 20% (9/45) of the studies documented the duration of women's UI symptoms before their inclusion in the study. Of these, Asklund et al [5] required participants to have had symptoms for at least 6 months as part of their inclusion criteria. The remaining 18% (8/45) of the studies [38,40,43,46,51,61,64,65] showed variable durations, ranging from <3 months to 26 years (Multimedia Appendix 3).

Stage in Life

A total of 17 (38%) of the 45 studies reported life stage parameters: 29% (5/17) recruited women in the postpartum period [67-69,75,77]; 18% (3/17) recruited pregnant women [47,51,60]; 12% (2/17) included both pregnant and postpartum women [52,66]; 6% (1/17) included postmenopausal women [59]; and 24% (4/17) reported including a mixture of premenopausal, perimenopausal, postmenopausal, lactating, and postpartum participants [45,46,58,62]. A case study [76] included women reported as parous and Campbell et al [42] recruited athletic women for their RCT.

Cultural Context

Some studies developed the DTs for use by women in their specific countries (eg, Sweden [5], Japan [68], and Germany [70]), and the Tät has been translated into a number of different languages [56,63,78].

One group conducted a systematic review to explore variables that may influence adherence to PFMT DTs, which led to the development of the iPelvis app [11,57]. The authors emphasized the importance of considering ethnicity as part of a woman's individuality, and as such, the avatar character within the iPelvis app can be ethnically matched to the woman by altering features such as skin color, the flag of the country, and cultural costumes, as well as age and stage in life (eg, pregnant or older adult). This concept was supported by Han et al [72], who stated that the information in apps needs to be formatted in a culturally relevant way to ensure that it is effective. The importance of ongoing research to evaluate apps in different and diverse cultural contexts was acknowledged in 7% (3/45) of the studies [72,73,79].

DTs in the Included Studies

Overview

Among the 45 primary studies, data related to DTs and PFMT were not extracted from 4 (9%)—1 (25%) [55] analyzed data collected from 3 previous RCTs [5,6,64], and 3 (75%) qualitative studies [46,77,80] took a broad approach without focusing on a specific technology for the delivery of PFMT. Therefore, the data and information in the following sections were derived from 91% (41/45) of the studies, some of which used the same DTs (eg, Tät, Leva, and Pen Yi Kang; Table 1).

Table 1. Summary of digital technologies (DTs) and their features.

Study[a]	DT	BF[b]	Mobility requirements	EB[c]	DE[d]	EF[e]	R and R[f]	Social media features	Self-monitoring	Gamification	Training or support in use of DT
Anglès-Acedo et al [37,38]	Mobile app—WOMEN UP	Yes	Internet; Bluetooth	Yes	Yes	Yes	Yes; NI[g]	NI	Yes	Yes	NI
Araujo et al [39]	Mobile app—Diário Saúde	No	Internet	NI	Yes	No	Yes; yes	No	NI	No	NI
Asklund et al [5]; Asklund and Samuelsson [66]; Nyström et al [73]; Rygh et al [63]; Samuelsson et al [50]	Mobile app—Tät	No	Internet	Yes	No[h]	Yes	Yes; yes	No	Yes	No	Yes
Wadensten et al [64]	Mobile app—Tät II	No	Internet	Yes	No	Yes	Yes; yes	NI	Yes	No	Yes
Bokne et al [41]	Internet-based program—Tät	No	Internet	Yes	No	Yes	No; no	No	NI	No	NI
Firet et al [56]	Internet-based program—Tät	No	Internet	Yes	Yes	Yes	NI; yes	NI	Yes	No	Yes
Sjöström et al [6]	Internet-based program—Tät	No	Internet	NI	No	Yes	No; yes	No	Yes	No	Yes
Barbato et al [40]	Internet-based program	No	Internet	NI	No	Yes	No; no	No	No	No	No
Campbell et al [42]	Mobile app—Squeezy App	No	Internet; Bluetooth	NI	Yes	NI	No; yes	No	Yes	No	NI
Robson [74]	Mobile app—Squeezy App	No	Internet	Yes	No	Yes	Yes; yes	NI	Yes	No	NI
Carrión Pérez et al [43]	Telerehabilitation device and vaginal probe	Yes	Internet; Bluetooth	NI	Yes	NI	NI; NI	No	NI	No	Yes
Coggins et al [44]	Mobile app and vaginal device—Elvie	Yes	Bluetooth	NI	Yes	NI	Yes; yes	NI	NI	No	NI
Conlan et al [45]	Telehealth	No	Internet	NA	No	No	No; no	No	No	No	NI
Cornelius [71]	Mobile app and vaginal probe—PeriCoach	Yes	Bluetooth; internet	NI	Yes	NI	Yes; yes	NI	Yes	No	NI
Shelly [76]	Mobile app and vaginal probe—PeriCoach	Yes	Bluetooth; internet	NI	Yes	NI	Yes; NI	NI	Yes	No	Yes
Smith [53]	Mobile app and vaginal probe—PeriCoach	Yes	Bluetooth; internet	NI	NI	NI	NI; NI	NI	NI	No	Yes
Dufour et al [67]	Mobile app and vaginal device—iBall	Yes	Bluetooth	NI	Yes	NI	Yes; NI	Yes	Yes	Yes	Yes
Goode et al [58]	Web-based—MyHealtheBladder	No	Internet	Yes	Yes	Yes	Yes; yes	NI	Yes	No	NI
Han et al [72]	Mobile app—Bwom	No	Internet	NI	NI	Yes	NI; NI	NI	NI	NI	NI
Hui et al [59]	Telemedicine continence program (videoconferencing)	No	Internet	NI	No	Yes	No; no	No	No	No	Yes
Jaffar et al [60]	Mobile app—KEPT[i]-app	No	Internet	Yes	Yes	Yes	Yes; yes	NI	Yes	NI	Yes
Kinouchi and Ohashi [68]	Smartphone-based reminder system	No	Internet	NI	Yes	No	NI; yes	NI	NI	No	Yes

Study[a]	DT	BF[b]	Mobility requirements	EB[c]	DE[d]	EF[e]	R and R[f]	Social media features	Self-monitoring	Gamification	Training or support in use of DT
Fischer Blosfield et al [57]	Mobile app—iPelvis	No	Internet	Yes	Yes	Yes	Yes; yes	Yes	Yes	Yes	Yes
Moossdorff-Steinhauser et al [52]	Mobile app—iPelvis	No	Internet	NI	NI	Yes	Yes; yes	NI	Yes	NI	NI
Li et al [69]	Mobile app and audio guidance—Pen Yi Kang	No	Internet	Yes	Yes	NI	NI; NI	NI	NI	NI	Yes
Wang et al [51]	Mobile app and audio guidance—Pen Yi Kang	No	Internet; Bluetooth	Yes	Yes	NI	Yes; NI	NI	NI	No	Yes
Li et al [47]	Mobile app—UIW[j]	No	Internet	Yes	Yes	Yes	Yes; yes	NI	Yes	NI	Yes
Loohuis et al [61]; Wessels et al [65]	Mobile app—UrinControl	No	Internet	Yes	Yes	Yes	NI; yes	NI	Yes	No	Yes[k]
Moretti [36]	Mobile app, vaginal probe, and surface electrodes—MyoPelvic	Yes	Bluetooth	Yes	Yes	NI	NI; NI	NI	Yes	Yes	Yes
Pedofsky et al [48]	Mobile app and intravaginal pressure sensor array—FemFit	Yes	Bluetooth	NA	NI	Yes	NI; NI	NI	NI	Yes	No
Pla et al [49]	Mobile app and vaginal device—Birdi	Yes	Bluetooth; internet	NI	Yes	NI	Yes; NI	NI	Yes	No	Yes
Pulliam et al [62]	Mobile app and vaginal insert—Leva Pelvic Digital Health System	Yes	Bluetooth	NI	Yes	NI	Yes; NI	NI	Yes	No	Yes
Weinstein et al [54]	Mobile app and vaginal insert—Leva Pelvic Digital Health System	Yes	Bluetooth; internet	NI	Yes	Yes	Yes; NI	NI	NI	No	Yes
Saboia et al [75]	Mobile app—Continence App	No	Internet	Yes	Yes	Yes	Yes; yes	NI	Yes	NI	NI
von Au et al [70]	Mobile app—Pelvina	No	Internet	NI	Yes	Yes	NI; NI	NI	NI	No	NI

[a]Studies are ordered alphabetically but grouped by app where relevant.
[b]BF: biofeedback.
[c]EB: evidence base (provide some evidence base or previous testing of the DTs).
[d]DE: data extraction (capacity to extract data).
[e]EF: educational features.
[f]R and R: reinforcements and reminders.
[g]NI: not indicated.
[h]With the exception of the study by Asklund and Samuelsson [66], in which women had the choice to use the statistics function.
[i]KEPT: Kegel Exercise Pregnancy Training.
[j]UIW: Urinary Incontinence for Women.
[k]Not indicated in the study by Wessels et al [65].

Types of DTs (Consensus on Exercise Reporting Template for PFMT Item 1)

A total of 28 types of DTs were used across the 41 studies. Just over 40% (12/28, 43%) of these were solely mobile apps, 4% (1/28) trialed a smartphone-based messaging system, 11% (3/28) were internet-based programs, and 7% (2/28) were dedicated to videoconferencing. A total of 32% (9/28) of the technologies involved the use of a portable vaginal biofeedback or accelerometer-based device [36-38,43,44,48,49,53,54,62,67,71,76] that provided real-time feedback transmitted via Bluetooth to a mobile app or, in one instance, to a computer application [43]. In addition to the vaginal device, electromyographic data were integrated from surface electrodes attached to the abdominal muscles [37,38] or PFMs [36].

Evidence Base or Previous Testing of DTs

A total of 54% (22/41) of the studies provided some evidence base for or testing of the DTs [5,41,50,51,56,58,61,63-66,69,73,74], which had been either undertaken in the development stage or as an iterative process (eg, Tät) or was one of the specific purposes of the study [11,36-38,47,48,60,75]. Some studies (4/41, 10%) implemented a design framework (eg, the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework [47] or the Fit Between Individuals, Task, and Technology framework [56]) in their trials. The development of the technologies generally involved collaboration, testing, and input from IT experts (such as hardware or software engineers) and HCPs or researchers with relevant clinical expertise (eg, obstetricians, women's health physiotherapists, and nurses) and feedback from women, the end users of the product. One app (1/12, 8%) was developed based on 12 key variables identified through a systematic review of the literature [57].

Capacity to Extract Data

Approximately 60% of the studies (25/41, 61%) reported the capacity to extract data to monitor women's progress; approximately 30% did not (12/41, 29%), and this was not indicated in 10% (4/41) of the studies. Data were extracted directly from the DTs in 20% (5/25) [39,43,49,61,65], and in 1 (4%) of the 25 studies, data could be emailed to the researchers via the app [36]. In other studies (19/25, 76%), data were transmitted from an app and stored or accessed on an associated web platform or server [37,38,42,44,47,56-58,68-71,75,76] or uploaded from the app to a cloud-based storage system [51,54,60,62,67]. Of the suite of studies that used Tät, 11% (1/9) indicated the collection of user statistics from the internet-based program [56], and 22% (2/9) reported that women could voluntarily choose to use the statistics function and submit their user statistics at follow-up [5,64].

Educational Features (Consensus on Exercise Reporting Template for PFMT Item 10)

Educational information was incorporated into 61% (25/41) of the studies, with most including a combination of topics such as education about the anatomy and function of the PFMs, PFMT, stress UI, and related lifestyle advice (eg, weight management, physical activity, and fluid management). A total of 2% (1/41) of the studies provided holistic advice on breathing, posture, and movement [40] and used videos to deliver this information, an approach also adopted by another 7% (3/41) of the studies [57,72,81], with another incorporating audio fragments [56]. In a videoconferencing study, education was provided by a nurse specialist across a series of talks rather than being integrated into the technology itself [59]. Of the studies that did not incorporate education, 6% (1/16) involved telerehabilitation [43], and 12% (2/16) were mobile apps [39,68], with 6% (1/16) solely using a smartphone reminder system [68].

Reinforcements, Reminders, and Self-monitoring (Consensus on Exercise Reporting Template for PFMT Items 5 and 6)

A variety of reinforcements were used across 59% (24/41) of the studies, but the most common was the provision of visual (eg, graphics) or audiovisual feedback to guide women on the performance of PFMT, with the inclusion of voice [52,57] or sound [47] commands or accompanying music [39]. Goode et al [58] included storytelling; another study had an exercise module with a timer and score board [60].

In just over 50% (21/41, 51%) of the studies, reminder systems were incorporated into or complemented the DTs. In most cases (13/21, 62%), the reminders were customizable and were set by the women; in 24% (5/21) of the studies, push notifications were sent by the researcher or HCPs [21,47,52,60,68], and in 14% (3/21) of the studies, women were emailed a reminder (internet-based programs) [6,56,58].

Self-monitoring was a feature in 61% (25/41) of the studies. Apps commonly enabled tracking of exercise progress by women, including a statistical function (eg, Tät) or graphs or the capacity to record exercise adherence over time (eg, number and level of exercises). This function was also available through a web portal [37,38,76], or training diaries were completed and sent via email [6]. Some technologies also included a bladder diary [58,60,71] to monitor urinary symptoms.

Social Media and Gamification

In total, 7% (2/28) of the DTs had the capacity for social media forums: the iPelvis, which included a website and Facebook page [11,57], and the iBall [67], which enabled women to connect with others in a web-based community (but it was disabled for the purpose of the study, as it was only available in Chinese).

A total of 21% (6/28) of the DTs incorporated gamification [36-38,48,57,67], which, with 1 exception [57], was used in conjunction with biofeedback. Descriptions of gamification included "serious games" [37,38], games or activities (eg, weight lifting room and flying arena) [67], and gaming and virtual reality mediated by a comic character [11,57] or a cyclist [36] with built-in scoring systems.

Technical Support

A total of 32% (13/41) of the studies offered instructions (eg, handouts and instructions via email) on how to download and install the app or use and effectively care for the equipment (eg, vaginal probes) [5,47,49,51,53,54,57,60,61,64,67-69].

Follow-up technical support was offered in 15% (6/41) of the studies by a research assistant [6,47,51,54,56,60] using encrypted email or via the app. A total of 10% (4/41) of the studies included in-person sessions with supervision or testing of the technology [36,43,62,76] by physiotherapists [36,43,76] or an unspecified individual [62].

PFMT in the Included Studies

Evidence Base

Just over half (21/41, 51%) of the studies provided some evidence base for the PFMT program that was being delivered via the DTs; in the remaining studies, this was not indicated or was unclear. Evidence for PFMT varied, ranging from existing programs tested in RCTs, including the seminal publication by Bø et al [119] and others later (eg, the *Group Rehabilitation Or IndividUal Physiotherapy for Urinary Incontinence in Aging*

Women [GROUP] trial [120]), to expert opinion [121], guidelines (eg, the National Institute for Health and Care Excellence) [122], and the Dutch clinical practice guidelines for the physiotherapy management of stress UI [123], with enhancements made based on feedback from clinicians, women, and researchers (eg, Tät).

Delivery of PFMT (Consensus on Exercise Reporting Template for PFMT Items 1, 2, 3, 4, and 12)

On the whole, women engaged with the DTs at home on an individual basis, with 22% (9/41) of the studies including exercise both at home and in a clinical setting [39,42,43,51,52,57,62,76] or community center [59] (Table 2; please note that, in some cases, the same DT was used across multiple studies). In 10% (4/41) of the studies, women also attended supervised group sessions once a week to undertake PFMT via teleconferencing [59], in person [52,57] with a maximum of 4 women per group [52], or specific to one of the study arms (app plus physiotherapy group) [57]. In the study by Pla et al [49], Skype was the medium for supervision of group-based hypopressive abdominal exercises 3 times per week (5-9 women per group), with women also receiving monthly individual videoconferencing sessions to check progress.

In addition to the 10% (4/41) of studies that provided supervision for women in a group setting [49,52,57,59], 34% (14/41) supported women on an individual basis. Examples included confirming a PFM contraction or PFMT practice [39,51] and checking women's adherence to the program [39]; providing a set number of supervised sessions over the duration of the program (which ranged between 1 and 12) either in person [42,76] or remotely via email [6,41], phone call [54], or videoconferencing [45]; and more intense supervision, such as five 30-minute sessions over 2 weeks [43] and daily sessions 5 days per week [62]. If women required extra support with PFMT-related content, this was offered through email [45,49,56] or the chat function on an app [54], with Anglès-Acedo et al [37,38] noting that their web platform enabled "personalised supervision."

Details of the personnel providing supervision or support for PFMT were reported in 18 studies, with 8 (44%) referring to a physiotherapist [49,57] who was specialized in women's or pelvic health [39,42,43,45,52,76]. In others, a nurse specialist [59], therapist [37,38], urogynecologist [61], urotherapist [6,41], general practitioner [56], trained researcher [51], trained research assistant [62], or trained study staff member [54] provided supervision or support.

Table 2. Summary of pelvic floor muscle training (PFMT) delivery and content.

Study[a]	DT[b]	PFMT evidence base	Individual or group PFMT and setting	Supervision of PFMT and qualifications	Confirmation of voluntary PFM[c] contraction	PFMT parameters	Duration of program
Anglès-Acedo et al [37,38]	Mobile app—WOMEN UP	NI[d]	Individual; home	Web platform; therapist	Biofeedback	NI	3 months
Araujo et al [39]	Mobile app—Diário Saúde	NI	Individual; home and clinic	In person, monthly; specialist women's health physiotherapist	Digital assessment; specialist physiotherapist	8-second hold, 8-second relaxation followed by 3 phasic contractions 8 times, 2 times per day (sitting, lying down, or standing)	3 months
Asklund et al [5]; Asklund and Samuelsson[e] [66]; Nyström et al[f] [73]; Rygh et al[f] [63]; Samuelsson et al[f] [50]	Mobile app—Tät	Yes	Individual; home	No; —[g]	No	Progressive PFMT (6 basic and 6 advanced levels), different combinations and repetitions of PFM contractions (strength and endurance, quick contractions, and the "knack"); advanced phase incorporates different positions (standing, lifting, and walking); strengthening: from a 5-second hold, 5-second relaxation 2 times (basic) to a 7-second hold, 7-second relaxation 40 times (advanced); endurance: from a 14-second hold once (basic) to a 59-second hold, 59-second relaxation 2 times (advanced); quick: from a 3-second hold, 3-second relaxation 5 times (end of the basic phase) to a 3-second hold, 3-second relaxation 20 times (advanced) 3 times daily	3 months
Wadensten et al [64]	Mobile app—Tät II	Yes	Individual; home	No; —	No	Progressive, 4 different PFM exercises are included across 8 modules based on the Tät [5], from 3 times per day for 2 minutes (module 1) to 3 times per day for 3-4 minutes (module 4) and 3 times per day for 12 minutes (module 8)	15 weeks
Bokne et al [41]; Sjöström et al [6]	Internet-based program—Tät	Yes	Individual; home	Email once per week plus support as needed; urotherapist	No	Progressive, tailored (in part) program with 8 levels, including the "knack"; strength: hold maximal contractions for 8 seconds, 8-10 repetitions, 3 times per day; endurance: hold submaximal contractions for 15-90 seconds, 1 repetition, 3 times per day; quick contractions: hold for 3 seconds, 8-10 repetitions, 2-3 times per day	3 months
Firet et al [56]	Internet-based program—Tät	Yes	Individual; home	Email support as needed; GP[h] in training or researcher	No	Progressive, 4 different PFM exercises are included across 8 modules based on the Tät [5], from 3 times per day for 2 minutes (module 1) to 3 times per day for 3-4 minutes (module 4) and 3 times per day for 12 minutes (module 8)	3 months
Barbato et al [40]	Internet-based program	NI	Individual; home	No; —	No	"Self-paced" PFMT, 10-15 minutes daily	3 weeks

Study[a]	DT[b]	PFMT evidence base	Individual or group PFMT and setting	Supervision of PFMT and qualifications	Confirmation of voluntary PFM[c] contraction	PFMT parameters	Duration of program
Campbell et al [42]	Mobile app—Squeezy App	NI	Individual; home and clinic	In person, ≤7 appointments (45-60 minutes) over 6 months depending on women's needs; specialist pelvic health physiotherapist	Digital assessment (crook lying) and specialist physiotherapist; biofeedback (standing)	Progressive, tailored PFMT (in different functional positions) for strength, power, endurance, and relaxation, and the "knack"; no specific details of the PFMT program	Phase 2: 6 months[i]
Robson [74]	Mobile app—Squeezy App	NI	Individual; home	—	—	NI	Survey, open for 3 months
Carrión Pérez et al [43]	Telerehabilitation device and vaginal probe	NI	Individual; home and clinic	In person, 5 times for 30 minutes over 2 weeks plus monthly follow-up; pelvic floor expert physiotherapist	Biofeedback	PFMT: five 30-minute sessions in the clinic (over 2 weeks) plus home exercise program; daily	3 months
Coggins et al [44]	Mobile app and vaginal device—Elvie	NI	Individual; home	No	Biofeedback	NI	NI
Conlan et al [45]	Telehealth	NI	Individual; home	In person, initial 1-hour session plus email support over 6 weeks[j]; continence physiotherapist	NI	Individualized PFMT	6 weeks
Cornelius [71]	Mobile app and vaginal probe—PeriCoach	NI	Individual; home	NI; pelvic floor clinicians (for some participants)	Digital palpation[k]; biofeedback	Dosage[k]: contraction, relaxation 5 times for 5 seconds, 10 repetitions, 4 times per day, 5 times per week	8 weeks
Shelly [76]	Mobile app and vaginal probe—PeriCoach	NI	Individual; home and clinic	In person, 6 sessions over 8 weeks; pelvic floor physiotherapy specialist	Digital assessment and specialist physiotherapist; biofeedback	Progressive, tailored, starting with contraction, relaxation 3 times for 8 seconds, 8 repetitions (20-25 repetitions per day; week 1); 5 times for 7 seconds, 8 repetitions (40-50 repetitions per day; week 2); 6 times for 3 seconds, 15 repetitions (week 5); 10 times for 3 seconds, 15 repetitions (week 8); supine, then standing; functional training with forward bending and during ADLs[l] 2 times per day, 5 times per week	8 weeks
Smith [53]	Mobile app and vaginal probe—PeriCoach	NI	Individual; NI	NI; NI	NI	NI	20 weeks
Dufour et al [67]	Mobile app and vaginal device—iBall	Yes	Individual; home	No; —	Digital palpation; specialist pelvic health practitioner	3 times, 10 sets of 10 exercises 3-4 times per week	16 weeks
Goode et al [58]	Web-based—My-HealtheBladder	Yes	Individual; home	No; —	No	Progressive, from contraction, relaxation 2 times for 4 seconds (week 1) to 5 times for 5 seconds (week 4), 9 times for 9 seconds, and 10 times for 10 seconds (week 8) plus bladder control strategies	8 weeks

Study[a]	DT[b]	PFMT evidence base	Individual or group PFMT and setting	Supervision of PFMT and qualifications	Confirmation of voluntary PFM[c] contraction	PFMT parameters	Duration of program
Han et al [72]	Mobile app—Bwom	NI	Individual; home	No; N/A[m]	No	Progressive, "personalized" exercise plans, each with 6-12 exercises, with a new exercise each week	2 weeks
Hui et al [59]	Videoconferencing	NI	Individual and group; home and community center	Weekly videoconferencing; nurse specialist assisted by a research assistant (registered nurse)	Digital assessment and nurse specialist; biofeedback	1 videoconferencing session per week	8 weeks
Jaffar et al [60]	Mobile app—KEPT[n]-app	Yes	Individual; home and clinic	No; —	No	Progressive, 3 training skills and modes (different positions): beginner (2-second hold), intermediate (6-second hold), and advanced (10-second hold); 10 repetitions, 3 times per day; adherence phase: once they can perform PFMT confidently, maintain 10 cycles, 3 times per day	At least 16 weeks[o]
Kinouchi and Ohashi [68]	Smartphone-based messaging system	Yes	Individual; home	No; —	No	Hold 3-6 seconds, 3 sets of 6 contractions per day; different positions (standing, bent-knee lying, and 4-point kneeling)	8 weeks
Fischer Blosfield et al [57]	Mobile app—iPelvis	Yes	Individual and group; home and clinic (depending on study group allocation)[p]	12 sessions once a week, in person, in a group; physiotherapist	All participants had "physical examination"; women who had difficulty contracting PFM had a vaginal examination, with instruction	App+physiotherapy: PFMT in a group once per week plus app at home App only: PFMT at home. Progressive and tailored, including strength, explosive strength, endurance, timing, precontraction, exhaustibility, coordination, and functional exercises (eg, sneezing and coughing); PFM contraction and relaxation in different positions, situations, and activities, 6 phases, each with a 15-day duration	3 months[p]
Mooss-dorff-Steinhauser et al [52]	Mobile app—iPelvis (in conjunction with the Motherfit program)	Yes	Individual and group; home and clinic	8 sessions in person (60 minutes) in a group (maximum of 4 women); specialist pelvic physiotherapist	Observation and digital assessment, supine; pelvic specialist physiotherapist	Progressive group program, including strength and endurance, speed, and functional exercises, and the "knack"; NI if the home program was the same; build up to 8-12 contractions, 6-8–second hold plus 3-4 fast contractions; strength and endurance: 3 times per day, daily (minimum of 3-4 times week); different positions (lying down, sitting, kneeling, and standing); after 6 months of training: maintenance 2 times per week; speed: fast repetitions, build up to 10 sets of 3 quick contractions and 10 sets of 5 quick contractions 3 times per day	8 weeks, continuing past 6 months of home training
Li et al [69]	Mobile app and audio guidance—Pen Yi Kang	NI	Individual; home	No; —	Digital assessment; experienced physiotherapist	NI	6 weeks

Study[a]	DT[b]	PFMT evidence base	Individual or group PFMT and setting	Supervision of PFMT and qualifications	Confirmation of voluntary PFM[c] contraction	PFMT parameters	Duration of program
Wang et al [51]	Mobile app and audio guidance—Pen Yi Kang	NI	Individual; home	In-person initial 45-minute session plus phone contact once a month; trained researcher	Digital palpation; surface EMG[q], supine, and hips and knees bent	Progressive, different positions (sitting, standing, and lying down); 3-second hold, 2-6–second relaxation for 15 minutes, 2 times per day or 150 contractions per day	3 months
Li et al [47]	Mobile app—UIW[r]	Yes	Individual; home	No; —	Perineum palpation, supine, and surface EMG in lithotomy position; experienced obstetrician	Adapted from Tät [5]; progressive, 2 basic and 4 advanced levels, including different combinations and repetitions of 4 commonly used contraction types: test contraction, strength contraction, endurance contraction, and quick contraction; up to each woman to determine use (frequency and duration)	8 weeks
Loohuis et al [61]	Mobile app—URinControl	Yes	Individual; home	No; —	Assessed according to the ICS[s]; urogynecologist	Progressive program, directed to appropriate part of the app to start training; no further information provided	4 months
Wessels et al [65]	Mobile app—URinControl	Yes	Individual; home	No; —	No	Progressive program, directed to appropriate part of the app to start training; no further information provided	—
Moretti [36]	Mobile app, vaginal probe, and surface electrodes—MyoPelvic	Yes	—	—	Biofeedback, maximal voluntary contraction, and supine; researcher	Phasic fibers: contract <4 seconds, relax for twice the duration of the contraction, 12 repetitions maximum (as dictated by the game); tonic (slow) fibers: contract 4-10 seconds, relax for the same duration, 12 repetitions maximum (as dictated by the game); 1-2–minute rest between games recommended but not enforced; muscle coordination training (not specified)	—
Pedofsky et al [48]	Mobile app and intravaginal pressure sensor array—FemFit	Yes	—	—	—	Progressive, graduated exercise; no further information provided	12 weeks
Pla et al [49]	Mobile app and vaginal device—Birdi	NI	Individual and group; home	Videoconferencing, 2 initial individual sessions 3 times per week in a group and monthly individual session plus email or phone support; physiotherapist	Measured by the device	Daily PFMT, tailored	2 months
Pulliam et al [62]	Mobile app and vaginal insert—Leva Pelvic Digital Health System	NI	Individual; home and clinic	In person, once a day, 5 times per week over 6 weeks; trained research assistant	Accelerometer-based system	15-second PFM contraction 5 times, 15-second relaxation, 2 times per day in standing position	6 weeks
Weinstein et al [54]	Mobile app and vaginal insert—Leva Pelvic Digital Health System	NI	Individual; home	3 phone calls in first 2 weeks plus support via the chat function; trained study staff	No	5 cycles of squeeze and lift, and 15 seconds of rest for 15 seconds each, 2.5 minutes 3 times per day	8 weeks

Study[a]	DT[b]	PFMT evidence base	Individual or group PFMT and setting	Supervision of PFMT and qualifications	Confirmation of voluntary PFM[c] contraction	PFMT parameters	Duration of program
Saboia et al [75]	Mobile app—Continence App	Yes	—	—	—	NI	12 weeks
von Au et al [70]	Mobile app—Pelvina	NI	Individual; home	NI; NI	No	NI	Survey, open for approximately 1 year

[a]Studies are ordered alphabetically by first author but grouped by app where relevant.
[b]DT: digital technology.
[c]PFM: pelvic floor muscle.
[d]NI: not indicated.
[e]Survey, open for 10 months.
[f]3 months of app use.
[g]No information provided.
[h]GP: general practitioner.
[i]3 phases to this study.
[j]None of the participants used the email option.
[k]As reported in the study by Starr et al [83].
[l]ADL: activities of daily living.
[m]N/A: not applicable.
[n]KEPT: Kegel Exercise Pregnancy Training.
[o]Start at 28-week gestation until delivery at 36 weeks and continue PFMT until 8 weeks post partum.
[p]6-month program=12 phases as per Latorre et al [11].
[q]EMG: electromyography.
[r]UIW: Urinary Incontinence for Women.
[s]ICS: International Continence Society.

Content of PFMT (Consensus on Exercise Reporting Template for PFMT Items 7, 8, 9, 13, 14, and 15)

Confirmation of a voluntary PFM contraction was undertaken in just under half (19/41, 46%) of the studies and was either not included in the study design in 41% (17/41) of the studies or not indicated or appropriate (5/41, 12%). Confirmation was obtained through digital assessment [39,57,61,69], which was also used to teach a correct contraction or relaxation of the PFM [52,67]; biofeedback or the digital device [36-38,43,44,49,62]; or a combination of both [42,47,51,59,71,76]. Digital assessment was undertaken by a specialized HCP, although it was not indicated who performed this procedure in 7% (3/41) of the studies [51,57,71]. In the RCTs that used digital assessment, this was provided to all women across all the study groups.

A total of 49% (20/41) of the studies offered PFMT programs that were progressive in terms of content or position. Although most programs were generic, 10% (2/20) were tailored to the individual by provision of supervision [49,76], and another 15% (3/20) provided some indication or tailoring of the starting level for PFMT [61,64,72].

The details of the PFMT content were not indicated in 34% (14/41) of the studies, whereas the rest used a combination of strength, endurance, and power exercises; relaxation; or a combination of these. Direct PFMT, together with functional PFMT (eg, the "knack"), was included in 29% (12/41) of the studies [5,6,41,42,50,52,57,63,64,73,76,78], and it is likely to have been integrated into another 5% (2/41) of the studies that were based on the Tät [47,56].

Of the 19 studies that provided details about the prescribed dose, 14 (74%) recommended PFMT 3 times a day [5,6,41,47,50,52,54,56,60,63,64,68,73,78], 4 (21%) recommended PFMT twice a day [39,51,62,76], and 1 (5%) recommended it 4 times a day [71]. The most common program duration was 3 months (17/41, 41% of the studies) [5,6,37-39,41,43,50,51,56,57,63,66,73-75,84], followed by 2 months [47,49,52,54,58,59,68,71,76], with others spanning 2 to 6 weeks [40,45,62,69,72] or >15 weeks up to 6 months [42,53,60,61,64,67] to 1 year [70]. It should be noted that women in the study by Moossdorff-Steinhauser et al [52] continued exercising at home for at least 6 months after the end of the 8-week group exercise PFMT, and the 16 weeks specified by Jaffar et al [60] were the minimum program duration.

Adverse Events (Consensus on Exercise Reporting Template for PFMT Item 11)

A total of 20% (8/41) of the studies documented adverse events [6,36,37,43,52,54,62,64]. Adverse effects were reported in 12% (5/41) of the studies, none of which were deemed serious. Some examples include vaginal discomfort, infection, or allergic reactions related to the use of the vaginal device [37,43]; lower abdominal pain related to PFMT [6]; and increased spontaneous urine leakage [64].

Treatment Fidelity (Consensus on Exercise Reporting Template for PFMT Item 16)

A total of 5% (2/41) of the studies, both of which were protocols [47,60], assessed the implementation of the intervention. The methods used included monitoring participant activity and training time through the app [60], tracking technical support provided, consultation support, and reminders sent to women who had not used the app over the previous week [47].

UI Outcomes

UI outcomes following DTs were presented in 56% (23/41) of the studies (Table S1 in Multimedia Appendix 4 [5,6,39-41,43-45,49-51,53,57-59,61-64,67,68,73,76]), with improvements reported across measures in all but 1 study (within groups; 22/23, 96% [68]). All studies (23/23, 100%) analyzed changes in UI symptoms and severity or general or UI-specific QoL as measured through self-reported questionnaires, UI episode frequency, pad weight tests, or UI aid use.

Adherence to the Program

A total of 61% (25/41) of the studies indicated methods for evaluating adherence, with some (9/25, 36%) using more than one approach (Table S2 in Multimedia Appendix 4 [5,6,37,39,41-44,47,50-52,54,56,58,60,61,63,64,67-69,71,73,74]). In total, 72% (18/25) of the studies gathered data from the DTs themselves [5,37,39,42,43,47,51,52,54,56,58,60,61,64,67-69,79], 48% (12/25) used self-report via email [6] or a web- or app-based questionnaire [5,41,44,50-52,61,63,64,74,79], and 4% (1/26) included in-person appointments [42].

A total of 68% (17/25) of the studies provided data following completion of the program. Women in 76.4% (13/17, 50%) of these studies provided a self-report of adherence to the prescribed PFMT, which was measured and reported in a variety of ways—visual analogue scale [39], validated questionnaire to assess efficacy [51], number of exercises completed over specified time points (eg, in the last month or last week, daily, weekly, or monthly) [5,41,44,50,58,63,74,79], or percentage of women who performed PFMT [37,68] or adhered to the program [58] or some of the program [85]. Self-reported daily PFMT for women using DTs ranged from 23.4% to 41% over 3 months [5,63].

Performance of PFMT was also captured via the DTs in some studies (9/25, 36%), including measures of how often the app was used (eg, never, once a week, and >3 times per week) [50,61,63,64,69], mean number of exercises performed per day (1.6 [5]), median number of days PFMT was performed per week (4.9 [43]), and percentage of women who completed at least 75% of the study requirements for the program (14.4% [85]).

Two studies compared adherence to the DTs with a control group. Adherence was significantly higher (P<.001) in the group using the DTs at 1, 2, and 3 months [39], but there was no difference (P=.40) reported between the groups in the study by Carrión Pérez et al [43].

Satisfaction With DTs and Outcomes

A total of 63% (17/41) of the studies considered satisfaction with the DTs, including reporting on experiences with specific aspects (eg, PFMT, exercise logs, reminder features, and ease of accessing videos or instructions) [5,36,37,39,40,44,45,52,54,59,60,62,64,67,72,74,75] (Table S2 in Multimedia Appendix 4). Although a range of different outcome measures was used, it appears that most participants were satisfied with the DTs and would recommend them to others. An exception was the study by Dufour et al [67], in which only 18.2% (2/11) of the women would recommend the mHealth device or consider using it again; however, this response increased to 63.6% (7/11) if the device were to be modified.

Satisfaction with the program as a whole, or self-reported improvement, was reported in 27% (11/41) of the studies [5,6,40,43,44,49,52,54,58,64,71]. Responses varied with respect to overall satisfaction, ranging from not satisfied (6% in the study by Goode et al [58] and 33% in the study by Wadensten et al [64]) to somewhat satisfied (75% in the study by Goode et al [58]) and completely satisfied and symptom-free (7% in the study by Wadensten et al [64]). Although satisfaction in the study by Sjöström et al [6] was higher in the intervention (app) group at 4 months, there was no significant difference at the 1- and 2-year follow-ups; no difference between groups was reported by Carrión Pérez et al [43]. Self-reported improvement in symptoms was variable, with <25% of the women in 18% (2/11) of the studies [40,44] reporting that they were much or very much better (10.3%-23.5%) with respect to symptoms, but more women (55.7%) reported the same for self-perceived improvement in PFM strength [5].

Qualitative Synthesis: Experiences of Women and HCPs

The summary data from the completed qualitative or mixed methods studies (11/45, 24%) are presented in Multimedia Appendix 5 [31,38,42,46-48,65,67,77,78,80,86,87]. In almost all the studies, the same factors were presented as both facilitators and barriers. This demonstrates that preference for or against any given DT may be related to the individual and that personal preferences can change over time.

The results demonstrated some overarching facilitators. Participants liked the anonymity that DTs provide for the treatment of UI symptoms [31,38,46,65,86,87]. Several studies (6/11, 55%) discussed that UI is still considered a socially “taboo subject,” a topic that women can find difficult to acknowledge to themselves, let alone discuss with an HCP [31,65,77,80,86,87]. Being judged or feeling embarrassed to discuss UI was a common finding, and the opportunity to access DTs provided women with a viable, accessible, and potentially less time-consuming alternative means of seeking support [31,38,46,48,65,77,78,80,86,87]. Furthermore, the ability to use an app in the convenience of their own environment facilitated empowerment, confidence, and self-efficacy regarding the ability to manage their UI symptoms with PFMT exercises [31,38,46,48,65,77,78,86,87].

All studies (11/11, 100%) demonstrated that the knowledge content across the various apps was helpful. Knowledge included gaining a better understanding of the causes of UI, where PFMs are and what is their function, and the fact that UI is a common problem. In total, 18% (2/11) of the studies [65,87] reported that women felt less isolated after learning about the prevalence of UI.

DTs were considered successful by participants if an improvement in UI symptoms was observed [31,67,78,86,87]. Participants were reported to be more adherent when UI symptoms were more severe [31,86,87] and less adherent to PFMT as UI symptoms improved [31,65,78,86,87]. However, other personal and technological factors also influenced adherence to the various PFMT programs and, thus, UI symptom outcomes. These included the needs of other family members, especially for new mothers; concomitant health issues; and other life events [31,65,78,80,86,87].

Establishing a routine; the use of reminders, journals, and diaries; and family support went some way toward mitigating these barriers [31,46,77,78,86,87]. Although some HCPs expressed concerns about the ability of older women to use DTs [80], the studies in this review suggest that the competing time pressures experienced by women with young families, especially if they were working, were more of a barrier [31,65,67,87]. Culture was not discussed in any of the studies, so it is unclear whether the same facilitators and barriers apply across all cultures and ethnic groups.

Another common finding across the studies was concern about the ability to "correctly" contract the PFMs. Although the concept of an internal exam to determine a "correct" contraction was not always appealing [65], being unsure of whether the exercises were being performed correctly was a barrier to adherence [31,46,65,77,78,80,86,87]. Consequently, several studies (8/11, 73%) [31,46,48,65,77,78,80,87] suggested that engagement with HCPs, perhaps for an initial assessment and then for progression at a later point in time, was an important facilitator. Other studies found that HCP consultations were required to support adherence and provide encouragement and progression of PFMT in addition to the benefits of DTs [31,86]. Both consultations with HCPs and DTs (if from a recognized institution, such as a university) reassured participants that the information they received and the PFMT program they were trying were from a credible source [31,77].

As per the results in Multimedia Appendix 5, technology that was easy to set up, insert (if applicable), comfortable, and portable was more acceptable to participants [31,38,46,48,67,77,78,87].

Discussion

Principal Findings

This systematic scoping review was undertaken to explore the range and features of DTs available for managing UI. Specifically, we sought to determine whether the PFMT embedded in DTs follows best-practice guidelines, is designed for women at specific stages in life, and considers cultural contexts and the experiences of women and other relevant stakeholders.

It is evident that the medium of DT for the conservative management of UI is prevalent and continually expanding, with rapid growth apparent particularly over the last 10 years. In total, 89 studies were included in this scoping review—51% (45/89) were primary studies and 49% (44/89) supplementary papers—which is larger than the number (between 3 and 10 papers) included in several recent narrative and systematic reviews in this field [7,8,18-21,124,125]. This difference likely reflects variations in inclusion and exclusion criteria, which in this study were intentionally broad so as to encompass a range of sources, including qualitative research.

The WHO global strategy on digital health stipulates that DTs should be "people-centred, trust-based, evidence-based, effective, efficient, sustainable, inclusive, equitable and contextualised" [126]. In terms of the evidence-based dimension, it is encouraging that over half of the DTs (22/41, 54%) were developed based on evidential research or testing. The means of achieving this varied across the studies, but most adopted an iterative process of continuous testing, implementation, and refinement. IT input is obviously integral to the development of DTs, but importantly, a number of studies in this review took a user-centered approach by seeking the opinions of women with or without UI and, in some cases, HCPs who may be involved in a woman's care. Considering users' opinions, needs, and expectations at all stages of DT design is not only endorsed by the WHO [126] but is also vital in optimizing the usability and acceptability of the DTs and their adaptation to ensure effectiveness in outcomes [127]. Some studies (4/41, 10%) adopted a theoretical user-centered framework to guide the design of the DTs [47,48,56,60], and standardization and use of such frameworks by future developers will assist in continued improvements in the quality of DT apps specifically for PFMT, which could ultimately enhance the conservative management of UI.

Free and commercial PFMT apps are readily available for download from app stores, but only some are clinically sound from a PFMT perspective [128], with many lacking in terms of accuracy, content, quality, and functionality [10,128-130]. Just over half (21/41, 51%) of studies documented that the PFMT programs were drawn or adapted from a known evidence base, which suggests that they are in line with the recommendations of PFM exercise theory that lead to improvements in UI symptoms [122,131]. However, there was a large variation in the PFMT reported, including the type of exercise, dose, frequency, progression, and supervision, and some PFMT details were often incompletely reported (particularly in abstracts, which is to be expected). In addition to details about PFMT, other items in the Consensus on Exercise Reporting Template for PFMT guidelines [28] were also inconsistently adopted across the studies—less than half incorporated confirmation of a voluntary PFM contraction (19/41, 46%) or reported on adverse events (8/41, 20%) or treatment fidelity (2/41, 5%), whereas just over half used reminder systems available with the DTs (21/41, 51%). From a technological perspective, some of these items, such as reminders (eg, individualized push notifications), and other features, such as social media and gamification (used in 2/28, 7% and 6/28, 21% of the studies, respectively), are suggested to be important in supporting adherence to mHealth [11,132] and are worth considering for future DTs.

As shown across a range of studies, using DTs to deliver PFMT can be effective in improving UI symptoms and QoL. In the 56% (23/41) of the studies that reported outcome measures, improvements were seen across most outcome measures for women using DTs and, in the case of comparison groups, often for those who were receiving PFMT via an alternative method (eg, pamphlet or usual care). Many of the outcome measures were self-reported, which is appropriate, as the lived experience of women is of interest. As the qualitative data show, aspects such as convenience and reduction in symptoms were of most relevance, which reinforces the need for future studies to include qualitative components to determine relevance to the primary end user. Women's satisfaction with the program as a whole, as documented in 27% (11/41) of the studies, was variable in terms of outcome measures and data but was likely closely connected with UI outcomes. For example, in an RCT [6], the satisfaction of the women using the app was higher than that in the control group (printed PFMT) at 4 months, aligning with a significant improvement in UI symptoms; however, there was no difference between the groups at the 1- and 2-year follow-ups [88], when the effectiveness of the intervention had also waned, as had adherence to the prescribed intervention program. These findings suggest that PFMT delivered via DTs is promising as a first-line conservative management for UI, but more high-quality research, which includes long-term follow-up, is required.

There was heterogeneity in the definitions of adherence used by the studies included in this review and the methods (eg, DTs and web-based questionnaires) and measurements used to monitor this. In addition, reporting of adherence data was variable with little standardization, making comparison difficult. Among the 2 RCTs that measured adherence, in 1 (50%; 21 women), adherence was significantly better in those who used an app in the short term (up to 3 months) [39]. However, no difference was found in UI symptoms between groups, consistent with the findings of the other RCT that compared telerehabilitation and control [43]. A known problem with app use is attrition after they have been downloaded. Examples from other areas of health research suggest that approximately 20% to 25% of apps are used only once or infrequently, with use dramatically reducing to <5% over a short period (eg, 8 sessions or 15 days) [132,133]. The self-reported daily PFMT for women using DTs ranged from 24.3% to 41% over 3 months [63,79], but no long-term data were available to determine whether this followed a downward trend. There is a plethora of research that demonstrates that managing a long-term condition with regular commitment to exercise is difficult irrespective of the condition [134,135]. Therefore, factoring this typical type of human behavior into PFMT programs delivered via DT, providing reassuring statements regarding the fact that this is typical, being kind to oneself, and knowing how to start again, would be beneficial.

Other suggested benefits of using remote or app-based technologies to deliver PFMT include helping women overcome their embarrassment about seeking help for UI, improving access to health services in remote or underdeveloped areas, and enhancing cost-effectiveness [46,61,82,136]. Although using DTs in isolation may be beneficial, personal or HCP support is also recommended [11,132]. This approach aligns with best-practice guidelines for effective PFMT [122], with supervision provided to support the behavioral aspect of exercise. In this review, many studies (18/41, 44%) incorporated HCP or researcher support either synchronously (eg, in person or remotely) or asynchronously (eg, email contact), ranging from confirmation of a PFM contraction to constant monitoring of progress across the course of the program. A notable feature from the synthesis of findings from the included qualitative studies was that engagement with an HCP was an important facilitator, not only to support adherence and progression of exercises but also because women valued knowing that they were performing the PFM exercises correctly and expressed concern if they were unsure about their technique [31,46,48,65,77,78,80,86,87]. This concern is valid as inadvertently performing an incorrect PFM contraction, such as the Valsalva maneuver, could result in an increase in intra-abdominal pressure, leading to depression of the levator ani muscle and weakening of the surrounding connective tissues, which may inadvertently increase UI [137].

Interestingly, group-based supervised PFMT (either in person or remotely) was offered in 10% (4/41) of the studies [49,52,57,59]. Although results related to improvements in UI outcomes in these studies were mixed, a recent large RCT has shown that group-based PFMT is not inferior to individually supervised PFMT in older women in the treatment of UI, with both groups also undertaking a home exercise program [138]. It is known that peer support is a key strategy to help with long-term self-management as it can facilitate individual problem-solving and goal setting, which can aid with self-efficacy [139,140]. This indicates that a group-based approach to exercise likely offers further advantages to women, such as enhanced motivation to perform PFMT and reduced stigma and feelings of isolation [141]. Given the large variation in the types and levels of support and supervision currently provided for PFMT delivered via DTs, further information is needed to establish what represents best practice in terms of integrating supervision to optimize women-centered care and UI outcomes.

Culture plays a role in how women interact with DTs [15,16], perceive UI [17,142,143], and engage with PFMT and should be taken into account when designing mHealth interventions to encourage use and enhance motivation [16]. Incorporating cultural characteristics into DTs includes considering not only the user's needs and preferences related to functionality (eg, color, typeface, and layout) but also more implicit aspects such as values, health beliefs, religion, social practices, and language [144,145]. In this scoping review, most DTs originated in high-income countries such as the United Kingdom and the United States and most likely targeted the dominant culture. This is also exemplified by the finding that only 4% (2/45) of the primary papers were written in a language other than English [35,36]. However, some apps (the Tät in particular) have been translated into a number of different languages, and research teams have also sought user input to refine them further [56,63,78], processes that are some of several different methods to enable cultural relevance [144]. The iPelvis app [11,57] explicitly incorporates culturally relevant elements, and although these may be features of other DTs included in this scoping review, they were not described. It cannot be assumed that PFMT DTs developed in one culture and translated for use in another will be successful without consulting the cultural context of the women who will use it [146], meaning that user engagement is successful in its success. Therefore, to meet the remit of inclusive and equitable DTs [126] and reach women in low- and middle-income and remote countries, more understanding is needed of what culturally related insights are required to increase the acceptability of and engagement with these technologies [146].

Many studies (28/45, 62%) did not explicitly document information related to the delivery of PFMT via DTs for women at a specific stage in life. Of those that did, most focused on pregnancy or the postpartum period, a time when UI is highly prevalent, with a risk that it could persist and become a long-term condition in some women [1]. During the childbearing years, women experience competing interests for their energy and time, such as preparing for or caring for their new baby, which means that it is vital that they receive sufficient support to adopt and maintain PFMT [67,147]. Engaging with an HCP in conjunction with using DTs was identified as an important facilitator to support PFM exercise (physical and behavioral aspects) [31,46,48,65,77,78,80,87], and there is evidence demonstrating that starting PFMT in early pregnancy may reduce the risk of UI later in pregnancy or up to 6 months post partum [4]. However, pregnant or postpartum women might not seek help from an HCP as they may feel embarrassed about their UI symptoms [148] or think that UI is a “normal” occurrence before and soon after childbirth [149]. In these instances, DTs provide a convenient tool that can support and motivate women to exercise [11,132] in the comfort of their own environment, facilitating empowerment, confidence, and self-efficacy with PFMT [31,38,46,48,65,77,78,86,87]. An additional avenue for support could be further developing and integrating social media into DTs, enabling pregnant and postpartum women to connect with each other as well as with HCPs. In general, more evidence is required to establish the acceptability, design, development, and effectiveness of PFMT DTs across various age ranges, including both adolescent and older women, to ensure that the programs meet women’s needs and circumstances. However, HCPs should have some confidence integrating DTs for PFMT into their practice as, in partnership with a clinician, this may offer women another tool in the management of UI symptoms.

Limitations

As this scoping review included a wide range of studies and a variety of DTs, heterogeneity was evident across many study parameters, including the PFMT programs and UI outcomes, and the duration of the trials was relatively short, demonstrating the need for longer follow-up and high-quality data in this developing field of research. Biofeedback is broadly considered a DT; however, we only included studies that provided feedback to women via an app, meaning that we did not capture valuable data from trials of biofeedback that did not have this feature [150,151]. Many studies were from high-middle–income urban settings, which restricts the diversity of the target populations despite one of the benefits of mHealth being its ability to reach a range of people, including those in remote areas [61,82,136]. This review considered women with stress UI and, therefore, did not explore the impacts of PFMT DTs on other conditions or populations, such as urge UI or pelvic organ prolapse or men. As described previously, owing to the large volume of data, we were unable to implement some elements of our a priori protocol, such as synthesizing data from systematic reviews and rating the quality of the apps used in the included studies. Owing to space limitations, we were only able to present the themes most coherently relevant to the scoping review objectives, and in our synthesis, we did not consider how the quality ratings (high, fair, and poor) influenced the data.

Conclusions

Evidence related to PFMT delivered via DTs for the conservative management of UI continues to grow exponentially. The development of DTs specifically for this purpose is increasingly based on evidential research or testing, including the exploration of the perspectives and experiences of women and HCPs. Although large variation exists in the reported PFMT parameters, PFMT delivered via DTs is promising in terms of improving UI symptoms and QoL. To further optimize UI outcomes and promote long-term adaptation of PFMT, incorporating technological features such as reminders, social media, and gamification, together with other facilitators such as support from HCPs, could be beneficial for women with UI. A greater understanding is required of how women from different cultures and stages in life regard the acceptability, design, development, and effectiveness of PFMT DTs. This is essential to ensure that the quality and content are appropriate and inclusive so that all women and clinicians can have confidence in using these technologies.

Acknowledgments

The authors wish to sincerely thank Mrs Thelma Fisher for her assistance in developing the systematic search strategy and the Health Research Council of New Zealand (Health Delivery Research Activation Grant; reference 20/1325) for its support in funding this project.

Authors' Contributions

SJW, MDB, ARC, and JK were responsible for the conceptualization and design of this study. BS developed the search strategy and performed the database searches. Literature screening, data extraction, and quality ratings were completed by MDB, ARC, JK, BM, BS, and SJW; data synthesis and analyses were performed by BM, BS, MAP, and SJW. SJW drafted the manuscript with contributions from BM and MAP. All authors critically revised and approved the final manuscript.

Conflicts of Interest

JK is the chief executive officer of Junofem, which has developed FemFit, one of the apps reviewed in this study. She had no role in the data extraction or analyses of data related to FemFit. The authors have no other conflicts of interest to declare.

Multimedia Appendix 1
Search strategy.
[DOCX File , 40 KB - mhealth_v11i1e44929_app1.docx]

Multimedia Appendix 2
Summary of the included primary and supplementary papers (N=89) and summary of the inclusion and exclusion criteria for the primary and supplementary papers.
[DOCX File , 59 KB - mhealth_v11i1e44929_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies and participants.
[DOCX File , 51 KB - mhealth_v11i1e44929_app3.docx]

Multimedia Appendix 4
Outcomes related to urinary incontinence symptoms and satisfaction with and adherence to the pelvic floor muscle training program delivered via digital technologies.
[DOCX File , 56 KB - mhealth_v11i1e44929_app4.docx]

Multimedia Appendix 5
Summary of main themes and facilitators and barriers from the qualitative studies (N=11).
[DOCX File , 34 KB - mhealth_v11i1e44929_app5.docx]

References

1. Milsom I, Altman D, Cartwright R, Lapitan M, Nelson R, Sjostrom S. Epidemiology of urinary incontinence (UI) and other lower urinary tract symptoms (LUTS), pelvic organ prolapse (POP) and anal incontinence (AI). In: Abrams P, Cardozo L, Wagg A, Wein AJ, editors. Incontinence: 6th International Consultation on Incontinence, Tokyo, September 2016. Bristol, UK: International Continence Society (ICS) and International Consultation on Urological Diseases (ICUD); 2017:1-142.
2. Wilson L, Brown JS, Shin GP, Luc KO, Subak LL. Annual direct cost of urinary incontinence. Obstet Gynecol 2001 Sep;98(3):398-406. [doi: 10.1016/s0029-7844(01)01464-8] [Medline: 11530119]
3. Dumoulin C, Cacciari LP, Hay-Smith EJ. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women. Cochrane Database Syst Rev 2018 Oct 04;10(10):CD005654 [FREE Full text] [doi: 10.1002/14651858.CD005654.pub4] [Medline: 30288727]
4. Woodley SJ, Lawrenson P, Boyle R, Cody JD, Mørkved S, Kernohan A, et al. Pelvic floor muscle training for preventing and treating urinary and faecal incontinence in antenatal and postnatal women. Cochrane Database Syst Rev 2020 May 06;5(5):CD007471 [FREE Full text] [doi: 10.1002/14651858.CD007471.pub4] [Medline: 32378735]
5. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Jun;36(5):1369-1376. [doi: 10.1002/nau.23116] [Medline: 27611958]
6. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: a randomised controlled study with focus on pelvic floor muscle training. BJU Int 2013 Aug;112(3):362-372 [FREE Full text] [doi: 10.1111/j.1464-410X.2012.11713.x] [Medline: 23350826]
7. Bernard S, Boucher S, McLean L, Moffet H. Mobile technologies for the conservative self-management of urinary incontinence: a systematic scoping review. Int Urogynecol J 2020 Jun;31(6):1163-1174. [doi: 10.1007/s00192-019-04012-w] [Medline: 31267139]
8. Bertuit J, Barrau M, Huet S, Rejano-Campo M. Interest of mobile and internet applications in the management of stress urinary incontinence in women. A systematic review. Prog Urol 2020 Dec;30(16):1022-1037 [FREE Full text] [doi: 10.1016/j.purol.2020.09.013] [Medline: 33008716]
9. Barnes KL, Dunivan G, Jaramillo-Huff A, Krantz T, Thompson J, Jeppson P. Evaluation of smartphone pelvic floor exercise applications using standardized scoring system. Female Pelvic Med Reconstr Surg 2019 Jul;25(4):328-335. [doi: 10.1097/SPV.0000000000000563] [Medline: 29489554]
10. Ho L, Macnab A, Matsubara Y, Peterson K, Tsang B, Stothers L. Rating of pelvic floor muscle training mobile applications for treatment of urinary incontinence in women. Urology 2021 Apr;150:92-98. [doi: 10.1016/j.urology.2020.08.040] [Medline: 32890617]

11. Latorre GF, de Fraga R, Seleme MR, Mueller CV, Berghmans B. An ideal e-health system for pelvic floor muscle training adherence: systematic review. Neurourol Urodyn 2019 Jan;38(1):63-80. [doi: 10.1002/nau.23835] [Medline: 30375056]
12. Loohuis AM, Wessels NJ, Jellema P, Vermeulen KM, Slieker-Ten Hove MC, van Gemert-Pijnen JE, et al. The impact of a mobile application-based treatment for urinary incontinence in adult women: design of a mixed-methods randomized controlled trial in a primary care setting. Neurourol Urodyn 2018 Sep;37(7):2167-2176. [doi: 10.1002/nau.23507] [Medline: 29392749]
13. Casey EK, Temme K. Pelvic floor muscle function and urinary incontinence in the female athlete. Phys Sportsmed 2017 Nov;45(4):399-407. [doi: 10.1080/00913847.2017.1372677] [Medline: 28845723]
14. Cultural contexts of health and well-being, no.1: beyond bias: exploring the cultural contexts of health and well-being measurement. World Health Organization, Regional Office for Europe. 2015. URL: https://apps.who.int/iris/handle/10665/182731 [accessed 2021-10-06]
15. Alsswey AH, Al-Samarraie H, El-Qirem FA, Alzahrani AI, Alfarraj O. Culture in the design of mHealth UI: an effort to increase acceptance among culturally specific groups. Electron Libr 2020;38(2):257-272 [FREE Full text] [doi: 10.1108/el-04-2019-0097]
16. Kongjit C, Nimmolrat A, Khamaksorn A. Mobile health application for Thai women: investigation and model. BMC Med Inform Decis Mak 2022 Jul 30;22(1):202 [FREE Full text] [doi: 10.1186/s12911-022-01944-0] [Medline: 35907950]
17. Lara C, Nacey J. Ethnic differences between Maori, Pacific Island and European New Zealand women in prevalence and attitudes to urinary incontinence. N Z Med J 1994 Sep 28;107(986 Pt 1):374-376. [Medline: 7619104]
18. da Mata KR, Costa RC, Carbone É, Gimenez MM, Bortolini MA, Castro RA, et al. Telehealth in the rehabilitation of female pelvic floor dysfunction: a systematic literature review. Int Urogynecol J 2021 Feb;32(2):249-259 [FREE Full text] [doi: 10.1007/s00192-020-04588-8] [Medline: 33175229]
19. Huang Z, Wu S, Yu T, Hu A. Efficacy of telemedicine for urinary incontinence in women: a systematic review and meta-analysis of randomized controlled trials. Int Urogynecol J 2020 Aug;31(8):1507-1513. [doi: 10.1007/s00192-020-04340-2] [Medline: 32476050]
20. Hou Y, Feng S, Tong B, Lu S, Jin Y. Effect of pelvic floor muscle training using mobile health applications for stress urinary incontinence in women: a systematic review. BMC Womens Health 2022 Oct 03;22(1):400 [FREE Full text] [doi: 10.1186/s12905-022-01985-7] [Medline: 36192744]
21. Leme Nagib AB, Riccetto C, Martinho NM, Camargos Pennisi PR, Blumenberg C, Paranhos LR, et al. Use of mobile apps for controlling of the urinary incontinence: a systematic review. Neurourol Urodyn 2020 Apr;39(4):1036-1048. [doi: 10.1002/nau.24335] [Medline: 32187704]
22. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. London, UK: JBI Publication; 2020.
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Woodley SJ, Bussey MD, Clark A, Sangelaji B, Kruger J. Digital technology for women's pelvic floor muscle training in the context of pelvic floor dysfunction across the life course: a scoping review of quantitative and qualitative evidence. Open Science Framework Registries. 2021. URL: https://osf.io/7hect [accessed 2021-02-04]
25. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019 Jun 06. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2021-01-25]
26. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
27. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
28. Slade SC, Morris ME, Frawley H, Hay-Smith J. Comprehensive reporting of pelvic floor muscle training for urinary incontinence: CERT-PFMT. Physiotherapy 2021 Sep;112:103-112. [doi: 10.1016/j.physio.2021.03.001] [Medline: 34062452]
29. Critical appraisal tools. Joanna Briggs Institute. URL: https://jbi.global/critical-appraisal-tools [accessed 2023-12-20]
30. Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore, Singapore: Springer; 2019:843-860.
31. Firet L, Teunissen TA, Kool RB, van Doorn L, Aourag M, Lagro-Janssen AL, et al. Women's adoption of a web-based intervention for stress urinary incontinence: a qualitative study. BMC Health Serv Res 2021 Jun 12;21(1):574 [FREE Full text] [doi: 10.1186/s12913-021-06585-z] [Medline: 34118900]
32. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
33. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/260480 [accessed 2021-08-09]

34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Loohuis A. App for treating incontinence. Huisarts Wet 2015;58(11):610.
36. Moretti EC. Desenvolvimento de um jogo virtual destinado a aparelhos com sistema operacional ANDROID para conscientização e fortalecimento da musculatura do assoalho pélvico. Universidade Federal de Pernambuco. 2016 Jun 29. URL: https://repositorio.ufpe.br/handle/123456789/18733 [accessed 2021-11-09]
37. Anglès Acedo S, Kastelein AW, Ros C, Raatikainen K, Alonso Lopez JF, Pagès Raventos A, et al. Information and communication technologies (ICT) self-management system for pelvic floor muscle training: a pilot study in women with stress urinary incontinence. Int Urogynecol J 2018 Jun 30;29(Supplement 1):S75 [FREE Full text]
38. Anglès-Acedo S, López-Frías L, Soler V, Alonso JF, Kastelein AW, Graaf BC, et al. The WOMEN-UP solution, a patient-centered innovative e-health tool for pelvic floor muscle training: qualitative and usability study during early-stage development. Int J Environ Res Public Health 2021 Jul 23;18(15):7800 [FREE Full text] [doi: 10.3390/ijerph18157800] [Medline: 34360093]
39. Araujo CC, Marques AD, Juliato CR. The adherence of home pelvic floor muscles training using a mobile device application for women with urinary incontinence: a randomized controlled trial. Female Pelvic Med Reconstr Surg 2020 Nov;26(11):697-703. [doi: 10.1097/SPV.0000000000000670] [Medline: 30624250]
40. Barbato KA, Wiebe JW, Cline TW, Hellier SD. Web-based treatment for women with stress urinary incontinence. Urol Nurs 2014 Sep;34(5):252-257. [Medline: 26298935]
41. Bokne K, Sjöström M, Samuelsson E. Self-management of stress urinary incontinence: effectiveness of two treatment programmes focused on pelvic floor muscle training, one booklet and one internet-based. Scand J Prim Health Care 2019 Sep;37(3):380-387 [FREE Full text] [doi: 10.1080/02813432.2019.1640921] [Medline: 31317808]
42. Campbell KG, Batt ME, Drummond A. A feasibility study of the physiotherapy management of urinary incontinence in athletic women: trial protocol for the POsITIve study. Pilot Feasibility Stud 2020 Jul 16;6:103 [FREE Full text] [doi: 10.1186/s40814-020-00638-6] [Medline: 32695435]
43. Carrión Pérez F, Rodríguez Moreno MS, Carnerero Córdoba L, Romero Garrido MC, Quintana Tirado L, García Montes I. Telerehabilitation to treat stress urinary incontinence. Pilot study. Med Clin (Barc) 2015 May 21;144(10):445-448. [doi: 10.1016/j.medcli.2014.05.036] [Medline: 25087210]
44. Coggins J, Thomson HR, Cartwright R. Self-reported changes in pelvic floor training frequency and incontinence symptoms with the intravaginal Elvie device. Neurourol Urodyn 2017 Sep 13;36(Supplement 3):S162-S164 [FREE Full text]
45. Conlan L, Thompson J, Fary R. An exploration of the efficacy of telehealth in the assessment and management of stress urinary incontinence among women in rural locations. Aust N Z Continence J 2016;22(3):58-64 [FREE Full text] [doi: 10.1891/9780826123091.0018]
46. Firet L, Teunissen D, Verhoeks C, Lagro-Janssen A. Expectations regarding eHealth among women with stress urinary incontinence. Int Urogynecol J 2019 Nov;30(11):1955-1963 [FREE Full text] [doi: 10.1007/s00192-018-3849-2] [Medline: 30594948]
47. Li T, Chen X, Wang J, Chen L, Cai W. Mobile app-based intervention for pregnant women with stress urinary incontinence: protocol for a hybrid effectiveness-implementation trial. JMIR Res Protoc 2021 Mar 10;10(3):e22771 [FREE Full text] [doi: 10.2196/22771] [Medline: 33688842]
48. Pedofsky L, Nielsen PM, Budgett D, Nemec K, Dumoulin C, Kruger J. Using codesign to develop a mobile application for pelvic floor muscle training with an intravaginal device (femfit®). Neurourol Urodyn 2021 Nov;40(8):1900-1907. [doi: 10.1002/nau.24775] [Medline: 34464005]
49. Pla AA, Povar AA, Caussa JE, Tallón VB, Rexachs D, Luque E. Telemedicine program for management and treatment of stress urinary incontinence in women: design and pilot test. In: Theng LB, editor. Assistive Technologies for Physical and Cognitive Disabilities. Pennsylvania, PA, USA: IGI Global; 2014:56-77.
50. Samuelsson E, Nyström E, Söderström L. Treatment for stress urinary incontinence with the support of a mobile application is effective when implemented for free use. Neurourol Urodyn 2016;35(S4):S92-S94 [FREE Full text]
51. Wang X, Xu X, Luo J, Chen Z, Feng S. Effect of app-based audio guidance pelvic floor muscle training on treatment of stress urinary incontinence in primiparas: a randomized controlled trial. Int J Nurs Stud 2020 Apr;104:103527. [doi: 10.1016/j.ijnurstu.2020.103527] [Medline: 32058140]
52. Moossdorff-Steinhauser HF, Bols EM, Spaanderman ME, Dirksen CD, Weemhoff M, Nieman FH, et al. Long-term effects of motherfit group therapy in pre-(MOTHERFIT1) and post-partum women (MOTHERFIT2) with stress urinary incontinence compared to care-as-usual: study protocol of two multi-centred, randomised controlled trials. Trials 2019 Apr 25;20(1):237 [FREE Full text] [doi: 10.1186/s13063-019-3331-6] [Medline: 31023381]
53. Smith D. A prospective multicentre randomised controlled trial comparing PeriCoach systemassisted pelvic floor exercises with clinician guided pelvic floor exercises in the management of female stress urinary incontinence. BJU Int 2017 Mar;119:51. [doi: 10.1016/j.jsxm.2017.04.024]

54. Weinstein MM, Pulliam SJ, Richter HE. Randomized trial comparing efficacy of pelvic floor muscle training with a digital therapeutic motion-based device to standard pelvic floor exercises for treatment of stress urinary incontinence (SUV trial): an all-virtual trial design. Contemp Clin Trials 2021 Jun;105:106406. [doi: 10.1016/j.cct.2021.106406] [Medline: 33866003]
55. Åström Y, Asklund I, Lindam A, Sjöström M. Quality of life in women with urinary incontinence seeking care using e-health. BMC Womens Health 2021 Sep 20;21(1):337 [FREE Full text] [doi: 10.1186/s12905-021-01477-0] [Medline: 34544393]
56. Firet L, Teunissen TA, van der Vaart H, Assendelft WJ, Notten KJ, Kool RB, et al. An electronic health intervention for Dutch women with stress urinary incontinence: protocol for a mixed methods study. JMIR Res Protoc 2019 Jul 11;8(7):e13164 [FREE Full text] [doi: 10.2196/13164] [Medline: 31298222]
57. Fischer Blosfeld CE, de Fraga R, Regina Seleme M, Berghmans B. Assessment of the pelvic floor exercises with iPelvis® app. for treatment of women with urinary incontinence: a randomized controlled trial. Pelviperineology 2021 Jul 04;40(3):134-144 [FREE Full text] [doi: 10.34057/ppj.2021.40.02.2021-7-4]
58. Goode PS, Markland AD, Echt KV, Slay L, Barnacastle S, Hale G, et al. A mobile telehealth program for behavioral treatment of urinary incontinence in women veterans: development and pilot evaluation of MyHealtheBladder. Neurourol Urodyn 2020 Jan;39(1):432-439. [doi: 10.1002/nau.24226] [Medline: 31774200]
59. Hui E, Lee PS, Woo J. Management of urinary incontinence in older women using videoconferencing versus conventional management: a randomized controlled trial. J Telemed Telecare 2006;12(7):343-347. [doi: 10.1258/135763306778682413] [Medline: 17059650]
60. Jaffar A, Mohd Sidik S, Foo CN, Muhammad NA, Abdul Manaf R, Fadhilah Ismail SI, et al. Protocol of a single-blind two-arm (waitlist control) parallel-group randomised controlled pilot feasibility study for mHealth app among incontinent pregnant women. Int J Environ Res Public Health 2021 Apr 30;18(9):4792 [FREE Full text] [doi: 10.3390/ijerph18094792] [Medline: 33946203]
61. Loohuis AM, Wessels NJ, Dekker JH, van Merode NA, Slieker-Ten Hove MC, Kollen BJ, et al. App-based treatment in primary care for urinary incontinence: a pragmatic, randomized controlled trial. Ann Fam Med 2021 Mar;19(2):102-109 [FREE Full text] [doi: 10.1370/afm.2585] [Medline: 33685871]
62. Pulliam S, Rosenblatt P, Igleseas R. Interactive pelvic floor muscle training for female urinary incontinence. Int Urogynecol J 2018;29(Supplement 1):S179.
63. Rygh P, Asklund I, Samuelsson E. Real-world effectiveness of app-based treatment for urinary incontinence: a cohort study. BMJ Open 2021 Jan 04;11(1):e040819 [FREE Full text] [doi: 10.1136/bmjopen-2020-040819] [Medline: 33397664]
64. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A mobile app for self-management of urgency and mixed urinary incontinence in women: randomized controlled trial. J Med Internet Res 2021 Apr 05;23(4):e19439 [FREE Full text] [doi: 10.2196/19439] [Medline: 33818395]
65. Wessels NJ, Hulshof L, Loohuis AM, van Gemert-Pijnen L, Jellema P, van der Worp H, et al. User experiences and preferences regarding an app for the treatment of urinary incontinence in adult women: qualitative study. JMIR Mhealth Uhealth 2020 Jun 12;8(6):e17114 [FREE Full text] [doi: 10.2196/17114] [Medline: 32530431]
66. Asklund I, Samuelsson E. The use of an app with a PFMT programme among pregnant and postnatal women for preventive use and treatment of urinary incontinence. Neurourol Urodyn 2019;38(Supplement 3):S452-S453 [FREE Full text]
67. Dufour S, Fedorkow D, Kun J, Deng SX, Fang Q. Exploring the impact of a mobile health solution for postpartum pelvic floor muscle training: pilot randomized controlled feasibility study. JMIR Mhealth Uhealth 2019 Jul 11;7(7):e12587 [FREE Full text] [doi: 10.2196/12587] [Medline: 31298221]
68. Kinouchi K, Ohashi K. Smartphone-based reminder system to promote pelvic floor muscle training for the management of postnatal urinary incontinence: historical control study with propensity score-matched analysis. PeerJ 2018 Feb 09;6:e4372 [FREE Full text] [doi: 10.7717/peerj.4372] [Medline: 29441244]
69. Li J, Sun X, Wang C, Zhang Z, Xie Z. A mobile application penyikang applied in postpartum pelvic floor dysfunction: a cross-sectional study to analyze the factors influencing postpartum pelvic floor muscle strength and women's participation in treatment. Biomed Res Int 2020 Jul 28;2020:4218371 [FREE Full text] [doi: 10.1155/2020/4218371] [Medline: 32775419]
70. von Au A, Wallwiener S, Matthies LM, Friedrich B, Keim S, Wallwiener M, et al. The burden of incontinence in a real-world data environment-insights from a digital patient companion. Int Urogynecol J 2022 Apr;33(4):977-984 [FREE Full text] [doi: 10.1007/s00192-021-04683-4] [Medline: 33704535]
71. Cornelius C. Pelvic floor muscle exercise technique guidance using PeriCoach system. BJU Int 2018;121(Supplement 1):8-9.
72. Han MN, Grisales T, Sridhar A. Evaluation of a mobile application for pelvic floor exercises. Telemed J E Health 2019 Feb;25(2):160-164. [doi: 10.1089/tmj.2017.0316] [Medline: 29792574]
73. Nyström E, Söderström L, Samuelsson E. Factors associated with completing self-management and achieving improvement with a free mobile app for urinary incontinence. Neurourol Urodyn 2018 Jul;37(Supplement 5):S64-S65 [FREE Full text]
74. Robson M. The Squeezy pelvic floor muscle exercise app: user satisfaction survey. J Pelvic Obstet Gynaecol Physiother 2017 Sep;75(9):64-68 [FREE Full text] [doi: 10.1073/pnas.75.9.4422]

75. Saboia DM, Vasconcelos CT, Oriá MO, de C Bezerra K, Vasconcelos Neto JA, de M Lopes MH. Continence app: construction and validation of a mobile application for postnatal urinary incontinence prevention. Eur J Obstet Gynecol Reprod Biol 2019 Sep;240:330-335. [doi: 10.1016/j.ejogrb.2019.07.026] [Medline: 31382145]
76. Shelly B. Pelvic muscle exercises using a home trainer for pelvic muscle dysfunction: a case report. Urol Nurs 2016 Mar;36(2):82-87. [Medline: 27281865]
77. Grant A, Currie S. Qualitative exploration of the acceptability of a postnatal pelvic floor muscle training intervention to prevent urinary incontinence. BMC Womens Health 2020 Jan 17;20(1):9 [FREE Full text] [doi: 10.1186/s12905-019-0878-z] [Medline: 31952500]
78. Asklund I, Samuelsson E, Hamberg K, Umefjord G, Sjöström M. User experience of an app-based treatment for stress urinary incontinence: qualitative interview study. J Med Internet Res 2019 Mar 14;21(3):e11296 [FREE Full text] [doi: 10.2196/11296] [Medline: 30869644]
79. Nyström E, Asklund I, Sjöström M, Stenlund H, Samuelsson E. Treatment of stress urinary incontinence with a mobile app: factors associated with success. Int Urogynecol J 2018 Sep;29(9):1325-1333 [FREE Full text] [doi: 10.1007/s00192-017-3514-1] [Medline: 29222718]
80. Firet L, de Bree C, Verhoeks CM, Teunissen DA, Lagro-Janssen AL. Mixed feelings: general practitioners' attitudes towards eHealth for stress urinary incontinence - a qualitative study. BMC Fam Pract 2019 Jan 26;20(1):21 [FREE Full text] [doi: 10.1186/s12875-019-0907-x] [Medline: 30684962]
81. Sidik SM, Jaffar A, Foo CN, Muhammad NA, Abdul Manaf R, Ismail SI, et al. KEPT-app trial: a pragmatic, single-blind, parallel, cluster-randomised effectiveness study of pelvic floor muscle training among incontinent pregnant women: study protocol. BMJ Open 2021 Jan 12;11(1):e039076 [FREE Full text] [doi: 10.1136/bmjopen-2020-039076] [Medline: 33436465]
82. Sjöström M, Lindholm L, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a cost-effectiveness analysis. J Med Internet Res 2017 May 08;19(5):e154 [FREE Full text] [doi: 10.2196/jmir.7383] [Medline: 28483745]
83. Starr JA, Drobnis EZ, Cornelius C. Pelvic floor biofeedback via a smart phone app for treatment of stress urinary incontinence. Urol Nurs 2016 Mar;36(2):88-97. [Medline: 27281866]
84. Pedofsky L, Nielsen PM, Budgett D, Nemec K, Dumoulin C, Kruger J. Urine luck! Design and development of a mobile application for an intra-vaginal device intended for pelvic floor muscle training. Urol Nur 2021 Aug 31;40(8):1900-1907 [FREE Full text] [doi: 10.1002/nau.24775]
85. Cornelius C, Monsour M, Noursalehi M. PeriCoach® 8-week structured programming real-world data and impact of clinical engagement. Int Urogynecol J 2019;30(1 Supplement):S157-S158.
86. Björk AB, Sjöström M, Johansson EE, Samuelsson E, Umefjord G. Women's experiences of internet-based or postal treatment for stress urinary incontinence. Qual Health Res 2014 Apr;24(4):484-493 [FREE Full text] [doi: 10.1177/1049732314524486] [Medline: 24598777]
87. Wessels NJ, Loohuis AM, van der Worp H, Abbenhuis L, Dekker J, Berger MY, et al. Barriers and facilitators associated with app-based treatment for female urinary incontinence: mixed methods evaluation. JMIR Mhealth Uhealth 2021 Sep 17;9(9):e25878 [FREE Full text] [doi: 10.2196/25878] [Medline: 34533466]
88. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: 1- and 2-year results of a randomized controlled trial with a focus on pelvic floor muscle training. BJU Int 2015 Dec;116(6):955-964 [FREE Full text] [doi: 10.1111/bju.13091] [Medline: 25683075]
89. Juliato C, Araujo C, Marques A. Mobile app increases the adherence of pelvic floor muscles training for women with urinary incontinence. Int Urogynecol J 2018;29(Supplement 1):S49 [FREE Full text]
90. Nyström E, Asklund I, Sjöström M, Stenlund H, Samuelsson E. High expectations for pelvic floor muscle training with mobile application predicts successful treatment in women with stress urinary incontinence. Neurourol Urodynam 2015;34:S448-S449 [FREE Full text] [doi: 10.1002/nau.22830]
91. Hoffman V, Söderström L, Samuelsson E. Self-management of stress urinary incontinence via a mobile app: two-year follow-up of a randomized controlled trial. Acta Obstet Gynecol Scand 2017 Oct;96(10):1180-1187 [FREE Full text] [doi: 10.1111/aogs.13192] [Medline: 28718223]
92. Joob B, Wiwanitkit V. Treatment of stress urinary incontinence with a mobile app. Int Urogynecol J 2018 Apr;29(4):613. [doi: 10.1007/s00192-018-3590-x] [Medline: 29508042]
93. Nyström E, Asklund I, Sjöström M, Stenlund H, Samuelsson E. Re: treatment of stress urinary incontinence with a mobile app: factors associated with success. Int Urogynecol J 2018 Jun;29(6):925. [doi: 10.1007/s00192-018-3632-4] [Medline: 29594320]
94. Coggins J, Cartwright R, Bergmann J. Measures of pelvic floor strength by age and parity using the Elvie device. Eur Urol Suppl 2017 Mar;16(3):e1915-e1916 [FREE Full text] [doi: 10.1016/s1569-9056(17)31149-1]
95. Cornelius C, Monsour M, Noursalehi M. PeriCoach clinical study and real-world data insights. Int Urogynecol J 2019;30(1 Supplement):S332.
96. Dufour S, Fedorkow D, Fang Q. The use of mobile health technology to support post-partum pelvic health: a randomized mixed methods pilot study. Neurourol Urodyn 2018;37(Supplement 5):S70-S71 [FREE Full text]
97. Han MN, Grisales T, Sridhar A. Evaluation of a mobile application for pelvic floor exercises: a pilot study. Female Pelvic Med Reconstr Surg 2019 Feb;23(5 Supplement 1):S91-S92 [FREE Full text] [doi: 10.1089/tmj.2017.0316]

98. Loohuis AM, Wessels NJ, Dekker JH, van Merode NA, Slieker-ten Hove MC, Kollen BJ, et al. App-based treatment for urinary incontinence in women: a pragmatic, randomized, controlled, non-inferiority trial in primary care setting. Neurourol Urodyn 2019;38(Supplement 3):S363-S365.
99. Heidelbaugh JJ. Self-directed technology to improve urinary symptoms. Ann Fam Med 2021 Mar;19(2):100-101 [FREE Full text] [doi: 10.1370/afm.2659] [Medline: 33685870]
100. Loohuis AM, Van Der Worp H, Wessels NJ, Dekker JH, Slieker-Ten Hove MC, Berger MY, et al. One year effectiveness of an app-based treatment for urinary incontinence in comparison to care as usual in Dutch general practice: a pragmatic randomised controlled trial over 12 months. BJOG 2022 Aug;129(9):1474-1480 [FREE Full text] [doi: 10.1111/1471-0528.16875] [Medline: 34437756]
101. Moretti E, de Moura Filho AG, de Almeida JC, Araujo CM, Lemos A. Electromyographic assessment of women's pelvic floor: what is the best place for a superficial sensor? Neurourol Urodyn 2017 Sep;36(7):1917-1923. [doi: 10.1002/nau.23212] [Medline: 28220534]
102. Silva MB, Silva CM, Lucena RJ, Alves GJ, Cavalcante EL, Moretti EC, et al. Rehabilitation physiotherapy of the perineum muscle through virtual reality. In: Proceedings of the XXVI Brazilian Congress on Biomedical Engineering. 2019 Presented at: CBEB '18; October 21-25, 2018; Armação de Buzios, Brazil p. 875-880 URL: https://link.springer.com/chapter/10.1007/978-981-13-2517-5_136 [doi: 10.1007/978-981-13-2517-5_136]
103. Cacciari LP, Kruger J, Goodman J, Budgett D, Dumoulin C. Reliability and validity of intravaginal pressure measurements with a new intravaginal pressure device: the FemFit®. Neurourol Urodyn 2020 Jan;39(1):253-260. [doi: 10.1002/nau.24179] [Medline: 31588623]
104. Rosenblatt PS, Pulliam SJ, McKinney JL, Sutherland R, Iglesias R. The leva incontinence system for the treatment of mild to moderate urinary incontinence: a pilot study. Female Pelvic Med Reconstr Surg 2018;24(5 Supplement 1):S121-S122.
105. Rosenblatt P, McKinney J, Rosenberg RA, Iglesias RJ, Sutherland RC, Pulliam SJ. Evaluation of an accelerometer-based digital health system for the treatment of female urinary incontinence: a pilot study. Neurourol Urodyn 2019 Sep;38(7):1944-1952 [FREE Full text] [doi: 10.1002/nau.24097] [Medline: 31310369]
106. Bohorquez J, McKinney J, Keyser L, Sutherland R, Pulliam SJ. Development of a wireless accelerometer-based Intravaginal device to detect pelvic floor motion for evaluation of pelvic floor dysfunction. Biomed Microdevices 2020 Mar 17;22(2):26 [FREE Full text] [doi: 10.1007/s10544-020-00479-3] [Medline: 32185505]
107. Nyström E, Söderström L, Samuelsson E. Self-management of incontinence using a free mobile app: factors associated with improvement. Int Urogynecol J 2022 Apr;33(4):877-885 [FREE Full text] [doi: 10.1007/s00192-021-04755-5] [Medline: 33825925]
108. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: a randomised controlled study. Neurourol Urodyn 2012;31(6):734-736. [doi: 10.1111/j.1464-410x.2012.11713.x]
109. Griebling TL. Re: internet-based treatment of stress urinary incontinence: a randomised controlled study with focus on pelvic floor muscle training. J Urol 2014 Sep 01;192(3):863-865 [FREE Full text] [doi: 10.1016/j.juro.2014.06.042]
110. Sjöström M, Umefjord G, Lindholm L, Samuelsson E. Cost-effectiveness of an internet-based treatment program for stress urinary incontinence. Neurourol Urodyn 2015 Mar;34(3):244-250 [FREE Full text] [doi: 10.1002/nau.22540] [Medline: 24347521]
111. Lindh A, Sjöström M, Stenlund H, Samuelsson E. Non-face-to-face treatment of stress urinary incontinence: predictors of success after 1 year. Int Urogynecol J 2016 Dec;27(12):1857-1865 [FREE Full text] [doi: 10.1007/s00192-016-3050-4] [Medline: 27260323]
112. Wein AJ. Re: cost-effectiveness of an internet-based treatment program for stress urinary incontinence. J Urol 2016 Aug;196(2):498-499. [doi: 10.1016/j.juro.2016.05.072] [Medline: 27479399]
113. Wilson Edwards A, Sherburn M, Rane A, Boniface R, Smith D, O'Hazy J. A prospective multicentre randomised controlled trial comparing unassisted pelvic floor exercises with PeriCoach system-assisted pelvic floor exercises in the management of female stress urinary incontinence. BJU Int 2016;117:14-15. [doi: 10.1016/j.jsxm.2017.04.024]
114. Cabral C, Von Au A, Wallwiener M, Matthies L, Friedrich B. The burden of incontinence in a real-world data environment - Insights from a digital prevention companion. Value in Health 2019 Nov;22(Supplement 3):S919. [doi: 10.1016/j.jval.2019.09.2719]
115. Friedrich B, Von Au A, Wallwiener M, Matthies L, Cabral C. The effectiveness of a digital pelvic floor training program in the treatment of urinary incontinence - a real world evaluation. Value Health 2019 Nov;22(Supplement 3):S917-S918 [FREE Full text] [doi: 10.1016/j.jval.2019.09.2709]
116. Wadensten T, Nyström E, Franzén K, Stenzelius K, Malmberg L, Samuelsson E. Development and use of an algorithm for identifying women with urgency or mixed urinary incontinence suitable for e-health treatment. Neurourol Urodyn 2018;37(Supplement 5):S71-S72 [FREE Full text]
117. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A smartphone app for self-management of urgency and mixed urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2021 Apr 5;38:S361-S363 [FREE Full text] [doi: 10.2196/19439]
118. Wessels N, Hulshof L, Loohuis A, Jellema P, Blanker M. APP-based treatment for women with urinary incontinence: what do patients experience and prefer? Neurourol Urodyn 2018;37(Supplement 5):S64-S65 [FREE Full text]

119. Bø K, Hagen RH, Kvarstein B, Jørgensen J, Larsen S, Burgio KL. Pelvic floor muscle exercise for the treatment of female stress urinary incontinence: III. Effects of two different degrees of pelvic floor muscle exercises. Neurourol Urodyn 1990;9(5):489-502 [FREE Full text] [doi: 10.1002/nau.1930090505]
120. Dumoulin C, Morin M, Mayrand MH, Tousignant M, Abrahamowicz M. Group physiotherapy compared to individual physiotherapy to treat urinary incontinence in aging women: study protocol for a randomized controlled trial. Trials 2017 Nov 16;18(1):544 [FREE Full text] [doi: 10.1186/s13063-017-2261-4] [Medline: 29145873]
121. Bø K, Berghmans B, Mørkved S, Van Kampen M. Evidence-Based Physical Therapy for the Pelvic Floor: Bridging Science and Clinical Practice. 2nd edition. London, UK: Churchill Livingstone; 2015.
122. NICE guideline: urinary incontinence and pelvic organ prolapse in women: management. National Institute for Health and Care Excellence. 2019 Jun 24. URL: https://www.nice.org.uk/guidance/ng123 [accessed 2022-11-09]
123. Bernards AT, Berghmans BC, Slieker-Ten Hove MC, Staal JB, de Bie RA, Hendriks EJ. Dutch guidelines for physiotherapy in patients with stress urinary incontinence: an update. Int Urogynecol J 2014 Feb;25(2):171-179 [FREE Full text] [doi: 10.1007/s00192-013-2219-3] [Medline: 24081496]
124. Chen X, Hu Y, Chen L, Zhang Y, Huang S, Cai W. Development of an intra-pregnancy pelvic floor muscle training program using mobile health technologies . Chinese Gen Pract 2020;23(30):3806-3811 [FREE Full text] [doi: 10.12114/j.issn.1007-9572.2020.00.137]
125. Jaffar A, Tan CE, Mohd-Sidik S, Admodisastro N, Goodyear-Smith F. Persuasive technology in an mHealth app designed for pelvic floor muscle training among women: systematic review. JMIR Mhealth Uhealth 2022 Mar 22;10(3):e28751 [FREE Full text] [doi: 10.2196/28751] [Medline: 35315777]
126. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-11-09]
127. Good A, Omisade O. Linking activity theory with user centred design: a human computer interaction framework for the design and evaluation of mHealth interventions. Stud Health Technol Inform 2019 Jul 30;263:49-63. [doi: 10.3233/SHTI190110] [Medline: 31411152]
128. Hay-Smith J, Peebles L, Farmery D, Dean S, Grainger R. Apps-olutely fabulous?-The quality of PFMT smartphone app content and design rated using the mobile app rating scale, behaviour change taxonomy, and guidance for exercise prescription. Neurourol Urodyn 2019;38(Supplement 3):S449-S450 [FREE Full text]
129. Dantas LO, Carvalho C, Santos BL, Ferreira CH, Bø K, Driusso P. Mobile health technologies for the management of urinary incontinence: a systematic review of online stores in Brazil. Braz J Phys Ther 2021;25(4):387-395 [FREE Full text] [doi: 10.1016/j.bjpt.2021.01.001] [Medline: 33468447]
130. Sudol NT, Adams-Piper E, Perry R, Lane F, Chen KT. In search of mobile applications for patients with pelvic floor disorders. Female Pelvic Med Reconstr Surg 2019 May;25(3):252-256. [doi: 10.1097/SPV.0000000000000527] [Medline: 29219859]
131. Dumoulin C, Adewuyi T, Booth J, Bradley C, Burgio B, Hagen S. Adult conservative management. In: Abrams P, Cardozo L, Wagg A, Wein AJ, editors. Incontinence: 6th International Consultation on Incontinence, Tokyo, September 2016. Bristol, UK: International Continence Society; 2017:1443-1628.
132. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
133. Fuente-Vidal A, Guerra-Balic M, Roda-Noguera O, Jerez-Roig J, Montane J. Adherence to eHealth-delivered exercise in adults with no specific health conditions: a scoping review on a conceptual challenge. Int J Environ Res Public Health 2022 Aug 17;19(16):10214 [FREE Full text] [doi: 10.3390/ijerph191610214] [Medline: 36011856]
134. Marcus B, Lewis B. Physical activity and the stages of motivational readiness for change model. President's Council on Physical Fitness and Sports. 2003 Mar. URL: https://files.eric.ed.gov/fulltext/ED474896.pdf [accessed 2022-12-01]
135. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6 [FREE Full text] [doi: 10.1080/17437199.2012.654964]
136. Addressing mobile health. World Health Organization. URL: https://www.who.int/activities/Addressing-mobile-health [accessed 2020-06-26]
137. Thompson JA, O'Sullivan PB. Levator plate movement during voluntary pelvic floor muscle contraction in subjects with incontinence and prolapse: a cross-sectional study and review. Int Urogynecol J Pelvic Floor Dysfunct 2003 Jun;14(2):84-88. [doi: 10.1007/s00192-003-1036-5] [Medline: 12851748]
138. Dumoulin C, Morin M, Danieli C, Cacciari L, Mayrand MH, Tousignant M, Urinary Incontinence and Aging Study Group. Group-based vs individual pelvic floor muscle training to treat urinary incontinence in older women: a randomized clinical trial. JAMA Intern Med 2020 Oct 01;180(10):1284-1293 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2993] [Medline: 32744599]
139. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 01;40:127-146 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]

140. Du S, Yuan C, Xiao X, Chu J, Qiu Y, Qian H. Self-management programs for chronic musculoskeletal pain conditions: a systematic review and meta-analysis. Patient Educ Couns 2011 Dec;85(3):e299-e310. [doi: 10.1016/j.pec.2011.02.021] [Medline: 21458196]
141. Cook T. Group treatment of female urinary incontinence: literature review. Physiother 2001 May;87(5):226-234 [FREE Full text] [doi: 10.1016/s0031-9406(05)60783-3]
142. Chaliha C, Stanton SL. The ethnic cultural and social aspects of incontinence--a pilot study. Int Urogynecol J Pelvic Floor Dysfunct 1999;10(3):166-170. [doi: 10.1007/s001920050039] [Medline: 10430009]
143. Laganà L, Bloom DW, Ainsworth A. Urinary incontinence: its assessment and relationship to depression among community-dwelling multiethnic older women. ScientificWorldJournal 2014;2014:708564 [FREE Full text] [doi: 10.1155/2014/708564] [Medline: 24982981]
144. Dobson R, Whittaker R, Bartley H, Connor A, Chen R, Ross M, et al. Development of a culturally tailored text message maternal health program: TextMATCH. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e49 [FREE Full text] [doi: 10.2196/mhealth.7205] [Medline: 28428159]
145. Khaddam I, Vanderdonckt J. Towards a culture-adaptable user-interface architecture. Romanian J Hum Comput Interact 2014;7(2):161-194 [FREE Full text]
146. Müller AM. Behavioural mHealth in developing countries: what about culture? Eur Health Psychol 2016;18(6):294-296 [FREE Full text]
147. Woodley SJ, Hay-Smith EJ. Narrative review of pelvic floor muscle training for childbearing women-why, when, what, and how. Int Urogynecol J 2021 Jul;32(7):1977-1988. [doi: 10.1007/s00192-021-04804-z] [Medline: 33950309]
148. Strahle A, Stainton MC. Women's experience of revealing perinatal bladder function--implications for midwifery care. Women Birth 2006 Mar;19(1):17-21. [doi: 10.1016/j.wombi.2006.01.001] [Medline: 16792000]
149. Buurman MB, Lagro-Janssen AL. Women's perception of postpartum pelvic floor dysfunction and their help-seeking behaviour: a qualitative interview study. Scand J Caring Sci 2013 Jun;27(2):406-413. [doi: 10.1111/j.1471-6712.2012.01044.x] [Medline: 22924517]
150. Huebner M, Riegel K, Hinninghofen H, Wallwiener D, Tunn R, Reisenauer C. Pelvic floor muscle training for stress urinary incontinence: a randomized, controlled trial comparing different conservative therapies. Physiother Res Int 2011 Sep;16(3):133-140. [doi: 10.1002/pri.489] [Medline: 20848671]
151. Hagen S, Bugge C, Dean SG, Elders A, Hay-Smith J, Kilonzo M, et al. Basic versus biofeedback-mediated intensive pelvic floor muscle training for women with urinary incontinence: the OPAL RCT. Health Technol Assess 2020 Dec;24(70):1-144 [FREE Full text] [doi: 10.3310/hta24700] [Medline: 33289476]

Abbreviations

DT: digital technology
GROUP: *Group Rehabilitation Or IndividUal Physiotherapy for Urinary Incontinence in Aging Women*
HCP: health care provider
mHealth: mobile health
PFM: pelvic floor muscle
PFMT: pelvic floor muscle training
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
QoL: quality of life
RCT: randomized controlled trial
UI: urinary incontinence
WHO: World Health Organization

Edited by L Buis; submitted 09.12.22; peer-reviewed by M Blanker, J Hay-Smith; comments to author 25.01.23; revised version received 21.03.23; accepted 21.04.23; published 05.07.23.

Please cite as:
Woodley SJ, Moller B, Clark AR, Bussey MD, Sangelaji B, Perry M, Kruger J
Digital Technologies for Women's Pelvic Floor Muscle Training to Manage Urinary Incontinence Across Their Life Course: Scoping Review
JMIR Mhealth Uhealth 2023;11:e44929
URL: https://mhealth.jmir.org/2023/1/e44929
doi:10.2196/44929
PMID:37405818

Review

Parents' Perceptions of Children's and Adolescents' Use of Electronic Devices to Promote Physical Activity: Systematic Review of Qualitative Evidence

María Eugenia Visier-Alfonso[1], PhD; Mairena Sánchez-López[2], PhD; Beatriz Rodríguez-Martín[3], PhD; Abel Ruiz-Hermosa[4], PhD; Raquel Bartolomé-Gutiérrez[5], PhD; Irene Sequí-Domínguez[6], PhD; Vicente Martínez-Vizcaíno[1,6], MD

[1]Faculty of Nursing, University of Castilla-La Mancha, Cuenca, Spain
[2]School of Education, University of Castilla-La Mancha, Ciudad Real, Spain
[3]Faculty of Occupational Therapy, Logopedia and Nursing, University of Castilla-La Mancha, Toledo, Talavera de la Reina, Spain
[4]Department of Didactics of Musical, Plastic and Body Expression, Faculty of Sports and Sciences, University of Extremadura, Cáceres, Spain
[5]Department of Psychology, University of Castilla-La Mancha, Albacete, Spain
[6]Health and Social Research Center, University of Castilla-La Mancha, Cuenca, Spain

Corresponding Author:
María Eugenia Visier-Alfonso, PhD
Faculty of Nursing
University of Castilla-La Mancha
Camino de Nohales 4
Cuenca, 16071
Spain
Phone: 34 630872012
Email: mariaeugenia.visier@uclm.es

Abstract

Background: The use of physical activity (PA) electronic devices offers a unique opportunity to engage children and adolescents in PA. For this age group (2-17 years), parents play a key role in promoting healthy lifestyles and regulating the use of electronic devices. Therefore, parents' perceptions of the use of electronic devices for PA in children and adolescents are critical for efficient intervention.

Objective: The aim of this qualitative systematic review was to improve the understanding of parents' perceptions of the use of electronic devices for PA in children and adolescents.

Methods: A systematic search of electronic databases (Medline/PubMed, SPORTDiscus, Web of Science, Scopus, OpenGrey, and Deep Blue) was conducted. Studies from inception (2010) to May 2022 were identified. Qualitative studies on the perceptions of healthy children's and adolescents' (aged 2-17 years) parents regarding PA interventions performed on electronic devices were included according to the Cochrane Qualitative and Implementation Methods Group Guidance Series and the Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) statement. The Joanna Briggs Institute Qualitative Assessment and Review Instrument was used for methodological validity.

Results: In total, 18 studies with 410 parents, mostly mothers, were included. Parents' perceptions were grouped into 4 categories: usefulness, advantages, general perceptions (electronic devices for health promotion, preferences for real-life PA, and concerns), and acceptability (barriers and facilitators) of electronic devices for PA. Parents perceived electronic devices as useful for increasing PA, learning new skills, and increasing motivation for PA and valued those devices that promoted socialization and family and peer bonding. In terms of general perceptions, parents had positive attitudes toward PA electronic devices; however, they preferred outdoor and real-life PA, especially for preschoolers and children. Concerns, such as physical and psychological harm, addiction, conflicts, and compliance difficulties, were found. Facilitators were identified as ease of use, appropriate feedback, promotion of socialization, and motivational strategies, such as rewards, challenges, and attractiveness. Barriers, such as discomfort, price, and difficulties in using or understanding electronic devices, were also identified. For older children and adolescents, parents were more concerned about high levels of screen time and setting limits on electronic devices and therefore preferred PA electronic devices rather than traditional ones.

Conclusions: Overall, the participants had positive attitudes toward electronic devices for PA and perceived them as an effective way to promote PA in children and adolescents. They also perceived several benefits of using electronic devices, such as health promotion, increased awareness and motivation, and socialization, as well as barriers, facilitators, and age differences. The results of this study could provide researchers with insights into designing more effective, age-appropriate PA electronic devices for children and adolescents and improving adherence to their use.

Trial Registration: PROSPERO CRD42021292340; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=292340

(JMIR Mhealth Uhealth 2023;11:e44753) doi:10.2196/44753

KEYWORDS

physical activity; electronic devices; eHealth; parents' perceptions; children; adolescents; systematic review; qualitative

Introduction

Currently, smartphones, tablets, computers, and apps that run on electronic devices have become part of the everyday life of children and adolescents [1]. Most parents allow their children to use their smartphones to play games or watch videos, and almost all children start handling electronic devices before the age of 1 year [2]. In addition, 73% of parents with children aged 9-11 years say that their children use a computer, 68% say that they use gaming devices, 67% say that they use a smartphone, and 78% say that they use a tablet [1]. There are substantial age differences in the use of electronic devices, and usage increases with age, being higher in adolescents, with most of them reporting using electronic devices daily or almost all the time [3]. Traditionally, research on the use of electronic devices has focused on its association with sleep problems, sedentarism, and overweight/obesity [4]. However, with the growth in technology, the use of eHealth (ie, electronic devices with health-related purposes [5], including physical activity [PA] and fitness apps), has increased [6].

Some advantages of using electronic devices to implement PA interventions are that these programs are more flexible, can be tailored to individual needs, and can be delivered anywhere at any time compared to traditional PA interventions [7]. Moreover, electronic devices might make PA more attractive to children and adolescents [8], as well as having other advantages, such as low cost, empowerment of participants, exposure to new information, increased opportunities for social contact, and new opportunities to access health promotion programs [9]. The potential role of apps in improving PA across children and adolescents has been suggested [10], but evidence of the efficacy of PA apps for this age group is still scarce [10,11]. Thus, more research on electronic devices to promote PA in children and adolescents is needed.

Furthermore, early habits track from childhood through adolescence to adulthood [12], making early childhood a crucial period for the acquisition of habits, such as PA. In addition, parents' behaviors related to PA have been shown to be associated with their children's health behaviors [13]. Previous research indicates that PA programs that include families are more effective in increasing PA in children [14,15]. Moreover, a meta-analysis by Hammersley et al [16] suggested that eHealth interventions might be more successful when parents are involved as agents of change. Not only health-related behaviors but also screen time and electronic device access and use depend on the individual's family [17]. Additionally, parents' attitudes toward electronic devices are associated with different regulation practices, depending on age and the time spent using electronic devices from childhood through adolescence [18]. All these results recommend parents' involvement in eHealth interventions [19], with the family being a key intervention target [20]. Finally, from a qualitative perspective, Burrows et al [21] found that most parents are interested in an online eHealth family program and that they feel that important features of the program should be easy to use, engaging, and endorsed by a reputable source and should involve their children directly.

To examine the feasibility of PA interventions delivered through electronic devices, before implementing the interventions, it is critical to understand parents' perceptions of the interventions because parents' engagement in these activities is a key factor for their success in children [21] and in the regulation and mediation practices that control electronic device use in adolescents [22]. However, to date, no reviews have focused on parents' opinions and perceptions of eHealth to promote PA in children and adolescents, although this knowledge might be relevant for the design of both PA electronic devices and effective interventions. The aim of this systematic review of qualitative evidence is to increase the understanding of parents' perceptions of electronic device–based PA interventions in children and adolescents.

Methods

Overview

This review was conducted according to the Cochrane Qualitative and Implementation Methods Group Guidance Series [23] and the Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) statement [24]. The review protocol was registered in PROSPERO (CRD42021292340).

Eligibility Criteria

Studies were eligible for inclusion if they reported qualitative research analyses of the use of electronic devices for PA in healthy children and adolescents. In this study, electronic devices were defined as tools that can receive, store, process, or send digital information, including computers, tablets, smartphones, smart or electronic watches, and virtual reality devices [25]. Studies using qualitative designs with any of the following data collection procedures were eligible for inclusion: interviews, focus groups, or other qualitative data collection procedures, such as observation. Mixed methods studies were included when quantitative and qualitative data were separately reported; however, only data on qualitative analyses were considered. There are different types of electronic devices (ie, activity trackers, video games, smartphone apps) for direct use by children, for use by parents to enhance their children's PA, or for use by both together.

Studies were excluded if (1) parents were not directly asked; (2) PA interventions referred participants to rehabilitation programs or facilities; (3) populations had developmental disabilities, developmental delays, or cognitive impairment; (4) the electronic device was not designed for use by children or adolescents or for interactive use by parents and children (eg, electronic devices for parents' use only); and (5) the study was a protocol, review, or meta-synthesis.

Search Strategy

Two authors (MVA and ARH) independently identified qualitative studies published from the beginning (in 2010) up to May 2022, reporting parents' perceptions of PA electronic devices. The research objective was addressed with the question framework PerSPecTIF proposed by Booth et al [26]. Both authors systematically searched Medline/PubMed, SPORTDiscus, Web of Science, and Scopus using a search strategy that combined 5 different concepts: "electronic devices," "physical activity," "parents," "qualitative research," and "children and adolescents." The free-text terms and Medical Subject Headings (MeSH) terms used to search were restricted to titles/abstracts. Searches for gray literature (eg, unpublished studies) were conducted using OpenGrey and Deep Blue. In addition, the 2 authors screened the reference lists of the papers included. The complete search strategy is presented in Multimedia Appendix 1.

Study Selection

Search terms were entered into each database, and duplicates were removed. The titles and abstracts retrieved were independently assessed for eligibility for inclusion in the review by 2 authors (MVA and ARH) and coded as "yes," "no," or "maybe." The 2 authors were trained regarding study inclusion/exclusion criteria before completing the coding of abstracts. Any disagreements between the 2 authors were resolved through discussion, and if disagreement persisted, a third author (MSL) was consulted.

Assessment of Methodological Quality

Papers selected for inclusion were assessed by 2 authors (MVA and MSL) using the 10-item checklist of the Johanna Briggs Institute Qualitative Assessment and Review Instrument (JBI-QARI) [27] for methodological validity prior to inclusion in the review. All items in the checklist were ranked as "yes," "no," or "unclear." Finally, each study was rated overall as "included," "excluded," or "seeking further info" [27]. Studies meeting more than 7 items were rated as "included," studies with items rated as "no" or "unclear" were rated as "seeking further info" and protocols, and corresponding authors were consulted. Studies meeting less than 5 items were rated as "excluded" and removed from the study. Any disagreements between the 2 authors were resolved through discussion, and a third author (BRM) was consulted if disagreement persisted.

Data Abstraction

Qualitative data were extracted by 2 independent authors (MVA and MSL). Both authors read the papers and extracted key themes and concepts. These were compared, and any differences were resolved through discussion. The following data were extracted from all eligible papers: authors and context, year of publication, location, paradigmatic approach, method of data collection and analysis, data analysis software, participants' background, sample size and age, recruitment location and method, study aims, intervention or exposure, and main results.

Data Analysis and Synthesis

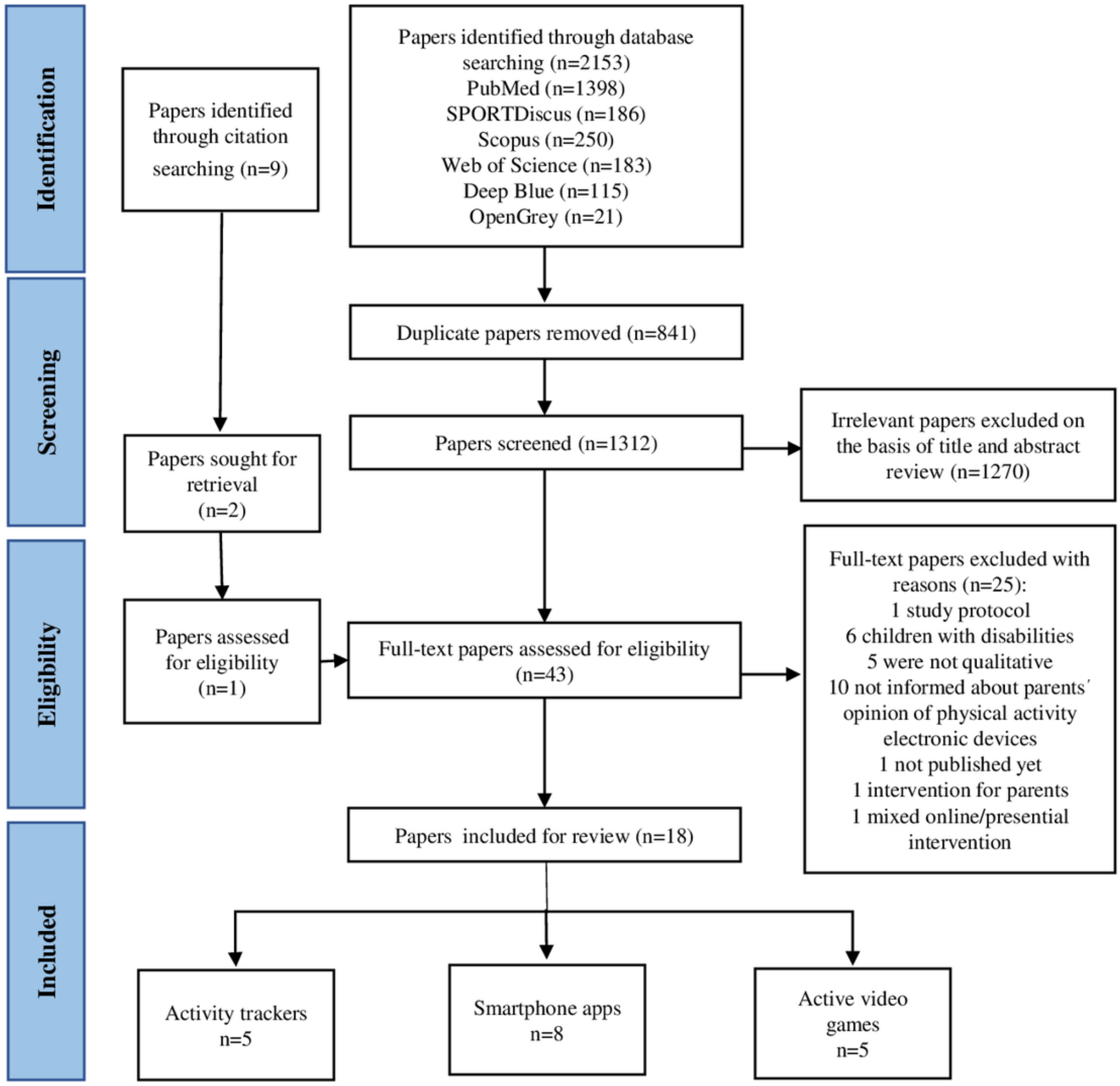
First, 2 authors (MVA and MSL) read the papers, extracted key themes and proofs (transcriptions of parents' verbalizations), and generated categories. A third author (BRM) was consulted if discrepancies arose. Differences were solved through discussion until agreement was reached. To identify common themes and analyze meanings, the meta-aggregation approach [28] was used. This process identifies meanings and common themes in qualitative studies using different methodologies and further extracts those meanings into categories that are then synthesized [29]. Next, MVA synthesized the key themes, meanings, and proofs (transcriptions of parents' verbalizations) into tables.

Results

Study Selection and Characteristics

The electronic search retrieved 2153 records. After the removal of duplicate studies, 1312 (60.9%) papers were reviewed based on the title and abstract. Following this, the full texts of 43 (3.3%) studies were reviewed; 1 (0.1%) additional study was identified after screening the reference lists of eligible papers. Finally, 18 (41%) eligible papers were included using the selection process shown in Figure 1.

Figure 1. Flowchart of search and selection process.



The 18 studies selected were published between 2010 and May 2022 and included 410 parents, mostly mothers, of 2-17-year-old children and adolescents (Tables 1-3). Of the 18 studies, 5 (28%) analyzed preschool children, 7 (39%) analyzed school children, 3 (17%) analyzed adolescents, and 3 (17%) did not provide separate results for children and adolescents. For data collection, 12 (67%) studies [30-41] used focus groups with semistructured interviews, 7 (39%) [20,30,42-46] used individual interviews, and 1 (6%) [41] used nonparticipant observation. Regarding the electronic devices analyzed, 5 (28%) studies [20,30,38,41,42] used smartphone apps, 2 (11%) [37,40] used the *Pokémon GO* mobile game, 1 (6%) [45] used mobile text messages, 5 (28%) [31-33,39,44] used activity trackers, 4 (22%) [34-36,43] used active video games, and 1 (6%) [46] used virtual reality.

Table 1. Characteristics of included studies (preschoolers).

Author, country	Method of data collection	Method of analysis (software); paradigmatic approach	Participants' details (background, age, parents' details)	Place and methods of recruitment
McCloskey et al [20], United States	Individual semistructured telephonic and face-to-face interviews	Thematic analysis, inductive approach (NVivo v.11, QSR International); N/I[a]	• Background: low-income families in rural areas • Age=3-5 years • Parents (telephonic interviews): n=29, mean age N/I, 93% (27/29) mothers • Parents (face-to-face interviews): n=31, mean age N/I, 77% (24/31) mothers	Purposive sampling (preschool centers, letters)
Alexandrou et al [30], Sweden	Focus groups, individual interviews	Thematic analysis, inductive approach; N/I	• Background: socioeconomically diverse district • Age=2.5-3 years • Somali parents: n=5, mean age 34 (SD 6.6) years; 100% (5/5) mothers • Arabic parents: n=4, mean age 31.2 (SD 2) years, 100% (4/4) mothers • Swedish parents: n=6, mean age 35.8 (SD 4.7) years, 83% (5/6) mothers	Purposive sampling (health care centers)
Costa et al [31], United Kingdom	Focus groups, semistructured interviews	Thematic analysis (NVivo v.9); N/I	• Background: low socioeconomic status • Age=2-3 years • Asian and White European parents: n=17, mean age 30.36 SD (6.9) years, 100% (17/17) mothers	Purposive sampling (children's centers)
Phillips et al [32], United Kingdom	Focus groups, semistructured interviews	Thematic analysis, inductive approach; N/I	• Background: highly deprived areas • Age=3-4 years • Parents: n=11, mean age 29 (SD N/I) years, 100% (11/11) mothers	Purposive sampling (children's centers, nurseries, preschools)
Ek et al [42], United States	Individual semistructured interviews	Thematic analysis, inductive approach; N/I	• Background: urban preschools • Age=3-4 years • Parents: n=10, mean age 38.9 (SD 5.2) years, 91% (9/10) mothers	Purposive selection of schools (posters)

[a]N/I: not informed.

Table 2. Characteristics of included studies (children).

Author, country	Method of data collection	Method of analysis (software); paradigmatic approach	Participants' details (background, age, parents' details)	Place and methods of recruitment
Creaser et al [33], United Kingdom	Focus groups, semistructured interviews	Thematic analysis, inductive approach (NVivo, QSR International); N/I[a]	• Background: families from different ethnicities • Age=5-9 years • Parents: n=36, mean age 38 (SD 7.7) years, 67% (24/36) mothers	Purposive sampling (social media)
Coknaz et al [34], Germany	Focus groups, semistructured interviews	Thematic analysis, inductive approach (NVivo v.10); N/I	• Background: public primary schools • Age=8-13 years • Parents: n=N/I, mean age N/I	Purposive sampling (from a clinical trial)
De Vet et al [35], the Netherlands	Focus groups, semistructured interviews	Content analysis (ATLAS.ti v 5.2); N/I	• Background: primary schools • Age=8-12 years • Parents: n=19, mean age 42.3 (SD 4.1) years, 95% (18/19) mothers	Purposive sampling (letter)
Dixon et al [36], New Zealand	Focus groups	Inductive approach; N/I	• Background: different ethnicity and socioeconomic groups in urban communities • Age=10-14 years • Maori parents: n=8, mean age N/I • Pacific parents: n=24, mean age N/I • Others: n=7, mean age N/I	Purposive sampling (community and church)
Lindqvist et al [37], United States	Focus groups, semistructured interviews	Latent content analysis; N/I	• Background: families • Age=7-12 years • Parents: n=9, mean age 38.7 (SD N/I), 78% (7/9) mothers	Purposive sampling
Rossi et al [38], Italy	Focus groups, semistructured interview	Content analysis (NVivo), community-based participatory action research	• Background: mothers • Age=0-14 years[b] • Parents: n=5, mean age N/I, 100% (5/5) mothers	Purposive sampling (public health local program)
Sharaievska et al [39], United States	Semistructured group interviews	Open, axial, selective coding techniques, grounded theory	• Background: families in rural communities • Age=7-13 years • Parents: n=N/I, mean age N/I	Purposive sampling
Sobel et al [40], United States	Nonparticipant observations and semistructured interviews	Inductive-deductive approach; N/I	• Background: families playing *Pokémon GO* in public locations • Age=2-17 years[b] • Parents: n=87, mean age 42 (SD 7.2) years, 70% (61/87) mothers	Purposive sampling (parks, shopping centers, events, online platforms)
Barnett et al [43], Australia	In-depth semistructured telephonic interviews	Thematic analysis (NVivo), descriptive qualitative approach	• Background: N/I • Age=9-10 years • Parents: n=29, mean age N/I	Purposive sampling (from a clinical trial)
Mackintosh et al [44], Australia	Web-based and face-to-face semistructured interviews	Thematic analysis, inductive approach (NVivo v.12); N/I	• Background: families • Age=7-12 years • Parents (web interview): n=25, mean age N/I, 84% (21/25) mothers • Parents (face-to-face interviews): n=10, mean age N/I, 100% (10/10) mothers	Purposive sampling (email)

[a]N/I: not informed.

[b]Some studies mixed ages in the sample and did not provide a separate analysis by age.

Table 3. Characteristics of included studies (adolescents).

Author, country	Method of data collection	Method of analysis (software); paradigmatic approach	Participants' details (background, age, parents' details)	Place and methods of recruitment
Carrion et al [41], Spain	Focus groups	Content analysis, phenomenological approach	• Background: parents from public or charter schools • Age=13-15 years • Parents: n=10, mean age N/I, 50% (5/10) mothers	Purposive sampling (schools)
Lindqvist [45], Sweden	Individual semistructured interview	Latent content analysis (NVivo, QSR International), empowerment	• Background: families of a municipality of North Sweden • Age=13-15 years • Parents: n=10, mean age N/I, 60% (6/10) mothers	Purposive sampling (from an intervention)
McMichael et al [46], United Kingdom	Semistructured interview	Framework analysis, Medical Research Council *(*MRC) framework	• Background: families • Age=13-17 years • Parents: n=18, mean age 53 (SD 3) years, 72% (13/18) mothers	Purposive sampling (social media, schools, university, emails, and posters)

Study Quality

The assessment of the 18 studies included in this systematic review is presented in Multimedia Appendix 2. Only 1 (6%) study [31] met all 10 items in the JBI-QARI checklist, 8 (44%) studies [20,30,32,33,40-42,46] met 9 items, 8 (44%) [35-39,43-45] met 8 items, and 1 (6%) [34] met 5 items. No studies were rated as "excluded"; thus, none was excluded based on methodological quality. The main weaknesses were a lack of clarity and a lack of reporting on the researcher's influence on the study and vice versa [20,30,34,35,39,40,43-45]. Other limitations were that participants and their voices were not adequately represented in 3 (17%) studies [34,36,38] and that there was no congruity between the stated philosophical perspectives and the research questions or methodology [34,35].

Synthesized Findings

We identified 4 main themes (Textbox 1) in terms of parents' perceptions of PA electronic devices: usefulness, advantages, general perceptions, and acceptability (barriers and facilitators). The main results are shown in Table 4, and proofs are shown in Multimedia Appendix 3.

Textbox 1. Themes and subthemes describing parents’ perceptions of physical activity (PA) electronic devices.

Usefulness of PA electronic devices

- PA promotion and PA in special moments
- Learning of skills and transferability to real life

Advantages of PA electronic devices

- Increase in motivation
- Awareness of behaviors
- Family bonding
- Socialization with peers

General perceptions

- Electronic devices for health promotion
- Preferences for real-life activities or active screen time
- Concerns: content, addiction, negative emotions, isolation, conflicts, limits

Acceptability (barriers and facilitators)

- Lack of time and stress
- Price
- Lack of space at home
- Discomfort/discomfort
- Difficulties with electronic devices or understanding feedback given by the app
- No new activities/suggestions
- Lack of use/interest after novelty
- Attractiveness (high technology, good graphs, good quality, videos)
- Gamification (competition, challenges, goals, and rewards) and fun
- Teacher and school support
- Ease of use
- Durability
- Integrated into daily routines

Table 4. Summary of findings.

Participants included, author, country	Area of inquiry/aims	Intervention/exposure	Main results
Preschoolers			
McCloskey et al [20], United States	To explore parents' beliefs about preschoolers' use of mobile devices and the acceptability and perceptions of a PA[a] intervention	Jungle Gym: a mobile app to encourage PA, focused on movement, motor skills (running, jumping, leaping, etc), and interactions with parents/children	Parents supported the use of mobile apps for PA and reported that they were useful in various situations (eg, on bad-weather days). Parents also expressed concerns about the apps.
Alexandrou et al [30], Sweden	To explore needs and concerns among Somali, Arabic, and Swedish parents regarding a PA app	MINISTOP 1.0 mobile app: a 6-month program to support parents in promoting PA	Parents found the app useful. Insights into their needs and important features were obtained.
Costa et al [31], United Kingdom	To assess mothers' opinions about the feasibility and acceptability of using an activity tracker	ActiGraph GT3Xþ, Actiheart (CamNtech Ltd), ActivPAL3 (PAL Technologies Ltd): 3 activity trackers	Children were most comfortable with ActiGraph and least comfortable with Actiheart. Problems with the devices were the possibility of children taking them off, allergic skin reactions, or discomfort.
Phillips et al [32], United Kingdom	To examine parents' acceptability and feasibility of measurement tools to assess PA	ActiGraph GT3X+, ActivPAL4 micro, Actical (Philips Respironics Inc): 3 accelerometers	Parents reported that ActivPAL was the least preferred electronic device (children's opposition to wearing it on their chest, skin irritation). ActiGraph was the most accepted.
Ek et al [42], United States	To explore parents' needs and perceptions of a PA app in a school setting	Mobile phone app to promote PA in a school setting	Parents reported the need for interactive features, problem-solving tasks, creativity, and music and dance activities and had a positive attitude toward the app. Children found activities more fun when adults participated.
Children			
Creaser et al [33], United Kingdom	To examine parents' acceptability of using wearables in a family setting	Fitbit Alta HR for 4 weeks, ActiGraph GT3X+	Fitbit was considered easy and enjoyable to use, but its perceived impact on PA was mixed. Most parents were willing to purchase a wearable.
Coknaz et al [34], Germany	To analyze the feelings and perspectives of parents about active video games	Nintendo Wii® sports (boxing, tennis, golf, baseball, bowling, skiing, aerobics, running, water skiing, etc) for 50-60 minutes, 3 days/week, 12 weeks	Parents believed that active video games might help in physical changes, socializing, and intellectual and personal development of children.
De Vet et al [35], the Netherlands	To explore parents' perceptions and opinions about active video games	Active video games	Parents had a positive attitude toward active and interactive video games. Some parents were less restrictive with them.
Dixon et al [36], New Zealand	To explore parents' perceptions of active video games and the probability of sustained engagement	Active video games (eg, EyeToy™, Dance Mat)	Parents supported active video games. They preferred nonviolent and sporty video games. Benefits, such as increased PA, improved fitness, and increased socializing, were reported.
Lindqvist et al [37], United States	To explore parents' perceptions of playing *Pokémon GO*	A gamification-inspired program using the *Pokémon GO* mobile game	Parents found that the game promotes PA. They were less likely to limit the time spent on this game. They suggested new features and concerns about safety.
Rossi et al [38], Italy	To explore parents' perceptions of a mobile app	Multimodal app for parents' mobile phones to promote children's health, including PA	Mothers had a positive attitude toward the app and made suggestions (feedback, geolocalization, and attractive features).
Sharaievska et al [39], United States	To explore the perception of a PA tracker	PA-tracking electronic device (Fitbit Zip), which each family member was asked to wear for 2 weeks	Parents reported minimal changes in PA because of a lack of interest or an already active lifestyle. The electronic device provided more awareness.

Participants included, author, country	Area of inquiry/aims	Intervention/exposure	Main results
Sobel et al [40], United States	To explore parents' perceptions of an app that promotes outdoor PA and to explore how they play with children	*Pokémon GO*	Parents reported an increased level of PA and valued how play led to family bonding. Concerns about safety and limits of game-play emerged.
Barnett et al [43], Australia	To identify parents' perceptions of active video games for development of movement skills	Active video games	Parents were skeptical of the capacity of video games to contribute to skill development and preferred real sports.
Mackintosh et al [44], Australia	To explore parents' perceptions of the acceptability and usability of wearable activity trackers to monitor PA	KidFit (X-Doria International) worn by each child for 4 weeks	Parents reported that the activity tracker is easy and useful. Barriers (lack of real-time feedback and difficulties in interpreting information) and suggestions (visual display, self-monitor activity, goal setting, and challenges) were identified.
Adolescents			
Carrion et al [41], Spain	To explore parents' perceptions, values, and preferences regarding mobile apps to promote PA	PEGASO Fit for Future: a mobile app to promote a healthy lifestyle, including PA, through gamification and family connections	Parents valued mobile apps for health promotion. They preferred apps that promote activity and interactions and include gamification and rewards.
Lindqvist [45], Sweden	To describe parents' perceptions of an empowerment-inspired PA intervention via mobile phones	Empowerment-based intervention via Short Messaging Service (SMS)	Parents found that children felt involved in the process and reported that social support and encouragement had an impact on PA. Goals and rewards could be motivating for PA.
McMichael et al [46], United Kingdom	To understand parents' views of PA, gaming, and virtual reality in PA interventions	vEngage project active virtual reality	Parents had a negative perception of gaming and preferred real-world PA. They reported the benefits of active games (socializing, motor skills, moving) and concerns (eg, addiction).

[a]PA: physical activity.

Parents' Perceptions of the Usefulness of PA Electronic Devices

The first theme reported was the main usefulness that eHealth technologies might have. The core concepts that support this theme included PA promotion and the learning of skills.

Parents perceived electronic devices as useful for increasing PA levels [34,35,37,39,40,44,45]; for example, parents reported that the *Pokémon GO* mobile game encourages children to be more active and promotes taking long walks through the neighborhood [37,40]. Alternatively, PA is not possible in specific moments when outdoors, for example, on bad-weather days [20]. Regarding activity trackers, parents reported that wearing the electronic device makes the children more motivated to accomplish daily step recommendations or take walks [33,39,44]. However, some parents said that their children, especially younger children, were physically active enough and so did not benefit much from the apps [43].

Regarding motor skills, such as balance or hand-eye coordination, some studies [20,35,40,43] reported that children show improvement and that those skills can be transferable to real sports [43]. In addition, they could learn how to score and follow the rules of some sports [43]. Furthermore, some parents found that eHealth might improve other skills, such as logical thinking and cognitive development [34,35]. In contrast, other parents were skeptical of the transferability of skills learned in video games to a real-life context, and they felt that it is unlikely that their children would benefit from learning skills from virtual apps [43].

Parents' Perceptions of the Advantages of PA Electronic Devices

The advantages of PA electronic devices that parents reported included an increase in motivation for engaging in real-life sports [39,41,43], more awareness, family bonding, and socialization with peers. For example, playing video games motivated children to engage in real-life sports [33,35,37,40,43,44]. Moreover, eHealth apps were useful for parents to become aware of their own levels of PA [39,44], and this, in turn, promoted changes in their attitude toward PA and increased their own PA levels [39]. In addition, parents said that using activity trackers made them aware of other interesting habits of their children, such as sleep or heart rate [30,33,44].

Another advantage of some electronic devices that parents highlighted is that they promote socialization [34,35,37,40,46] and cooperation and competition [37,40,45] with peers and family [20,35,37,39,40,44]. Parents also reported that active video games are suitable for playing with the family and an enjoyable activity to do together, reinforcing their bonds [20,35,37,39,40,44]. Other games promoted social interactions by providing users with something in common to talk about [39,40,44-46] or by enabling them to play interactively with others [35,40,46]; these features were particularly important for adolescents. Thus, parents reported how cooperation and social interaction were important factors in continuing to use the apps, since they found the apps fun and motivating [37,39].

Parents' General Perceptions of PA Electronic Devices

The general perceptions of parents about PA electronic devices were grouped into 3 key concepts: attitudes about electronic devices for health promotion, preference for real-life sports or active electronic devices, and concerns about the use of electronic devices.

Generally, parents were prone to using technology for health and educational purposes [20,42,46]. Furthermore, parents reported the desirability of apps being targeted not only at children but also at parents [30]. They suggested tracking their health lifestyles to be important, such as having an agenda or a reminder and the inclusion of health information [30]. Additionally, parents reported a preference for active and social video games or the active use of screens over passive screen time [35,36,46]. For example, active video games, such as Nintendo Wii, were perceived as a healthier alternative to passive screen time [35]. However, parents distinguished between real-life sports and virtual worlds, showing preferences toward playing outside rather than virtual PA [17,36,40,43,47].

In contrast, they also highlighted several concerns and dangers. Many of the parents were worried about violent content in video games, the appropriateness of content for different ages [35,46], concerns about children playing with strangers, safety [40,46], and physical accidents resulting from walking with the phone in hand [37]. In addition, psychological effects, such as anger, frustration, isolation, or addiction, were also reported [35,37,40,46]. Other common issues highlighted were conflicts when playing video games [37] and difficulties in establishing time limits, which increased with age. In that respect, although parents were more positive toward active video games and active screen use, setting limits and supervising screen use were important issues [20,35,37,40,46].

Parents' Perceptions of the Acceptability of PA Electronic Devices

Barriers

Some barriers to using PA electronic devices were found. Commonly, parents reported a lack of time to engage in eHealth activities because of their work or children's schedules [33,39,45]. Others found difficulties in managing extensive health information and reported feeling stressed by trying to follow all the recommendations [30]. Still, others highlighted the high prices of video games and electronic devices [35,36], and some were annoyed by the noise and space the devices occupy at home while playing [36,42,46].

Regarding the physical characteristics of activity trackers, the main issues raised included unsuitability, discomfort caused by a large size, drawbacks of wearable devices, children trying to remove electronic devices [31,32,44], and difficulties with batteries and syncing [44]. The size of the electronic device was especially important for younger children [31,32]. Other issues were difficulties in using activity trackers or understanding the information provided [33]. Several other factors impacted the use and wearability of activity trackers, including forgetting to wear them, having to remove them for certain sports, the lack of real-time feedback [44], and the lack of interest by parents [33,39]. In this sense, some parents said that activity trackers did not promote any new activity [39]. They also highlighted concerns about the lack of use of the electronic devices once they lost their novelty [33,36] and a lack of long-term wear compliance [44].

Facilitators

Parents reported several facilitators of the use of PA electronic devices. For example, they showed a preference for cheaper games that they could afford [35]. Other factors that facilitated engagement were the attractiveness of the game or electronic device, whether it uses high-level technology or appealing graphics [33,46], or the inclusion of videos [30,32,35].

Parents also reported that 2 important facilitators that ensure long-term engagement are gamification and fun [32,33,35,37,42,44]. Teacher support was found to be an important factor in engagement [44,45]. Parents said that goals [31,45] and rewards and new challenges [38,39,43,47] are important features—for example, different levels and new challenges to accomplish [47]. In that sense, many of the parents reported that an important feature is for an app to be fun [39,42,43]. To make apps appealing to children, parents recommended including reinforcement, such as treasure hunts or challenges, which might make the apps motivating. Regarding goal setting, the possibility of establishing goals with others, such as family members, peers, or classmates, was also recommended [31,45]. Furthermore, parents suggested that apps provide interaction with professionals, such as online forums [30,38], and be linked to the school curriculum [44], and teacher support was found to be an important factor in engagement [44,45]. Other ideas were links with sports associations and outdoor activities, such as events, active commuting, and geolocalization [38].

For activity trackers, parents reported some important characteristics that facilitate engagement. Most of them highlighted the importance of comfort [31-33,44], considering that an activity tracker should be worn all the time [32], and ease of use so that the children can understand and handle the device on their own [33,44] with an easy-to-use app [33]. Parents also reported the importance of considering the durability and damage resistance of electronic devices, since younger children might break them [32], and the integration of eHealth with their daily routines [33]. Other suggestions for activity trackers were real-time feedback and a complete dashboard showing information about scores, steps with good graphs, and demonstrations [32]. Features such as competition with others, options for new activities, and high-level technology were perceived as important.

Age Group Differences

Of the 18 studies, 5 (28%) [20,30-32,42] analyzed the opinions of the parents of preschoolers' (<5 years old). Generally, parents were less worried about their children's PA [30] because they perceived them as spontaneously active and preferred outside PA [20,42]. For preschoolers, most parents tried to limit technology as much as they could [20,42] and used PA apps when real PA was not possible [20,30]. Regarding activity trackers, the problems of wearability due to the size of the devices were highlighted [32].

Furthermore, 10 (55%) studies [33-40,43,44] analyzed schoolchildren between 7 and 12 years old. Parents of children in this age group also showed preferences for real PA [43], although they preferred PA apps over passive screen use [35,36]. Parents were worried about content and addiction and the necessity to set limits on screen time [35-37], and they more frequently reported interactive uses of PA electronic devices with peers and family [35,37]. Regarding activity trackers, parents highlighted the requirement of usefulness for children [44] and the importance of PA electronic devices and activity trackers to be designed specifically for children's use [33].

In addition, 3 (17%) studies [41,45,46] analyzed samples of parents of adolescents and showed that technology could be an effective strategy to connect with adolescents and help them acquire healthier habits [41]. Regarding this age group, parents were more worried about screen time, the time spent in gaming, and the time spent in sedentary pursuits and preferred technology uses that promote health, education, or socializing [45,46]. They perceived technology as unavoidable and reported difficulties in limiting screen time [46].

Discussion

Principal Findings

To the best of our knowledge, this is the first study that systematically reviews qualitative research that explores parents' perceptions of electronic devices that promote PA in children and adolescents. Overall, parents perceived electronic devices as useful for PA promotion. Moreover, they found other advantages, such as health promotion, awareness of health behaviors, learning of motor and cognitive skills, increased motivation for PA, and promotion of family and social interactions. Parents also valued some of the features of electronic devices, such as being comfortable, easy to use, active, challenging, and fun. However, some barriers and concerns, such as the risk of addiction, safety issues, or difficulties in setting limits, emerged. Preschoolers' parents found it less necessary to promote PA and preferred that their children spend time in outdoor activities. In contrast, in the case of older children and adolescents, when screen time increased, parents reported more advantages of using active electronic devices that promote PA.

A previous qualitative study that asked parents about their attitudes toward the use of electronic devices and media reported that parents are concerned about the total amount of time that children engage with electronic devices; specifically, they said that engaging with electronic devices prevents children from being physically active [47]. Additionally, other studies have reported positive attitudes of parents toward the use of electronic devices in children, as parents perceive them as a reality in children's and adolescents' lives [48], especially for educational and health purposes [49,50]. Similarly, in our study, parents had positive attitudes toward the use of technology for health purposes, such as promoting PA, and they preferred active electronic devices and dance- or sports-based video games rather than traditional sedentary screens [35] because parents perceive active electronic devices as a healthier alternative to passive screen time. Nevertheless, they preferred real PA or outdoor PA over PA on an electronic device [20,35,46]; thus, PA apps do not substitute but complement traditional forms of PA.

Other concerns that parents had, in addition to the high amount of time spent on electronic devices by children and adolescents, were the risk of addiction; the lack of skills; the emergence of negative emotions, such as anger; and violent or sexual content. These concerns are similar to those shown by previous studies, where parents reported being worried about access to inappropriate content, addiction, and negative emotions [9,47,51,52]. In this study, as in previous studies [47,52], parents perceived difficulties in setting limits on the time spent on electronic devices. Their concerns led them to implement different mediation strategies, such as couse, supervision, active mediation, restrictive mediation, and monitoring, depending on positive or negative attitudes toward media [53]. Along this line, parents reported being less restrictive in the case of active electronic devices, rather than passive ones, that promoted social interactions. Regarding social bonds, strong social and family bonds play a large role in controlling the overuse of electronic devices [52]. In this study, parents liked electronic devices that promoted family interactions to play together or that promoted peer interactions, as they believed that games that promote interactions might mitigate the lack of skills and isolation arising from the overuse of electronic devices.

Regarding age, as in a previous study [54], some differences were found, since electronic device usage and social, cultural, and cognitive experiences are vastly different between a 3-year-old child, an older child, and a teenager. In this study, parents of preschool children found no necessity for PA promotion since they perceived that their children were naturally active and used as few electronic devices as possible. In contrast, a study that analyzed general attitudes toward the use of electronic devices and media exposure in young children found that most parents have positive attitudes toward electronic devices, not only for educational purposes but also for entertainment [48]. This difference might be because our study analyzed only PA electronic devices and parents showed a general tendency to overestimate their children's PA [55], and thus, they perceived a low necessity of electronic devices to increase PA in their children. As children grow older, parents show increasing concerns about the amount of time spent using electronic devices, due to a substantial increase in hours using electronic devices with age [56]. In older children and adolescents, parents report more conflicts and difficulties in limiting electronic device use, consistent with previous studies [18] in which parents of adolescents have reported that setting limits on electronic device use is often confrontational and frequently escalates into arguments and shouting [57]. Therefore, parents implement different mediation practices [58] to regulate the use of electronic devices according to age, as the needs of children and adolescents change with development. Regarding gender differences, only 1 study showed that girls might engage in different challenges and games than boys [46]; congruently, a previous study found limited evidence of children's gender differences that precluded us from drawing conclusions [54], suggesting that differences in electronic device use and preferences might be considered in further studies.

Finally, parents reported some barriers that need to be considered in further studies, such as lack of time, stress, and high prices of electronic devices. Specifically for activity trackers, comfort, ease of use, difficulties in understanding the apps, or difficulties in understanding the feedback provided were the most common barriers. Conversely, facilitating factors for engagement included the attractiveness of the app, comfort, and children's self-efficacy in using the electronic device, similar to a previous study of eHealth programs [21]. Some suggestions provided by parents for new PA electronic devices included goal setting and rewards, usability, comfort, real-time feedback, and activities that promote interactions with friends and family, similar to a previous study [8]. In addition, parents had a favorable attitude toward the promotion of technology-based PA strategies in school contexts, and some also considered the involvement of schools and teachers in interventions and connection with the community [42,44,59].

Strengths

To the best of our knowledge, this is the first systematic review to synthesize findings from qualitative studies examining parents' perceptions of PA electronic devices. To ensure that the search process was systematic, an exhaustive search was carried out in specialized databases and gray literature by multiple researchers. This search was reported accurately according to the ENTREQ statement [24]. The meta-aggregation approach [29] was used to extract key themes and proofs, which enhanced the reliability of the data. In addition, data were meticulously documented in a matrix, and an assessment of the methodological strength of the analyzed papers was performed.

Limitations

This review has some limitations that should be acknowledged. First, there was high heterogeneity in the studies regarding the type of electronic device (mobile phones, activity trackers, exergames, virtual reality), data collection methods, location, duration of interventions, sample recruitment strategies, and the age of users. Along this line, studies considering differences between preschoolers, children, and adolescents are needed because these 3 age groups have different lifestyles, interests, and needs. Furthermore, gender differences between boys and girls were considered only in 1 study [46], which might be a source of bias since girls and boys have different levels of PA and different uses and preferences of technology. Second, most participants in the included studies were mothers, which might be due to mothers still parenting more than fathers; however, further studies considering fathers' opinions are recommended. Finally, some studies did not include an adequate description of the theoretical paradigm and did not provide information about how the researchers' background was managed.

Conclusion

This review explored the perceptions of children's and adolescents' parents regarding the use of electronic devices for PA enhancement. Parents reported that PA electronic devices could be an effective way to promote PA in children and adolescents and to overcome barriers, such as bad weather, lack of motivation, or the high rate of sedentarism in this population. In addition, parents prefer games and apps that require PA over traditionally passive games and apps. Parents also reported negative attitudes toward the use of technology in terms of addiction, safety problems, and difficulties in establishing limits, which should be considered in future interventions. These insights might provide researchers with more knowledge of how parents manage, promote, and regulate the use their children make of PA eHealth, the acceptability of interventions, and how they use PA eHealth at home. Some important features to consider in the development of new PA apps and technology-based interventions are the developmental stage, ease of use, appropriate feedback, promotion of socialization, and motivating strategies, such as rewards, challenges, and an appealing appearance.

Authors' Contributions

MVA contributed to writing the original draft, project administration, and visualization; VMV contributed to conceptualization and supervision; MSL and MVA performed investigation and formal analysis; ARH contributed to validation and data curation; BRM and RBG conducted supervision, methodology, and review and editing; and ISD contributed to data curation and review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 18 KB - mhealth_v11i1e44753_app1.docx]

Multimedia Appendix 2
Methodological quality of included studies.
[DOCX File , 21 KB - mhealth_v11i1e44753_app2.docx]

Multimedia Appendix 3
Findings extracted from included studies, with verbalization of parents' responses by theme.
[DOCX File , 29 KB - mhealth_v11i1e44753_app3.docx]

References

1. Auxier B, Anderson M, Perrin A, Turner E. Children's engagement with digital devices, screen time. Pew Research Center. 2020. URL: https://www.pewresearch.org/internet/2020/07/28/childrens-engagement-with-digital-devices-screen-time/ [accessed 2023-04-13]
2. Kabali HK, Irigoyen MM, Nunez-Davis R, Budacki JG, Mohanty SH, Leister KP, et al. Exposure and use of mobile media devices by young children. Pediatrics 2015 Dec 02;136(6):1044-1050. [doi: 10.1542/peds.2015-2151] [Medline: 26527548]
3. Smahel D, MacHackova H, Mascheroni G. EU Kids Online 2020: survey results from 19 countries. EU Kids Online. 2020. URL: https://www.eukidsonline.ch/files/Eu-kids-online-2020-international-report.pdf [accessed 2023-04-14]
4. Radesky JS, Weeks HM, Ball R, Schaller A, Yeo S, Durnez J, et al. Young children's use of smartphones and tablets. Pediatrics 2020 Jul 01;146(1):e20193518 [FREE Full text] [doi: 10.1542/peds.2019-3518] [Medline: 32482771]
5. EHealth. World Health Organization. 2021. URL: http://www.emro.who.int/health-topics/ehealth/ [accessed 2022-11-02]
6. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
7. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
8. Carlin A, Murphy MH, Gallagher AM. Current influences and approaches to promote future physical activity in 11-13 year olds: a focus group study. BMC Public Health 2015 Dec 21;15(1):1270 [FREE Full text] [doi: 10.1186/s12889-015-2601-9] [Medline: 26689932]
9. Reid Chassiakos YL, Radesky J, Christakis D, Moreno MA, Cross C, Council on Communications and Media. Children and adolescents and digital media. Pediatrics 2016 Nov;138(5):e20162593 [FREE Full text] [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
10. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]
12. Birch LL, Anzman SL. Learning to eat in an obesogenic environment: a developmental systems perspective on childhood obesity. Child Dev Perspect 2010;4(2):138-143. [doi: 10.1111/j.1750-8606.2010.00132.x]
13. Natale R, Messiah S, Asfour L, Uhlhorn S, Delamater A, Arheart L. Role modeling as an early childhood obesity prevention strategy: effect of parents and teachers on preschool children's healthy lifestyle habits. J Dev Behav Pediatr 2014;35(6):378-387. [doi: 10.1097/dbp.0000000000000074]
14. Rhodes RE, Blanchard CM, Quinlan A, Naylor P, Warburton DE. Family physical activity planning and child physical activity outcomes: a randomized trial. Am J Prev Med 2019 Aug;57(2):135-144. [doi: 10.1016/j.amepre.2019.03.007] [Medline: 31248744]
15. Chen J, Weiss S, Heyman MB, Cooper B, Lustig RH. The efficacy of the web-based childhood obesity prevention program in Chinese American adolescents (Web ABC study). J Adolesc Health 2011 Aug;49(2):148-154 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.11.243] [Medline: 21783046]
16. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity ehealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
17. Hoyos Cillero I, Jago R. Systematic review of correlates of screen-viewing among young children. Prev Med 2010 Jul;51(1):3-10. [doi: 10.1016/j.ypmed.2010.04.012] [Medline: 20417227]
18. Sanders W, Parent J, Forehand R, Sullivan AD, Jones DJ. Parental perceptions of technology and technology-focused parenting: associations with youth screen time. J Appl Dev Psychol 2016 May;44:28-38 [FREE Full text] [doi: 10.1016/j.appdev.2016.02.005] [Medline: 27795603]
19. Council on Communications and Media. Media and young minds. Pediatrics 2016 Nov;138(5):e20162591. [doi: 10.1542/peds.2016-2591] [Medline: 27940793]
20. McCloskey ML, Thompson DA, Chamberlin B, Clark L, Johnson SL, Bellows LL. Mobile device use among rural, low-income families and the feasibility of an app to encourage preschoolers' physical activity: qualitative study. JMIR Pediatr Parent 2018 Dec 06;1(2):e10858 [FREE Full text] [doi: 10.2196/10858] [Medline: 31518295]
21. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition interventions for prevention and management of childhood obesity: what do parents want from an eHealth program? Nutrients 2015 Dec 15;7(12):10469-10479 [FREE Full text] [doi: 10.3390/nu7125546] [Medline: 26694456]

22. Young R, Tully M. Autonomy vs. control: associations among parental mediation, perceived parenting styles, and U. S. adolescents' risky online experiences. Cyberpsychology 2022 Apr 14;16(2):Article 5. [doi: 10.5817/cp2022-2-5]
23. Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J, et al. Chapter 21: qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). London, UK: Cochrane; 2022.
24. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12(1):181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
25. Children and digital devices. UNICEF. 2020 May. URL: https://www.unicef.org/albania/media/2881/file/Children%20and%20the%20digital%20devices.pdf [accessed 2023-04-10]
26. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. BMJ Glob Health 2019 Jan 25;4(Suppl 1):e001107 [FREE Full text] [doi: 10.1136/bmjgh-2018-001107] [Medline: 30775019]
27. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
28. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. JBI Reports 2004 Mar;2(2):45-64. [doi: 10.1111/j.1479-6988.2004.00008.x]
29. Florczak KL. Meta-aggregation: just what is it? Nurs Sci Q 2019 Jan 13;32(1):11. [doi: 10.1177/0894318418807933] [Medline: 30798762]
30. Alexandrou C, Müssener U, Thomas K, Henriksson H, Löf M. Adapting a parental support app to promote healthy diet and physical activity behaviors (MINISTOP) for a multi-ethnic setting: a qualitative study on the needs and preferences of parents and nurses within Swedish child health care. Nutrients 2021 Jun 25;13(7):2190 [FREE Full text] [doi: 10.3390/nu13072190] [Medline: 34202326]
31. Costa S, Barber SE, Griffiths PL, Cameron N, Clemes SA. Qualitative feasibility of using three accelerometers with 2?3-year-old children and both parents. Res Q Exerc Sport 2013 Sep;84(3):295-304. [doi: 10.1080/02701367.2013.812002] [Medline: 24261008]
32. Phillips SM, Summerbell C, Hesketh KR, Saxena S, Hillier-Brown FC. Parental views on the acceptability and feasibility of measurement tools used to assess movement behaviour of pre-school children: a qualitative study. Int J Environ Res Public Health 2022 Mar 21;19(6):3733 [FREE Full text] [doi: 10.3390/ijerph19063733] [Medline: 35329419]
33. Creaser AV, Hall J, Costa S, Bingham DD, Clemes SA. Exploring families' acceptance of wearable activity trackers: a mixed-methods study. Int J Environ Res Public Health 2022 Mar 15;19(6):3472 [FREE Full text] [doi: 10.3390/ijerph19063472] [Medline: 35329166]
34. Coknaz D, Mirzeoglu AD, Atasoy HI, Alkoy S, Coknaz H, Goral K. A digital movement in the world of inactive children: favourable outcomes of playing active video games in a pilot randomized trial. Eur J Pediatr 2019 Oct 30;178(10):1567-1576. [doi: 10.1007/s00431-019-03457-x] [Medline: 31471690]
35. De Vet E, Simons M, Wesselman M. Dutch children and parents' views on active and non-active video gaming. Health Promot Int 2014 Jun 02;29(2):235-243. [doi: 10.1093/heapro/das064] [Medline: 23208151]
36. Dixon R, Maddison R, Ni Mhurchu C, Jull A, Meagher-Lundberg P, Widdowson D. Parents' and children's perceptions of active video games: a focus group study. J Child Health Care 2010 Jun 04;14(2):189-199. [doi: 10.1177/1367493509359173] [Medline: 20203134]
37. Lindqvist A, Castelli D, Hallberg J, Rutberg S. The praise and price of Pokémon GO: a qualitative study of children's and parents' experiences. JMIR Serious Games 2018 Jan 03;6(1):e1 [FREE Full text] [doi: 10.2196/games.8979] [Medline: 29298750]
38. Giorgi Rossi P, Ferrari F, Amarri S, Bassi A, Bonvicini L, Dall'Aglio L, Childhood Obesity Prevention Working Group. Describing the process and tools adopted to cocreate a smartphone app for obesity prevention in childhood: mixed method study. JMIR Mhealth Uhealth 2020 Jun 08;8(6):e16165 [FREE Full text] [doi: 10.2196/16165] [Medline: 32357123]
39. Sharaievska I, Battista RA, Zwetsloot J. Use of physical activity monitoring devices by families in rural communities: qualitative approach. JMIR Pediatr Parent 2019 Feb 20;2(1):e10658 [FREE Full text] [doi: 10.2196/10658] [Medline: 31518327]
40. Sobel K, Bhattacharya A, Hiniker A, Lee J, Kientz J, Yip J. It wasn't really about the Pokémon: parents' perspectives on a location-based mobile game. 2017 Presented at: CHI '17: 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 1483-1496. [doi: 10.1145/3025453.3025761]
41. Carrion C, Arroyo Moliner L, Castell C, Puigdomènech E, Felipe Gómez S, Domingo L, et al. Use of the smartphone to promote healthy habits among teen-agers, Spain. Rev Esp Salud Publica 2016 Nov 03;90:e1-e11 [FREE Full text] [Medline: 27808084]
42. Ek A, Sandborg J, Delisle Nyström C, Lindqvist A, Rutberg S, Löf M. Physical activity and mobile phone apps in the preschool age: perceptions of teachers and parents. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e12512 [FREE Full text] [doi: 10.2196/12512] [Medline: 30994465]

43. Barnett LM, Ridgers ND, Hanna L, Salmon J. Parents' and children's views on whether active video games are a substitute for the 'real thing'. Qual Res Sport Exerc Health 2013 Jun 26;6(3):366-381. [doi: 10.1080/2159676x.2013.809379]
44. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
45. Lindqvist A. Physiotherapists enabling school children's physical activity using social cognitive theory, empowerment and technology. Eur J Physiother 2017 May 10;19(3):147-153. [doi: 10.1080/21679169.2017.1322141]
46. McMichael L, Farič N, Newby K, Potts HWW, Hon A, Smith L, et al. Parents of adolescents perspectives of physical activity, gaming and virtual reality: qualitative study. JMIR Serious Games 2020 Aug 25;8(3):e14920 [FREE Full text] [doi: 10.2196/14920] [Medline: 32840487]
47. Dorey E, Roberts V, Maddison R, Meagher-Lundberg P, Dixon R, Ni Mhurchu C. Children and television watching: a qualitative study of New Zealand parents' perceptions and views. Child Care Health Dev 2010 May;36(3):414-420. [doi: 10.1111/j.1365-2214.2009.01031.x] [Medline: 19961500]
48. Vittrup B, Snider S, Rose KK, Rippy J. Parental perceptions of the role of media and technology in their young children's lives. J Early Child Res 2014 Apr 15;14(1):43-54. [doi: 10.1177/1476718X14523749]
49. Ihmeideh F, Alkhawaldeh M. Teachers' and parents' perceptions of the role of technology and digital media in developing child culture in the early years. Child Youth Serv Rev 2017 Jun;77:139-146. [doi: 10.1016/j.childyouth.2017.04.013]
50. Lupton D. Young people's use of digital health technologies in the global north: narrative review. J Med Internet Res 2021 Jan 11;23(1):e18286 [FREE Full text] [doi: 10.2196/18286] [Medline: 33427684]
51. Maxwell J, Kamp J, Cullen T. Parent perceptions of technology use in K-12 classrooms. SRATE Journal 2021;30(1):1-11.
52. Buabbas A, Hasan H, Shehab AA. Parents' attitudes toward school students' overuse of smartphones and its detrimental health impacts: qualitative study. JMIR Pediatr Parent 2021 May 20;4(2):e24196 [FREE Full text] [doi: 10.2196/24196] [Medline: 33878024]
53. Nikken P, Schols M. How and why parents guide the media use of young children. J Child Fam Stud 2015 Feb 24;24(11):3423-3435 [FREE Full text] [doi: 10.1007/s10826-015-0144-4] [Medline: 26472932]
54. Goh WL, Bay S, Chen VH. Young school children's use of digital devices and parental rules. Telematics and Informatics 2015 Nov;32(4):787-795. [doi: 10.1016/j.tele.2015.04.002]
55. Kippe K, Marques A, Martins J, Lagestad PA. Parents' inadequate estimate of their children's objectively physical activity level. Children (Basel) 2022 Mar 10;9(3):392 [FREE Full text] [doi: 10.3390/children9030392] [Medline: 35327764]
56. Shalani B, Azadfallah P, Farahani H. Correlates of screen time in children and adolescents: a systematic review study. J Mod Rehabil 2021 Nov 09;15(4):187-208. [doi: 10.18502/jmr.v15i4.7740]
57. Hattersley LA, Shrewsbury VA, King LA, Howlett SA, Hardy LL, Baur LA. Adolescent-parent interactions and attitudes around screen time and sugary drink consumption: a qualitative study. Int J Behav Nutr Phys Act 2009 Sep 09;6(1):61 [FREE Full text] [doi: 10.1186/1479-5868-6-61] [Medline: 19740410]
58. Livingstone S, Helsper EJ. Parental mediation of children's internet use. J Broadcast Electron Media 2008 Nov 07;52(4):581-599. [doi: 10.1080/08838150802437396]
59. Rioux C, Konkin A, MacKinnon AL, Cameron EE, Tomfohr-Madsen LM, Watts D, et al. Parent preferences for peer connection in eHealth programs. iproc 2022 Jun 15;8(1):e39278. [doi: 10.2196/39278]

Abbreviations

ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
JBI-QARI: Johanna Briggs Institute Qualitative Assessment and Review Instrument
PA: physical activity

Edited by L Buis; submitted 01.12.22; peer-reviewed by W Perveen, W Wang, A Videira-Silva; comments to author 14.03.23; revised version received 17.04.23; accepted 09.06.23; published 20.07.23.

Please cite as:
Visier-Alfonso ME, Sánchez-López M, Rodríguez-Martín B, Ruiz-Hermosa A, Bartolomé-Gutiérrez R, Sequí-Domínguez I, Martínez-Vizcaíno V
Parents' Perceptions of Children's and Adolescents' Use of Electronic Devices to Promote Physical Activity: Systematic Review of Qualitative Evidence
JMIR Mhealth Uhealth 2023;11:e44753
URL: https://mhealth.jmir.org/2023/1/e44753
doi:10.2196/44753
PMID:37471127

Original Paper

Current Status and Trends in mHealth-Based Research for Treatment and Intervention in Tinnitus: Bibliometric and Comparative Product Analysis

Yuanjia Hu[1], BSc; Yang Lu[1], BSc; Chenghua Tian[1*], MM; Yunfan He[2], MSc; Kaiyi Rong[3], MBBS; Sijia Pan[1], BSc; Jianbo Lei[4,5,6*], MD

[1]School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hang Zhou, China

[2]School of Public Health, Zhejiang University, Hangzhou, China

[3]The Second School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China

[4]Center for Medical Informatics, Peking University, Beijing, China

[5]Institute of Medical Technology, Peking University, Beijing, China

[6]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

[*]these authors contributed equally

Corresponding Author:
Chenghua Tian, MM
School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University
No. 548 Binwen Road
Hang Zhou, 310053
China
Phone: 86 (0571) 86613786
Fax: 86 (0571) 86613729
Email: 20071044@zcmu.edu.cn

Abstract

Background: As a global medical problem, tinnitus can seriously harm human health and is difficult to alleviate, ranking among the top 3 complex diseases in the otolaryngology field. Traditional cognitive behavioral therapy and sound therapy require offline face-to-face treatment with medical staff and have limited effectiveness. Mobile health (mHealth), which, in recent decades, has been greatly applied in the field of rehabilitation health care, improving access to health care resources and the quality of services, has potential research value in the adjunctive treatment of tinnitus.

Objective: This study aimed to understand the research trends, product characteristics, problems, and research transformation of tinnitus treatment software by analyzing the research progress of mHealth for tinnitus treatment based on the literature and related marketed apps.

Methods: Bibliometric methods were used to describe the characteristics of the relevant literature in terms of the number and topics of publications, authors, and institutions. We further compared the features and limitations of the currently available tinnitus treatment software.

Results: Data published until February 28, 2022, were collected. Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standardized screening process, 75 papers were included. The country with the highest number of publications was Germany, followed by the United Kingdom and the United States, whereas China had only a single relevant study. The most frequently found journals were the *American Journal of Audiology* and the *Journal of the American Academy of Audiology* (18/75, 24%). With regard to publication topics, cognitive behavioral therapy started to become a hot topic in 2017, and research on mHealth apps has increased. In this study, 28 tinnitus treatment apps were obtained (n=24, 86% from product data and n=4, 14% from literature data); these apps were developed mainly in the United States (10/28, 36%) or China (9/28, 32%). The main treatment methods were sound therapy (10/28, 36%) and cognitive behavioral therapy (2/28, 7%). Of the 75 publications, 7 (9%) described apps in the market stage. Of the 28 apps, 22 (79%) lacked literature studies or evidence from professional bodies.

Conclusions: We found that, as a whole, the use of mHealth for treatment and intervention in tinnitus was showing a rapid development, in which good progress had been made in studies around sound therapy and cognitive behavioral therapy, although most of the studies (50/75, 67%) focused on treatment effects. However, the field is poorly accepted in top medical journals, and the majority are in the research design phase, with a lack of translation of the literature results and clinical validation of the marketed apps. Furthermore, in the future, novel artificial intelligence techniques should be used to address the issue of staged monitoring of tinnitus.

(JMIR Mhealth Uhealth 2023;11:e47553) doi:10.2196/47553

KEYWORDS

tinnitus; mobile health; mHealth; internet; application; software; bibliometrics; mobile phone

Introduction

Background

The American Academy of Otolaryngology-Head and Neck Surgery defines tinnitus as "a sound perceived by the patient in the absence of an external sound source" [1]. Several studies have shown that the worldwide prevalence rate of tinnitus ranges from 5.1% to 42.7% [2]. Xu et al [3] showed that 9.6% of the global adult population has experienced tinnitus in the past 12 months. Of these, 36% had persistent tinnitus, and 27% had tinnitus for >15 years. Persistent tinnitus can lead to severe psychological disorders, such as depression, anxiety, mania, and other affective disorders. Suicide and personal injury rates are significantly increased in patients with tinnitus [4]. As a matter of concern, the decreased availability of effective treatments for tinnitus greatly increases the socioeconomic burden of this disease in medical care [5].

Tinnitus lacks effective standardized and individualized treatments. Drugs, surgery, electrical stimulation, and psychophysiological integrative therapy are traditional treatment modalities that have serious adverse effects. Moreover, their long-term safety is unknown. Mobile health (mHealth) apps are a novel tool expected to be effective in alleviating tinnitus. The popularity of smartphones and the inherent immediacy, accuracy, and low costs of the mobile internet have contributed to the unique advantages of mHealth apps in providing health interventions and other outcomes. As of 2020, there were >300,000 (this number is rapidly growing) mHealth apps available in mobile app stores [6]. These apps help users to monitor multiple physiological indicators and provide relevant health knowledge and services to address a range of health issues [7], including tinnitus. Several studies have found that mHealth apps can provide continuous and remote monitoring of tinnitus as well as diagnostic and intervention services for patients with the condition. This can, in turn, be effective in alleviating or resolving this disorder. Most Chinese researchers provide sound therapy for patients with tinnitus through software. There are also attempts to use herbal medicine, acupuncture, and electrical stimulation for the same purpose. He et al [8] used audition software for sound therapy. Cai et al [9] integrated personalized music into tinnitus software, in combination with cognitive behavioral therapy, to innovate a tinnitus treatment intervention with substantial results.

Objectives

Nevertheless, most studies have focused only on monotherapeutic interventions for tinnitus. As such, there is a lack of a systematic analysis of tinnitus treatment software studies and marketed apps [10]. Therefore, using bibliometric and comparative product analysis methods, this study provides a complete and detailed description of the field of tinnitus monitoring, diagnosis, and intervention with the help of mHealth apps. Our work is the first bibliometric study to comprehensively analyze trends in publication, national and institutional distribution, core journals and highly productive authors, and research topic hot spots in this area. This study aimed to understand the research trends in tinnitus treatment software, including the characteristics and limitations of the apps available in the market. The outputs of our research can help patients to choose the right tinnitus treatment software and provide suggestions to manufacturers regarding potential ways to improve the quality and use of tinnitus treatment software.

Methods

Data Retrieval and Filtering

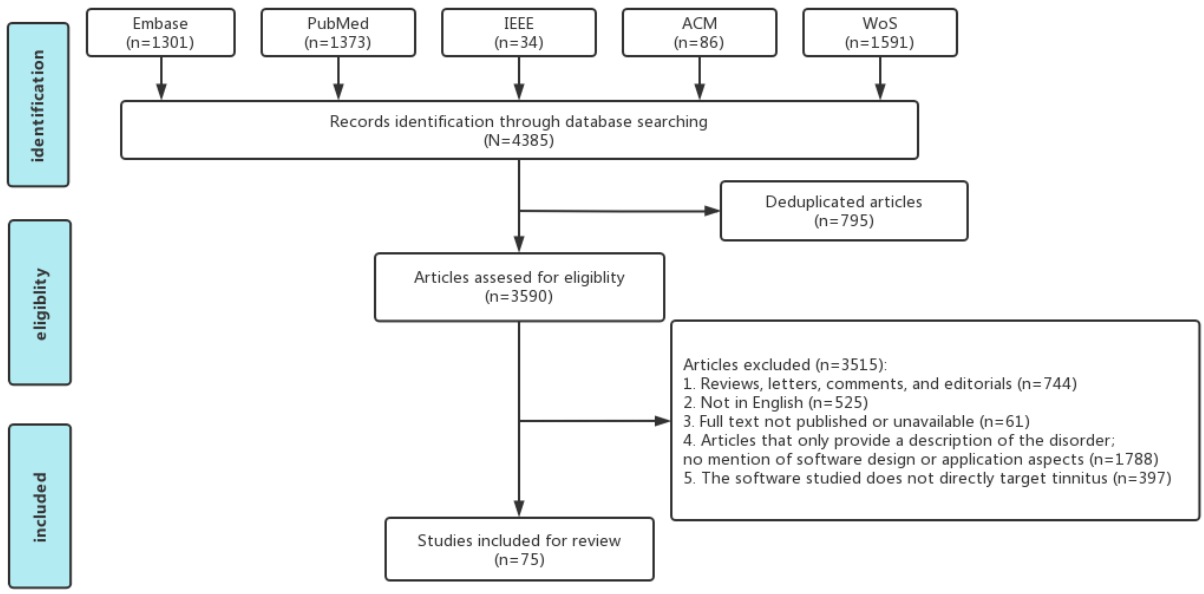
Regarding the research literature, to ensure completeness, our study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principle and searched databases in 5 different fields [11,12]. In the medical field, the PubMed and Embase databases were used. In the general science field, the Web of Science core collection was chosen. In the computer field, the IEEE and ACM databases were chosen. The search queries and strategies used for the different databases are listed in Multimedia Appendix 1. The inclusion and exclusion criteria for this study are shown in Multimedia Appendix 2.

A total of 4385 papers were retrieved from 5 databases (n=1301, 29.67% from Embase; n=34, 0.78% from IEEE; n=1373, 31.31% from PubMed; n=1591, 36.28% from Web of Science; and n=86, 1.96% from ACM). After literature deduplication, of the 4385 papers, 3590 (81.87%) remained. Two trained postgraduate students screened the retrieved papers according to the established inclusion and exclusion criteria (Cohen κ=0.91), and, of the 3590 papers, 75 (2.09%) were included in the final literature pool, as illustrated in Figure 1.

After a preliminary market survey, we identified 2 sources of information in which to search for product data: the Android platform and the Apple App Store (China and the United States), for a total of 7 official mobile software marketplaces, with the Android platform accounting for 5 (71%) mobile software marketplaces (Google Play Store as well as the mobile software marketplaces operated by Huawei, Vivo, Oppo, and Xiaomi). To obtain relevant search results for mHealth apps on the Qimai app data analysis platform [13] (accessed in March 2021), we used the following search terms: "tinnitus diagnosis," "tinnitus intervention," and "tinnitus treatment" A total of 2384 mHealth apps were retrieved, with 1812 (76%) stand-alone mHealth apps remaining after data deduplication.

In this study, mHealth apps related to tinnitus treatment were screened from 75 literature reports and 1812 independent software reports, using established inclusion and exclusion criteria. A total of 28 tinnitus treatment apps were obtained (n=24, 86% from product data and n=4, 14% from literature data). The specific inclusion and exclusion criteria applied to select the target apps are listed in Textbox 1. A screening flowchart of the mHealth apps is shown in Figure 2.

Figure 1. Flowchart of literature data screening. WoS: Web of Science.



Textbox 1. Inclusion and exclusion criteria applied to select the target apps.

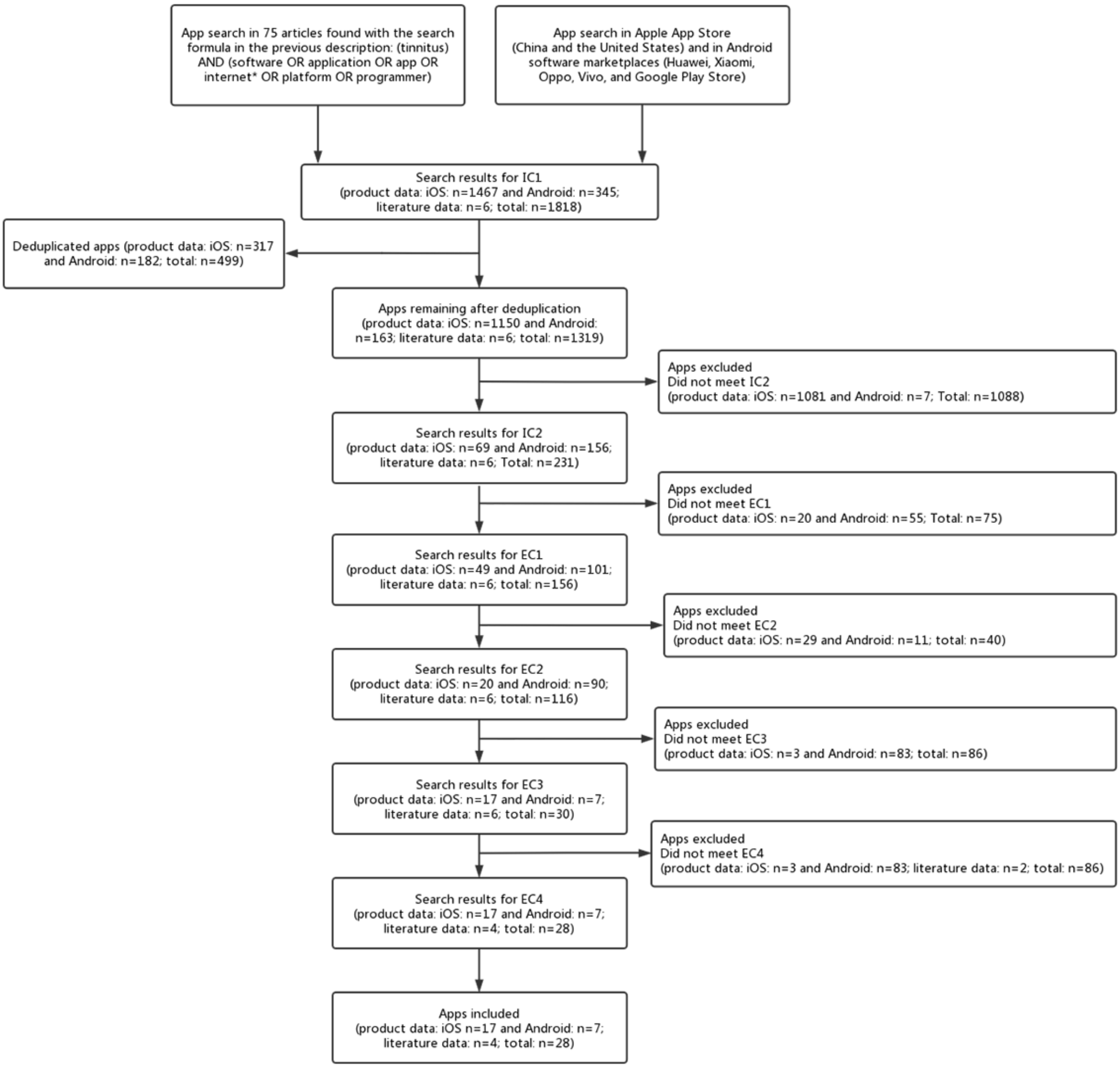
Inclusion criteria

- Inclusion criterion (IC) 1: apps identified with the search terms "tinnitus," "hearing test," "sound therapy," "cognitive behavioural therapy," "cbt," "perceptual mask," "perceptual masking," "retraining therapy," and "sleep"
- IC2: apps with either a health purpose or a medical purpose

Exclusion criteria

- Exclusion criterion (EC) 1: Apps consisting only of sleep aid, relaxation, or meditation software; pure mood journals; or diary functionalities (with or without cognitive behavioral therapy)
- EC2: apps consisting only of mobile hearing test, aid, or debugging software, pure mobile sound level meter software, or pure speech therapy software (not related to tinnitus)
- EC3: health management software for other diseases (not related to tinnitus)
- EC4: insufficient software information

Figure 2. The screening flowchart of the mobile health apps. EC1: exclusion criterion 1; EC2: exclusion criterion 2; EC3: exclusion criterion 3; EC4: exclusion criterion 4; IC1: inclusion criterion 1; IC2: inclusion criterion 2.



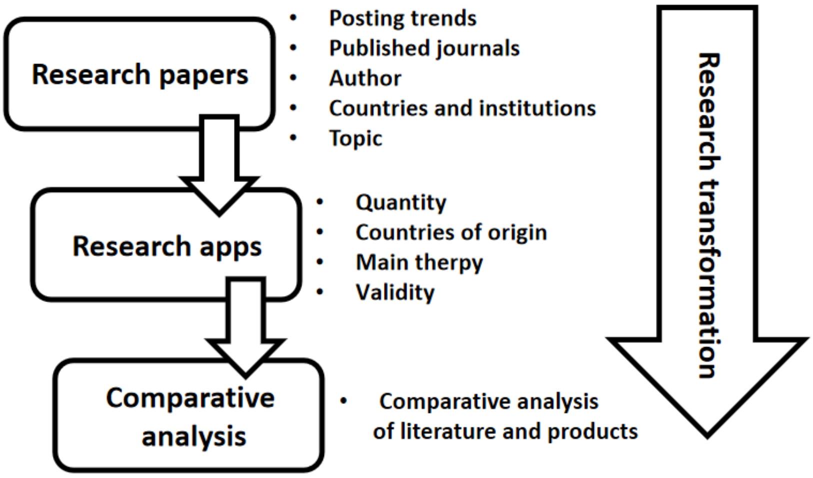
Extraction of App Information

In this study, we extracted general and specific information about 28 apps, using the Qimai app data analysis platform. General information about the apps included their names, reviews, and ratings. The specific information obtained on the apps was analyzed from 3 perspectives: time, space, and content. From a temporal perspective, the fields extracted from the specific information included the dates of product release and last update, as well as in-app, off-app, and unreleased stores. With regard to the spatial perspective, the specific information included the product's country of origin and the name, location, and size of the product manufacturer. From the content perspective, the specific information included type, purpose, input and output information, clinical role, main treatments related to the software, their effectiveness, and software scenario.

Data Analysis Methods

To review and summarize the current status of, and research trends in, mHealth treatments and interventions for tinnitus, we used a bibliometric and comparative product analysis approach, based on a *literature output-product output–literature versus product* chain. We used the *bibliometrix* package in R for data analysis and the *ggplot2* package for mapping. First, a bibliometric approach was used to analyze the literature data and obtain the current publication outputs. Second, product data were analyzed to obtain information about product outputs and applications. Finally, we compared the development of literature and product components. The purpose was to obtain the differences between the current state of research and its actual use, as well as to gain a preliminary insight into the translation of research in this field. The analysis chain and process are shown in Figure 3.

Figure 3. Analysis chain and process.



Results

Analysis of the Literature Output

Regarding the literature output, we analyzed the retrieved articles using bibliometric methods concerning 5 aspects: publication trends, journals, authors, countries and institutions, and themes.

Analysis of Publication Trends

As of February 2022, a total of 162 authors had published 75 relevant articles in 44 different journals. Among the 162 authors, 46 (28.4%) were lead authors. In addition, there were an average of 0.11 (SD 0.98) relevant articles per journal per year and an average of 0.46 (SD 1.27) articles per author. The evolution in the annual number of relevant studies published over the years is shown in Multimedia Appendix 3. With regard to the 75 articles included in this study, there was an overall increasing trend in the annual number of literature publications. There was an average annual increase rate of 24.46% (SD 0.69%), which resulted in 16 articles published in 2021. This trend demonstrates that an increasing amount of research is being carried out on this topic, indicating the increasing need for research in this field.

There was a substantially increased number of articles published in 2015 and 2021, in comparison with the previous and subsequent years. The authors of the 9 articles published in 2015 were mainly from Sweden and the United Kingdom, accounting for 3 (33%) and 2 (22%) articles, respectively. There were 4 articles by Gerhard Andersson as the lead author, from Linköping University (Linköping, Sweden), whose primary research interests were psychopathology and psychotherapy [14-17]. Of the 16 articles published in 2021, a total of 9 (56%) were from the United Kingdom, including 4 (44%) articles by Eldre Beukes, from Anglia Ruskin University, whose main research interest was audiology.

Analysis of the Major Publishing Journals

According to the Bradford law of scattering [18], all publications in a specific field, released during a given period, can be divided into core, related, or peripheral zones, according to the number of articles the zones include. The ratio of the number of publications in the 3 zones is 1:a:a^2, with the approximate value of a being 5. Core publications account for approximately 1 of 31 of the total number of publications. The core publications were calculated to correspond to the top 2 journals, which accounted for 36% (18/75) of the published articles: the *American Journal of Audiology* (12/75, 16%) and the *Journal of the American Academy of Audiology* (6/75, 8%). In 2021, the journal impact factors of these 2 journals were 1.636 and 1.249, respectively, and the journal citation reports divisions were Q4 and Q3, respectively. The remaining publications revealed minor differences in terms of the number of published articles.

Analysis of Article Authors

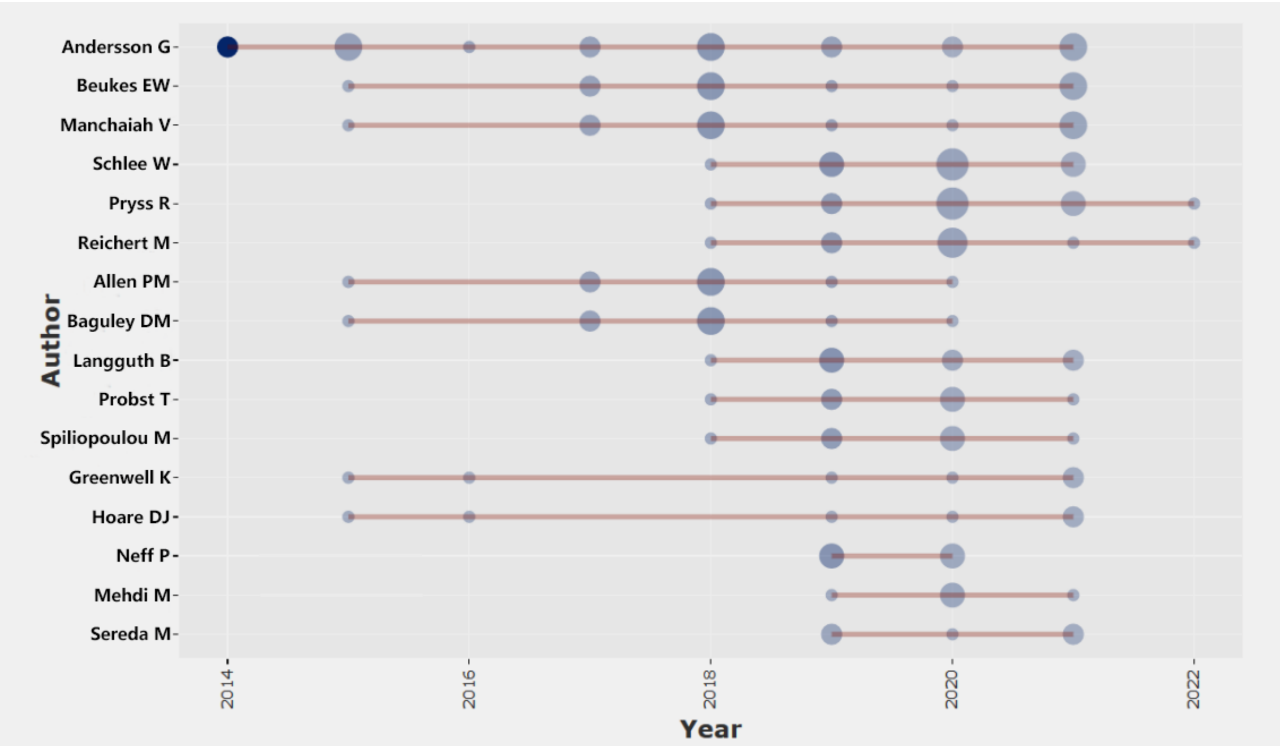
There were 162 authors in the included literature. According to the Lotka law [19], which states that authors with >0.749 times the square root of the number of papers published by the most prolific scientists, this study defined authors with >4 publications as high-producing authors, accounting for 9.9% (16/162). As can be seen in Table 1, Andersson G was the most prolific author, with 22 publications.

Next, we performed an analysis of the posting trends from highly productive authors. A graph of the annual posting trends of highly productive authors, based on their annual posting data and yearly citation frequencies, is shown in Figure 4. Here, the size of the circles indicates the annual posting volume, and the color shade indicates the annual citation frequency. Of the 16 most productive authors, 15 (94%) were found to have started focusing on mHealth-based treatments and interventions for tinnitus only after 2015. Of these 15 authors, 9 (60%) started publishing after 2018, and they have produced a steady annual output since then.

Table 1. Ranking of authors with >4 publications.

Rank by number of articles	Author's name	Articles (n=153), n (%)
1	Andersson G	22 (14.4)
2	Beukes EW	14 (9.2)
2	Manchaiah V	14 (9.2)
3	Schlee W	13 (8.5)
4	Pryss R	12 (7.8)
5	Reichert M	10 (6.5)
6	Allen PM	9 (5.9)
6	Baguley DM	9 (5.9)
7	Langguth B	8 (5.2)
8	Probst T	7 (4.6)
8	Spiliopoulou M	7 (4.6)
9	Greenwell K	6 (3.9)
9	Hoare DJ	6 (3.9)
9	Neff P	6 (3.9)
10	Mehdi M	5 (3.3)
10	Sereda M	5 (3.3)

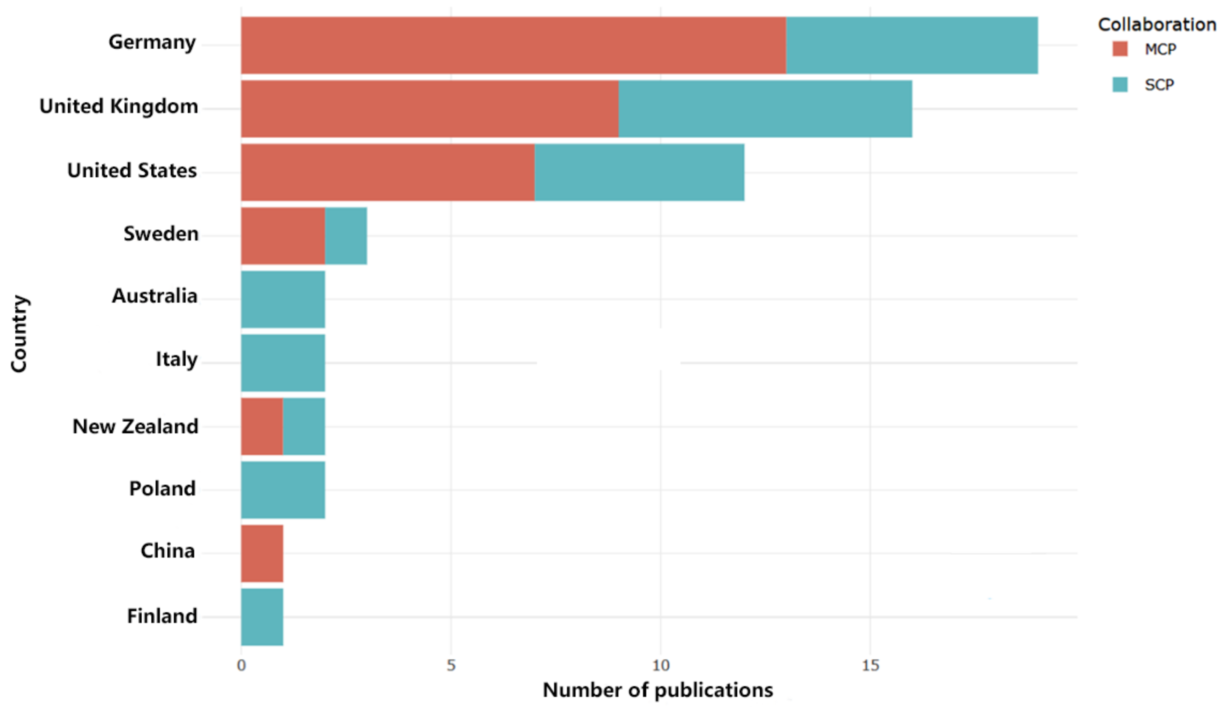
Figure 4. Productivity of the top authors over time.



Analysis of the Publication Country of Origin

The 162 authors in our literature pool were from 15 countries. The top 10 countries (corresponding authors' countries) are shown in Figure 5. Germany has published the largest number of articles in this field (19/75, 25%) and also has the largest number of international collaborations, followed by the United Kingdom (16/75, 21%) and the United States (12/75, 16%). The numbers of multicountry publications and single country publications for the top 10 high-output countries are shown in Figure 5.

Figure 5. Articles by country of corresponding author. MCP: multicountry publication; SCP: single country publication.



Analysis of Posting Themes

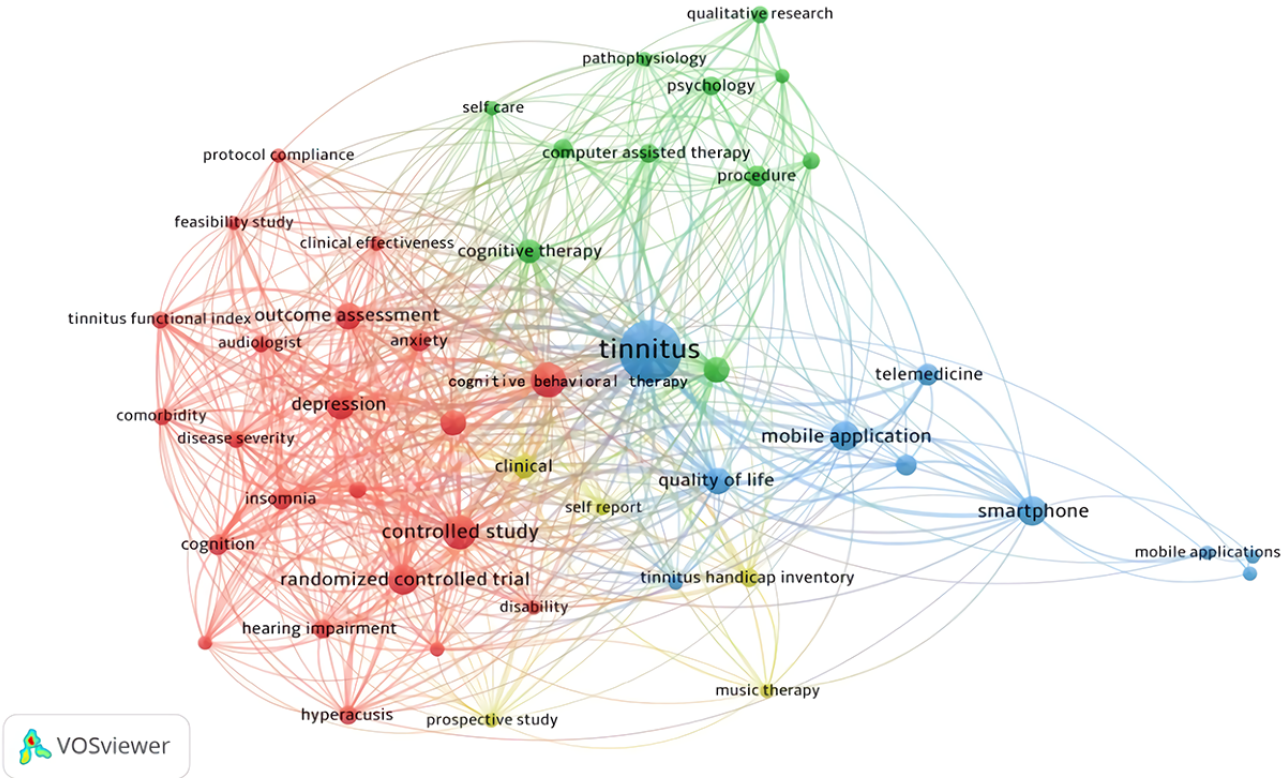
VOSviewer Keyword Clustering Analysis

In this study, we analyzed the themes of the publications in the field from 3 aspects. First, highly frequent keywords were extracted from the abstracts. These keywords referred to literature data, common aspects of highly frequent keywords, and their co-occurrence network. The main research themes were identified from the clustering results. Second, to explore current research hot spots, the thematic map method, proposed by Cobo et al [20], was used to cluster and map themes according to their density and centrality. Finally, trends in the evolution of themes in the field were analyzed from a time-course perspective.

Next, we conducted a visual analysis of highly frequent keyword co-occurrence networks. To obtain all keywords from the relevant studies, the abstracts of the included literature articles were divided, deactivated, and lexically normalized using Python (Python Software Foundation). As shown in Figure 6, a high-frequency keyword co-occurrence network was produced using VOSviewer [21]. As shown in this figure, clustering has divided all keywords into 4 main domains, which could be further combined into 3 categories: red and yellow, green, and blue. The red and yellow categories represent the main treatments currently available for tinnitus (including *cognitive behavioral therapy* and *music therapy*), as well as clinical trials and the main feelings experienced by patients with tinnitus. Clinical trial methods include *randomized controlled trials* and *clinical trials*. The subjective feelings of patients with tinnitus include *anxiety*, *depression*, and *insomnia*. The green category contains computer-related keywords such as *computer-assisted therapy and procedure*. By contrast, the blue category includes keywords such as *mobile app*, *smartphone*, and *telemedicine* for mHealth and mobile devices.

It was particularly interesting to note that *cognitive behavioral therapy* was used as a link between the words *disorder* and *computer*. The words linking *mHealth* and *computer* were mainly basic concepts such as *procedure*. The words linking *mHealth* and *disease* were mainly research methods such as *clinical study*, which were slightly less connected than the other 2 groups.

Figure 6. Keyword co-occurrence network diagram produced using VOSviewer [21].

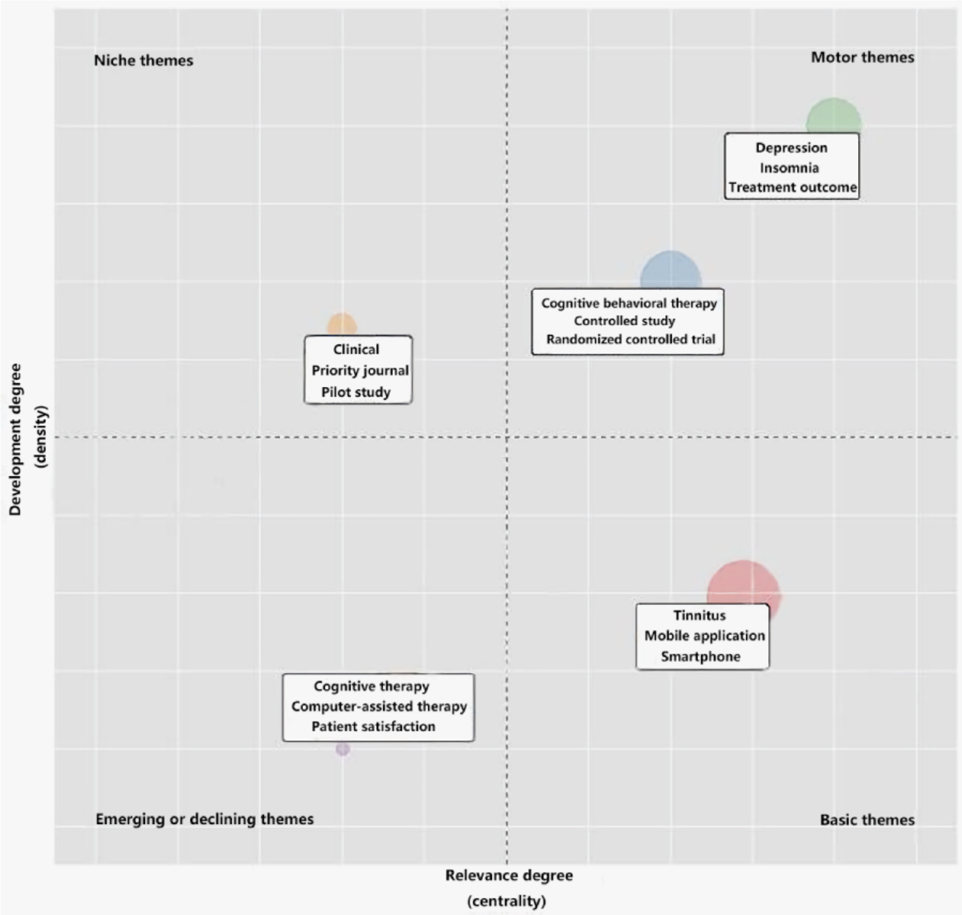


Thematic Map Analysis

Cobo et al [20] proposed the thematic map method, which uses the division of themes into quadrants to analyze the *hotness* and importance of a topic. Topics located in the first quadrant (niche themes) are well developed but relatively less important. Those in the second quadrant (motor themes) are well developed and essential. Topics located in the third quadrant (emerging or declining themes) are not well developed and relatively less compelling. The topics found in the fourth quadrant (basic themes) are not well developed but necessary and generally refer to basic concepts.

By calculating the density and centrality of the already clustered coword matrix, each of the 5 categories was visualized within the 2D coordinates, as shown in Figure 7. We found that, in the first quadrant, the vocabulary of research methods (terms such as *clinical* and *pilot study*) was predominantly used, as well as concepts that have developed well and are currently being phased out. In the second quadrant, we observed that cognitive behavioral therapy and treatment outcomes were essential and well-studied topics. In the third quadrant, we observed that computer-assisted therapy was a topic that has not been well developed yet. In the fourth quadrant, as basic technologies in the field, the concepts of tinnitus, smartphones, and mobile apps were predominantly used.

Figure 7. Thematic map with keywords divided into 4 domains.

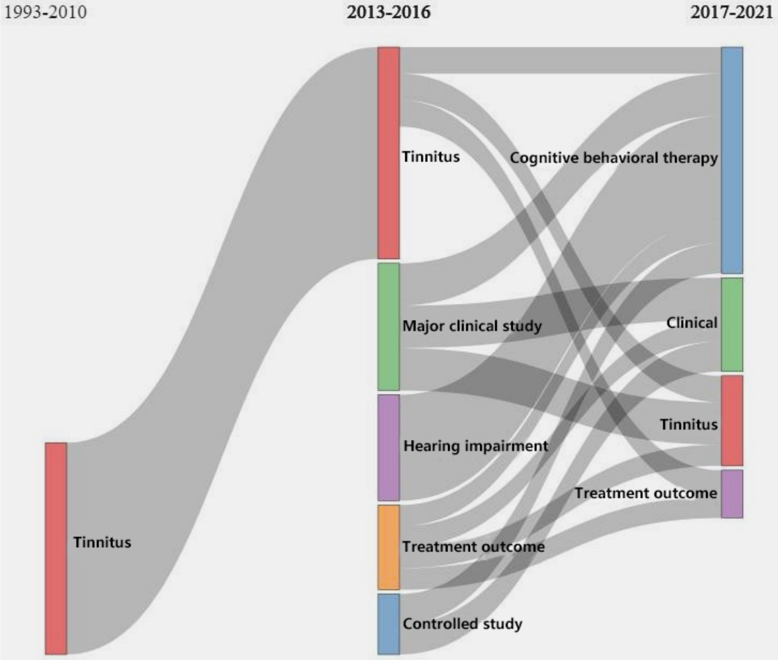


Sankey Diagram Analysis

A thematic evolution analysis was carried out using coword network analysis and clustering. This analysis incorporated the time dimension to analyze the evolution of the research themes from 1993 to 2021 and to create a Sankey diagram (Figure 8).

From this observation, it was clear that there were both changing and unchanging themes. Researchers from as early as the 1993-2010 period focused on tinnitus as a disorder, without introducing further meaningful concepts. During the 2013-2016 period, the traditional research themes did not entirely dissipate, but several research methods were added to these clinical studies. During the 2017-2021 period, research moved from primary to applied, with emerging concepts such as cognitive behavioral therapy. Since 2022, cognitive behavioral therapy has remained the main research topic, whereas the concept of mHealth has not received much attention.

Figure 8. Sankey diagram of trends in the evolution of research themes.



Analysis of the Software Development Stages Described in the Literature

In terms of the software development stages described in the literature, of the 75 papers, 49 (65%) referred to the design stage, 19 (25%) pertained to the clinical validation stage, and 7 (9%) concerned the marketing stage. This distribution reflects the negligible development of software for tinnitus treatment. Preliminary research found that the actual demand for such software was notably high.

Analysis of the Core Treatment Techniques Described in the Literature

Upon further analysis, it was found that 50 (67%) of the 75 papers primarily described treatments that used specific software. Of the remaining 25 studies, 12 (48%) mentioned only internet-based or web-based assessment tools for tinnitus, without any specific treatment associated. Of these 12 publications, 6 (50%) described mobile transient tinnitus assessment, and 4 (33%) used machine learning for predicting tinnitus, whereas 1 (8%) described a repository of tinnitus information, and 1 (8%) described a method for characterizing tinnitus heterogeneity. Of the 50 articles describing treatments, the mentioned treatments were mainly in the categories of cognitive behavioral therapy and sound therapy, with 30 (60%) and 13 (26%) papers, respectively. Of the remaining 7 studies, 1 (14%) included usual service therapy using artificial intelligence (AI) to expose the patient to the tinnitus environment, 1 (14%) used a tinnitus e-plan (an intervention coding method), 1 (14%) used visualized mobile electroencephalogram (EEG) for tinnitus detection, 1 (14%) used meditation (to help patients relieve tinnitus), 1 (14%) used both cognitive behavioral therapy and sound therapy, and 2 (29%) used a game training method related to the localization and perception of a sound source as well as auditory attention.

We compared the stage of software development reported in the literature with the therapies related to the respective software. We found that only a single software using cognitive behavioral therapy and sound therapy was in the market stage. With regard to software in the clinical validation stage, 9 (32%) of the 28 apps used cognitive behavioral therapy tools, 3 (12%) used sound therapy, and 2 (17%) used game training methods related to sound source localization perception and auditory attention. Of the 29 apps described in the literature that were in the design phase, 18 (62%) used cognitive behavioral therapy, 7 (24%) used sound therapy, 1 (3%) used habituation therapy (in which AI exposes the patient to the tinnitus environment), 1 (3%) used meditation, 1 (3%) used visualized mobile EEG, and 1 (3%) used an intervention coding approach.

Analysis of the Listed Apps

Analysis of the Number of Apps

Mobile tinnitus treatment software is currently in high demand. In this study, 28 apps were identified from software data sources (Figure 2), 4 of which were from software with specific names in the literature. These 28 apps consisted of 18 (64%) iOS apps and 10 (36%) Android apps (Multimedia Appendix 4). Most of these apps were developed by institutions (22/28, 79%); a few were developed by individuals (2/28, 7%). Among the 24 apps obtained from product data, 13 (54%) were medical apps, and 9 (38%) were health apps.

Regarding the software providers, of the 28 apps, 22 (79%) were available on the Apple App Store (China), and 21 (75%) were available on the Apple App Store (United States). In addition, 5 (18%) were available on the Google Play Store, and 5 (18%) were available on the Huawei, Xiaomi, Oppo, and Vivo marketplaces. Software activity can be reflected at the time of listing and at the time of the latest update. After data collection, we found that in 23 (82%) of the 28 apps, time information was complete.

Analysis of the Distribution of Source Countries

Multimedia Appendix 5 depicts the origin of the software developers. Of the 28 apps, 10 (36%) were developed in the United States, 8 (29%) in China, and 9 (32%) in other countries (n=1, 11% in the United Kingdom; n=3, 33% in Germany; n=3, 33% in Denmark; n=1, 11% in Poland; and n=1, 11% in New Zealand). Of the 28 apps, there was 1 (4%) whose country of origin could not be found. Regarding specific locations, the US software developers hailed from 7 states (3/10, 30% from Minnesota; 2/10, 20% from Texas; 1/10, 10% from Colorado; 1/10, 10% from Wisconsin; 1/10, 10% from California; 1/10, 10% from Florida; and 1/10, 10% from Oregon). Of the 8 software developers based in China, 7 (88%) were from regions within mainland China (n=3, 43% from Beijing; n=2, 29% from Jiangsu; n=1, 14% from Shanghai; and n=1, 14% from Guangdong), and 1 (13%) was from Hong Kong, China.

Analysis of the Functional Division of the Apps

When analyzing the 28 tinnitus treatment–related apps, this study divided them into three main categories: (1) screening and evaluation, (2) intervention and rehabilitation, and (3) education and information. Of the 28 software programs, 28 (100%) were related to intervention and rehabilitation, 11 (39%) were related to screening and evaluation, and 5 (18%) were related to education and information. Regarding the users of the software, of the 28 apps, 26 (93%) were intended to be used by patients or their families, and 2 (7%) were intended to be used by medical staff. In this study, we considered whether a professional was required to assist patients with the software and found that, in the case of 7 (25%) of the 28 apps, a professional was needed to assist the patient. By contrast, the remaining apps (21/28, 75%) could be used by patients without assistance from a professional.

When analyzing the specific functions of the apps, this study divided them into the following 6 areas: assessment, advice, detection, counseling, treatment, and relief. Of the 28 apps, 6 (21%) were dedicated to patient assessment, 4 (16%) could provide advice based on the actual patient situation, 5 (18%) could perform a preliminary test of the patient's tinnitus or hearing condition, another 5 (18%) had a web-based counseling function, and 13 (46%) could provide the user with a treatment plan for the different tinnitus conditions. The primary product information is described in Multimedia Appendix 6.

Analysis of Treatments Provided by the Apps

After examining the treatments mentioned in the app descriptions, we found that, of the 28 apps, 7 (25%) used masking therapy, 2 (7%) used habituation therapy, and 2 (7%) used cognitive behavioral therapy. Moreover, of the 28 apps, 1 (4%) used sound therapy, 1 (4%) used neuromusic therapy, 1 (4%) used play therapy and frequency discrimination therapy, 1 (4%) used transcallosal vagus nerve microcurrent stimulation, and 1 (4%) used neuromodulation therapy. The remaining apps (12/28, 43%) did not specify the type of treatment used.

Analysis of App Validity

Regarding the validity and reliability of the apps, we found that some of the apps (6/28, 21%) were created by developers in collaboration with specific authorities. The 6 partner organizations are the American Tinnitus Association; the Eye, Ear, Nose, and Throat Hospital affiliated to Fudan University (Shanghai, China); a team of doctors from Tsinghua University (Beijing, China); a team of tinnitus specialists from well-known tertiary hospitals and first-line tinnitus experts; a Canadian team with Food and Drug Administration (FDA)–approved tinnitus rehabilitation core technology; and the University of California (Los Angeles, California, United States).

Comparative Analysis of Literature and Apps

Regarding the target audience, the software described in the literature was mostly consistent with the marketed apps. Most of the apps (26/28, 93%) focused on patients and their families, whereas a few (2/28, 7%) targeted medical staff. Most of the apps (20/28, 71%) provided different types of sounds to relieve tinnitus. A few of the apps (5/28, 18%) required the use of hearing aids, being aimed at medical staff use.

In the comparative analysis of the app scenarios, there was a strong consistency between the 2 types of apps. Most of the apps focused on the home scenario (22/28, 79%), whereas a few of the apps (6/28, 21%) are meant to be used in hospitals or other institutions dedicated to professional audiology. The tinnitus treatment apps currently available for download and use are mostly rudimentary. Most of them (22/28, 79%) can only provide initial relief to the patients or provide an assessment of their condition

In a comparative analysis of the treatment techniques, the software described in the literature was predominantly based on sound therapy. This was a preferred approach, probably because, on the one hand, it was easier to implement in combination with mHealth software and, on the other hand, it was effective in tinnitus treatment. Half of the apps (16/28, 57%) listed in the app stores were based on acoustic and cognitive behavioral therapy, with significant user feedback available. Taking advantage of the psychoemotional characteristics of patients with tinnitus, a potential treatment method for tinnitus that involves transcranial magnetic electrical stimulation (containing 2 models: one for transcaudal vagus nerve microcurrent stimulation technology and one for neuromodulation therapy) needs to be developed. The other half (12/28, 43%) of the available apps did not describe the treatment method and lacked user evaluations. This was probably because of the relatively early stage of development of such apps and their limited user results.

A comparative analysis from a clinical standpoint showed that the apps mentioned in the literature were in the theoretical design stage. Relevant studies involving the application to human patients generally need ethics review committee approval, have to meet high product quality requirements, and have to sustain through long development and validation pathways, as well as overcome low translation rates. Individual apps are not marketed because they are not currently registered, despite having been rigorously designed at the theoretical level. Among the marketed apps, the proportion of those in the therapeutic intervention stage is high (5/7, 71%), with a relatively strong practicality and a modest performance.

Discussion

Current Status and Trends in Research

Literature Information

The number of studies in the field of tinnitus apps has increased yearly, with a rapid growth after 2015 and with the annual publication volume peaking in 2021. The articles were published in journals of moderate quality. Nevertheless, the acceptance rates of articles in the field by top medical journals were low. Of the 75 research articles in this field, 51 (68%) were included in the Web of Science core collection and had an average journal impact factor of 5.12 (spanning between 1.245 and 25.617 in the field of psychotherapy and psychosomatics). The studies were mainly published in the *American Journal of Audiology* and in the *Journal of the American Academy of Audiology*, with 18 (n=12, 67% and n=6, 33%, respectively) articles published as of February 2022, representing 24% (18/75) of the total number of articles. Nevertheless, the acceptance rates of articles in this field in top medical journals were poor, with no relevant studies published in the *New England Journal of Medicine*, *The Lancet*, the *Journal of the American Medical Association*, or *The BMJ*. The annual output of the most productive authors was the highest in 2021. In accordance with the Lotka law, this study identified 16 highly productive authors in the field, all of whom have published consistently in the last 3 years. This suggests that this area is steadily and rapidly growing. The authors were mainly from Germany (19/75, 25%), the United Kingdom (16/75, 21%), and the United States (12/75, 16%), with fewer authors from other countries (27/75, 36%), and only a single author from China (1/75, 1%). Of the 75 articles, 47 (63%) were published by authors from Germany, the United Kingdom, and the United States.

The current hot spot in this field is the research on sound and cognitive behavioral therapy, focusing on the outcomes for the treatment of patients with tinnitus. Thematic map analysis showed that the terms in the second quadrant of the thematic map included cognitive behavioral therapy, sound therapy, and tinnitus treatment outcomes. This indicates that research on cognitive behavioral therapy and sound therapy (with a focus on tinnitus treatment outcomes) is a current hot spot in the field of mHealth-based software interventions for tinnitus treatment. This area of research is well developed, with a focus on the assessment of tinnitus, but its importance is not sufficiently recognized yet. Our results are in agreement with those of relevant studies conducted in recent years [22].

App Information

Most of the apps (16/28, 57%) in this field have been newly developed and provide functional support for therapeutic interventions for tinnitus. Nevertheless, software development in this field lacks a research basis. Such software often has high requirements for ancillary apps [23], such as the sound quality and calibration of headphones. Cognitive behavioral therapy relies on professional tinnitus specialists for guidance. By contrast, it is difficult to achieve an accurate grasp of a user's criteria for the use of such apps with mobile phone sensors or wearable apps alone [24]. As a result, and in comparison with other fields, the number of diagnostic apps is relatively small.

The adoption of apps under development has been poor, with the vast majority of the available apps used in very few or no real-life scenarios. The results showed that only 28 designed apps were available for download from app stores. This limitation suggests that, although many tinnitus treatment–related apps have been developed by experts and academics in recent years, most of them still lack practical use. It also suggests that these apps have a very homogeneous collection of patient information and a single-center validation of their effectiveness; in addition, they lack validity studies for clinical retesting as well as user feedback [25]. The reasons behind these aspects could be related to other attributes of the apps, such as their effectiveness and ease of use. To develop and validate apps that can be used in the clinical setting, researchers should also focus on other attributes, such as the ease of use of the apps.

Translation Status

For many years, there has been a lack of translation of the results of tinnitus treatment interventions using mHealth software and apps. This lack of translation is consistent with the results of our study on the analysis of product applications in this area. The vast majority of these apps (22/26, 85%) require both research and development, with scientific validity and effectiveness in a state of urgent need, and are far from truly achieving their goal of an effective tinnitus treatment intervention [26].

Problems and Potential

The Overall Picture of Tinnitus Treatment Apps Is Unsatisfactory

Scoring Analysis of Tinnitus Treatment Apps

Regarding the ratings of the included 28 tinnitus treatment–related apps (n=24, 86% from product data and n=4, 14% from literature data) on the Qimai app data analysis platform, 17 (61%) had ratings as high as 5.00 out of 5.00 and as low as 1.00 out of 5.00 (with an average rating of 3.84 out of 5.00). Of these 17 rated apps, China and the United States accounted for 7 (41%) and 5 (29%), respectively. The 5 apps developed in the United States had ratings of 4.40, 4.80, 4.80, 4.90, and 5.00 (average rating: 4.78), whereas the 7 apps developed in China had ratings of 1.00, 1.80, 3.80, 4.00, 4.10, 4.30, and 4.70 (average rating: 3.39).

Evaluation Analysis of Tinnitus Treatment Apps

With regard to the information on the 28 tinnitus treatment–related apps provided on the Qimai app data analysis platform, 10 good reviews were collected over 1 year. All these reviews referred to 6 (21%) of the 28 apps and were about their usefulness and usability. After analyzing the geographical regions of origin, our study found that the developers of 5 (83%) of these 6 apps were from China, with 5 positive reviews and 3 negative reviews.

Traditional Treatment Methods Are Far From Satisfactory

There is currently no international uniform treatment protocol for tinnitus. Moreover, evidence-based medical research has not found a single method with definitive efficacy for all types of tinnitus. The available methods include medications, repetitive transcranial magnetic stimulation, electrical stimulation of the tympanic capsule, cochlear implants, electrical stimulation of the vagus nerve, and acupuncture [27]. The treatment of tinnitus relies on obtaining an accurate hearing profile of the patient as well as indicators of the central frequency and intensity of the sound produced by tinnitus. After excluding the etiologies associated with tinnitus, counseling, cognitive behavioral therapy, and sound therapy have become the main treatment approaches [28]. Other approaches include relaxation and sequential therapy [29]. The lack of adjunctive therapies to improve outcomes in complex conditions is a common problem in tinnitus treatment, recognized both domestically and internationally.

Clear Differences in Tinnitus Treatment Apps at the Research Level Versus the App Level

Current research has uncovered the lack of connection between mHealth apps and their applications. In recent years, national and international research scholars have recognized the roles of neural synchronization and remodeling the mechanisms of action in tinnitus [30]. On the basis of these findings, adjunctive sound therapy involves the administration of appropriate acoustic stimulation (of a specific frequency and duration) to break abnormal nerve synchronization and remodeling. This procedure seeks to form a new auditory center, thereby reducing or eliminating tinnitus [31]. The use of mHealth apps to provide appropriate auditory training for patients with tinnitus has proven effective at the research level. Nevertheless, important aspects are lacking and need to be addressed. These include a scientific basis for software function, the standardization and calibration of the sounds emitted, real-time guidance from experts during the use of the apps by patients, clinical validation and feedback from patients on product effectiveness, and the connection between research and apps in this field.

Significant Differences in the Status of Relevant Research and Apps Between China and Other Countries

Our study showed that, although most research and apps in this area were developed outside China, there are still important limitations. The sample size of research and apps in this field, both domestically and abroad, was small. The final number of apps found to meet all inclusion criteria was only 28, of which only 4 (14%) were documented in the literature. This limitation has a particular impact on the results of our study. Moreover, it also reflects the decreased volume of research in this field, which strengthens the innovation and value of our work.

Tinnitus Is Difficult to Treat and Requires Urgent Assistance From Emerging Technologies

With the increased stress levels in modern work and life, tinnitus does not appear in isolation. Nevertheless, it is a combination of symptoms closely related to the etiological mechanisms and aggravating factors of several systemic diseases. More than 50% of patients with tinnitus experience sleep disturbances [32]. In the United Kingdom, the prevalence rate of tinnitus in the adult population is 10.1%, with the disorder severely affecting the quality of daily life [33]. In the United States, the prevalence rate of tinnitus in adolescents is from 7% to 9%, and in older people, it is approximately 25% [34]. In Japan, the prevalence rate of tinnitus in the adult population is 11.9% [35]. Although there are no large-scale epidemiological data on the prevalence rates of tinnitus in China, available studies have reported its prevalence rate in the adult population to be approximately 10%. Approximately 5% of the patients with tinnitus seek medical treatment, and approximately 2% of the patients have tinnitus that severely affects their life, sleep, the ability to concentrate and work, and social activities [36]. The complex and heterogeneous etiology of tinnitus as well as its unclear pathogenesis have contributed to the poor treatment outcomes for this disorder. Nevertheless, as industry and technology continue to develop and society ages, the prevalence of tinnitus is gradually increasing, making its early detection, diagnosis, and effective treatment urgent.

The evidence-based guidelines for clinicians managing patients with tinnitus [1] recommend an acoustic therapy approach, in which any sound can be used to alter the perception of, and response to, tinnitus, for the patient's benefit. This therapy can be performed with environmental optimization devices such as hearing aids, sound generators, and tinnitus hybrid acoustic therapy devices, as well as commonly used devices such as mp3 players, smartphones, and radios. Tinnitus rehabilitation approaches must not disregard the integration of traditional methods with innovative technologies [37]. With the expansion and increased application of the mobile internet in various industries, health care–focused mobile internet technology has been receiving increased interest owing to its rapid development. Therefore, the combination of sound therapy and cognitive behavioral therapy for tinnitus has led to the development of an mHealth app model with this purpose.

Future Developments and Outlook

The Theory of mHealth for Adjunctive Treatment and Intervention in Tinnitus Is Logically Clear and Scientifically Solid

To address the ineffectiveness of traditional treatments for tinnitus, one of the 3 most difficult conditions plaguing human health in the otolaryngology field (the other 2 being deafness and vertigo), researchers have been exploring interdisciplinary avenues [38]. The evidence-based guidelines for clinicians managing patients with tinnitus recommend cognitive behavioral therapy for patients with chronic compensatory tinnitus, as well as effective and feasible acoustic treatment options provided by physicians or hearing aids for patients with tinnitus with hearing loss [39]. mHealth is unique in providing behavioral health interventions. One of its scientific theories is cognitive behavioral therapy. The future integration of mHealth for tinnitus treatment and intervention is in line with the recommendations of professional guidelines and has a sound scientific and theoretical basis.

The Pathway for mHealth to Be Used as an Adjunctive Treatment and Intervention in Tinnitus Is Clear and Highly Feasible

One of the theoretical foundations of mHealth is cognitive behavioral therapy. As of 2016, approximately 500 mHealth apps based on cognitive behavioral theory were available in the Apple App Store and Google Play Store. Current mobile apps are relatively well established in the use of cognitive behavioral therapy interventions in psychosocial health services to relieve anxiety, depression, and stress as well as treat other physical ailments and improve maladaptive behavioral problems (such as eating disorders and substance abuse). As a result, the product experience in mHealth is relatively mature. More than half of all patients with tinnitus have comorbid psychological problems such as anxiety, depression, and sleep disorders [40]. On the basis of the existing experience with mHealth, there is a clear and feasible pathway for adjunctive treatment and intervention in tinnitus.

The Use of mHealth for Adjunctive Treatment and Intervention in Tinnitus Is Effective and Has Good Generalizability

Our findings regarding mHealth use in mental health, disease management, and health behavior promotion indicate that mHealth apps, combined with cognitive behavioral therapy, were highly effective in reducing tinnitus pain, anxiety, and depression as well as in improving the patients' quality of life. In 2019, Aazh et al [41] found that the use of cognitive behavioral therapy for tinnitus was more effective when used 8 to 24 times per week for 60 to 120 minutes each time. The treatment could be remotely guided through an mHealth App to accommodate more patients and promote research progress on effective tools for treating tinnitus [42].

Limitations

Our study has some limitations. First, the data used were mainly from published literature (which included only scientific literature). Second, this study performed a simplified analysis of the product translation. We only analyzed the overall conversion rate and did not analyze specific apps and specific application scenarios. Finally, owing to the difficulty of obtaining complete data on apps and scientific research funds, we have only preliminarily explored the input-output relationship of some of the apps.

Conclusions

The field of monitoring, diagnosing, and treating tinnitus disorders using mHealth apps has shown rapid overall growth during recent years. Accordingly, there has been an increase in the number of relevant publications per year. Progress has been made in research on cognitive behavioral therapy, with a focus on the improvement of tinnitus symptoms and on the development of apps to support tinnitus monitoring and interventions. Nevertheless, the number of studies in this area is still low, international collaborations are lacking, the acceptance rates of articles in this field in refereed medical journals are poor, and most of the developed apps are not used in real-world settings. Future research should address the need for increasing tinnitus awareness and strengthening international collaborations to achieve an improved monitoring of tinnitus, using novel AI tools. In addition to the validation of tinnitus treatment apps, future research should also focus on the app properties that can promote the application of such apps in the real world.

Acknowledgments

The authors are grateful to the many students and research assistants at Peking University, Zhejiang University, and Zhejiang Chinese Medicine University who collaborated with us. This research was supported by the 2022 National College Student Innovation and Entrepreneurship Training Program (202210344001) and the 2022 Zhejiang Provincial Science and Technology Innovation Activity Plan for College Students (2022R410A045).

Data Availability

This study was conducted using public databases. Users can download relevant data free of charge.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Query strategy and results.
[DOCX File , 13 KB - mhealth_v11i1e47553_app1.docx]

Multimedia Appendix 2
Inclusion and exclusion criteria applied to select the target literature.
[DOCX File , 13 KB - mhealth_v11i1e47553_app2.docx]

Multimedia Appendix 3
The evolution in the annual number of relevant studies published over the years.

[PNG File , 44 KB - mhealth_v11i1e47553_app3.png]

Multimedia Appendix 4
Description of basic app information.
[PNG File , 88 KB - mhealth_v11i1e47553_app4.png]

Multimedia Appendix 5
Distribution of software sources.
[PNG File , 9 KB - mhealth_v11i1e47553_app5.png]

Multimedia Appendix 6
Distribution of source countries.
[PNG File , 103 KB - mhealth_v11i1e47553_app6.png]

References

1. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ERJ, et al. Clinical practice guideline: tinnitus. Otolaryngol Head Neck Surg 2014 Oct;151(2 Suppl):S1-40. [doi: 10.1177/0194599814545325] [Medline: 25273878]
2. Barozzi S, Del Bo L, Crocetti A, Dyrlund O, Passoni S, Zolin A, et al. A comparison of nature and technical sounds for tinnitus therapy. Acta Acust United Acust 2016 May;102(3):540-546 [FREE Full text] [doi: 10.3813/aaa.918971]
3. Xu Y, Wang M, Ren F. Acoustic therapy combined with psychological counselling for chronic tinnitus. J Audiol Speech Disord 2015;23(4):358-360. [doi: 10.3969/j.issn.1006-7299.2015.04.007]
4. Zhang Y, Zhang J, Li M. Relationship between the psychoacoustic characteristics of tinnitus and the effect of general environmental sound masking of tinnitus. J Audiol speech disord 2015;23(1):69-72. [doi: 10.3969/j.issn.1006-7299.2015.01.017]
5. Stein A, Wunderlich R, Lau P, Engell A, Wollbrink A, Shaykevich A, et al. Clinical trial on tonal tinnitus with tailor-made notched music training. BMC Neurol 2016 Mar 17;16(1):38 [FREE Full text] [doi: 10.1186/s12883-016-0558-7] [Medline: 26987755]
6. Grand View Research. 2022. URL: https://www.grandviewresearch.com/industry-analysis/mhealth-App-market [accessed 2022-08-22]
7. Castelnuovo G, Simpson S. Ebesity - e-health for obesity - new technologies for the treatment of obesity in clinical psychology and medicine. Clin Pract Epidemiol Ment Health 2011 Mar 04;7(1):5-8 [FREE Full text] [doi: 10.2174/1745017901107010005] [Medline: 21559215]
8. He J. An attempt to treat tinnitus with Audition software. 2017 Presented at: Proceedings of the Inaugural Meeting of the Chinese Medicine Ear, Nose and Throat Professional Committee of Beijing Huaxia Chinese Medicine Development Foundation and the 2017 Symposium on Standardized Treatment of Chinese Medicine Ear, Nose and Throat; 2017; China p. 24-26 URL: https://d.wanfangdata.com.cn/conference/ChZDb25mZXJlbmNlTmV3UzIwMjMwMjI3Egc5OTI1MjM1GggzdGpmMml1dA%3D%3D/>
9. Cai Y, Sun Y, Yang H, Chen L, Huang X, Zhao F, et al. Preliminary observations on personalised notching music for subjective tinnitus. Chin J Otol 2017;15(4):465-470. [doi: 10.3969/j.issn.1672-2922.2017.04.016]
10. Baguley D, McFerran D, Hall D. Tinnitus. Lancet 2013 Nov 09;382(9904):1600-1607 [FREE Full text] [doi: 10.1016/S0140-6736(13)60142-7] [Medline: 23827090]
11. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR mHealth uHealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
12. LeBlanc ES, Patnode CD, Webber EM, Redmond N, Rushkin M, O'Connor EA. Behavioral and pharmacotherapy weight loss interventions to prevent obesity-related morbidity and mortality in adults: updated evidence report and systematic review for the US preventive services task force. JAMA 2018 Sep 18;320(11):1172-1191. [doi: 10.1001/jama.2018.7777] [Medline: 30326501]
13. Qimai Data (formerly ASO100)-professional mobile product business analysis platform: ASO-ASM optimization. Qimai Data. URL: https://www.qimai.cn/ [accessed 2022-08-24]
14. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
15. Vlaescu G, Carlbring P, Lunner T, Andersson G. An e-platform for rehabilitation of persons with hearing problems. Am J Audiol 2015 Sep;24(3):271-275. [doi: 10.1044/2015_AJA-14-0083] [Medline: 26649528]
16. Kaldo-Sandström V, Larsen HC, Andersson G. Internet-based cognitive-behavioral self-help treatment of tinnitus: clinical effectiveness and predictors of outcome. Am J Audiol 2004 Dec;13(2):185-192. [doi: 10.1044/1059-0889(2004/023)] [Medline: 15903144]

17. Andersson G, Lunner T, Laplante-Lévesque A, Preminger JE. Internet and audiology: a review of the first international meeting. Am J Audiol 2015 Sep;24(3):269-270. [doi: 10.1044/2015_AJA-15-0033] [Medline: 26649527]
18. Vickery BC. Bradford's law of scattering. J Doc 1948;4(3):198-203 [FREE Full text]
19. de Solla Price DJ. Little Science, Big Science. New York, NY: Columbia University Press; 1965.
20. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. J Informetr 2011 Jan;5(1):146-166 [FREE Full text] [doi: 10.1016/j.joi.2010.10.002]
21. VOSviewer Visualizing scientific landscapes. 2023. URL: https://www.vosviewer.com/ [accessed 2023-08-22]
22. Papatheodoropoulos C, Kouvaros S. High-frequency stimulation-induced synaptic potentiation in dorsal and ventral CA1 hippocampal synapses: the involvement of NMDA receptors, mGluR5, and (L-type) voltage-gated calcium channels. Learn Mem 2016 Aug 16;23(9):460-464 [FREE Full text] [doi: 10.1101/lm.042531.116] [Medline: 27531836]
23. Blase KL, van Dijke A, Cluitmans PJ, Vermetten E. Efficacy of HRV-biofeedback as additional treatment of depression and PTSD. Article in Dutch. Tijdschr Psychiatr 2016;58(4):292-300 [FREE Full text] [Medline: 27075221]
24. Huang X, Yang H, Xiong H. The use of full-parameter biofeedback indicators in patients with subjective tinnitus. Chinese J Hear Speech Rehabilitation Sci 2014;281:12. [doi: 10.3969/j.issn.1672-4933.2014.04.010]
25. Chinese Journal of Otolaryngology-Head and Neck Surgery Editorial Committee Otology Professional Group. 2012 tinnitus expert consensus and interpretation. J Otolaryngol Head N 2012 Mar;47(09):709-712 [FREE Full text] [doi: 10.3760/cma.j.issn.1673-0860.2012.09.002]
26. Ruo J, Bo L. The application of emotional cognitive theory to the study of tinnitus. Journal of Clinical Otolaryngology-Head and Neck Surgery 2015:691. [doi: 10.13201/j.issn.1001-1781.2015.08.005]
27. Manabe Y, Sato K, Fukuda S, Miyashita T. Unpleasantness of amplified environmental sound used in tinnitus sound therapy: a preliminary study of clinical assessment. J Int Adv Otol 2019 Apr;15(1):94-98 [FREE Full text] [doi: 10.5152/iao.2019.6581] [Medline: 30924776]
28. Asadpour A, Jahed M, Mahmoudian S. Brain waves evaluation of sound therapy in chronic subjective tinnitus cases using wavelet decomposition. Front Integr Neurosci 2018 Sep 19;12:38 [FREE Full text] [doi: 10.3389/fnint.2018.00038] [Medline: 30283307]
29. Searchfield GD, Linford T, Durai M. Sound therapy and aural rehabilitation for tinnitus: a person centred therapy framework based on an ecological model of tinnitus. Disabil Rehabil 2019 Aug;41(16):1966-1973. [doi: 10.1080/09638288.2018.1451928] [Medline: 29571274]
30. Zhou X, Qian X, Gu L. Analysis of the efficacy of compound sound with the addition of background tones in the treatment of loss of compensatory tinnitus at different frequencies. J Otolaryngol Ophthalmol 2019;33(5):37-93. [doi: 10.16542/j.cnki.issn.1007-4856.2020.05.012]
31. Sereda M, Xia J, El Refaie A, Hall DA, Hoare DJ. Sound therapy (using amplification devices and/or sound generators) for tinnitus. Cochrane Database Syst Rev 2018 Dec 27;12(12):CD013094 [FREE Full text] [doi: 10.1002/14651858.CD013094.pub2] [Medline: 30589445]
32. Hofmeier B, Wertz J, Refat F, Hinrichs P, Saemisch J, Singer W, et al. Functional biomarkers that distinguish between tinnitus with and without hyperacusis. Clin Transl Med 2021 May;11(5):e378 [FREE Full text] [doi: 10.1002/ctm2.378] [Medline: 34047478]
33. Tyler RS. Tinnitus Handbook. California, CA: Singular Publishing Goup; 2000.
34. Loprinzi PD, Lee H, Gilham B, Cardinal BJ. Association between accelerometer-assessed physical activity and tinnitus, NHANES 2005-2006. Res Q Exerc Sport 2013 Jun;84(2):177-185. [doi: 10.1080/02701367.2013.784840] [Medline: 23930543]
35. Fujii K, Nagata C, Nakamura K, Kawachi T, Takatsuka N, Oba S, et al. Prevalence of tinnitus in community-dwelling Japanese adults. J Epidemiol 2011;21(4):299-304 [FREE Full text] [doi: 10.2188/jea.je20100124] [Medline: 21646745]
36. Wang H. New advances in the diagnosis and treatment of tinnitus. J Clin Med 2005;21(2):114-116. [doi: 10.3969/j.issn.1006-5725.2005.02.003]
37. Jiang T, Gong S, Wang J. Tinnitus audiology treatment: theoretical framework and clinical applications. J Clin Otolaryngol Head Neck Surg 2014;28(4):225-228. [doi: 10.13201/j.issn.1001-1781.2014.04.004]
38. Duan W, Zhang J, He L. Theory and effectiveness evaluation of mHealth. Adv Psychol Sci 2022 Jul 13;26(7):1232-1243. [doi: 10.3724/SP.J.1042.2018.01232]
39. Stein A, Engell A, Junghoefer M, Wunderlich R, Lau P, Wollbrink A, et al. Inhibition-induced plasticity in tinnitus patients after repetitive exposure to tailor-made notched music. Clin Neurophysiol 2015 May;126(5):1007-1015. [doi: 10.1016/j.clinph.2014.08.017] [Medline: 25441152]
40. Brueggemann P, Mebus W, Boecking B, Amarjargal N, Niemann U, Spiliopoulou M, et al. Dimensions of tinnitus-related distress. Brain Sci 2022 Feb 16;12(2):275 [FREE Full text] [doi: 10.3390/brainsci12020275] [Medline: 35204037]
41. Aazh H, Landgrebe M, Danesh AA, Moore BC. Cognitive behavioral therapy for alleviating the distress caused by tinnitus, hyperacusis and misophonia: current perspectives. Psychol Res Behav Manag 2019 Oct 23;12:991-1002 [FREE Full text] [doi: 10.2147/PRBM.S179138] [Medline: 31749641]

42. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]

Abbreviations

AI: artificial intelligence
EEG: electroencephalogram
FDA: Food and Drug Administration
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 24.03.23; peer-reviewed by S Cao, J Liang; comments to author 17.04.23; revised version received 28.05.23; accepted 21.07.23; published 24.08.23.

Please cite as:
Hu Y, Lu Y, Tian C, He Y, Rong K, Pan S, Lei J
Current Status and Trends in mHealth-Based Research for Treatment and Intervention in Tinnitus: Bibliometric and Comparative Product Analysis
JMIR Mhealth Uhealth 2023;11:e47553
URL: https://mhealth.jmir.org/2023/1/e47553
doi:10.2196/47553
PMID:37616044

Review

The Use of Patient-Oriented Mobile Phone Apps in Oral Health: Scoping Review

Elina Väyrynen[1*], DDS; Sanna Hakola[1*], BDS; Anniina Keski-Salmi[1*], BDS; Hannaleena Jämsä[1], DDS; Raija Vainionpää[1], DDS, PhD; Saujanya Karki[1], BDS, MPH, PhD

Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland

[*]these authors contributed equally

Corresponding Author:
Saujanya Karki, BDS, MPH, PhD
Research Unit of Population Health
Faculty of Medicine
University of Oulu
Aapistie 3
Oulu, 90220
Finland
Phone: 358 294485643
Email: saujanya.karki@oulu.fi

Abstract

Background: Oral health is a significant part of general health. Poor oral health can influence an individual's appearance, self-esteem, eating, and speaking. The use of mobile phone apps has been growing in the field of medicine, including dentistry. However, to date, there is no evidence related to the availability of mobile apps focusing on various branches of dentistry.

Objective: The aim of this study was to review the scientific literature on the use of patient-oriented mobile phone apps in oral health and summarize the key findings.

Methods: A scoping review of published scientific literature on the use of patient-oriented mobile phone apps in oral health was conducted in accordance with the Joanna Briggs Institute. A search was performed in PubMed and Scopus for studies published between January 2000 and June 2021 that were written in English. All study types except for those reporting developmental protocols were included in this review. In total, 2 reviewers independently screened the studies using the eligibility criteria. The study protocol was registered in the Open Science Framework registries in June 2021.

Results: The initial search yielded a total of 977 studies, 45 (4.6%) of which met the inclusion criteria. All the studies (45/45, 100%) were published after 2009. Most studies (31/45, 69%) concerned oral health promotion using mobile phone apps, followed by behavior management (5/45, 11%). More than half (23/45, 51%) of the included studies were conducted in Asian countries. Overall, 31% (14/45) of the studies focused on adolescents. A total of 51% (23/45) of the studies were randomized controlled trials (RCTs). Approximately 39% (9/23) of the included RCT studies reported a substantial reduction in dental plaque, and 26% (6/23) of the studies reported significant improvement in gingival health. Regarding dental anxiety management, 13% (3/23) of the RCT studies reported a significant decrease in mean heart rate and lower Facial Image Scale scores.

Conclusions: According to the literature, the use of mobile apps in oral health is increasing among patients, mainly children and adolescents. Many studies that have used mobile apps have focused on promoting oral health. However, other areas such as diagnostic and remote consultations (teledentistry) have until recently been neglected despite their great potential.

(JMIR Mhealth Uhealth 2023;11:e46143) doi:10.2196/46143

KEYWORDS

oral health; dentistry; mobile apps; mobile health; mHealth; mobile phone

Introduction

Background

Oral health is a significant part of general health. Poor oral health can influence an individual's appearance, self-esteem, eating, and speaking [1]. Poor oral health is also one of the leading causes of loss of work productivity, increased school absenteeism, and reduced academic performance [2,3].

The risk factors for the most common oral diseases, such as dental caries and periodontal diseases, include physical, biological, environmental, behavioral, and lifestyle-related factors [4,5]. Most of the behavioral and lifestyle-related factors responsible for oral diseases are also common to various systemic diseases [6]. For instance, frequent consumption of foods and drinks containing free sugar can lead to dental caries as well as obesity [7]. Similarly, tobacco use is a risk factor for periodontal diseases, cardiovascular disease, respiratory diseases, and cancer [6,8]. In addition, poor oral health has an adverse effect on individuals' general health. For example, periodontitis is considered one of the risk factors in the pathogenesis of diabetes mellitus, cardiovascular disease, kidney disease, and recurrent pneumonia [7,9]. Despite being preventable, >3.5 billion people have been affected by oral diseases worldwide, predominantly dental caries and periodontal diseases [10]. Although the concept of managing oral diseases has shifted toward prevention at the individual level, actions are being taken to move toward a more patient-centered management of oral diseases, focusing on promoting and maintaining good oral health in partnership with the patient [11].

Health promotion has been a key component in disease prevention. Behavior change approaches, such as communicating disease risk information and the self-monitoring of one's own health as part of health promotion, have been successful in modifying the health behavior of individuals. Behavior change approaches refer to the specific strategies used in interventions to promote behavior change [12,13]. Behavioral interventions involve people's decision-making process about their health.

Recently, the use of mobile technologies in the medical field (also known as mobile health [mHealth]) has been growing. With their many features and high use by consumers, mobile phones are one of the devices suitable for accessing health information, improving the quality and coverage of health care, and promoting health [14,15]. In addition, the use of mobile phone apps has been shown to have positive outcomes in general health. For instance, mobile apps are used for remote consultation [16], disease diagnosis [17], reminders for patients [18], and behavior modifications [19].

As of 2017, there were >325,000 mobile apps available on the Google Play Store and Apple App Store platforms that mainly focused on health and well-being [20]. Similarly, in dentistry, a total of 1075 oral hygiene apps were available as of 2018 [21]. Evidence on the availability of mobile apps focusing on various branches of dentistry is still unclear. In addition, with mHealth being popular, it is important for health professionals to know the availability of different kinds of mobile apps and be assured that the mobile apps are disseminating fact-based information [22]. Therefore, the aim of this study was to examine the main findings in the literature on the use of patient-oriented mobile apps in oral health.

Review Questions

The primary question is as follows: What are the main findings in the literature related to the use of patient-oriented mobile apps in oral health? The secondary questions are as follows: What types of evidence are available in the literature related to the primary question? and How has this research been conducted?

Methods

Information Sources and Search Strategies

This scoping review was conducted in accordance with the Joanna Briggs Institute methodology for scoping reviews [23]. A preliminary search of MEDLINE, the Cochrane Database of Systematic Reviews, and *JBI Evidence Synthesis* was conducted on January 22, 2021, and no current or ongoing systematic or scoping reviews on the topic were identified. The text words contained in the titles and abstracts of relevant articles and the index terms used to describe the articles were used to develop a full search strategy for PubMed and Scopus. The search strategy, including all the identified keywords and index terms, was adapted for each database included. The reference lists of all included sources of evidence were screened for additional studies. The search terms used were "Mobile Applications" (Medical Subject Headings [MeSH]) OR "Mobile Application*" OR "Mobile App*" (text word) OR "Mobile health" (text word) OR "mHealth" (text word) OR "Health app*" (text word) OR "smartphone app*" (text word) AND "Dentistry" (MeSH) OR "dental*" (text word) OR "Oral Health" (MeSH) OR "Oral Hygiene" (text word) OR "Oral Medicine" (text word).

The study protocol was registered in the Open Science Framework registries in June 2021 [24].

Eligibility Criteria

Overview

This scoping review considered all study types (experimental and quasi-experimental, cohort, case-control, cross-sectional, case series, case reports, descriptive cross-sectional studies, systematic reviews, and both qualitative and quantitative studies) that were published between January 2000 and June 2021 in English. Studies focusing on the development of mobile apps related to oral health were excluded. The participants, concept, and context guides for this scoping review are as follows.

Participants

This study included all age groups.

Concept

Patient-oriented mobile apps that are used in oral health were included.

Context

All areas, cultures, and sexes were included.

Study and Source of Evidence Selection

Following the search, all identified citations were collated and uploaded into the Covidence data screening and extraction tool, and duplicates were removed. To ensure the calibration, a pilot test was arranged. A total of 25 samples of titles and abstracts were selected for the pilot test by one of the team members (SK), 2 reviewers (AK-S and SH) conducted the pilot test using the eligibility criteria, and the agreement was 84% [23]. Following the pilot test, titles and abstracts were screened by 2 independent reviewers (AK-S and SH) for assessment against the inclusion criteria. Full texts of potentially relevant sources were retrieved, and their citation details were imported into the Mendeley Reference Manager (Mendeley Ltd). The full texts of selected citations were assessed in detail against the inclusion criteria by 2 independent reviewers. The reason for excluding sources was that the full text did not meet the inclusion criteria of this scoping review. Any disagreements between the reviewers at each stage of the selection process were resolved with an additional reviewer (SK). The results of the search and the study inclusion process were reported in full in the final scoping review and presented in a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram [25].

Data Extraction, Data Analyses, and Risk-of-Bias Assessment

The extracted data included specific details about the participants, concept, context, study methods, and key findings relevant to the review questions. On the basis of the extracted data, the proportions of year of publication, subgroups based on focus areas, and study types were calculated and presented in a table.

No risk-of-bias assessment was performed in accordance with the Joanna Briggs Institute methodology for scoping reviews [23].

Results

Selection of Studies

A total of 977 studies were found in the initial search, as shown in Figure 1. After removing duplicates, 64.4% (629/977) of the studies were screened. Studies that did not meet the inclusion criteria were excluded. The most common reasons for exclusion were that the study was not related to oral health, it did not include a mobile app, the full text was not available, or the study was not written in English. A total of 45 studies were included in the scoping review.

Data were extracted from the included studies (n=45), and their general characteristics are presented in Table 1. All the included studies (45/45, 100%) were categorized into different groups depending on the focus area of the mobile apps. The categories are oral health promotion, diagnostic, orthodontic, behavior management, trauma, and other.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of study selection progress.

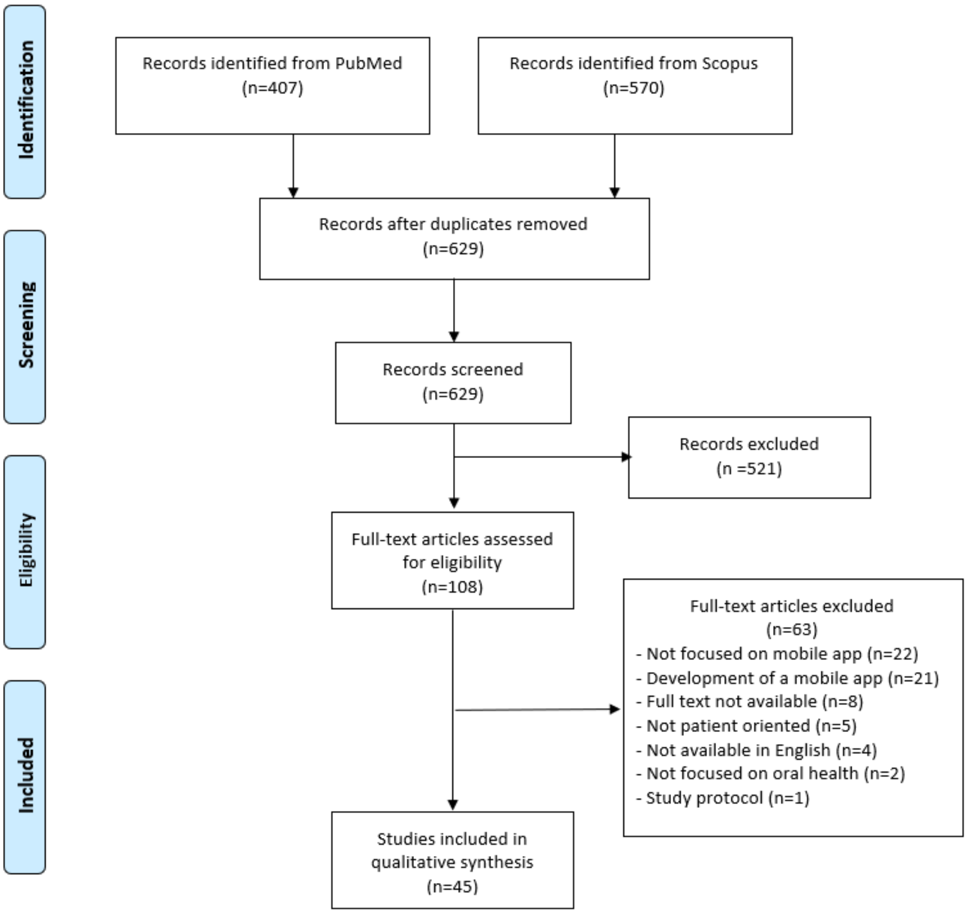


Table 1. General characteristics of the included studies (n=45).

Characteristic	Studies, n (%)
Year of publication	
2010-2019	22 (49)
2020-2021[a]	23 (51)
Category	
Oral health promotion	31 (69)
Diagnostic	1 (2)
Orthodontic	5 (11)
Behavior management	5 (11)
Trauma	2 (4)
Others	1 (2)
Study type	
Randomized controlled trial	23 (51)
Nonrandomized controlled trial and quasi-experimental study	5 (11)
Cross-sectional study	6 (13)
Longitudinal study	6 (13)
Case report	1 (2)
Systematic review	4 (9)

[a]June 2021.

Characteristics of the Included Studies

Most studies (31/45, 69%) were categorized under oral health promotion. All studies (45/45, 100%) were published between 2009 and 2021. One-third (14/45, 31%) of the studies focused on adolescents aged 10 to 19 years, and 24% (11/45) of the studies focused on children aged <10 years. More than half (23/45, 51%) of the included studies were randomized controlled trials (RCTs). More than half (23/45, 51%) of the included studies were from Asia (Figure 2).

Table 2 shows the main findings of the included studies. Most of the included studies (21/45, 47%) focused on improving oral hygiene through effective toothbrushing. A total of 20% (9/45) of the studies focused on knowledge, attitudes, and practice [26-34]. In total, 11% (5/45) of the studies were based on treatment outcome [35-39], and 9% (4/45) of the studies reported on the use of mobile apps in the management of dental anxiety [40-43]. Only 2% (1/45) of the studies were based on disease diagnosis [44], and 2% (1/45) of the studies were based on remote dental care services [45]. Among the 45 studies, 4 (9%) systematic reviews were also included [46-49].

Approximately 39% (9/23) of the included RCT studies reported a significant reduction in dental plaque [26,46,53-57,59,63]. Similarly, 26% (6/23) of the studies reported significant improvement in gingival health measured via bleeding on probing [51,53-55,57,59], and only 4% (1/23) of the studies reported a reduction in enamel caries [59]. Regarding dental anxiety management, 75% (3/4) of RCT studies reported a significant decrease in mean heart rate and lower image scale scores compared with the controls [41-43].

Figure 2. Geographical areas of the included studies (n=45).

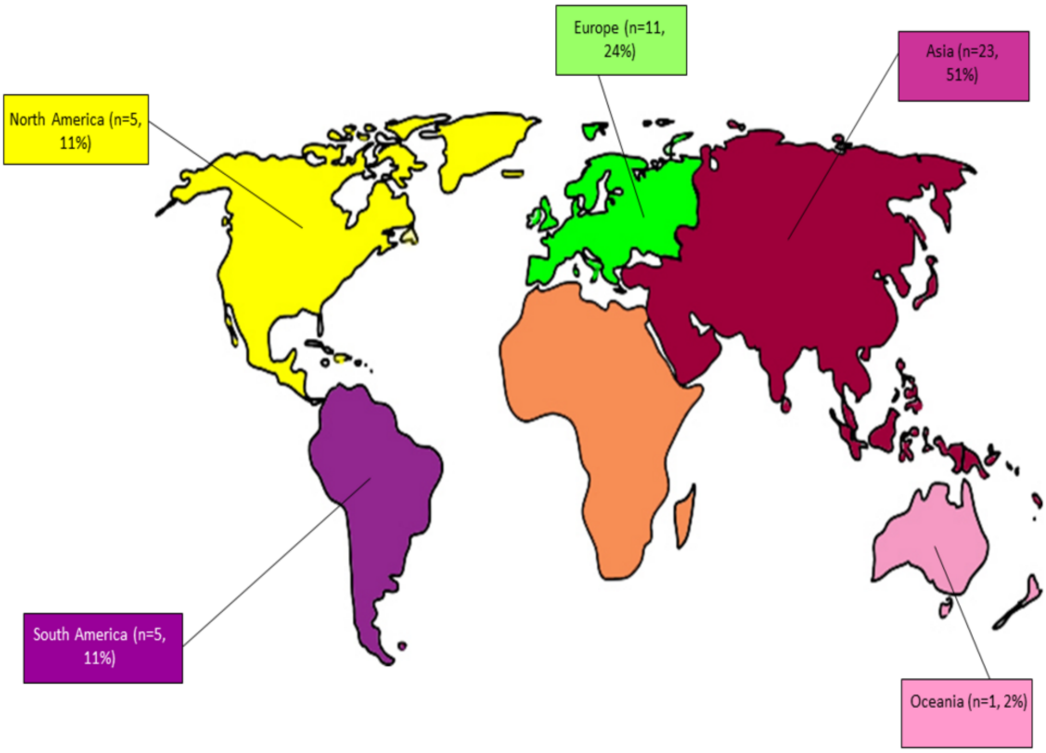


Table 2. Main findings of the included studies (n=45).

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Alkilzy et al [50], 2019	Germany	Randomized controlled trial	60 recruited; 49 completed	Children aged 5-6 y	Investigate the efficacy of a manual toothbrush with a gravity sensor and mobile app for improving manual toothbrushing	At the 6- and 12-week follow-ups, the test group showed statistically significantly better oral health indexes than the controls. After the 6-week follow-up, the Quigley-Hein Plaque Index was 0.8 (SD 0.5) for the test group and 1.88 (SD 0.9) for the control group (*P*<.001), and the Papillary Bleeding Index was 0.08 (SD 0.1) for the test group and 0.26 (SD 0.2) for the control group (*P*<.001). After the 12-week follow-up, the Quigley-Hein Plaque Index was 0.44 (SD 0.5) for the test group and 1.49 (SD 0.7) for the control group (*P*<.001), and the Papillary Bleeding Index was 0.05 (SD 0.18) for the test group and 0.21 (SD 0.1) for the control group (*P*<.001).
Chang et al [51], 2021	Taiwan	Randomized controlled trial	150 recruited; 88 completed	26-77 y	Investigate the effectiveness of a mobile app (OSCA[a]) in improving OHS[b] and OHBs[c]	After the 4- to 8-week follow-ups, no significant difference in OHS improvement measured using the O'Leary PCR[d] was found between the control and intervention groups (mean OHS improvement was 17.0, SD 18.84 in the control group and 26.17, SD 21.76 in the intervention group; *P*=.06). OHBs (frequency of toothbrushing, duration of toothbrushing, interdental cleaning, and tongue cleaning) in both the control and intervention groups improved significantly (*P*=.007 and *P*<.001, respectively).
Alkadhi et al [52], 2017	Saudi Arabia	Randomized controlled trial	44	≥12 y	Investigate the impact of using mobile app active reminders to improve oral hygiene compared with verbal oral hygiene instructions	Both PI[e] and GI[f] significantly decreased after 4 weeks of using active reminders of oral hygiene instructions on a mobile app compared with verbal oral hygiene instructions (*P*=.04 and *P*=.02, respectively). For mobile app users, mean PI was 0.8007 (SD 0.4062) at baseline and 0.6677 (SD 0.3146) after 4 weeks, and mean GI was 0.3450 (SD 0.2955) at baseline and 0.2273 (SD 0.2256) after 4 weeks. For verbal oral hygiene instruction recipients, mean PI was 0.8959 (SD 0.4824) at baseline and 0.9891 (SD 0.5244) after 4 weeks, and mean GI was 0.4927 (SD 0.3005) at baseline and 0.5941 (SD 0.5679) after 4 weeks.
Farhadifard et al [53], 2020	Iran	Randomized controlled trial	120	15-25 y	Evaluate the efficacy of a smartphone app (Brush DJ) for oral hygiene compliance of patients with fixed orthodontic appliances	Significant improvements in PI and GI were observed in the group using the smartphone app (Brush DJ) compared with conventional oral hygiene instruction recipients after 4 weeks (T1), 8 weeks (T2), and 12 weeks (T3; *P*<.001). For mobile app users, mean PI was 75.21 (SD 13.36) at baseline, 73.39 (SD 12.50) at T1, 69.18 (SD 11.84) at T2, and 67.84 (SD 12.33) at T3, whereas the mean GI was 1.29 (SD 0.49) at baseline, 1.20 (SD 0.04) at T1, 1.04 (SD 0.04) at T2, and 1.00 (SD 0.05) at T3. For conventional oral hygiene instruction recipients, mean PI was 76.59 (SD 12.76) at baseline, 76.89 (SD 11.11) at T1, 78.90 (SD 8.89) at T2, and 80.82 (SD 10.05) at T3, whereas the mean GI was 1.49 (SD 0.59) at baseline, 1.35 (SD 0.04) at T1, 1.41 (SD 0.04) at T2, and 1.37 (SD 0.05) at T3.
Deleuse et al [54], 2020	Belgium	Randomized controlled trial	44 recruited; 38 completed	12-18 y	Compare the use of an oscillating electrical toothbrush with an internet-based oscillating electrical toothbrush connected to a brushing aid app in adolescent patients treated with fixed multibracket orthodontic appliances	PI and GI decreased significantly in both the control and intervention groups. WSL[g] score was stable for both groups. PI was significantly lower in the app group than in the control group (*P*=.01) after 12 weeks. GI decreased significantly in each group (control group: *P*=.003; test group: *P*=.001), and no difference was observed between the 2 groups. WSL scores remained stable in each group (control group: *P*=.07; test group: *P*=.73), and no difference was observed between the 2 groups (*P*=.28)

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Desai et al [55], 2021	India	Randomized controlled trial	247	4-6 y	Test the impact of a mobile app (Brush Up) on OHB in children	After the 1-month follow-up, plaque scores measured using the visible biofilm index was lower in the mobile app group (mean rank scores were 112.5 and 70.9 at baseline and follow-up, respectively) than those of the video demonstration group (mean rank scores were 114.8 and 112.5 at baseline and follow-up, respectively) and the manual demonstration group (mean rank scores were 134.8 and 176.8 at baseline and follow-up, respectively; *P*<.001). There was also significant change in the frequency and duration of toothbrushing, cleaning of the lingual surfaces of the teeth, and tongue cleaning in all groups (*P*<.001).
Kay and Shou [56], 2019	United Kingdom	Randomized controlled trial	108 recruited; 103 completed	18-69 y	Investigate the effectiveness and acceptability of a smartphone app used in conjunction with a movement sensor toothbrushing attachment (device) for reducing plaque levels	After the 4-week follow-up, mean full mouth plaque scores declined from 40.1 to 11.7 in the test group compared with a reduction from 29.1 to 20.5 in the control group (*P*<.001).
Marchetti et al [26], 2018	Brazil	Randomized controlled trial	291 recruited; 263 completed	14-19 y	Study the effectiveness of an app associated with common education methods in adolescents' oral health	After the 1-month follow-up, a significant difference in mean KS[h] between the adolescents who used the app (mean 4.77, SD 0.52) and those who did not (mean 4.35, SD 0.66; *P*<.001) was observed, and OHI-S[i] decreased significantly among app users (mean OHI-S scores were 1.31, SD 0.37 at phase I and 0.24, SD 0.18 at phase IV, with *P*<.001, for the oral guidance plus app arm and 1.21, SD 0.39 at phase I and 0.23, SD 0.22 at phase IV, with *P*<.001, for the video guidance plus app arm). GBI[j] decreased significantly among app users (mean GBI scores were 11.57, SD 5.09 at phase I and 2.03, SD 1.56 at phase IV, with *P*<.001, for the oral guidance plus app arm and 9.76, SD 4.07 at phase I and 1.87, SD 2.23 at phase IV, with *P*<.001, for the video guidance plus app arm).
Shida et al [57], 2020	Japan	Randomized controlled trial	118 recruited; 112 completed	≥18 y	Compare the effectiveness between brushing teeth with the help of a mobile app and usual brushing instructions	The mean 6-point PCR score at week 4 was 45.05% in the intervention group and 49.65% in the control group. The change of PCR score from baseline was −20.46% in the intervention group and −15.77% in the control group, indicating no statistically significant difference (95% CI −0.70 to 10.07; *P*=.09).
Zotti et al [58], 2016	Italy	Randomized controlled trial	80	Control group: mean age 13.6 y; study group: mean age 14.1 y	Study the effectiveness of the use of a mobile app via chat room participation (WhatsApp) to improve oral hygiene in adolescents wearing fixed appliances	After 6, 9, and 12 months, study group patients had significantly lower scores for both PI and GI and a lower incidence of new WSs[k] compared with the control group. Mean PI scores at 12 months were 1.06 (SD 0.47) for the study group and 1.79 (SD 0.54) for the control group (*P*<.001). Mean GI scores at 12 months were 0.67 (SD 0.36) for the study group and 1.40 (SD 0.57) for the control group (*P*<.001). The number of patients with WSs was 7 for the study group and 16 for the control group (*P*<.05).
Scheerman et al [59], 2020	The Netherlands	Randomized controlled trial	132 recruited; 124 completed	12-16 y	Study the effectiveness of the use of a mobile app (WhiteTeeth) to improve oral hygiene in adolescent patients with fixed orthodontic appliances	At the 6-week follow-up, the intervention led to a significant decrease in gingival bleeding (B=−3.74, 95% CI −6.84 to −0.65; *P*=.02) and an increase in the use of fluoride mouth rinse (B=1.93, 95% CI 0.36-3.50; *P*=.02). At the 12-week follow-up, dental plaque accumulation (B=−11.32, 95% CI −20.57 to −2.07; *P*=.02) and the number of sites covered with plaque (B=−6.77, 95% CI −11.67 to −1.87; *P*=.007) had decreased significantly more in the intervention group than in the control group.

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Patil et al [46], 2021	Saudi Arabia	Systematic review	N/A[l]	N/A	Review the effectiveness of the use of mobile apps to improve oral hygiene in patients with orthodontic appliances	Mobile apps have a significant short-term effect for improving oral hygiene when measuring using PI and GI scores. The intervention groups (62%) had a lower level of plaque at a 12-week interval as compared with the control group (72%).
Underwood et al [60], 2015	United Kingdom	Cross-sectional survey	189	All age groups	Evaluate the user experience of using a mobile app (Brush DJ) to provide a basis for future research	A total of 70% of respondents reported that their teeth felt cleaner since using the app. A total of 88% of respondents reported that the app motivated them to brush their teeth for longer. A total of 92.3% of respondents would recommend the app to their friends and family.
Toniazzo et al [47], 2019	Brazil	Systematic review and meta-analysis	N/A	N/A	Assess the effectiveness of the use of mobile apps and SMS text messages and compare them with conventional oral hygiene instructions to improve oral hygiene	The use of mobile apps and SMS text messages significantly improved oral health compared with conventional oral hygiene instructions. The pooled SMD[m] for the PI was −9.43 (95% CI −14.36 to −4.495; I^2=99%; *P*<.001), and that of gingival bleeding was −8.54 (95% CI −13.16 to −3.91; I^2=99%; *P*<.001).
Rasmus et al [61], 2021	Finland	Longitudinal study (follow-up)	36	Children aged 4-12 y and their parents	Investigate the acceptability of a mobile app (Denny the Tooth and Denny the Timer) and evaluate OHB change	After the 5-week follow-up, most of the children considered the Denny the Tooth app clear (n=34), amusing (n=31), and useful (n=29). Denny the Timer was useful, and the odds for toothbrushing frequency significantly increased (OR[n] 8.9, 95% CI 1.29-60.60; *P*=.03).
Scheerman et al [62], 2020	Iran	Randomized controlled trial	791 recruited; 718 completed	12-17 y	Study the effectiveness of a mobile app (Telegram) to promote and improve OHB	Increases in adolescent toothbrushing at the 1- and 6-month follow-ups in both intervention groups (adolescents only [A] and adolescents and their mothers [M+A]) compared with the control group were observed (1-month follow-up: B=3.74, SE 0.28, and *P*<.001 for the M+A group and B=2.64, SE 0.29, and *P*<.001 for the A group; 6-month follow-up: B=3.90, SE 0.27, and *P*<.001 for the M+A group and B=2.78, SE 0.29, and *P*<.001 for the A group). Adolescents in both intervention groups showed a significantly greater improvement in their VPI[o] scores than adolescents in the control group at the 1- and 6-month follow-ups (*P*<.01; 1-month follow-up: B=−0.60, SE 0.05, and *P*<.001 for the M+A group and B=−0.29, SE 0.07, and *P*<.001 for the A group; 6-month follow-up: B=−0.64, SE 0.08, and *P*<.001 for the M+A group and B=−0.33, SE 0.06, and *P*<.001 for the A group).
Humm et al [63], 2020	Switzerland	Randomized controlled trial	20	≥18 y	Determine whether a smartphone app used with an electric toothbrush improves plaque removal compared with the use of an electric toothbrush without the app; in addition, the compliance and consideration of user-friendly were evaluated	No relevant difference in plaque score was found between the test and control groups (*P*=.39). However, PI improved by 8.5% (*P*=.10) in the intervention group compared with 4.7% (*P*=.56) in the control group.
Alqarni et al [27], 2018	Saudi Arabia	Longitudinal study	120	Parents of infants to adolescents aged 15 y	Develop a mobile app (Your child's smile) and evaluate its efficacy in improving the dental health knowledge of parents	After the 15-day follow-up, most responders showed highly significant (*P*<.01) or significant (*P*<.05) improvement in their knowledge on tooth development (8.33%-40%), importance of deciduous teeth (25%-33%), importance of regular dental checkups (20%-34%), pit and fissure sealants (24%-32%), and consequences of early loss of deciduous teeth (19%-37%). A total of 75% of parents favored the use of mobile apps as an effective child dental health knowledge tool.

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Zahid et al [64], 2020	Saudi Arabia	Nonrandomized quasi-experimental study	271 recruited; 234 completed	Mean age 16.6 (SD 0.96) y	Compare the impact of a mobile app (Brush DJ) and an educational lecture on knowledge and behavior regarding oral health	After the 3-month follow-up, both the mobile app and educational lecture groups showed significant improvements on knowledge and behavior regarding oral health except for the frequency and duration of toothbrushing in the app group.
Krishnan et al [65], 2021	India	Nonrandomized controlled trial	60	13-17 y	Evaluate the effectiveness of visual cards and a mobile-based app (Brush Up) on oral health education among adolescents with ASD[p]	At 6 and 12 weeks after the intervention, no statistically significant difference in PI (1.023 and 0.812; *P*=.91) and GI scores (0.264 and 0.283; *P*=.93) between visual pedagogy (group A) and the Brush Up mobile app (group B) was observed. For group A, mean plaque scores were 2.02 (SD 0.06) at baseline, 1.00 (SD 0.00) at 6 weeks, and 0.45 (SD 0.13) at the 12-week follow-up. Mean GI scores were 1.05 (SD 0.13) at baseline, 0.61 (SD 0.14) at 6 weeks, and 0.28 (0.10) at the 12-week follow-up. For group B, mean plaque scores were 2.01 (SD 0.06) at baseline, 1.00 (SD 0.00) at 6 weeks, and 0.46 (SD 0.11) at the 12-week follow-up. Mean GI scores were 1.03 (SD 0.17) at baseline, 0.58 (SD 0.16) at 6 weeks, and 0.24 (SD 0.11) at the 12-week follow-up.
Setijanto et al [28], 2021	Indonesia	Longitudinal study	47	17-45 y	Evaluate the effectiveness of a mobile app to increase knowledge about oral health among pregnant women	There was a significant (*P*<.001) improvement in the knowledge about oral health at posttest measurement (87%) compared with at pretest measurement (56%).
Fernández et al [48], 2021	Chile	Systematic review and meta-analysis	N/A	N/A	Determine the effect of teledentistry on oral health promotion and prevention as compared with other conventional strategies	Teledentistry was found effective—mostly mHealth[q] (messages and apps)—when compared with conventional strategies. SMD for PI was −1.18 (95% CI −1.54 to −0.82; I^2=92%; low certainty). SMD for GI was −2.17 (95% CI −3.15 to −1.19; I^2=97%; moderate certainty). Risk ratio for WSLs was 0.48 (95% CI 0.35-0.66; I^2=0%; moderate certainty).
Bohn et al [29], 2018	United States	Cross-sectional study	25	22-89 y	To assess preferences and perceptions regarding the use of apps in dental care	Participants believed that apps should be used in conjunction with a dentist's explanation about a procedure. Participants felt that the apps would be more beneficial if they could be customized to individual dental needs. Participants favored esthetic images of teeth that did not show structural anatomy. Participants preferred the internet-based apps.
Rahaei et al [66], 2022	Iran	Randomized controlled trial	158	10-12 y	Compare the effectiveness of an educational mobile app (My Tooth) with conventional oral health education among elementary school students	Before the intervention, the mean scores of behavior were 13.69 (SD 3.89; intervention group) and 13.93 (SD 3.02; control group). After the intervention, mean scores of behavior increased significantly in the intervention group (16.02, SD 3.48; *P*<.001).
Nayak et al [30], 2018	India	Randomized controlled trial	159 recruited; 150 completed	18-24 y	Evaluate the feasibility and effectiveness of a mobile app (WhatsApp) to improve knowledge about oral cancer among college students	After the 1-month intervention, a statistically significant increase in KS was observed in both groups, with highly significant improvement in the intervention group (mean KS was 28.72, SD 5.9 at baseline and 49.56, SD 5.4 after the intervention; *P*<.001).

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Zolfaghari et al [67], 2021	Iran	Randomized controlled trial	58	Children aged 1-6 y and their parents	Design a gamified mobile app and evaluate its effectiveness on educating mothers about their children's oral health	The mean KS of mothers in the pretest was 10.5 (SD 2.1) in the simple app group and 11.3 (SD 11.9) in the gamified app group, which changed to 13.1 (SD 1.6) and 14.3 (SD 2.0), respectively, in the posttest (*P*<.001). The mean practice score of mothers in the pretest was 4.4 (SD 2.4) in the simple app group and 4.8 (SD 3.2) in the gamified app group, which changed to 8.5 (SD 1.7) and 8.0 (SD 2.2), respectively, in the posttest (*P*<.001). The mean dental PI of children in the pretest was 0.8 (SD 0.4) in the simple app group and 1.0 (SD 0.3) in the gamified app group, which changed to 0.5 (SD 0.3) and 0.5 (SD 0.3), respectively, in the posttest. Children had better plaque control in the gamified app group (*P*<.001).
Panchal et al [68], 2017	India	Longitudinal study	150 recruited; 132 completed	Children aged 2-6 y and their parents	Study the efficiency of a mobile app (Cariometer) in monitoring diet and oral hygiene habits	There was a significant improvement in the dietary pattern followed by the patients at day 7 as compared with day 1 (mean dietary score was 1.51, SD 0.24 at day 1 and 0.07, SD 0.14 at day 7; *P*<.001). Approximately 90% of children brushed twice a day at day 7 after the use of the Cariometer app. There was an increase in the frequency of rinsing after meals at day 7 as compared with day 1 after the use of the Cariometer app.
Lotto et al [31], 2020	Brazil	Randomized controlled trial	104	Children aged 36-60 months and their parents	Assess the efficiency of educational messages via mobile app (WhatsApp) to control early childhood caries	Proportion of participants with the increment of maximum ICDAS[r] did not increase significantly in the intervention group (15.4%-23.1%; *P*=.13), differently from that observed in the control group (21.2%-36.5%; *P*=.008) between the 3- and 6-month follow-ups. eHEALS[s] scores increased significantly in the intervention group (+10.32%; *P*=.001) in contrast to a nonsignificant decrease observed in the control group (−2.65%; *P*=.38).
Wang et al [32], 2020	Taiwan	Nonrandomized controlled trial	120 recruited; 100 completed	48-66 y	Evaluate an educational mobile app regarding changes in the care needs and quality of life of patients with oral cancer	The overall improvement in quality of life was higher in the experimental group than in the control group (−7.24 vs −4.36; *P*=.22). The physiological care needs decreased in the experimental group compared with the control group (experimental group: 26.33 and control group: 21.33 before the intervention; experimental group: 20.67 and control group: 20.25 after the intervention; *P*<.02).
Lozoya et al [69], 2019	United States	Nonrandomized quasi-experimental study	33 recruited; 26 completed	Children (mean age 3.48, SD 0.93 y) and their parents	Evaluate the effect of a smartphone app on the OHBs of the parents of preschoolers	Parents' behavioral intentions or OHBs with their children did not significantly change from before to after the intervention (*P*>.05). SNs[t] and PBC[u] predicted behavioral intentions before the intervention and behavior change after the intervention. Thematic analysis revealed that parents' belief in the importance of establishing oral health habits and brushing reminders and videos delivered via a mobile app supported efforts to form oral health habits.
Jacobson et al [70], 2019	United States	Quasi-experimental study	34	Children aged 5-6 y and their parents	Evaluate a mobile app (Brush Up) to improve toothbrushing behaviors among children	After 7 days, toothbrushing duration increased significantly (*P*<.001). Mean time (in seconds) consumed toothbrushing was 46.2 (SD 31.3) at baseline and 69.4 (SD 30.4) after 7 days (*P*<.001). After 2 weeks (n=15), mean time (in seconds) consumed toothbrushing was 39.9 (SD 21.0) at baseline and 108.9 (SD 47.6) after 14 days (*P*<.001).
Tobias and Spanier [44], 2020	Israel	Longitudinal study	44	≥18 y	Classification of gum health based on the MGI[v] score using dental selfies via a mobile app (iGAM[w]).	The mobile app produced accurate classification of gum health based on the MGI. Area under the curve ranged between 1.0 and 0.84.

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Al-Moghrabi et al [35], 2020	United Kingdom	Randomized controlled trial	84 recruited; 64 completed	12-21 y	Assess the effectiveness of a mobile app (My Retainers) on objectively assessed TPR[w] wear time, stability, periodontal outcomes, patient experiences, and knowledge related to retainers	Use of the mobile app resulted in slightly higher median wear time (0.91 h/d, 95% CI −4.01 to 2.19; *P*=.56). No significant differences were found in terms of stability (β=.002, 95% CI −.03 to .04; *P*=.92), plaque levels (β=−.02, 95% CI −.07 to .03; *P*=.44), bleeding on probing (β=−.01, 95% CI −.05 to .03; *P*=.61), and probing depth (β=−.01, 95% CI −.09 to .07; *P*=.79). Similar levels of patient experiences (*P*=.94) and knowledge related to retainers (*P*=.26) were found.
Moylan et al [36], 2019	United States	Cross-sectional study	12	10-17 y	Study the reliability and accuracy of mobile app monitoring of tooth movement in patients with orthodontic appliances	The intercanine and intermolar measurement differences between intraoral video scans using the monitoring software's smartphone app (Dental Monitoring) and plaster models were on average 0.17 mm (90% CI 0.00-0.34) and −0.02 mm (90% CI −0.26 to 0.29), respectively.
Li et al [37], 2016	China	Randomized controlled trial	343 recruited; 224 completed	Mean age 17.6 (SD 5.7) y	Evaluate the effectiveness of a messaging mobile app (WeChat) in improving patients' compliance and decreasing the length of orthodontic treatment	Duration of orthodontic treatment in the WeChat group was shorter than that in the compared group (median 80.5, range 66-93 weeks vs median 84.5, range 75-103 weeks; *P*=.007). There was less failed attendance (3.1% vs 10.9%; *P*<.001), late attendance (20.1% vs 29.9%; *P*<.001), and bracket bond failure (11.8% vs 16.1%; *P*<.001) in the WeChat group than in the control group. There was no difference in orthodontic PI or MGI between the 2 groups before and after treatment.
Hannequin et al [38], 2020	France	Case report	N/A	21 y	Use dental monitoring software to manage aligner-mediated tooth movement on a woman aged 21 years treated with corticotomy-accelerated presurgical decompensation with Invisalign clear aligners	The software allowed for fewer chairside appointments and remote monitoring and allowed for early detection and correction of an error on the aligner. This case was managed in 6 months instead of 10-15 months for the presurgical decompensation phase and 3-4 months after surgery in conventional orthodontic treatments.
Henzell et al [39], 2013	New Zealand	Cross-sectional study	130	≥10 y	Investigate the use of internet-based social media sites by patients with orthodontic appliances and whether a web-based application or mobile app would be considered helpful in improving cooperation in orthodontic treatment	Internet-based social media sites were used by 80.8% of patients, with Facebook being the most popular. Approximately 13.3% of the sample had posted comments about braces on these social media sites, and only 6.7% had considered obtaining information about orthodontic treatment from internet-based social media sites, with most (81%) preferring to seek this information directly from their orthodontist. Nearly two-thirds of those who had difficulty remembering to wear their orthodontic appliances reported that a reminder app on their phone would be beneficial.
Al-Musawi et al [33], 2017	Kuwait	Longitudinal study	87	Not reported	Evaluate the effectiveness of a mobile app (Dental Trauma App) in delivering information to schoolteachers about the optimal emergency management of traumatic dental injuries and compare it with the traditional lecture-based method	Participants using the app only had a significantly higher mean score (12.72, SE 0.47) than participants receiving the lecture only (mean 11.20, SE 0.44; *P*=.02) and participants in the lecture and app group (mean 9.87, SE 0.50; *P*<.001).
Iskander et al [34], 2016	United States	Cross-sectional study	89	Not reported	Compare effectiveness and user preference of a mobile app and conventional poster for first aid to dental trauma	Individuals using the mobile app were more likely to select "put the tooth back in place" (71.1%) compared with those using the poster, who chose "put the tooth in milk" (56.8%; *P*=.004). Less educated individuals were willing to pay more for the app (*P*=.02) and were more likely to report being interested in receiving dental information through mobile technology in the future (*P*=.006). Most of the respondents preferred the mobile app.

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Abbasi et al [40], 2021	Pakistan	Randomized controlled trial	160	Children aged 4-10 y	Evaluate the effectiveness of a mobile app (Little Lovely Dentist), a dental song, and TSD[x] techniques in pediatric patients	A statistically significant difference in mean heart rate of participants was observed in the Little Lovely Dentist, dental song, and control groups, whereas no difference was observed in the TSD group. For the mobile "Little Lovely Dentist" app group, mean heart rate before and after the intervention was 107.9 (SD 8.2) and 104.9 (SD 6.8), respectively (*P*=.002). For the YouTube "dental video song" group, mean heart rate before and after the intervention was 106.6 (SD 6.1) and 104.0 (SD 7.6), respectively (*P*=.001). For the TSD group, mean heart rate before and after the intervention was 101.4 (SD 15.6) and 108.2 (SD 7.5), respectively (*P*=.68). For the control group, mean heart rate before and after the intervention was 102.8 (SD 5.3) and 107.5 (SD 5.9), respectively (*P*=.68). FIS[y] scores decreased significantly in participants in the Little Lovely Dentist and dental song groups, whereas the scores increased in the TSD and control groups. For the mobile "Little Lovely Dentist" app group, mean FIS scores before and after the intervention were 2.80 (SD 1.06) and 2.52 (SD 0.87), respectively (*P*=.03). For the YouTube "dental video song" group, mean FIS scores before and after the intervention were 1.80 (SD 0.96) and 2.47 (SD 0.84), respectively (*P*=.04). For the TSD group, mean FIS scores before and after the intervention were 2.60 (SD 0.74) and 3.30 (SD 0.96), respectively (*P*=.001). For the control group, mean FIS scores before and after the intervention were 2.91 (SD 0.92) and 3.52 (SD 1.24), respectively (*P*=.01).
Elicherla et al [41], 2019	India	Randomized controlled trial	50	7-11 y	Evaluate the effectiveness of a mobile app (Little Lovely Dentist) compared with the TSD technique in managing anxiety and fear in children during their first dental visit	A statistically significant difference in mean heart rate of the participants was observed in the Little Lovely Dentist group, whereas no difference was observed in the TSD group. For the mobile "Little Lovely Dentist" app group, mean heart rate before and after the intervention was 108.2 (SD 12.8) and 97.4 (SD 12.3), respectively (*P*<.001). For the TSD group, mean heart rate before and after the intervention was 95.9 (SD 10.0) and 97.2 (SD 12.3), respectively (*P*=.32). RMS pictorial scores decreased significantly in participants in both the Little Lovely Dentist and TSD groups. For the mobile "Little Lovely Dentist" app group, mean RMS pictorial scores before and after the intervention were 3.20 (SD 1.04) and 1.32 (SD 0.5), respectively (*P*<.001). For the TSD group, mean RMS pictorial scores before and after the intervention were 2.6 (SD 0.8) and 1.5 (SD 0.6), respectively (*P*<.001).
Kevadia et al [42], 2020	India	Randomized controlled trial	75	6-9 y	To evaluate the effectiveness of 3 different behavioral management techniques: a mobile app (My Little Dentist), Tell, Play, Do, and the film modeling technique	All the index scores were significantly lower among children who received the Tell, Play, Do (group II) intervention than in those who received the film modeling intervention (group I) and the mobile dental app (group III). Average heart rate was *P*=.03 for group II vs group I and *P*=.046 for group II vs group III. FIS was *P*=.03 for group II vs group I and *P*=.03 for group II vs group III. Venham pictorial index scores were *P*=.04 for group II vs group I and *P*=.045 for group II vs group III.
Zink et al [43], 2018	Brazil	Randomized controlled trial	40	9-15 y	Development and evaluation of a mobile app for patient-professional communication during dental visits of patients with ASD	The decayed, missing, and filled primary and permanent teeth index was similar for both groups (*P*=.60), being 1.5 (SD 3.0) for the app group and 0.7 (SD 1.3) for the Picture Exchange Communication System group. There were statistically significant differences in the number of attempts required for the pictures to acquire each skill proposed (room, ground, dentist, and 3-in-1 air and water syringe; *P*<.05) between the 2 groups, which were lower for the app group.

Study, year	Country of origin	Study design	Sample size	Age	Purpose	Key findings
Cunningham et al [49], 2021	United Kingdom	Systematic review	N/A	N/A	To examine and identify studies that apply virtual reality or bespoke smartphone apps in dentistry to decrease patient anxiety and to study the effectiveness of these apps	In total, 3 studies using virtual reality in a dental setting demonstrated decreased pain and anxiety compared with no intervention. A fourth study used a bespoke dental app and imagery to prepare patients with ASD for dental treatment, finding statistically significant decreases in both the number of appointments and number of attempts required to carry out a procedure.
Lin et al [45], 2014	Taiwan	Cross-sectional study	26 dentists and 32 patients	Not reported	To improve dental care services using a mobile app (Dental Calendar) combined with cloud service	Results of assessments through interviews and questionnaires indicated a significant increase (*P*<.05) in both dentists' and patients' overall experiences. Total mean increment in after and before test for dentists was 5.875 (95% CI 1.968-9.782; *P*=.009). Total mean increment in after-before test for patients was 18.500 (95% CI 13.625-23.375; *P*<.001). Patients' ability to reschedule appointments for sudden worse prostheses was 0.385 (95% CI 0.081-0.689; *P*=.02). Appointment reminding was 0.844 (95% CI 0.242-1.445; *P*=.007), appointment rescheduling for sudden worse prostheses was 0.781 (95% CI 0.204-1.359; *P*=.01), and dentist-patient relationship was 0.500 (95% CI 0.007-0.993; *P*=.047)

[a]OSCA: oral self-care mobile app.
[b]OHS: oral hygiene status.
[c]OHB: oral hygiene behavior.
[d]PCR: plaque control record.
[e]PI: Plaque Index.
[f]GI: Gingival Index.
[g]WSL: white spot lesion.
[h]KS: knowledge score.
[i]OHI-S: Simplified Oral Hygiene Index.
[j]GBI: Gingival Bleeding Index.
[k]WS: white spot.
[l]N/A: not applicable.
[m]SMD: standardized mean difference.
[n]OR: odds ratio.
[o]VPI: Visible Plaque Index.
[p]ASD: autism spectrum disorder.
[q]mHealth: mobile health.
[r]ICDAS: International Caries Detection and Assessment System.
[s]eHEALS: eHealth Literacy Scale.
[t]SN: social norm.
[u]PBC: perceived behavioral control.
[v]MGI: Modified Gingival Index.
[w]TPR: thermoplastic retainer.
[x]TSD: Tell, Show, Do.
[y]FIS: Facial Image Scale.

Discussion

Principal Findings

The aim of this study was to present the main findings in the literature on the use of mobile phone apps in oral health and evaluate the evidence available in the literature. This scoping review identified and reviewed 45 papers published between January 2000 and June 2021. Half (23/45, 51%) of the included studies were from Asian countries and focused on children and adolescents. In addition, most of the included studies (31/45, 69%) focused on oral health promotion using mobile phone apps, followed by behavior management. There was limited evidence of diagnostic and remote consultations using mobile phone apps.

A total of 47% (21/45) of the publications focused on improving oral hygiene through effective toothbrushing. Dental plaque is a biological component in the initiation of dental caries and periodontal diseases [71]. Therefore, it is important to remove dental plaque from tooth surfaces through daily toothbrushing with fluoridated toothpaste to prevent tooth decay and periodontal diseases [71]. This scoping review found that several of the RCT studies (9/23, 39%) reported a significant reduction in dental plaque among mobile phone app users compared with their controls. Hence, mobile phone apps can be used for oral health education and to teach patients good home care skills. A recent systematic review and meta-analysis showed good results in decreasing plaque, thereby improving gingival health and preventing the development of dental caries when both education and skills are incorporated [72].

In total, 11% (5/45) of the publications focused on behavior management, and in these publications, mobile apps were used to reduce the dental anxiety of patients. Dental anxiety usually develops in childhood, and its prevalence ranges from 5% to 20% among children [73,74]. Patients with dental anxiety can easily avoid or miss dental appointments and, therefore, might have poorer dental health [75]. New situations in dental clinics are unfamiliar to children. Mobile apps used beforehand are an effective tool to familiarize a child with dental appointments and instruments and help improve cooperation between a dentist and a child during dental treatments [43,49].

More than half (23/45, 51%) of the included studies were from Asia, especially from India. The population of Asia was 4.7 billion in 2022, and the 2 most populous countries in Asia are China (1.37 billion) and India (1.29 billion) [76]. Thus, the rate of mobile phone use is higher in Asia than in other continents. For the same reason, the production and use of mobile phone apps are also higher in Asia than in other continents. In addition, none of the included studies were published in Africa. This could be due to various challenges, such as infrastructure, scientific technologies, and resources in implementing mHealth interventions in Africa, as described by Kruse et al [77] in their systematic review. Regarding the European region, the General Data Protection Regulation law drafted and implemented by the European Union on May 25, 2018, may have affected the conduct of research that requires the processing of sensitive data (eg, oral health data) despite protecting patients' privacy [78]. Cagnazzo [79] also argued that the overlap of the General Data Protection Regulation and the European Clinical Trials Regulation is creating a state of confusion among the scientific community. Furthermore, the author has argued about the bureaucratic issues related to follow-up clinical studies [79]. However, these aspects need further research.

This review also found that most of the included studies (25/45, 56%) focused on children (aged <10 years) and their families as well as on adolescents aged 10 to 19 years. The studies in this review did not include older adults as a focus group, which may be explained by the fact that the younger population has grown up with smartphones and is more familiar with using them. Using smartphones is part of adolescents' daily routine, and smartphones tend to move around with them throughout the day. In addition, smartphones are highly valued by adolescents; they are easy to use and easily switched on [15]. Thus, approaching children and adolescents via smartphone is more convenient. In addition, the use of smartphones gives people the possibility to access information anywhere and anytime, and this increases patients' autonomy and comfort [80].

Despite excluding studies focused on the development of mobile apps in dentistry, half (23/45, 51%) of the included studies were published between 2020 and 2021. This implies that the field is constantly evolving and growing. Furthermore, the COVID-19 pandemic started at the end of 2019 and changed dental care in many aspects. The use of teledentistry increased during the pandemic, and dentists with a clinical practice considered teledentistry to be the most effective way to reschedule patients' appointment times and provide dental hygiene education and emergency advice [81,82]. In this scoping review, we found only 1 study [45] that focused on remote health care services. However, more studies on teledentistry and the use of artificial intelligence for diagnosis and remote consultations are necessary. In addition, topics related to the uses and importance of teledentistry and artificial intelligence must be added to the curriculum of undergraduate and postgraduate dental studies [83].

Limitations

This study has some limitations. First, only studies written in English were included. Second, the search was carried out in only 2 databases (PubMed and Scopus), therefore leaving out literature from other databases and gray literature. However, a recent meta-research study suggests that searching at least 2 databases is sufficient to increase coverage and decrease the risk of missing eligible studies [84]. In addition, an information specialist was consulted for searching the databases. Another limitation would be not undergoing any peer-review for electronic search strategies.

The field of mobile apps is still new and under development. New research must be conducted along with the development of new patient-oriented apps.

Conclusions

The use of mobile apps in oral health is increasing among patients, mainly children and adolescents. In the literature, there are many studies related to mobile apps that are focused on promoting oral health. Other areas such as diagnosis and remote consultations are neglected in the current studies. There are potential uses for improving oral hygiene, knowledge, and behavior via mobile apps, but more studies are required.

Acknowledgments

The authors would like to acknowledge Kirsi Vähäkangas, informatician, Oulu University Library, for her help. The European Regional Development Fund (project: *digileap of oral health toward virtual reception*; A76934) is gratefully acknowledged for supporting this work. AK-S and SK received a research grant from the Finnish Dental Society Apollonia to conduct this research.

Conflicts of Interest

None declared.

References

1. Isola G. The impact of diet, nutrition and nutraceuticals on oral and periodontal health. Nutrients 2020 Sep 06;12(9):2724 [FREE Full text] [doi: 10.3390/nu12092724] [Medline: 32899964]
2. Righolt AJ, Jevdjevic M, Marcenes W, Listl S. Global-, regional-, and country-level economic impacts of dental diseases in 2015. J Dent Res 2018 May;97(5):501-507. [doi: 10.1177/0022034517750572] [Medline: 29342371]
3. Ruff RR, Senthi S, Susser SR, Tsutsui A. Oral health, academic performance, and school absenteeism in children and adolescents: a systematic review and meta-analysis. J Am Dent Assoc 2019 Feb;150(2):111-21.e4. [doi: 10.1016/j.adaj.2018.09.023] [Medline: 30473200]
4. Bouchard P, Carra MC, Boillot A, Mora F, Rangé H. Risk factors in periodontology: a conceptual framework. J Clin Periodontol 2017 Feb;44(2):125-131. [doi: 10.1111/jcpe.12650] [Medline: 27862138]
5. Selwitz RH, Ismail AI, Pitts NB. Dental caries. Lancet 2007 Jan 06;369(9555):51-59. [doi: 10.1016/S0140-6736(07)60031-2] [Medline: 17208642]
6. Sheiham A, Watt RG. The common risk factor approach: a rational basis for promoting oral health. Community Dent Oral Epidemiol 2000 Dec;28(6):399-406. [doi: 10.1034/j.1600-0528.2000.028006399.x] [Medline: 11106011]
7. Delli Bovi AP, Di Michele L, Laino G, Vajro P. Obesity and obesity related diseases, sugar consumption and bad oral health: a fatal epidemic mixtures: the pediatric and odontologist point of view. Transl Med UniSa 2017 Jul 01;16:11-16 [FREE Full text] [Medline: 28775964]
8. Fiorillo L. Oral health: the first step to well-being. Medicina (Kaunas) 2019 Oct 07;55(10):676 [FREE Full text] [doi: 10.3390/medicina55100676] [Medline: 31591341]
9. Dörfer C, Benz C, Aida J, Campard G. The relationship of oral health with general health and NCDs: a brief review. Int Dent J 2017 Sep;67 Suppl 2(Suppl 2):14-18 [FREE Full text] [doi: 10.1111/idj.12360] [Medline: 29023744]
10. Kassebaum NJ, Smith AG, Bernabé E, Fleming TD, Reynolds AE, Vos T, GBD 2015 Oral Health Collaborators. Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990-2015: a systematic analysis for the global burden of diseases, injuries, and risk factors. J Dent Res 2017 Apr;96(4):380-387 [FREE Full text] [doi: 10.1177/0022034517693566] [Medline: 28792274]
11. Innes NP, Chu CH, Fontana M, Lo EC, Thomson WM, Uribe S, et al. A century of change towards prevention and minimal intervention in cariology. J Dent Res 2019 Jun;98(6):611-617. [doi: 10.1177/0022034519837252] [Medline: 31107140]
12. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
13. Dusseldorp E, van Genugten L, van Buuren S, Verheijden MW, van Empelen P. Combinations of techniques that effectively change health behavior: evidence from Meta-CART analysis. Health Psychol 2014 Dec;33(12):1530-1540. [doi: 10.1037/hea0000018] [Medline: 24274802]
14. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
15. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
16. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020 Feb;11:2150132720980612 [FREE Full text] [doi: 10.1177/2150132720980612] [Medline: 33300414]
17. Sheikhtaheri A, Kermani F. Telemedicine in diagnosis, treatment and management of diseases in children. Stud Health Technol Inform 2018;248:148-155. [Medline: 29726431]
18. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
19. Chu K, Matheny SJ, Escobar-Viera CG, Wessel C, Notier AE, Davis EM. Smartphone health apps for tobacco Cessation: a systematic review. Addict Behav 2021 Jan;112:106616 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106616] [Medline: 32932102]
20. Tiffany B, Blasi P, Catz SL, McClure JB. Mobile apps for oral health promotion: content review and heuristic usability analysis. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e11432 [FREE Full text] [doi: 10.2196/11432] [Medline: 30181114]
21. Parker K, Bharmal RV, Sharif MO. The availability and characteristics of patient-focused oral hygiene apps. Br Dent J 2019 Apr 26;226(8):600-604. [doi: 10.1038/s41415-019-0197-7] [Medline: 31028329]
22. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]

23. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Munn Z, editors. Joanna Briggs Institute Reviewer's Manual. Miami, FL: JBI; 2020.
24. The use of patient-orientated mobile phone applications in oral health: a scoping review. Open Science Framework. 2021 Jun 17. URL: https://osf.io/9jefu [accessed 2023-08-17]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Marchetti G, Fraiz FC, Nascimento WM, Soares GM, Assunção LR. Improving adolescents' periodontal health: evaluation of a mobile oral health app associated with conventional educational methods: a cluster randomized trial. Int J Paediatr Dent 2018 Jul;28(4):410-419. [doi: 10.1111/ipd.12371] [Medline: 29756308]
27. Alqarni AA, Alfaifi HM, Aseeri NA, Gadah T, Togoo RA. Efficacy of a self-designed mobile application to improve child dental health knowledge among parents. J Int Soc Prev Community Dent 2018 Sep;8(5):424-430 [FREE Full text] [doi: 10.4103/jispcd.JISPCD_195_18] [Medline: 30430070]
28. Setijanto RD, Bramantoro T, Ramadhani A, Setyaji AM, Rusyidina Z. Analysis of oral health knowledge improvement of pregnant mothers using oral health monitoring mobile application. J Int Oral Health 2021;13(2):169-174 [FREE Full text] [doi: 10.4103/jioh.jioh_234_18]
29. Bohn CE, McQuistan MR, McKernan SC, Askelson NM. Preferences related to the use of mobile apps as dental patient educational aids: a pilot study. J Prosthodont 2018 Apr;27(4):329-334. [doi: 10.1111/jopr.12667] [Medline: 28872732]
30. Nayak PP, Nayak SS, Sathiyabalan D, Aditya NK, Das P. Assessing the feasibility and effectiveness of an app in improving knowledge on oral cancer-an interventional study. J Cancer Educ 2018 Dec;33(6):1250-1254. [doi: 10.1007/s13187-017-1239-y] [Medline: 28612324]
31. Lotto M, Strieder AP, Ayala Aguirre PE, Oliveira TM, Andrade Moreira Machado MA, Rios D, et al. Parental-oriented educational mobile messages to aid in the control of early childhood caries in low socioeconomic children: a randomized controlled trial. J Dent 2020 Oct;101:103456 [FREE Full text] [doi: 10.1016/j.jdent.2020.103456] [Medline: 32827598]
32. Wang TF, Huang RC, Yang SC, Chou C, Chen LC. Evaluating the effects of a mobile health app on reducing patient care needs and improving quality of life after oral cancer surgery: Quasiexperimental study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18132 [FREE Full text] [doi: 10.2196/18132] [Medline: 32716303]
33. Al-Musawi A, Al-Sane M, Andersson L. Smartphone app as an aid in the emergency management of avulsed teeth. Dent Traumatol 2017 Feb;33(1):13-18. [doi: 10.1111/edt.12298] [Medline: 27381488]
34. Iskander M, Lou J, Wells M, Scarbecz M. A poster and a mobile healthcare application as information tools for dental trauma management. Dent Traumatol 2016 Dec;32(6):457-463. [doi: 10.1111/edt.12278] [Medline: 27140068]
35. Al-Moghrabi D, Pandis N, McLaughlin K, Johal A, Donos N, Fleming PS. Evaluation of the effectiveness of a tailored mobile application in increasing the duration of wear of thermoplastic retainers: a randomized controlled trial. Eur J Orthod 2020 Nov 03;42(5):571-579 [FREE Full text] [doi: 10.1093/ejo/cjz088] [Medline: 31799628]
36. Moylan HB, Carrico CK, Lindauer SJ, Tüfekçi E. Accuracy of a smartphone-based orthodontic treatment-monitoring application: a pilot study. Angle Orthod 2019 Sep;89(5):727-733 [FREE Full text] [doi: 10.2319/100218-710.1] [Medline: 30888840]
37. Li X, Xu ZR, Tang N, Ye C, Zhu XL, Zhou T, et al. Effect of intervention using a messaging app on compliance and duration of treatment in orthodontic patients. Clin Oral Investig 2016 Nov;20(8):1849-1859. [doi: 10.1007/s00784-015-1662-6] [Medline: 26631059]
38. Hannequin R, Ouadi E, Racy E, Moreau N. Clinical follow-up of corticotomy-accelerated Invisalign orthodontic treatment with dental monitoring. Am J Orthod Dentofacial Orthop 2020 Dec;158(6):878-888. [doi: 10.1016/j.ajodo.2019.06.025] [Medline: 33129633]
39. Henzell M, Knight A, Antoun JS, Farella M. Social media use by orthodontic patients. N Z Dent J 2013 Dec;109(4):130-133. [Medline: 24396951]
40. Abbasi H, Saqib M, Jouhar R, Lal A, Ahmed N, Ahmed MA, et al. The efficacy of little lovely dentist, dental song, and tell-show-do techniques in alleviating dental anxiety in paediatric patients: a clinical trial. Biomed Res Int 2021 May 23;2021:1119710 [FREE Full text] [doi: 10.1155/2021/1119710] [Medline: 34124238]
41. Elicherla SR, Bandi S, Nuvvula S, Challa RS, Saikiran KV, Priyanka VJ. Comparative evaluation of the effectiveness of a mobile app (Little Lovely Dentist) and the tell-show-do technique in the management of dental anxiety and fear: a randomized controlled trial. J Dent Anesth Pain Med 2019 Dec;19(6):369-378 [FREE Full text] [doi: 10.17245/jdapm.2019.19.6.369] [Medline: 31942452]
42. Kevadia MV, Sandhyarani B, Patil AT, Gunda SA. Comparative evaluation of effectiveness of tell-play-do, film modeling and use of smartphone dental application in the management of child behavior. Int J Clin Pediatr Dent 2020 Nov;13(6):682-687 [FREE Full text] [doi: 10.5005/jp-journals-10005-1857] [Medline: 33976496]
43. Zink AG, Molina EC, Diniz MB, Santos MT, Guaré RO. Communication application for use during the first dental visit for children and adolescents with autism spectrum disorders. Pediatr Dent 2018 Jan 01;40(1):18-22. [Medline: 29482677]
44. Tobias G, Spanier AB. Modified Gingival Index (MGI) classification using dental selfies. Appl Sci 2020;10(24):8923 [FREE Full text] [doi: 10.3390/app10248923]

45. Lin CY, Peng KL, Chen J, Tsai JY, Tseng YC, Yang JR, et al. Improvements in dental care using a new mobile app with cloud services. J Formos Med Assoc 2014 Oct;113(10):742-749 [FREE Full text] [doi: 10.1016/j.jfma.2014.02.009] [Medline: 24796822]
46. Patil S, Hedad IA, Jafer AA, Abutaleb GK, Arishi TM, Arishi SA, et al. Effectiveness of mobile phone applications in improving oral hygiene care and outcomes in orthodontic patients. J Oral Biol Craniofac Res 2021 Jan;11(1):26-32 [FREE Full text] [doi: 10.1016/j.jobcr.2020.11.004] [Medline: 33344158]
47. Toniazzo MP, Nodari D, Muniz FW, Weidlich P. Effect of mHealth in improving oral hygiene: a systematic review with meta-analysis. J Clin Periodontol 2019 Mar 19;46(3):297-309. [doi: 10.1111/jcpe.13083] [Medline: 30761580]
48. Fernández CE, Maturana CA, Coloma SI, Carrasco-Labra A, Giacaman RA. Teledentistry and mHealth for promotion and prevention of oral health: a systematic review and meta-analysis. J Dent Res 2021 Aug;100(9):914-927. [doi: 10.1177/00220345211003828] [Medline: 33769123]
49. Cunningham A, McPolin O, Fallis R, Coyle C, Best P, McKenna G. A systematic review of the use of virtual reality or dental smartphone applications as interventions for management of paediatric dental anxiety. BMC Oral Health 2021 May 07;21(1):244 [FREE Full text] [doi: 10.1186/s12903-021-01602-3] [Medline: 33962624]
50. Alkilzy M, Midani R, Höfer M, Splieth C. Improving toothbrushing with a smartphone app: results of a randomized controlled trial. Caries Res 2019;53(6):628-635. [doi: 10.1159/000499868] [Medline: 31132765]
51. Chang WJ, Wang YL, Chang YH, Lo SY. Effectiveness of an app-based mobile intervention for precision oral self-care in patients with periodontitis from initial therapy to re-evaluation. Appl Sci 2021 May 07;11(9):4229 [FREE Full text] [doi: 10.3390/app11094229]
52. Alkadhi OH, Zahid MN, Almanea RS, Althaqeb HK, Alharbi TH, Ajwa NM. The effect of using mobile applications for improving oral hygiene in patients with orthodontic fixed appliances: a randomised controlled trial. J Orthod 2017 Sep;44(3):157-163. [doi: 10.1080/14653125.2017.1346746] [Medline: 28705122]
53. Farhadifard H, Soheilifar S, Farhadian M, Kokabi H, Bakhshaei A. Orthodontic patients' oral hygiene compliance by utilizing a smartphone application (Brush DJ): a randomized clinical trial. BDJ Open 2020 Nov 20;6(1):24 [FREE Full text] [doi: 10.1038/s41405-020-00050-5] [Medline: 33298841]
54. Deleuse M, Meiffren C, Bruwier A, Maes N, Le Gall M, Charavet C. Smartphone application-assisted oral hygiene of orthodontic patients: a multicentre randomized controlled trial in adolescents. Eur J Orthod 2020 Feb 01:2020. [doi: 10.1093/ejo/cjz105] [Medline: 32006440]
55. Desai RV, Badrapur NC, Mittapalli H, Srivastava BK, Eshwar S, Jain V. "Brush up": an innovative technological aid for parents to keep a check of their children's oral hygiene behaviour. Rev Paul Pediatr 2021 Apr 02;39:e2020085 [FREE Full text] [doi: 10.1590/1984-0462/2021/39/2020085] [Medline: 33825795]
56. Kay E, Shou L. A randomised controlled trial of a smartphone application for improving oral hygiene. Br Dent J 2019 Apr;226(7):508-511. [doi: 10.1038/s41415-019-0202-1] [Medline: 30980006]
57. Shida H, Okabayashi S, Yoshioka M, Takase N, Nishiura M, Okazawa Y, et al. Effectiveness of a digital device providing real-time visualized tooth brushing instructions: a randomized controlled trial. PLoS One 2020 Jun 25;15(6):e0235194 [FREE Full text] [doi: 10.1371/journal.pone.0235194] [Medline: 32584893]
58. Zotti F, Dalessandri D, Salgarello S, Piancino M, Bonetti S, Visconti L, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016 Jan;86(1):101-107 [FREE Full text] [doi: 10.2319/010915-19.1] [Medline: 25799001]
59. Scheerman JF, van Meijel B, van Empelen P, Verrips GH, van Loveren C, Twisk JW, et al. The effect of using a mobile application ("WhiteTeeth") on improving oral hygiene: a randomized controlled trial. Int J Dent Hyg 2020 Feb;18(1):73-83 [FREE Full text] [doi: 10.1111/idh.12415] [Medline: 31291683]
60. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]
61. Rasmus K, Toratti A, Karki S, Pesonen P, Laitala ML, Anttonen V. Acceptability of a mobile application in children's oral health promotion-a pilot study. Int J Environ Res Public Health 2021 Mar 22;18(6):3256 [FREE Full text] [doi: 10.3390/ijerph18063256] [Medline: 33809845]
62. Scheerman JF, Hamilton K, Sharif MO, Lindmark U, Pakpour AH. A theory-based intervention delivered by an online social media platform to promote oral health among Iranian adolescents: a cluster randomized controlled trial. Psychol Health 2020 Apr;35(4):449-466. [doi: 10.1080/08870446.2019.1673895] [Medline: 31621423]
63. Humm V, Wiedemeier D, Attin T, Schmidlin P, Gartenmann S. Treatment success and user-friendliness of an electric toothbrush app: a pilot study. Dent J (Basel) 2020 Sep 01;8(3):97 [FREE Full text] [doi: 10.3390/dj8030097] [Medline: 32882808]
64. Zahid T, Alyafi R, Bantan N, Alzahrani R, Elfirt E. Comparison of effectiveness of mobile app versus conventional educational lectures on oral hygiene knowledge and behavior of high school students in Saudi Arabia. Patient Prefer Adherence 2020 Oct 13;14:1901-1909 [FREE Full text] [doi: 10.2147/PPA.S270215] [Medline: 33116434]
65. Krishnan L, Iyer K, Kumar PD. Effectiveness of two sensory-based health education methods on oral hygiene of adolescent with autism spectrum disorders: an interventional study. Spec Care Dentist 2021 Sep;41(5):626-633. [doi: 10.1111/scd.12606] [Medline: 34050975]

66. Rahaei Z, Moradian E, Falahati-Marvast F. Improving dental-oral health learning in students using a mobile application ("My tooth"): a controlled before and after study. Int J Dent Hyg 2022 Aug;20(3):512-518. [doi: 10.1111/idh.12504] [Medline: 33829640]
67. Zolfaghari M, Shirmohammadi M, Shahhosseini H, Mokhtaran M, Mohebbi SZ. Development and evaluation of a gamified smart phone mobile health application for oral health promotion in early childhood: a randomized controlled trial. BMC Oral Health 2021 Jan 07;21(1):18 [FREE Full text] [doi: 10.1186/s12903-020-01374-2] [Medline: 33413304]
68. Panchal V, Gurunathan D, Shanmugaavel AK. Smartphone application as an aid in determination of caries risk and prevention: a pilot study. Eur J Dent 2017 Oct;11(4):469-474 [FREE Full text] [doi: 10.4103/ejd.ejd_190_17] [Medline: 29279672]
69. Lozoya CJ, Giblin-Scanlon L, Boyd LD, Nolen S, Vineyard J. Influence of a smartphone application on the oral health practices and behaviors of parents of preschool children. J Dent Hyg 2019 Oct;93(5):6-14. [Medline: 31628171]
70. Jacobson D, Jacobson J, Leong T, Lourenco S, Mancl L, Chi DL. Evaluating child toothbrushing behavior changes associated with a mobile game app: a single arm pre/post pilot study. Pediatr Dent 2019 Jul 15;41(4):299-303 [FREE Full text] [Medline: 31439090]
71. Jepsen S, Blanco J, Buchalla W, Carvalho JC, Dietrich T, Dörfer C, et al. Prevention and control of dental caries and periodontal diseases at individual and population level: consensus report of group 3 of joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases. J Clin Periodontol 2017 Mar;44 Suppl 18:S85-S93. [doi: 10.1111/jcpe.12687] [Medline: 28266120]
72. Tsai C, Raphael S, Agnew C, McDonald G, Irving M. Health promotion interventions to improve oral health of adolescents: a systematic review and meta-analysis. Community Dent Oral Epidemiol 2020 Dec;48(6):549-560. [doi: 10.1111/cdoe.12567] [Medline: 32767825]
73. Kroniņa L, Rasčevska M, Care R. Psychosocial factors correlated with children's dental anxiety. Stomatologija 2017;19(3):84-90 [FREE Full text] [Medline: 29339671]
74. Seligman LD, Hovey JD, Chacon K, Ollendick TH. Dental anxiety: an understudied problem in youth. Clin Psychol Rev 2017 Jul;55:25-40. [doi: 10.1016/j.cpr.2017.04.004] [Medline: 28478271]
75. Armfield JM, Heaton LJ. Management of fear and anxiety in the dental clinic: a review. Aust Dent J 2013 Dec;58(4):390-531 [FREE Full text] [doi: 10.1111/adj.12118] [Medline: 24320894]
76. Asia population 2022. World Population Review. URL: https://worldpopulationreview.com/continents/asia-population [accessed 2022-09-30]
77. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
78. What is GDPR, the EU's new data protection law? General Data Protection Regulation, European Union. URL: https://gdpr.eu/what-is-gdpr/ [accessed 2022-09-30]
79. Cagnazzo C. The thin border between individual and collective ethics: the downside of GDPR. Lancet Oncol 2021 Nov;22(11):1494-1496. [doi: 10.1016/S1470-2045(21)00526-X] [Medline: 34735806]
80. Reid Chassiakos YL, Radesky J, Christakis D, Moreno MA, Cross C, Council on Communications and Media. Children and adolescents and digital media. Pediatrics 2016 Nov;138(5):e20162593. [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
81. Plaza-Ruíz SP, Barbosa-Liz DM, Agudelo-Suárez AA. Impact of COVID-19 on the knowledge and attitudes of dentists toward teledentistry. JDR Clin Trans Res 2021 Jul;6(3):268-278. [doi: 10.1177/2380084421998632] [Medline: 33632011]
82. Gurgel BC, Borges SB, Borges RE, Calderon PD. COVID-19: perspectives for the management of dental care and education. J Appl Oral Sci 2020 Sep 28;28:e20200358 [FREE Full text] [doi: 10.1590/1678-7757-2020-0358] [Medline: 32997092]
83. Nassani MZ, Al-Maweri SA, AlSheddi A, Alomran A, Aldawsari MN, Aljubarah A, et al. Teledentistry-knowledge, practice, and attitudes of dental practitioners in Saudi Arabia: a nationwide web-based survey. Healthcare (Basel) 2021 Dec 04;9(12):1682 [FREE Full text] [doi: 10.3390/healthcare9121682] [Medline: 34946408]
84. Ewald H, Klerings I, Wagner G, Heise TL, Stratil JM, Lhachimi SK, et al. Searching two or more databases decreased the risk of missing relevant studies: a metaresearch study. J Clin Epidemiol 2022 Sep;149:154-164 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.05.022] [Medline: 35654269]

Abbreviations

MeSH: Medical Subject Headings
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 31.01.23; peer-reviewed by JH Rajendran, X Qu, C Mbotwa; comments to author 24.05.23; revised version received 30.06.23; accepted 11.07.23; published 06.09.23.

Please cite as:
Väyrynen E, Hakola S, Keski-Salmi A, Jämsä H, Vainionpää R, Karki S
The Use of Patient-Oriented Mobile Phone Apps in Oral Health: Scoping Review
JMIR Mhealth Uhealth 2023;11:e46143
URL: https://mhealth.jmir.org/2023/1/e46143
doi:10.2196/46143
PMID:37672331

Review

Digital Health Reimbursement Strategies of 8 European Countries and Israel: Scoping Review and Policy Mapping

Robin van Kessel[1,2], PhD; Divya Srivastava[1], PhD; Ilias Kyriopoulos[1], PhD; Giovanni Monti[1], MBA; David Novillo-Ortiz[3], PhD; Ran Milman[4], MBA; Wojciech Wilhelm Zhang-Czabanowski[5], MSc; Greta Nasi[6], PhD; Ariel Dora Stern[7,8], PhD; George Wharton[1], MSc; Elias Mossialos[1,9], PhD

[1]LSE Health, Department of Health Policy, London School of Economics and Political Science, London, United Kingdom

[2]Department of International Health, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[3]Division of Country Health Policies and Systems, World Health Organisation Regional Office for Europe, Copenhagen, Denmark

[4]Digital Health Division, Israeli Ministry Of Health, Jerusalem, Israel

[5]Institute of Public Health, Jagiellonian University, Krakow, Poland

[6]Department of Social and Political Sciences, Bocconi University, Milan, Italy

[7]Harvard Business School, Harvard University, Boston, MA, United States

[8]Harvard-MIT Center for Regulatory Science, Harvard University, Boston, MA, United States

[9]Institute of Global Health Innovation, Imperial College London, London, United Kingdom

Corresponding Author:
Elias Mossialos, PhD
LSE Health
Department of Health Policy
London School of Economics and Political Science
Houghton Street
London, WC2A 2AE
United Kingdom
Phone: 44 7772 707841
Email: e.a.mossialos@lse.ac.uk

Abstract

Background: The adoption of digital health care within health systems is determined by various factors, including pricing and reimbursement. The reimbursement landscape for digital health in Europe remains underresearched. Although various emergency reimbursement decisions were made during the COVID-19 pandemic to enable health care delivery through videoconferencing and asynchronous care (eg, digital apps), research so far has primarily focused on the policy innovations that facilitated this outside of Europe.

Objective: This study examines the digital health reimbursement strategies in 8 European countries (Belgium, France, Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom) and Israel.

Methods: We mapped available digital health reimbursement strategies using a scoping review and policy mapping framework. We reviewed the literature on the MEDLINE, Embase, Global Health, and Web of Science databases. Supplementary records were identified through Google Scholar and country experts.

Results: Our search strategy yielded a total of 1559 records, of which 40 (2.57%) were ultimately included in this study. As of August 2023, digital health solutions are reimbursable to some extent in all studied countries except Poland, although the mechanism of reimbursement differs significantly across countries. At the time of writing, the pricing of digital health solutions was mostly determined through discussions between national or regional committees and the manufacturers of digital health solutions in the absence of value-based assessment mechanisms. Financing digital health solutions outside traditional reimbursement schemes was possible in all studied countries except Poland and typically occurs via health innovation or digital health–specific funding schemes. European countries have value-based pricing frameworks that range from nonexistent to embryonic.

Conclusions: Studied countries show divergent approaches to the reimbursement of digital health solutions. These differences may complicate the ability of patients to seek cross-country health care in another country, even if a digital health app is available in both countries. Furthermore, the fragmented environment will present challenges for developers of such solutions, as they look to expand their impact across countries and health systems. An increased emphasis on developing a clear conceptualization of digital health, as well as value-based pricing and reimbursement mechanisms, is needed for the sustainable integration of digital health. This study can therein serve as a basis for further, more detailed research as the field of digital health reimbursement evolves.

(JMIR Mhealth Uhealth 2023;11:e49003) doi:10.2196/49003

KEYWORDS

digital health; telehealth; telemedicine; reimbursement; policy; Europe; policy mapping; mapping; pricing; digital health app; application; health care ecosystem; framework; integration

Introduction

Background

The COVID-19 pandemic has significantly accelerated the digital transformation of the health care sector, providing an opportunity to leverage software to prevent, manage, or treat disease [1-6]. Digital health solutions encompass a broad range of technologies that promote, improve, or support health system functioning and the delivery of health care, including electronic health records, telemedicine, mobile health apps, health data analytics, and digital therapeutics [7,8]. These solutions can be used for a range of functions, including web-based consultations with medical professionals, tools and software for remote patient monitoring, real-time updates of algorithms based on patient data, and the delivery of health care interventions. Recent studies have indicated that combining medication with digital health solutions can lead to improved outcomes for a variety of chronic conditions such as type 2 diabetes, cardiovascular diseases, and psychiatric and mental health conditions [9,10].

The adoption of digital health solutions within health systems is subject to various factors, including pricing and reimbursement [8,9,11-13]. Pricing models in health comprise mechanisms such as cost-based pricing (ie, the price of a product is based on the cost of care provided to patients and allowable covered costs), the use of cost-effectiveness thresholds (ie, the price or quality-adjusted life-year is compared with a preset threshold), external price referencing (ie, the price of a product is based on the prices set in other countries), and value-based pricing (ie, the price is based on the value that an intervention adds to the health care process, such as improved health outcomes or reduced costs) [14,15]. The latter approach is recognized as a promising solution to optimize resource allocation and address numerous challenges faced by health systems [11]. National price regulations that incorporate value-based elements, such as the price regulations in Germany, have also been shown to more closely align the prices paid for medical products with their benefit to patients [16]. Value-based pricing also supports evidence-based decision-making in health care procurement by providing a benchmark for what constitutes a high-quality intervention [17]. However, applying a value-based framework to digital health care requires the use of comprehensive frameworks to determine the value of digital interventions and can be challenging because of the continuous improvement of digital health apps through performance data and patient feedback [11,12]. Finally, there is a dearth of data and assessment frameworks in use to evaluate the cost-effectiveness of digital health care [18,19] as well as a scarcity of established pricing models that can be used to streamline the introduction of digital health apps to the health care market [20].

Robust reimbursement mechanisms are vital in facilitating the access and affordability of new health technologies [15], although the reimbursement landscape for digital health in Europe remains underresearched and poorly characterized. Although various emergency reimbursement decisions were made during the COVID-19 pandemic to enable health care delivery through videoconferencing and asynchronous care [21,22], such tools represent only a small set of digital health approaches. Additonally, research so far has primarily focused on the policy innovations that facilitated this outside of Europe. As an example, studies have focused on the Australian government’s recent expansion of Medicare-subsidized telehealth services to facilitate the remote delivery of care and mitigate the risk of virus transmission. Consequently, telehealth services became eligible for reimbursement through the Australian Medicare system and eventually became subject to copayments [22,23]. In the United States, research has highlighted the rapid modification of coverage and payment parity policies by states in response to the COVID-19 pandemic to promote the adoption of telehealth and minimize physical contact, thereby overcoming a significant use barrier [21,24]. In comparison, developments in Europe, particularly in terms of the eligibility of digital health apps (ie, software designed to provide a specific form of therapy with or without the involvement of a health care professional) for reimbursement, remain largely unexplored (a notable exception being Germany, which created the first combined regulation and reimbursement pathway for digital health apps in 2019, which has been described in the literature) [25].

Objective

This study presents information on the reimbursement practices for digital health in 9 countries within the World Health Organization European Region (WHO/Europe): Belgium, France, Germany, Israel, Italy, the Netherlands, Poland, Sweden, and the United Kingdom. These countries were chosen based on the availability of information on the reimbursement of digital health and feedback received from the Data and Digital Health Unit at WHO/Europe [11,26-29]. We aim to map and compare four distinct reimbursement characteristics across the studied countries: (1) whether digital health solutions are recognized as a reimbursable form of health care, (2) what mechanisms are used to reimburse digital health solutions, (3) how digital health solutions are priced and whether value-based health care frameworks are embedded in that process, and (4) whether any funding is available to reimburse digital health solutions outside of public or private insurance policies.

Methods

Policy Mapping Framework

This study uses a policy mapping framework, which has been used and validated by previous research in the fields of autism, disability, and substance abuse policy [30-35]. The policy mapping framework is based on the foundation of a scoping review, which allows for the rapid mapping of the key concepts underpinning a broad research area that is particularly valuable for complex issues that have not been reviewed comprehensively to date [36,37]. The established framework is suited to analyzing the development of health and social policy over time and across multiple layers of governance. In the context of this study, we only sought to collect information on current digital health reimbursement practices; as such, the longitudinal aspect of the policy mapping framework was not applied. We further developed this policy mapping framework from a cross-country analysis lens by presenting both individual country information in tabulated form and cross-country differences narratively. This approach is also supported by previous policy mapping exercises [38-40].

Data Collection and Analysis

In line with the Joanna Briggs Institute Manual for Evidence Synthesis for Scoping Reviews [41], we searched MEDLINE (Ovid), Embase (Ovid), Global Health (Ovid), and Web of Science on January 20, 2023, and conducted a follow-up search on August 18, 2023, for articles addressing the reimbursement and financing of digital health apps in Belgium, France, Germany, Israel, Italy, the Netherlands, Poland, Sweden, and the United Kingdom. These databases were chosen to cover both the health-specific and interdisciplinary academic fields. To identify gray literature, Google Scholar (first 300 hits [42]) was used. To be eligible for inclusion, a record had to capture a part of the reimbursement or financing pathway of digital health apps in the studied countries. Only records from 2018 onward were eligible for inclusion as this timeframe captures the developments before and during the COVID-19 pandemic in terms of digital health reimbursement and financing as well as the launch of the first country-level reimbursement policy for digital health apps in Germany in late 2019 [25]. The policy mapping framework, in contrast to a traditional scoping review, takes a broader approach to the types of evidence that are eligible for inclusion. Specifically, after searching exclusively for original research and literature reviews, we identified only 14 articles eligible for inclusion, which was too scarce to provide information on existing digital health reimbursement pathways in the studied countries. As such, we expanded the eligibility criteria to include editorials, commentaries, viewpoints, and gray literature as these documents may provide important details of policy developments before these are more rigorously captured in empirical research. We did not directly search for policy repositories as is common practice in the policy mapping framework seeing as we aim to map reimbursement processes rather than the legal basis for reimbursement. Given our specific interest in policy developments in the WHO/Europe region, articles without a focus on the abovementioned countries were excluded.

Table 1 shows the build-up of the search strings for the academic database searches as well as the number of hits per query. The search string was reviewed and validated by an information specialist at the London School of Economics and Political Science Library. The search terms for the supplementary searches in Google Scholar consisted of the phrases "reimbursement of digital health," "reimbursement of digital therapeutics," "financing of digital health," "financing of digital therapeutics," "digital health tariff," "digital health pricing," "telehealth pricing," "telehealth tariff," and "financing of telehealth" combined with the respective country. This combination of keywords ensured that less complex forms of digital health solutions (eg, telehealth and telemedicine) and more complex forms (eg, digital health apps or digital therapeutics) were covered in the search string. If the dominant language was a language other than English, we used Google Translate to translate the search phrases into the desired language [43]. To minimize the potential bias introduced through machine translations, experts from the studied countries were asked to assist in searching and interpreting the digital health reimbursement landscape for their respective countries. Additional articles were identified through a review of the references of the studies included through database search. Policy contents and mechanisms were identified from the selected references and reviewed using thematic content analysis. After finalizing the data collection and analysis, country experts reviewed the collected information to validate the findings and, where necessary, add additional expertise and insights. Finally, individual country information was tabulated per studied reimbursement characteristic and the disparities between countries were narratively synthesized.

Table 1. Search queries for the respective databases.

Database	Query	Hits, n
• MEDLINE (Ovid) • Embase (Ovid) • Global Health (Ovid)	1. (digital adj3 (health or medicine or therapeutic? or care)).ti,ab. 2. (ehealth or e-health or mhealth or m-health or telehealth or telemedicine or "health app*"[a] or telecare or "virtual health" OR "mobile app*").ti,ab. 3. exp telemedicine/ 4. 1 or 2 or 3 5. Insurance, Health, Reimbursement/ 6. exp "Costs and Cost Analysis"/ 7. reimburse* OR financ* OR pricing OR price* OR tariff*).ti,ab. 8. 5 or 6 or 7 9. Belgium/OR France/OR Germany/OR Israel/OR Italy/OR Netherlands/OR Poland/OR Sweden/OR United Kingdom/ 10. (Belgi* OR France OR French OR German* IR Israel* IR Ital* OR Netherlands OR Dutch OR Poland OR Polish OR Sweden OR Swedish OR "United Kingdom" OR Engl* OR Wales OR Welsh OR Scot* OR "Northern Ir*").ti,ab. 11. 9 or 10 12. 4 and 8 and 11 13. limit 12 to yr="2018 -Current"	1004
• Web of Science	1. (TS=(digital NEAR/3 (health or medicine or therapeutic? or care)) OR TS=(ehealth or e-health or mhealth or m-health or telehealth or telemedicine or "health app*" or telecare or "virtual health" OR "mobile app*")) AND TS=(reimburse* OR financ* OR pricing OR price* OR tariff*) AND TS=(Belgi* OR France OR French OR German* IR Israel* IR Ital* OR Netherlands OR Dutch OR Poland OR Polish OR Sweden OR Swedish OR "United Kingdom" OR Engl* OR Wales OR Welsh OR Scot* OR "Northern Ir*")	231

[a]Searches were originally performed using straight quotations.

Ethical Considerations

This study has no inherent ethical implications or considerations.

Results

Search Results

Our search strategy yielded a total of 1536 records (1235 documents through academic database searching and 301 through supplementary searches). A further 23 records were identified through the country experts. After deduplication, 1266 documents were screened for eligibility, and 2.57% (40/1559) were ultimately included in the analysis. Most documents were excluded because they did not discuss the current reimbursement or financing pathways in place in the studied countries. We included 9 original research articles [11,20,44-50], 5 reviews [8,18,27,51,52], 4 conference abstracts [53-56], 18 gray literature sources [28,57-73], 3 reports [74-76], and 1 commentary [25]. In terms of country focus, 3 documents focused on Belgium [57-59], 8 on France [11,28,51,54-56,60,61], 6 on Germany [18,27,44,52,54,62], 3 on Israel [45,63,76], 4 on Italy [46-48,64], 5 on the Netherlands [8,20,49,65,75], 5 on Poland [50,66-68,74], 3 on Sweden [69-71], and 6 on the United Kingdom [28,53,54,56,72,73]. Figure 1 shows a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the data collection process. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist is shown in Multimedia Appendix 1. Table 2 shows the details of country-specific digital health reimbursement.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the screening process.

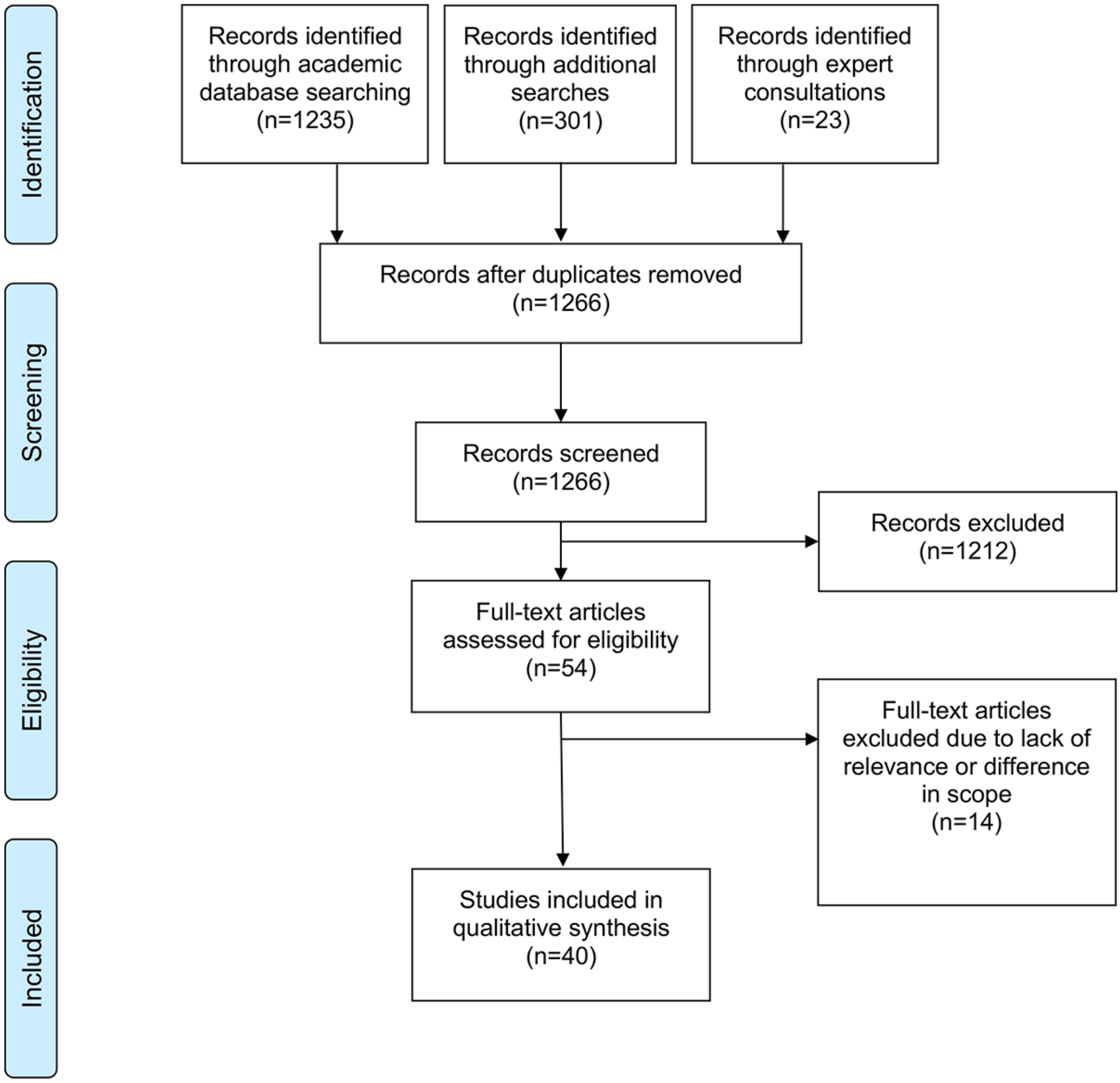


Table 2. Country-specific details on the reimbursement and financing of digital health solutions and apps.

Country	Reimbursement characteristics			
	Eligible for reimbursement	Mechanism of reimbursement	Pricing of digital health	Noninsurance reimbursement
Belgium	Yes	• Reimbursement of digital health solutions does not focus on reimbursing the solution itself. Rather, the Belgian reimbursement system covers a particular health care trajectory holistically and digital health solutions may be deployed by the practitioner as part of the health care process [57].	• Once a digital health solution has reached the highest level of the mHealth[a] validation pyramid, the pricing is done within the context of the health care process that the app will be integrated into. Each health care process requires its own price determination [58].	The TBM[b] program aims at contributing to the implementation of new therapies, diagnostic techniques, and preventive methods, which, without government funding, would not make it to the patient due to a lack of industrial interest [59].
France	Yes	• Reimbursement of digital health solutions can occur through the centralized pathway of medical devices. Connected medical devices have recently been added to the scope of the French National Authority for Health's directory of products that qualify for reimbursement within the statutory health insurance [28,54,55]. After the pricing process has been completed, the National Union of Health Insurers adds the digital health solution to a directory of reimbursable products and sets a reimbursement rate for the next 5 y that matches the digital health solution's clinical benefits assessment [11,54]. • Following the success of the ETAPES[c] experimental program, a reimbursement pathway was introduced in French law under the 2022 Social Security Act (Article 36). This pathway aims to deploy novel remote monitoring apps and requires an evaluation by the French National Authority for Health, much like the existing pathway of medical devices. In case digital health solutions also have therapeutic functions, this pathway can only be applied to the telemonitoring parts of the digital health solution [60]. • In 2023, the French Ministry of Health and Prevention introduced an early access pathway for the reimbursement of sufficiently developed digital medical devices (ie, digital health apps) through the PECAN[d] program [61]. The PECAN program allows 1 y of special coverage by the French health care system, enabling the manufacturer to be reimbursed while they finalize the demonstration of their clinical or organizational benefits.	• The Economic Committee for Health Products (CEPS[e]) negotiates the pricing of digital health solution within the statutory insurance with the manufacturer [11,51]. Currently, there are no specific value-based pricing guidelines for digital health solution available [51].	Until December 2022 inclusive, a second reimbursement pathway existed in the form of the experimental program for telemonitoring in France (ETAPES), which focused on the development of telemonitoring approaches in 5 health specialties: heart failure, kidney failure, respiratory failure, diabetes, and implantable cardiac devices [28,56].

Country	Reimbursement characteristics			
	Eligible for reimbursement	Mechanism of reimbursement	Pricing of digital health	Noninsurance reimbursement
Germany	Yes	• Digital health apps specifically can be reimbursed under the statutory health insurance as long as they are approved by the Federal Institute for Drugs and Medical Devices and listed in the national digital health directory [18,25,44,52,54]. It is prescribed on a fee-for-service basis [18,27,62].	• For the first 12 mo of being listed in the digital health app directory, the manufacturer is generally able to freely set the sales price and pricing model of their digital health app. After 12 mo, the set price is a negotiated price between the manufacturer and the National Association of Statutory Health Insurance Funds [18]. • Before the start of the pricing negotiations, the following details have to be clarified [18]: 1. The evidence on general requirements and positive health effects. 2. The results of the studies conducted as part of the possible trial phase. 3. Information on prices for self-payers; 4. Information on prices in other European countries. 5. The complete notification of the Federal Institute for Drugs and Medical Devices about the inclusion of the digital health app in the national directory. 6. The number of activation or prescription codes redeemed for the digital health app in the period from inclusion in the national directory to 5 d before submission.	Financial support for the development of digital health apps can be obtained from the German Innovation Fund, funded by the German Health Insurance (Gestzlicher Krankenversicherung) [27].
Israel	Yes	• Health maintenance organizations are obligated to provide the services described in the National List of Health Services to their insured population [63,76]. The list is updated annually by an appointed Public Committee, with additional budget allocated to these new additions. As the national list is part of the health insurance law, the technologies not considered as stand-alone technologies but as entitlement to the medical service provided by these technologies. The health maintenance organizations may choose to use digital health solutions to provide an existing entitlement, instead or in parallel to the more traditional methods. • Health maintenance organizations can also decide to purchase digital health solutions as part of a service that is not included in the national list. In this case, they fund it through their internal basket or through the complementary insurances [63].	• When a new technology is added to the national list, its price is estimated by subcommittee adjacent to the public committee and used for budgeting purposes [45,76]. However, the actual price of the technology is negotiated between each of the health maintenance organizations and the manufacturer [63].	• Other funds that are available to reimburse digital health solutions outside insurance policies mostly include Ministry of Health grant programs specifically aimed at supporting the development and implementation of digital health solutions. These funds are used to support the development of new digital health solutions or to help fund the adoption and implementation of existing solutions [63].

Country	Reimbursement characteristics			
	Eligible for reimbursement	Mechanism of reimbursement	Pricing of digital health	Noninsurance reimbursement
Italy	Yes	• Digital health solutions can be reimbursed under the national health system [46], although the reimbursement procedures of digital health apps represent an open challenge and is open to multiple approaches [47]. All regions in Italy adopted tariffs for telehealth and matching reimbursement procedures for all modes of service delivery (digitally supported or in person) [64]. These tariffs should follow a payment parity mechanism, indicating health care providers are paid a fixed amount/patient, regardless of the services provided and their mode of provision [64].	• Digital health solutions that enhance the current health care and therapies process may be merged and embedded into an updated price list of existing services. Pricing protocols for digital health solutions that require a fee-for-subscription attached to a drug or medical device have not yet been established [47].	• Pilot projects within the national health system promote the integration of digital health solutions into the delivery of health care through public health services [48].
Netherlands	Yes	• The Dutch Healthcare Authority has published guidance documents to help health professionals in the Netherlands distinguish between clinical medical apps and assistive health apps. This distinction is important in the Dutch health care context to determine whether a digital health solution has to be reimbursed under individual health insurance companies (in the case of use in primary care, home, or community settings) or whether they can be reimbursed by the basic health insurance package under the diagnosis-related groups (in the case of hospital-based specialist care) [65].	• Pricing of (digital) health solutions in the Netherlands currently involves a negotiation between the health insurers, health providers, the Dutch Healthcare Authority (NZa[f]), and the Dutch Health Institute (Zorginstituut Nederland), though no specific guidelines for the pricing of digital health solutions have been established [49,75].	• The Dutch Healthcare Authority offers a financing mechanism for “promising types of healthcare” until 2023 inclusive, which cover treatment options currently not yet covered by the basic health insurance [65]. • In addition, some digital health solutions are made available through sponsorships from nongovernmental organizations (eg, Alzheimer Netherlands sponsoring 2 digital health apps for dementia) [8,20].
Poland	No	• A reimbursement pathway within the National Health Fund for digitally delivered health services was created as part of an emergency COVID-19 pandemic policy. The scope of digital health in Poland is currently limited to tool of digital consultation between health care professionals and patients, consultations between health care professionals, as well as the e-prescription system [50]. This situation is further strengthened by the limitation that reimbursable digital health care should be performed by a health care professional whose services are already covered under the National Health Fund [66].		• Although no concrete alternative financing pathways are laid out, the Program for the Development of eHealth in Poland for the years 2022-2027 indicates that funds can be obtained for this purpose from programs, such as the Regional Operational Programme, Operational Programme Infrastructure and Environment, National Recovery Plan, and the Digital Europe Programme [68].

Country	Reimbursement characteristics			
	Eligible for reimbursement	Mechanism of reimbursement	Pricing of digital health	Noninsurance reimbursement
			• Health care services tariffing involves the President of the National Health Fund, the Tariffing Council, and the Minister of Health [74]. The Tariffing Council is responsible for providing an opinion on the determination of the tariff of the service, whereas the Minister of Health is responsible for approving the tariffing plan. The report on the determination of the tariff of the service includes a description of the health care service subject to tariffing, an analysis of demand and current and desired supply of the health care service subject to tariffing, a description of the manner and level of financing of the health care service subject to tariffing in other countries, an analysis of cost data, a draft tariff of the service, an analysis of the financial effects on the health care system, and other available data necessary to determine the tariff of the service. The Tariffing Council is required to issue an opinion on the determination of the tariff of the service within 30 d of receipt of the report. The Supreme Audit Office notes that the pricing of individual telemedicine services reimbursed by the National Health Fund usually takes much less time than the pricing of other services [67].	
Sweden	Yes	• For the purpose of streamlining the remuneration of digital health care, Sweden drafted recommendations in 2019 on what regions or local authorities should be reimbursed for if a citizen from another region seeks digital health care within their region [69].	• Each region is responsible for the price-setting of digital health care services and the corresponding copayments for patients. However, patients are not limited to only seek health care in their region of residence, resulting in a complicated system with different prices for digital health care outside the county and for physical care within the region [70].	• Public funding for early-stage innovations can be acquired through Vinnova, Sweden's innovation agency [71].
United Kingdom	Yes	• Once a digital health app receives a positive recommendation from the NICE[g], the app becomes eligible for purchase by the integrated care boards pending negotiations [28,56]. Delivery of digital health apps to patients is free of charge at the point of service [72].	• In their assessment, NICE provides recommendations on a value-based pricing for a digital health app. However, pricing negotiations occur individually with the 42 integrated care boards (the replacement of the clinical commissioning groups) [28,53,54,56].	

Country	Reimbursement characteristics			
	Eligible for reimbursement	Mechanism of reimbursement	Pricing of digital health	Noninsurance reimbursement
				• In England, the MedTech funding mandate can reimburse the costs of using digital health apps for a duration of up to 4 y, though a positive assessment of the NICE is required to be eligible for this funding scheme [73]. The NHS[h] Innovation Accelerator aims to fast-track digital innovations into the NHS by supporting high-impact, evidence-based interventions and providing bursaries for scaling across the NHS. The Innovation and Technology Tariff has been introduced in 2017 to foster the adoption and centralize funding of 6 innovations deemed suitable for NHS-scale deployment (one of which is a digital health intervention) [53].

[a]mHealth: mobile health.

[b]TBM: Toegepast Biomedisch onderzoek met een primair Maatschappelijke finaliteit.

[c]ETAPES: Expe´rimentations de Te´le´me´decine pour l'Ame´lioration des Parcours en Sante´.

[d]PECAN: Prise en Charge Anticipée Numerique des Dispositifs Médicaux.

[e]CEPS: Comité Economique des Produits de Santé.

[f]NZa: Nederlandse Zorgautoriteit.

[g]NICE: National Institute for Health and Care Excellence.

[h]NHS: National Health Service.

Eligibility and Mechanism of Reimbursement Through Insurance

All studied countries except Poland allowed for some reimbursement of digital health solutions. However, the reimbursement mechanisms differed substantially across countries, which can be attributed in part to the differences in how the health systems are financed. Two reimbursement archetypes could be derived from the studied countries: countries either reimbursed the digital health solution itself (France, Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom) or reimbursed the clinical pathway that the digital health solution is part of (Belgium and Israel). Further variation was observed across the studied countries in terms of how the digital health apps were classified. Belgium, France, and Israel classify digital health solutions under the traditional paradigm of medical devices, whereas Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom recognize digital health solutions as their own classification. Italy was the only country to implement an explicit system of parity for the reimbursement of digital health solutions. This means that health providers are paid a fixed fee regardless of whether health care is delivered face-to-face or through digital health.

Pricing of Digital Health Solutions or Apps

Pricing mechanisms were found to be at different stages of development across countries. France, Germany, Israel, and Poland each have a dedicated committee that decides on the price of digital health solutions or apps within their respective national insurance systems, although only Germany reports a concrete framework upon which the prices of digital health apps are based. In contrast, Sweden and the United Kingdom delegate pricing negotiations to the regional level, with each region being responsible for reaching an agreement with the digital health manufacturer. The Netherlands exhibited a combination of both approaches, depending on whether a specific digital health solution was included in their mandatory health insurance package or the optional insurer-determined insurance package. In contrast, Belgium and Italy determine the price of digital health solutions in the context of the health care process in which they are deployed. Italy reported having no concrete pricing framework in place for digital health apps even though they are deployed on a fee-for-service basis. Despite using differentiated approaches, all countries adopted a variation of cost-based pricing for digital health except for Belgium, where it is integrated in a value-based pricing system and Germany, where price negotiations need to be based on the demonstrated value of the digital health solution.

Financing Digital Health Solutions or Apps Outside Insurance

The studied countries reported an array of options for financing digital health solutions that were not reliant on their inclusion in health insurance packages. France offered an experimental program for digital health solutions in the fields of heart failure, kidney failure, respiratory failure, diabetes, and implantable cardiac devices until January 2023. Belgium, Germany, Israel, Italy, the Netherlands, Sweden, and the United Kingdom offer innovation grants that finance treatment options currently under development and are not yet covered by their respective national health insurance frameworks, which can cover digital health solutions. However, in Germany, digital health apps that are approved via Germany's fast-track process, which combines regulation and reimbursement, have to be directly reimbursed in the statutory health insurance system. Furthermore, certain nongovernmental organizations in the Netherlands offer access to disease-specific digital health solutions outside of insurance packages. In this scenario, they are purchased by the nongovernmental organization and distributed to its members free of charge. Poland was the only country that did not report any concrete funding mechanism for digital health solutions.

Discussion

Principal Findings

Although there is nascent literature on national reimbursement practices vis-à-vis digital health solutions, this study is, to the best of the authors' knowledge, the first to compare reimbursement pathways across countries in the WHO/Europe region with and without specified digital health pathways and to do so for a broader set of digital health tools and approaches. Our findings reveal that the reimbursement pathways for digital health are varied and that value-based pricing frameworks are rare. Although this can be partly attributed to the distinct systems for financing health care in the examined countries, it also emphasizes how the absence of a consistent definition and classification of digital health can contribute to disparate policy and implementation approaches. The present conceptualization of digital health solutions can encompass a range of meanings, including technology, user experience, individual service, product, or process, and can be viewed as part of the broader ecosystem of health services [2]. Although this expansive definition allows for significant flexibility in integrating digital health into existing health care delivery and reimbursement frameworks, it can also lead to uncertainty regarding where and how digital health is meant to fit within a health system. For instance, digital health solutions are regarded as stand-alone health care products in France, Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom, whereas they are framed as tools to deliver traditional health care in Belgium and Israel. Furthermore, Belgium, France, Germany, Italy, and the United Kingdom have specific policy frameworks in place that outline the concept of digital health apps, whereas the other countries categorize digital health apps under either the broader digitalization of health care or medical devices [18,28,56,60,65].

Overall, the observed differences may be explained by the novelty and unprecedented nature of digital health tools and approaches, especially in light of how change resistant the health care sector can be [77]. Another contributing factor may be the lack of digital health literacy among both clinicians and patient users and the necessity (but not sufficient) for adoption training of the health workforce and policy makers to understand the scope, potential benefits, and limitations of these digital transformations [5,78,79]. Furthermore, the acute need to act during the COVID-19 pandemic resulted in divergent policies being implemented across Europe [38,40]. When combined, these factors may help explain why disparate digital health reimbursement policies are currently in place across the studied countries.

In the specific context of digital health apps, we found 2 broad categories of approaches toward reimbursement in the studied countries. On the one hand, digital health apps may be reimbursed per use cycle (ie, the period in which the digital health app must be used to produce positive health benefits). This fee-for-service approach has historically been easy to develop and implement, reflects the actual number of services rendered, and can create an incentive for health care providers to increase access to and use of health care services [80]. However, this reimbursement approach has a significant and oft-cited disadvantage, namely providing health care professionals with the incentive to induce demand [81,82]. In the context of digital health care, which can function autonomously and asynchronously, this risk may be exacerbated as health professionals no longer have to consider their own time constraints in prescribing this form of health care (although it could be tempered when a digital health solution is expected to be a substitute for billable services by a clinician). Consequently, a digital health economic paradox can arise, namely that digital health care has the potential to reduce health care costs relative to in-person services, but can equally inflate health care spending if poorly implemented, although this is dependent on the reimbursement methods applied and whether digital health tools serve as a complement or substitute for in-person care—or a combination of the 2. On the other hand, reimbursement for digital health apps may be included as part of the remuneration of a larger health care process, which is more compatible with diagnosis-related groups or global budget approaches to paying for health care. Both these approaches may be more suited than fee-for-service for realizing the potential for digital health care to reduce costs, as they characteristically introduce incentives to improve efficiency in processes and in the allocation of resources.

Furthermore, reimbursement parity for digital health solutions was explicitly introduced in Italy, although the Belgian and Israeli reimbursement approaches could produce a degree of parity as well. Parity systems are considered enablers for the uptake of digital health [21,24], although concerns exist that parity systems could impede the development of innovative care delivery models, which could limit the potential of telehealth to address high health care costs and complicate the introduction of value-based models [24].

Among the countries examined, the pricing of digital health solutions and apps was mostly determined through discussions between committees at the national or regional level and the digital health app developers and manufacturers. However, it is presently unclear what criteria must be met to participate in these committees and whether the designated institutions have the necessary capabilities to assess the worth of digital health solutions or apps [11,83]. Only France, Germany and the United Kingdom have reported the use of value-based frameworks with clear information requirements to help stakeholders navigate the price negotiations for digital health apps. These 3 countries are at a more advanced stage of digital health implementation and both patients and professionals have more experience with digital health apps and their effects, which is an important foundation for the introduction of value-based reimbursement models [17].

We found that, except for Poland, there were many opportunities for financing digital health solutions and apps outside of insurance-based frameworks in the studied countries. However, there were discrepancies in the specificity of these funding opportunities, as it was unclear whether they targeted digital health specifically or health innovations more generally. Nonetheless, such funding initiatives can alleviate the financial risk for health insurers and allow patients, professionals, and insurers to become accustomed to the use and effects of digital health solutions and apps, which is a key factor in their uptake [84].

Limitations

Some limitations of this study need to be considered. As is common with scoping reviews, the quality of the included studies was not assessed, which should be considered when interpreting the results. However, as the aim of this study was not to validate methodological rigor to ascertain confidence in the data synthesis but rather to collect information about reimbursement processes in different countries, the absence of a quality assessment does not inhibit the validity of this study. In fact, the information collected did not solely rely on scientific articles as country experts also ensured that the collected information was complete and correct. We also recognized the possibility of selection bias and failed to capture all relevant studies as only 3 academic databases and Google Scholar were used, and the search strategy was not exhaustive. This study focused on 9 countries in the WHO/Europe region, which may not reflect the realities of other regions or countries. Furthermore, the lack of a uniform definition and classification of digital health across countries may have influenced the interpretation and comparison of the reimbursement pathways. Therefore, caution should be exercised when interpreting and generalizing the findings of this study. Moreover, the rapidly evolving nature of digital health and its reimbursement pathways means that the information presented in this study may become outdated relatively quickly. Finally, this study did not explore the potential impact of reimbursement policies on patient outcomes, which could be an area of future research.

We have identified several avenues for future research. First, the composition of committees that determine the price of the digital health solutions in the studied countries could be better characterized and assessed. Second, the expertise requirements to adequately assess the value and price of digital health solutions could be further investigated and potentially pooled into a comprehensive framework. Third, a comprehensive, value-based, and value-sensitive framework for the assessment of digital health apps should be developed. Although some research in this area has already been conducted [19,85-87], no comprehensive assessment of how digital health apps generate value across different levels of health systems has been conducted so far. Fourth, a comparative assessment of health expenditure (eg, in the management of a particular disease) after the introduction of digital health care in the studied countries can offer a more sophisticated insight into how digital health interacts with different reimbursement systems to affect overall health expenditure. Fifth, it is unclear whether these findings apply to the broader spectrum of novel health technologies. Digital health is a particular type of innovation in health care and future work can explore how these findings translate to, for instance, the growing role of artificial intelligence in health care, where artificial intelligence–based digital health solutions represent a unique subset of digital health products. Sixth, future research should investigate how digital health solutions or apps would be reimbursed in case of cross-country health service delivery and how these reimbursement practices interact with current international and European Union legislation. Finally, future research should investigate how digital health policy has evolved over time across multiple layers of governance (eg, international, national, and regional where applicable) and even before the COVID-19 pandemic to better understand the extent to which the pandemic accelerated the development of digital health policy holistically and how that affected the formation of reimbursement-specific policy for digital health.

Conclusions

The studied countries have been pursuing heterogeneous approaches to the reimbursement of digital health care. Although no approach is inherently superior, a fee-for-service approach might encourage more prescribing of certain digital health solutions or apps, which may be desirable to increase digital health use for purposes such as screening, yet it equally opens the possibility of supplier-induced demand and thus inflating health care expenditure. Ultimately, reimbursement policies should aim to stimulate the value-based integration of digital health into the health care ecosystem to promote equitable access to innovative health care technologies and improve health outcomes for patients. A clearer understanding of reimbursement and its accompanying incentives at present will help to shape more thoughtful, value-based reimbursement policies going forward. This study can therein serve as a basis for further, more detailed research as the field of digital health reimbursement evolves.

Data Availability

No primary data were collected for this study.

Authors' Contributions

RvK was involved in study conceptualization and methodology, software analysis, validation, formal analysis, investigation, data curation, writing the original draft including reviewing and editing, visualization, and project administration. DS and IK were involved in study conceptualization and methodology, software analysis, validation, formal analysis, investigation, writing the original draft including reviewing and editing, and visualization. GM was involved in validation, writing the original draft including reviewing and editing. DNO, WWZ-C, GN, ADS, and GW performed the data curation, validation, writing the original draft including reviewing and editing. EM was involved in study conceptualization and methodology, validation, formal analysis, writing the original draft including reviewing and editing, and supervising the study.

Conflicts of Interest

DNO is a staff member of the World Health Organization and is solely responsible for the views expressed in this publication, which do not necessarily represent the decisions or policies of the organization. ADS reports consulting income from the US Department of Health and Human Services outside of the context of this project and sits on the Scientific Advisory Board of the German Society for Digital Medicine and on the Strategic Advisory Board of HumanFirst.

Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 99 KB - mhealth_v11i1e49003_app1.pdf]

References

1. Wong BL, Maaß L, Vodden A, van Kessel R, Sorbello S, Buttigieg S, European Public Health Association (EUPHA) Digital Health Section. The dawn of digital public health in Europe: implications for public health policy and practice. Lancet Reg Health Eur 2022 Mar;14:100316 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100316] [Medline: 35132399]
2. van Kessel R, Hrzic R, O'Nuallain E, Weir E, Wong BL, Anderson M, et al. Digital health paradox: international policy perspectives to address increased health inequalities for people living with disabilities. J Med Internet Res 2022 Feb 22;24(2):e33819 [FREE Full text] [doi: 10.2196/33819] [Medline: 35191848]
3. Lazarus JV, Romero D, Kopka CJ, Karim SA, Abu-Raddad LJ, Almeida G, COVID-19 Consensus Statement Panel. A multinational Delphi consensus to end the COVID-19 public health threat. Nature 2022 Nov 03;611(7935):332-345 [FREE Full text] [doi: 10.1038/s41586-022-05398-2] [Medline: 36329272]
4. van Kessel R, Kyriopoulos I, Wong BL, Mossialos E. The effect of the COVID-19 pandemic on digital health-seeking behavior: big data interrupted time-series analysis of Google trends. J Med Internet Res 2023 Jan 16;25:e42401 [FREE Full text] [doi: 10.2196/42401] [Medline: 36603152]
5. Holly L, Wong BL, van Kessel R, Awah I, Agrawal A, Ndili N. Optimising adolescent wellbeing in a digital age. BMJ 2023 Mar 20;380:e068279 [FREE Full text] [doi: 10.1136/bmj-2021-068279] [Medline: 36940933]
6. van Kessel R, Kyriopoulos I, Mastylak A, Mossialos E. Changes in digital healthcare search behavior during the early months of the COVID-19 pandemic: a study of six English-speaking countries. PLOS Digit Health 2023 May 1;2(5):e0000241 [FREE Full text] [doi: 10.1371/journal.pdig.0000241] [Medline: 37126489]
7. European Observatory on Health Systems and Policies, Fahy N, Williams GA. Use of digital health tools in Europe: before, during and after COVID-19. World Health Organization. 2021. URL: https://apps.who.int/iris/handle/10665/345091 [accessed 2023-08-20]
8. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. J Med Internet Res 2023 Jul 25;25:e48000 [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]
9. Patel NA, Butte AJ. Characteristics and challenges of the clinical pipeline of digital therapeutics. NPJ Digit Med 2020 Dec 11;3(1):159 [FREE Full text] [doi: 10.1038/s41746-020-00370-8] [Medline: 33311567]
10. Fast N, van Kessel R, Humphreys K, Ward NF, Roman-Urrestarazu A. The evolution of telepsychiatry for substance use disorders during COVID-19: a narrative review. Curr Addict Rep 2023 Apr 25;10(2):187-197 [FREE Full text] [doi: 10.1007/s40429-023-00480-9] [Medline: 37266192]
11. Prodan A, Deimel L, Ahlqvist J, Birov S, Thiel R, Toivanen M, et al. Success factors for scaling up the adoption of digital therapeutics towards the realization of P5 medicine. Front Med (Lausanne) 2022 Apr 12;9:854665 [FREE Full text] [doi: 10.3389/fmed.2022.854665] [Medline: 35492346]
12. Kählke F, Buntrock C, Smit F, Ebert DD. Systematic review of economic evaluations for internet- and mobile-based interventions for mental health problems. NPJ Digit Med 2022 Nov 23;5(1):175 [FREE Full text] [doi: 10.1038/s41746-022-00702-w] [Medline: 36424463]
13. van Kessel R, Haig M, Mossialos E. Strengthening cybersecurity for patient data protection in Europe. J Med Internet Res 2023 Aug 24;25:e48824 [FREE Full text] [doi: 10.2196/48824] [Medline: 37616048]
14. WHO Centre for Health Development (Kobe, Japan) , Organisation for Economic Co-operation and Development, Barber SL, Lorenzoni L, Ong P. Price setting and price regulation in health care: lessons for advancing universal health coverage. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/325547 [accessed 2023-08-20]

15. Vogler S, Paris V, Ferrario A, Wirtz VJ, de Joncheere K, Schneider P, et al. How can pricing and reimbursement policies improve affordable access to medicines? Lessons learned from European countries. Appl Health Econ Health Policy 2017 Jun 6;15(3):307-321. [doi: 10.1007/s40258-016-0300-z] [Medline: 28063134]
16. Lauenroth VD, Kesselheim AS, Sarpatwari A, Stern AD. Lessons from the impact of price regulation on the pricing of anticancer drugs in Germany. Health Aff (Millwood) 2020 Jul 01;39(7):1185-1193. [doi: 10.1377/hlthaff.2019.01122] [Medline: 32634355]
17. European C. Defining value in "valuebased healthcare" report of the expert panel on effective ways of investing in health (EXPH). Publications Office of the European Union. Luxembourg: Publications Office of the European Union; 2019. URL: https://health.ec.europa.eu/system/files/2019-11/024_defining-value-vbhc_en_0.pdf [accessed 2023-08-20]
18. Gensorowsky D, Witte J, Batram M, Greiner W. Market access and value-based pricing of digital health applications in Germany. Cost Eff Resour Alloc 2022 Jun 13;20(1):25 [FREE Full text] [doi: 10.1186/s12962-022-00359-y] [Medline: 35698135]
19. Haig M, Main C, Chávez D, Kanavos P. A value framework to assess patient-facing digital health technologies that aim to improve chronic disease management: a Delphi approach. Value Health 2023 Jun 28:S1098-3015(23)03025-5 (forthcoming) [FREE Full text] [doi: 10.1016/j.jval.2023.06.008] [Medline: 37385445]
20. Christie HL, Boots LM, Hermans I, Govers M, Tange HJ, Verhey FR, et al. Business models of eHealth interventions to support informal caregivers of people with dementia in the Netherlands: analysis of case studies. JMIR Aging 2021 Jun 03;4(2):e24724 [FREE Full text] [doi: 10.2196/24724] [Medline: 34081009]
21. Medicare billing codes for digital therapeutics: a path forward. HealthLeaders. 2022 Aug 11. URL: https://www.healthleadersmedia.com/innovation/medicare-billing-codes-digital-therapeutics-path-forward [accessed 2022-10-20]
22. COVID-19 advice for health professionals. Australian Government Department of Health and Aged Care. URL: https://www.health.gov.au/topics/covid-19/health-sector/health-professionals [accessed 2022-10-21]
23. Savira F, Orellana L, Hensher M, Gao L, Sanigorski A, Mc Namara K, et al. Use of general practitioner telehealth services during the COVID-19 pandemic in regional Victoria, Australia: retrospective analysis. J Med Internet Res 2023 Feb 07;25:e39384 [FREE Full text] [doi: 10.2196/39384] [Medline: 36649230]
24. Erikson C, Herring J, Park YH, Luo Q, Burke G. Association between state payment parity policies and telehealth usage at community health centers during COVID-19. J Am Med Inform Assoc 2022 Sep 12;29(10):1715-1721 [FREE Full text] [doi: 10.1093/jamia/ocac104] [Medline: 35864736]
25. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3(1):94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
26. Marshall MN. Sampling for qualitative research. Fam Pract 1996 Dec;13(6):522-525. [doi: 10.1093/fampra/13.6.522] [Medline: 9023528]
27. Ju JH, Sim B, Lee J, Lee JY. Reimbursement of digital therapeutics: future perspectives in Korea. Korean Circ J 2022 Apr;52(4):265-279 [FREE Full text] [doi: 10.4070/kcj.2022.0014] [Medline: 35388995]
28. Recognising the value of digital health apps: an assessment of five European healthcare systems. MedTech Europe. 2021 Nov 15. URL: https://www.medtecheurope.org/resource-library/recognising-the-value-of-digital-health-apps-an-assessment-of-five-european-healthcare-systems/ [accessed 2023-01-15]
29. Chawla V. How to get your digital health app reimbursed in Europe? Start with Germany, Belgium and France. Research 2 Guidance. URL: https://research2guidance.com/how-to-get-your-digital-health-app-reimbursed-in-europe-start-with-germany-belgium-and-france/ [accessed 2022-11-11]
30. Neicun J, Steenhuizen M, van Kessel R, Yang JC, Negri A, Czabanowska K, et al. Mapping novel psychoactive substances policy in the EU: the case of Portugal, the Netherlands, Czech Republic, Poland, the United Kingdom and Sweden. PLoS One 2019 Jun 26;14(6):e0218011 [FREE Full text] [doi: 10.1371/journal.pone.0218011] [Medline: 31242225]
31. Bunt D, van Kessel R, Hoekstra RA, Czabanowska K, Brayne C, Baron-Cohen S, et al. Quotas, and anti-discrimination policies relating to autism in the EU: scoping review and policy mapping in Germany, France, Netherlands, United Kingdom, Slovakia, Poland, and Romania. Autism Res 2020 Aug 22;13(8):1397-1417 [FREE Full text] [doi: 10.1002/aur.2315] [Medline: 32441457]
32. van Kessel R, Hrzic R, Czabanowska K, Baranger A, Azzopardi-Muscat N, Charambalous-Darden N, et al. Autism and education-international policy in small EU states: policy mapping in Malta, Cyprus, Luxembourg and Slovenia. Eur J Public Health 2020 Dec 11;30(6):1078-1083 [FREE Full text] [doi: 10.1093/eurpub/ckaa146] [Medline: 32879964]
33. van Kessel R, Dijkstra W, Prasauskiene A, Villeruša A, Brayne C, Baron-Cohen S, et al. Education, special needs, and autism in the Baltic states: policy mapping in Estonia, Latvia, and Lithuania. Front Educ 2020 Sep 4;5. [doi: 10.3389/feduc.2020.00161]
34. van Kessel R, Steinhoff P, Varga O, Breznoščáková D, Czabanowska K, Brayne C, et al. Autism and education-teacher policy in Europe: policy mapping of Austria, Hungary, Slovakia and Czech Republic. Res Dev Disabil 2020 Oct;105:103734 [FREE Full text] [doi: 10.1016/j.ridd.2020.103734] [Medline: 32758993]
35. van Kessel R, Hrzic R, Cassidy S, Brayne C, Baron-Cohen S, Czabanowska K, et al. Inclusive education in the European Union: a fuzzy-set qualitative comparative analysis of education policy for autism. Soc Work Public Health 2021 Feb 17;36(2):286-299. [doi: 10.1080/19371918.2021.1877590] [Medline: 33535919]

36. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
38. van Kessel R, Forman R, Milstein R, Mastylak A, Czabanowska K, Czypionka T, et al. Divergent COVID-19 vaccine policies: policy mapping of ten European countries. Vaccine 2023 Apr 24;41(17):2804-2810 [FREE Full text] [doi: 10.1016/j.vaccine.2023.03.036] [Medline: 36967287]
39. Hargreaves S, Nellums L, Ravensbergen S, Friedland J, Stienstra Y, ESGITM Working Group on Vaccination in Migrants. Divergent approaches in the vaccination of recently arrived migrants to Europe: a survey of national experts from 32 countries, 2017. Euro Surveill 2018 Oct;23(41):1700772 [FREE Full text] [doi: 10.2807/1560-7917.ES.2018.23.41.1700772] [Medline: 30326996]
40. Dal-Ré R, Banzi R, Becker SL, Launay O, Pavli A. High heterogeneity on the accepted vaccines for COVID-19 certificates in European countries. Travel Med Infect Dis 2022;48:102321 [FREE Full text] [doi: 10.1016/j.tmaid.2022.102321] [Medline: 35390530]
41. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, South Australia: Joanna Briggs Institute; 2020.
42. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015 Sep 17;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
43. Groves M, Mundt K. Friend or foe? Google Translate in language for academic purposes. English Specif Purp 2015 Jan;37:112-121. [doi: 10.1016/j.esp.2014.09.001]
44. Gregor-Haack J. [Reimbursement of health apps by the German statutory health insurance]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar 26;61(3):328-333. [doi: 10.1007/s00103-018-2689-z] [Medline: 29374299]
45. Assor Y, Greenberg D. Public legitimacy of healthcare resource allocation committees: lessons learned from assessing an Israeli case study. BMC Health Serv Res 2022 Jun 02;22(1):737 [FREE Full text] [doi: 10.1186/s12913-022-07992-6] [Medline: 35655271]
46. Brunetti ND, Molinari G, Acquistapace F, Zimotti T, Parati G, Indolfi C, et al. 2019 Italian Society of Cardiology Census on telemedicine in cardiovascular disease: a report from the working group on telecardiology and informatics. Open Heart 2020 Mar 17;7(1):e001157 [FREE Full text] [doi: 10.1136/openhrt-2019-001157] [Medline: 32206315]
47. Tarricone R, Petracca F, Ciani O, Cucciniello M. Distinguishing features in the assessment of mHealth apps. Expert Rev Pharmacoecon Outcomes Res 2021 Aug 04;21(4):521-526. [doi: 10.1080/14737167.2021.1891883] [Medline: 33663324]
48. Pardolesi A, Gherzi L, Pastorino U. Telemedicine for management of patients with lung cancer during COVID-19 in an Italian cancer institute: SmartDoc project. Tumori 2022 Aug 10;108(4):357-363. [doi: 10.1177/03008916211012760] [Medline: 33971749]
49. Bally EL, Cesuroglu T. Toward integration of mHealth in primary care in the Netherlands: a qualitative analysis of stakeholder perspectives. Front Public Health 2019 Jan 15;7:407 [FREE Full text] [doi: 10.3389/fpubh.2019.00407] [Medline: 32010660]
50. Furlepa K, Śliwczyński A, Kamecka K, Kozłowski R, Gołębiak I, Cichońska-Rzeźnicka D, et al. The COVID-19 pandemic as an impulse for the development of telemedicine in primary care in Poland. J Pers Med 2022 Jul 18;12(7):1165 [FREE Full text] [doi: 10.3390/jpm12071165] [Medline: 35887661]
51. Mantovani A, Leopaldi C, Nighswander CM, Di Bidino R. Access and reimbursement pathways for digital health solutions and diagnostic devices: current scenario and challenges. Front Med Technol 2023 Feb 20;5:1101476 [FREE Full text] [doi: 10.3389/fmedt.2023.1101476] [Medline: 36891483]
52. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206 [FREE Full text] [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
53. Braileanu G, Tavella F, Rousseau B. PHP227 - the current reimbursement landscape for digital health in the UK: how far are we from a centralised, national system? Value Health 2018 Oct;21(Supplement 3):S188. [doi: 10.1016/j.jval.2018.09.1121]
54. Vollmer L, Foxon G, Danev V, Berard I, Benazet F, Walzer S. PMD17 comparison of market access routes of digital health applications in France, Germany and the UK. Value Health 2020 Dec;23(Supplement 2):S579. [doi: 10.1016/j.jval.2020.08.1050]
55. Berard I, Benazet F, Halimi M, Amory D, Bouchara G. PMD148 - the budget impact of global telemedicine solution in France. Value Health 2018 Oct;21(Supplement 3):S268. [doi: 10.1016/j.jval.2018.09.1597]
56. Barzey V, Brennan J, Tutt S, Cope N. Innovative access pathways for digital healthcare solutions: learnings from a European analogue analysis for Asian HTA bodies. Value Health 2018 Sep;21(Supplement 2):S73. [doi: 10.1016/j.jval.2018.07.552]
57. Financing. mHealthBelgium. URL: https://mhealthbelgium.be/financing [accessed 2022-11-11]
58. Validation Pyramid. mHealthBelgium. URL: https://mhealthbelgium.be/validation-pyramid [accessed 2022-11-11]

59. TBM projects (applied biomedical research with a primary social finality). Flanders. URL: https://www.fwo.be/en/fellowships-funding/research-projects/tbm-projects/ [accessed 2023-04-05]
60. LOI n° 2021-1754 du 23 décembre 2021 de financement de la sécurité sociale pour 2022 (1). Legifrance. URL: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044553428 [accessed 2023-08-23]
61. Early access to reimbursement for digital devices (PECAN). French Ministry of Health and Prevention. URL: http://gnius.esante.gouv.fr/en/financing/reimbursement-profiles/early-access-reimbursement-digital-devices-pecan [accessed 2023-05-18]
62. The Fast-Track process for digital health applications (DiGA) according to Section 139e SGB V. Federal Institute for Drugs and Medical Devices. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf?__blob=publicationFile&v=2 [accessed 2022-10-15]
63. International profiles of health care systems. The Commonwealth Fund. 2020 Dec. URL: https://www.commonwealthfund.org/sites/default/files/2020-12/International_Profiles_of_Health_Care_Systems_Dec2020.pdf [accessed 2023-08-21]
64. Approval of guidelines for telemedicine services - Functional requirements and service levels. (22A06184) (GU General Series n.256 of 02-11-2022). Gazzetta Ufficiale. 2022 Sep 21. URL: https://www.gazzettaufficiale.it/eli/id/2022/11/02/22A06184/sg [accessed 2023-03-23]
65. Wegwijzer bekostiging digitale zorg 2023. Nederlandse Zorgautoriteit. URL: https://puc.overheid.nl/nza/doc/PUC_655318_22/1/ [accessed 2022-11-12]
66. Rozporządzenie Ministra Zdrowia z dnia 24 września 2013 r. w sprawie świadczeń gwarantowanych z zakresu podstawowej opieki zdrowotnej. ISAP – Internetowy System Aktów Prawnych. URL: https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20130001248 [accessed 2023-08-24]
67. Wycena świadczeń opieki zdrowotnej finansowanych ze środków publicznych. Najwyzsza Izba Kontroli. 2021 Aug 24. URL: https://www.nik.gov.pl/kontrole/P/20/055/ [accessed 2023-08-24]
68. Program Rozwoju e-Zdrowia na lata 2022-2027. Ministerstwo Zdrowia. 2022 Oct 20. URL: https://www.gov.pl/web/zdrowie/program-rozwoju-e-zdrowia-na-lata-2022-2027 [accessed 2023-08-24]
69. Ersättningar, digitala vårdtjänster i primärvården. Sveriger Kommuner och Regioner. 2022 Mar 22. URL: https://skr.se/skr/halsasjukvard/utvecklingavverksamhet/ehalsa/dethargorskrinomehalsa/digitalavardtjanster/ersattningardigitalavardtjansteriprimarvarden.28836.html [accessed 2022-11-13]
70. Blix M. Regeringen lägger krokben för digitala psykologer. Timbro. 2022 Feb 22. URL: https://timbro.se/smedjan/regeringen-lagger-krokben-for-digitala-psykologer/ [accessed 2022-11-13]
71. Funding. Vinnova. URL: https://www.vinnova.se/en/apply-for-funding/ [accessed 2023-03-23]
72. NHS Scotland pioneers digital insomnia app. The Scotsman. 2022 Sep 15. URL: https://www.scotsman.com/health/nhs-scotland-pioneers-digital-insomnia-app-3844593 [accessed 2022-11-12]
73. MedTech funding mandate and MedTech support. National Health Service England. URL: https://www.england.nhs.uk/aac/what-we-do/how-can-the-aac-help-me/the-medtech-funding-mandate/ [accessed 2022-11-13]
74. Topór-Mądry R, Balwicki Ł, Kowalska-Bobko I, Włodarczyk CW. Poland. In: Rechel B, Maresso A, Sagan A, Hernández-Quevedo C, Williams G, Richardson E, et al, editors. Organization and Financing of Public Health Services in Europe: Country Reports. Copenhagen, Denmark: World Health Organization Regional Office for Europe; 2018.
75. Kroneman M, Boerma W, van den Berg M, Groenewegen P, de Jong J, van Ginneken E. Health system review: Netherlands. The European Observatory on Health Systems and Policies. 2006. URL: https://apps.who.int/iris/bitstream/handle/10665/330244/HiT-18-2-2016-eng.pdf?sequence=5&isAllowed=y [accessed 2023-08-26]
76. Rosen B, Samuel H. Health care systems in transition: Israel. The European Observatory on Health Care Systems. 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/107537/HiT-5-1-2003-eng.pdf?sequence=11&isAllowed=y [accessed 2023-08-26]
77. Glaser J, Shaw S. Digital transformation success: what can health care providers learn from other industries? NEJM Catal 2022 Mar 22 [FREE Full text] [doi: 10.1056/CAT.21.0434]
78. van Kessel R, Wong BL, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022 Mar;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
79. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, Secretariat of the LancetFinancial Times Commission. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. Lancet 2021 Nov 06;398(10312):1727-1776. [doi: 10.1016/S0140-6736(21)01824-9] [Medline: 34706260]
80. Langenbrunner J, Cashin C, O'Dougherty S. Designing and implementing health care provider payment systems how-to manuals. The World Bank. URL: https://documents1.worldbank.org/curated/en/577441468169753880/pdf/543180BRI0Box31ymentMethodsoverview.pdf [accessed 2023-08-21]
81. Dorn SD. Backslide or forward progress? Virtual care at U.S. healthcare systems beyond the COVID-19 pandemic. NPJ Digit Med 2021 Jan 08;4(1):6 [FREE Full text] [doi: 10.1038/s41746-020-00379-z] [Medline: 33420420]
82. Wharton GA, Sood HS, Sissons A, Mossialos E. Virtual primary care: fragmentation or integration? Lancet Digit Health 2019 Nov;1(7):e330-e331 [FREE Full text] [doi: 10.1016/S2589-7500(19)30152-9] [Medline: 33323207]

83. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HG. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
84. Busso M, Gonzalez MP, Scartascini C. On the demand for telemedicine: evidence from the COVID-19 pandemic. Health Econ 2022 Jul 08;31(7):1491-1505 [FREE Full text] [doi: 10.1002/hec.4523] [Medline: 35527351]
85. D'Anza B, Pronovost PJ. Digital health: unlocking value in a post-pandemic world. Popul Health Manag 2022 Feb 01;25(1):11-22. [doi: 10.1089/pop.2021.0031] [Medline: 34042532]
86. Lantzsch H, Panteli D, Martino F, Stephani V, Seißler D, Püschel C, et al. Benefit assessment and reimbursement of digital health applications: concepts for setting up a new system for public coverage. Front Public Health 2022 Apr 21;10:832870 [FREE Full text] [doi: 10.3389/fpubh.2022.832870] [Medline: 35530738]
87. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
WHO/Europe: World Health Organization European Region

Edited by L Buis; submitted 14.05.23; peer-reviewed by J Drott, P Mechael; comments to author 17.08.23; revised version received 03.09.23; accepted 08.09.23; published 29.09.23.

Please cite as:
van Kessel R, Srivastava D, Kyriopoulos I, Monti G, Novillo-Ortiz D, Milman R, Zhang-Czabanowski WW, Nasi G, Stern AD, Wharton G, Mossialos E
Digital Health Reimbursement Strategies of 8 European Countries and Israel: Scoping Review and Policy Mapping
JMIR Mhealth Uhealth 2023;11:e49003
URL: https://mhealth.jmir.org/2023/1/e49003
doi:10.2196/49003
PMID:37773610

Review

Evaluation of Mobile Health Technology Interventions for the Postdischarge Management of Patients With Head and Neck Cancer: Scoping Review

Yufei Li[1], PhD; Weihong Chen[2], MS; Yanjing Liang[1], MS; Ling Yang[2], MS; Lili Hou[3], PhD

[1]School of Nursing, Shanghai Jiao Tong University, Shanghai, China

[2]School of Nursing, Chengdu University of Traditional Chinese Medicine, Chengdu, China

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

Corresponding Author:
Lili Hou, PhD
Department of Nursing
Shanghai Ninth People's Hospital
Shanghai Jiao Tong University School of Medicine
Bldg 8, 2nd Fl
639 Zhizaoju Road, Huangpu District
Shanghai, 200011
China
Phone: 86 13816033620
Email: pisces_liz@163.com

Abstract

Background: Patients with head and neck cancer (HNC) often experience various types and degrees of complications and functional impairment following surgery or radiotherapy. Consequently, these patients require extensive postdischarge rehabilitation, either at home or in the community. Numerous studies have shown the advantages of mobile Health (mHealth) technology in assisting patients with cancer with self-management and rehabilitation during the postdischarge period. However, few reviews have focused on the intervention, management, and evaluation of mHealth technology in postdischarge patients with HNC.

Objective: This study aimed to conduct a scoping review of mHealth technology apps and interventions currently available to patients discharged from hospitals after receiving treatment for HNC. This study sought to identify and summarize the types and effectiveness of existing mHealth interventions as well as the differences in their outcome assessments.

Methods: The PubMed, Embase, Web of Science, and CINAHL databases were used to identify studies with no publication time limits. The keywords "mobile health technology" and "head and neck cancer" were combined to address the main concepts of the research questions.

Results: Of the 1625 papers identified, 13 (0.8%) met the inclusion and exclusion criteria. Most studies (n=8, 61.5%) were randomized controlled trials (RCTs) and cohort studies. These studies were conducted in 6 countries. The main aims of the mHealth interventions in these studies are as follows: (1) symptom monitoring and assessment, (2) rehabilitation training, (3) access to medical health information, (4) telehealth advisers, (5) peer communication and support, and (6) follow-up/review reminders. The outcome evaluations of the 13 included studies were grouped into 4 categories: (1) technology usability and patient satisfaction, (2) self-management of symptoms and patient-reported outcome–related indicators, (3) adherence, and (4) health-related quality of life.

Conclusions: A limited number of studies have investigated the use of mHealth technology in the postdischarge self-management of patients with HNC. The existing literature suggests that mHealth technology can effectively assist patients with HNC in self-management and postdischarge interventions. It plays an important role in addressing patients' health information needs, reducing both their somatic and psychological burdens, and improving their overall quality of life. Future research should prioritize conducting additional high-quality RCTs to evaluate the usability and analyze the cost-effectiveness of mHealth technology.

(JMIR Mhealth Uhealth 2023;11:e49051) doi:10.2196/49051

KEYWORDS
head and neck cancer; mobile health technology; postdischarge; self-management; rehabilitation

Introduction

Background

Malignant head and neck tumors, one of the most common malignancies, are classified according to the site of tumor origin into oral cavity, nasopharyngeal, oropharyngeal, and laryngeal cancers. Among these, head and neck squamous cell carcinoma (HNSCC) accounts for >90% of all head and neck tumors. The latest statistics of the International Agency for Research on Cancer (IARC) GLOBOCAN on global cancer incidence and mortality in 2020 revealed 890,000 new cases of HNSCC, with 450,000 resulting deaths annually [1]. The incidence of HNSCC is continuing to rise and is expected to increase by 30% by 2030, reaching an estimated 1.08 million new cases of HNSCC per year [1].

Surgical resection, radiation, chemotherapy, targeted therapy, and a combination of these therapies are the available treatment options for head and neck cancer (HNC). Patients with various tumor stages and locations receive individualized treatment approaches [2]. However, both surgical treatment and radiotherapy can result in various types and intensities of complications, which can negatively affect patients' somatic function, outward appearance, and psychological well-being. For example, physical dysfunctions, such as swallowing disorders, mouth-opening issues, and shoulder syndrome, as well as psychological difficulties, such as social and workplace reintegration due to an altered outward appearance, can result from both surgical treatment and radiotherapy [3].

Studies have shown that patients with HNC frequently experience increased functional impairment and negative side effects after surgery or radiotherapy. Short-term home rehabilitation after discharge is crucial for enhancing patients' function and long- and short-term quality of life (QoL) because brief rehabilitation care treatment during hospitalization is insufficient to assist patients in achieving full recovery. Although patients have continuous access to medical care and guidance from doctors and nurses while they are in the hospital, they must assume responsibility for their own functional rehabilitation and self-care after discharge, whether they return home or move to other facilities, such as community nursing homes. Most patients and their family caregivers lack a medical background, and despite receiving necessary verbal health education or health information booklets by doctors and nurses before discharge, forgetfulness inevitably occurs over time. As a result, survivors of HNC may continue to experience significant debilitating issues with swallowing, speech, hearing, and psychological effects due to loss of function and changes in their body image as a result of treatment. Survivors of HNC often experience a lower QoL [4-6]. Studies have shown that their QoL is lower compared to survivors of other cancer types [7]. Therefore, improving the ability of patients with HNC to manage their own care after leaving hospital is a challenging and complex research subject.

With the increasing application and popularity of mobile health (mHealth) technology, numerous studies have shown that mHealth technology has the potential to assist patients with cancer and other chronic diseases in self-management [8,9]. mHealth is defined as "medical and public health practice through the use of mobile devices such as mobile phones, patient monitoring devices, personal digital assistants (PDAs) and other wireless devices" [10]. Currently, mHealth technology is used in a variety of devices and formats, including telephones, mobile phones apps for calls or videos, web-based platforms, and tablets, which are more commonly used in home environments. Although existing studies have systematically evaluated the use of mHealth technology in patients with HNC, the primary focus of these studies is not on patients' self-management after hospital discharge. The included studies mainly include randomized controlled trials (RCTs), excluding other relevant studies [11]. Therefore, this study aimed to explore how patients can use mHealth technology for self-management and functional rehabilitation, along with assessing the associated outcome indicators, in both the short and the long term, following discharge from treatment.

Objectives

This study systematically reviewed the use of mHealth technology in the postdischarge self-management of patients with HNC. The review centered on 2 sections: intervention and outcome evaluation. The aims of this study were (1) to summarize the categories of mHealth interventions and their main types of functions/services for the postdischarge self-management of patients with HNC through a systematic review of the existing literature and (2) to examine how these mHealth interventions are evaluated and the differences that exist between outcome indicators across studies.

Methods

Study Design

This study used the 5-stage methodological framework outlined by Arksey and O'Malley [12] to define the scope of the review: (1) identifying the research question, (2) identifying relevant studies, (3) selecting relevant papers for the review, (4) charting the data, and (5) collating, summarizing, and reporting the results.

Stage 1: Identifying the Research Question

The study population included adult patients with HNC who had been discharged from the hospital after surgical treatment or radiotherapy and were recovering at home or in the community. The type of intervention involved the use of mHealth technology. The research questions were developed based on an initial literature search and further refined through iterative discussions within the research team. The research questions were as follows: (1) What mHealth technologies exist to support patients with HNC after hospital discharge? (2) How are these mHealth technologies used to implement interventions? (3) How are mHealth interventions evaluated?

Stage 2: Identifying Relevant Studies

A systematic search strategy was used to identify the literature related to the research questions. We combined the keywords “mobile health technology” and “head and neck cancer” according to the patient/problem, intervention, comparison, and outcome (PICO) principles of literature search, which identified the 2 main concepts of the research questions and summarized the subject terms and free words related to these 2 main concepts. Systematic searches were conducted in the PubMed, Web of Science, Embase, and Cumulative Index to Nursing & Allied Health (CINAHL) databases. Two independent researchers (authors LYF and CWH) searched the databases for references to identify papers published between the time of database creation and March 1, 2023. The search terms and strategy used are presented in Multimedia Appendix 1.

Stage 3: Selecting Relevant Papers for the Review

The inclusion and exclusion criteria for this review are presented in Textbox 1. The citations obtained from each database were imported into Endnote Reference Manager for bibliographic analysis. Duplicate papers were excluded. Each level was assessed by 2 reviewers, who independently considered studies based on the inclusion and exclusion criteria. The first screening process involved reviewing the titles and abstracts to make the following decisions: (1) if at least 1 reviewer agreed with the inclusion criteria or found the abstract or title inconclusive, the study was moved to the second level of screening, and (2) if both reviewers agreed to exclude a study, the study was excluded. Two independent reviewers evaluated the full texts at the second level of screening. Any disagreements between the reviewers were resolved through discussion or by a third reviewer.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- Adult patients including both males and females (≥18 years old)
- Patients with head and neck cancer (HNC) who have been discharged from the hospital to recover in a nonmedical setting, such as their homes or communities, after undergoing at least 1 surgical or radiotherapy treatment
- Use of mobile health (mHealth) technology to implement interventions
- English papers only
- Contains at least 1 quantitative result
- Selection of a paper by the same research team (same app) on a specific system for inclusion in the analysis

Exclusion criteria:

- Participants aged <18 years (pediatric, adolescent)
- Integrated interventions that do not use mHealth technology alone but in other collaborative ways, such as multidisciplinary cooperation
- Full-text documents not available, such as conference abstracts or protocols, as well as review papers
- No specific outcome evaluation indicators related to HNC

Stage 4: Charting the Data

We collaborated to develop a graphical form of the data and to identify variables to be extracted. Descriptive graphical information included (1) a general description of the paper (first author and year, study design, study site, and patient population) and (2) intervention-specific information (purpose of the intervention and mHealth app, key features, delivery methods, duration and follow-up period, data collected, outcomes measured, and findings); see Multimedia Appendix 2.

Stage 5: Collating, Summarizing, and Reporting the Results

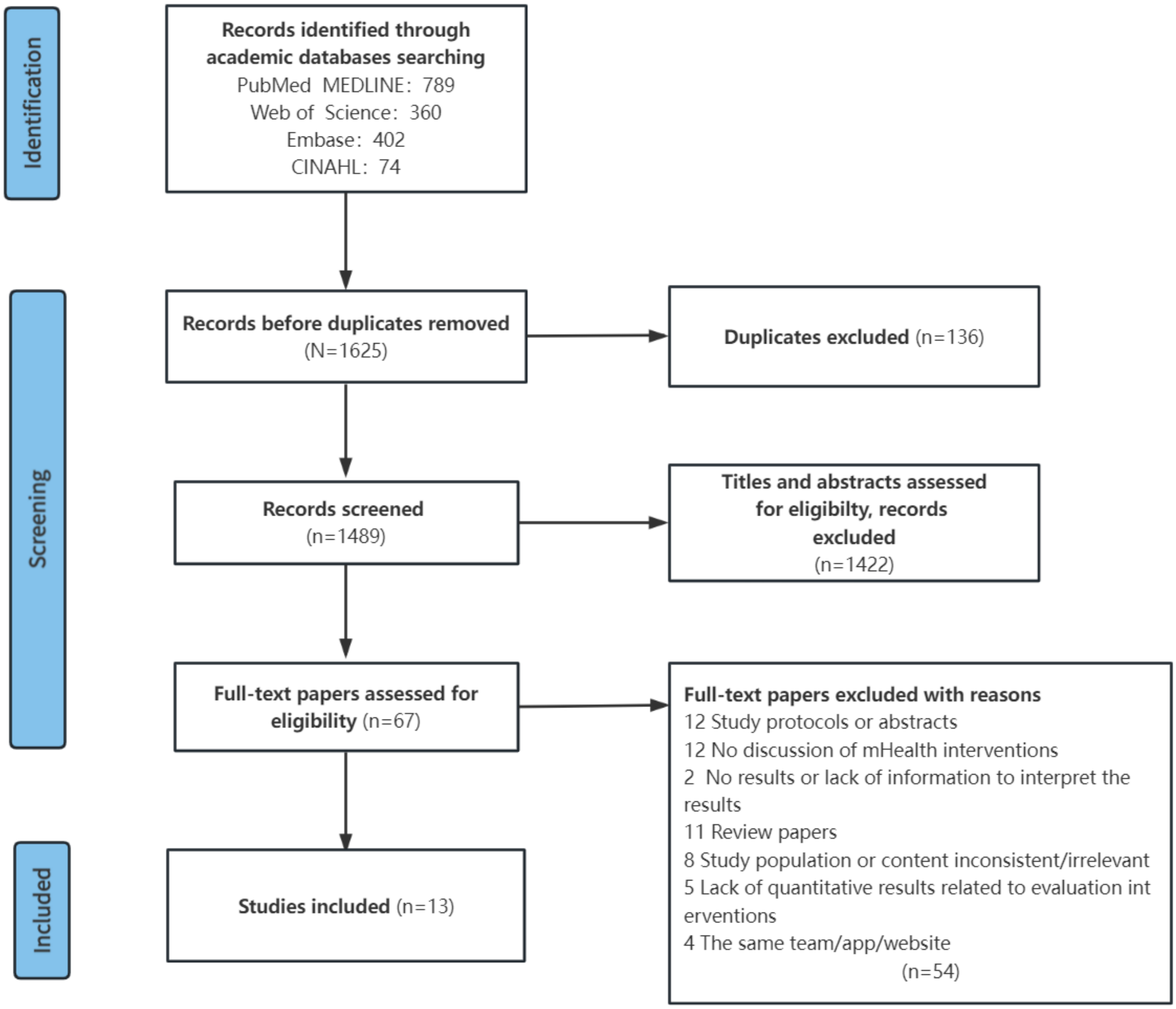
General descriptions of review papers were collated according to the descriptive characteristics. Following a concurrent review of the chart data, we conducted a thematic content analysis of the interventions and associated outcomes for each study. First, codes were developed and applied to analyze the data. The coding segments for all chart data were created using color-coded quotes. The code summaries were organized in a Microsoft Excel table for thematic content analysis. The table was sorted by code and density, looking for repeating patterns addressed by the included papers, including a comparison of the studies across the data set and within each study until key themes were identified. The results of this study summarize the objectives.

Results

Search Results

In total, 1625 papers were retrieved, of which 1489 (91.6%) different papers remained after removing duplicate data. Figure 1 displays the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the retrieved literature, the level of screening, and the included studies. Finally, 13 (0.9%) studies, projects, or reports were included in this review.

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



General Characteristics of the Included Studies

Of the 13 studies included in this review, RCTs (n=6, 46.2%) were the most common type [13-18], followed by 2 (15.4%) cohort studies [19,20] and 5 (38.5%) other types (n=4, 80%, quasi-experimental studies and n=1, 20%, mixed study) [21-25]. The quality of the included studies was assessed using the JBI critical appraisal tools, and detailed information is provided in Multimedia Appendix 3. The general characteristics of the included studies are presented in Table 1.

Table 1. General characteristics of the included studies (N=13).

Characteristics	Studies, n (%)
Study design	
RCT[a]	6 (46.2)
Cohort study	2 (15.4)
Others	5 (38.5)
Origin of study	
United States	5 (38.5)
Netherlands	2 (15.4)
Taiwan	3 (23.1)
China	1 (7.7)
Germany	1 (7.7)
Australia	1 (7.7)
Year of publication	
2007-2019	4 (30.8)
2020-2023	9 (69.2)

[a]RCT: randomized controlled trial.

Types of mHealth Interventions Currently Being Implemented

Current mHealth interventions and management services for the population with HNC postdischarge include 6 main categories: (1) symptom monitoring and assessment reports (n=8, 61.5% [15,16,19,20,22-25]), (2) home rehabilitation training (n=4, 30.8% [13,16-18]), (3) medical health information access (n=7, 53.8% [13,14,16,17,22-24]), (4) telehealth support (n=7, 53.8% [14,18-21,23,25]), (5) peer-to-peer communication (n=2, 15.4% [23,24]), and (6) follow-up/review reminders (n=2, 15.4% [14,22]). These mHealth interventions were delivered through various technology platforms, including smartphones (n=9, 69.2%), personal digital assistants (PDAs)/tablets (n=2, 15.4%), web-based platform (n=1, 7.7%), and home monitoring or telemonitoring unit or telemetry system (n=1, 7.7%). The results are summarized in Table 2.

Table 2. mHealth[a] intervention tools and categories.

Characteristics	Studies (N=13), n (%)
mHealth intervention tools	
Smartphone	9 (69.2)
Tablet or computer	2 (15.4)
Web-based portal	1 (7.7)
Home monitoring or telemonitoring unit	1 (7.7)
mHealth intervention categories	
Self-monitoring and report	8 (61.5)
Home practice	4 (30.8)
Provision of health information	7 (53.8)
Telemedical support	7 (53.8)
Communication community	2 (15.4)
Follow-up/review reminders	2 (15.4)

[a]mHealth: mobile Health.

Intervention evaluation indicators for patients with HNC adopting mHealth technology mostly included general patient characteristics, QoL indicators, adherence, symptom self-reporting, patient satisfaction, and evaluation of technology usability. The thematic content analysis identified 4 main themes for outcome evaluation: (1) technology usability and patient satisfaction, (2) indicators related to self-management of symptoms and patient-reported outcomes (PROs), (3) adherence, and (4) health-related quality of life (HRQoL).

This study found that mHealth technology plays a more beneficial role in the postdischarge management of patients with HNC. The mHealth program helped improve physical and psychological symptoms after discharge, enabled patients to gain more knowledge about health care, and enhanced their self-management of behaviors and functional rehabilitation exercises after leaving the hospital. Patients in these studies showed high levels of satisfaction with and acceptance of mHealth technologies.

Self-Management of Symptoms and PRO-Related Indicators

Of the 13 included studies, 8 (61.5%) focused on the impact of mHealth technology on symptoms and the validity of outcome reporting.

Of these 8 (61.5%) studies, 2 (25%) discussed the impact of mHealth technology on the effectiveness of PROs. Ma et al [20] found that using a fully automated interactive chatbot to provide support and assist patients in self-reporting outcomes during the postradiotherapy period for HNC treatment resulted in good consistency between PROs and clinician-reported outcomes (CROs). According to the study results [20], 61% of the patients felt that the chatbot helped with symptom self-management and reduced the need to call the care team, demonstrating that mHealth technology can be effective in helping patients with self-reporting outcomes.

Due to the significant symptom burden associated with HNC, mHealth technology may be beneficial in assisting patients with symptom management. Van Cleave et al [19] developed a web-based platform, Electronic Patient Visit Assessment (ePVA), to explore patients' reports of symptoms and functional limitations. ePVA can inquire about and document 21 common symptoms and functional limitations associated with HNC and generate an ePVA report. The HNC care team members can use the ePVA report to implement real-time clinical interventions based on their clinical judgment and the patients' knowledge. The study phase results showed that patients with the most symptoms and functional limitations reported a significantly poorer HRQoL, demonstrating that ePVA may be a proven mHealth tool that can be used as a real-time intervention for patient reporting.

Next, 6 (75%) studies discussed the effect of mHealth technology on patients' symptom self-management. The mobile app SwallowIT was tailored to provide telepractice-assisted therapy to patients with swallowing disorder during radiotherapy and after HNC treatment to support home practice of the pharyngeal program [13]. The program provides instructional videos, images, and text descriptions for each exercise, allowing patients to record the number of motor repetitions and the perceived effort as they complete each exercise. Results of the study during the outcome evaluation phase showed no significant differences between the conventional and experimental groups in terms of swallowing, nutrition, or functional measures. However, more patients (76%) preferred using SwallowIT to receive instructions compared to instructing themselves during exercises.

Lin et al [24] developed an oral care mobile app to provide medical health information and oral mucosal care advice to patients during radiotherapy for HNC. The results of the study showed that the Patient-Generated Subjective Global Assessment scores of the group using mHealth technology were significantly lower than those of the control group at all 3 nodes during the intervention. Additionally, the severity of oral mucositis grading was also significantly lower in all groups. These findings indicate that the use of mobile apps is effective in improving the nutritional status and reducing the side effects in patients with HNC treated with concurrent radiotherapy.

The mobile app Oncokompas, developed by Dutch academics, can support symptom self-management by providing medical information and a personalized overview of supportive care options aimed at reducing the symptom burden among survivors of cancer [16]. The results of the study indicated that the mHealth technology group showed statistically significant improvements in the clinical course and symptom burden scores for oral pain, social eating, swallowing, cough, and dental symptoms compared to the control group.

Wang et al [18] used telephone calls to help patients with oral cancer after radical surgery perform functional exercises for their trismus, monitor training progress, and obtain appropriate feedback. After the 12-week intervention period, the change in the maximum interincisional opening was 10.30 mm (95% CI 8.22-12.37) greater in the experimental group than in the active control group. The change in the mandibular functional impairment score was –0.36 (95% CI –0.44 to –0.28) greater in the experimental group than in the active control group. This study provides evidence supporting the effectiveness of the intervention program in reducing dental and mandibular functional impairments in patients undergoing radical oral cancer surgery.

Regarding complication management, an RCT [14] with patients with nasopharyngeal cancer who were discharged from the hospital after radiotherapy used a mobile app to provide the patients with information about the disease, including observation and treatment of radiotherapy complications and reminders for regular review. The aim was to enhance the self-management skills of the patients, enabling them to cope effectively with radiotherapy complications [14]. The results of this study showed that the incidence of oral mucositis, dry mouth syndrome, difficulty in opening the mouth, and nasal congestion was significantly lower in the intervention group than in the control group 5 months after discharge, suggesting that the intervention can effectively help patients improve their ability to cope with and manage complications and reduce their occurrence. In contrast, another retrospective controlled trial [25] was conducted in which patients were followed up over the phone within 72 h of discharge as well as by wound visits to answer their questions in order to reduce emergency department visits and readmission rates. The results of this study showed a statistically significant reduction in emergency department visits compared to the previous year, with no change in readmission rates, demonstrating the potential of telephone interventions in the early postoperative period to reduce unnecessary emergency department visits.

Technology Availability and Patient Satisfaction

Of the 13 included studies, 4 (30.8%) investigated technology usability and patient satisfaction in outcome evaluations. Patients demonstrated higher satisfaction with the mHealth technologies in all study outcomes.

In terms of technology usability, software usability refers to the extent to which a particular user can use a product in a specific context and achieve a particular goal effectively, efficiently, and satisfactorily. Patient satisfaction with a fully automated interactive chatbot [20] reached 89%, with 83% of patients finding it easy to use, 79% feeling confident in using the chatbot, 71% finding the chatbot functionality well integrated, and 86% feeling they did not need additional training or technical support (80%) to use the chatbot.

In the SwallowIT study [13], 76% of patients preferred SwallowIT or clinician guidance. In a study [21] on coping with somatic imagery, 89% of patients were very satisfied with BRIGHT, found its use effective, and would recommend it to other survivors of HNC. In a quasi-experimental study in Taiwan [22], the acceptability score of the mHealth app significantly improved (*P*<.05) in terms of the intention to use, perceived usefulness, and ease of use after 3 months of the intervention.

These data show that patients' use of and familiarity with mHealth apps reduce their uncertainty and improve their acceptance of the new technology. This, in turn, can promote better adoption and use of mHealth technology in the future and enhance its overall usability.

Health-Related Quality of Life

Most studies (n=9, 69.2%) measured and evaluated the HRQoL. Of these, 6 (66.7%) studies [14,19,21-24] used the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC-QLQ-C30) and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 35 and Head and Neck Module (EORTC-QLQ-H&N35), 2 (22.2%) studies [13,15] used the Functional Assessment of Cancer Therapy—Head and Neck (FACT-H&N) scale, 2 (22.2%) studies [16,19] used the HRQoL scale, and 1 (11.1%) study [17] used the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire and Provisional 25-item Information Module (EORTC-QLQ-INFO-25) to assess the QoL of survivors of HNC.

There were mixed results regarding the improvement in the QoL with mHealth technology in the studies. The mobile app SwallowIT was not associated with statistically significant improvements in the final overall QoL [13]. However, BRIGHT was associated with improvements in the social eating disorder and social contact disorder substructural domains of the EORTC-QLQ-H&N35 [21]. Additionally, a greater number of reported symptoms and functional limitations were associated with lower EORTC-QLQ-C30 overall QoL health scores [19]. Furthermore, 5 (55.6%) studies demonstrated higher overall improvements in the QoL in the experimental group using mHealth technology than in the control group [14,15,22-24], and 1 (11.1%) study reported that increased adherence was associated with a better patient-reported QoL [17].

Intervention Adherence

Of the 13 included studies, 5 (38.5%) reported on adherence to the use of mHealth technology. In the SwallowIT app [13], it was found that adherence decreased significantly over time for the entire cohort, with low adherence at 6 weeks in all groups (27%). However, adherence was relatively high in the group with clinician guidance and the group using SwallowIT [13].

The results of an RCT [14] showed that adherence to mouth-opening exercises and nasal rinsing was higher in the intervention group than in the control group at 3 and 6 months postdischarge (*P*<.05). However, changes in adherence within groups were not elaborated. The researchers concluded that mHealth technology can provide pictures and videos to be viewed repeatedly, and all these activities can be performed with the help of relevant videos as a means to improve patient adherence.

The ePVA study [19] showed that 59 of 64 (92.2%) patients completed the ePVA, with 1 or more follow-up visits within the 6-month study window, with high completion rates and adherence. The investigators concluded that the high completion rate results were related to the study team accommodating patients' needs regarding completion times during treatment, such as allowing patients to delay completion until a convenient time.

The HNC Virtual Coach study [17] conducted as a pre-radiotherapy prophylactic swallowing rehabilitation demonstrated that 80% of patients used the app and over 50% completed at least 1 swallowing exercise per day, with better, although not statistically significant, adherence in the experimental group. Although adherence declined in both groups during radiotherapy, the results showed that higher adherence was associated with a better patient-reported QoL.

Additional support for the intervention exercise program for patients with restricted mouth opening was provided by remote telephone support [18], which showed that at week 12, the experimental group had 299.67 minutes (95% CI 223.44-357.89) more intervention exercise time than the control group. From baseline to week 12, the change in the maximum interincisional opening was 10.30 mm (95% CI 8.22-12.37) greater in the experimental group than in the active control group. This demonstrates that the use of mHealth technology can help enhance patient compliance with intervention programs and, in turn, improve functional impairment outcomes.

Discussion

Principal Findings

In this review, 13 papers were analyzed. Some studies have indicated that patients with HNC frequently experience increased functional impairment after surgery. Brief rehabilitation during hospitalization may not be adequate to help patients recover, while long-term home-based rehabilitation after discharge plays a crucial role in improving patients' function and long- and short-term QoL. Therefore, this study focused on assessing how patients use mHealth technology for self-management, functional rehabilitation, and related outcome evaluation indicators after hospital discharge.

This scoping review summarized the existing literature on mHealth technology for the postdischarge self-management of patients with HNC and the intervention tools, intervention methods, and types of outcome evaluations of mHealth technology in the currently available literature, following a generalization and thematic summary.

Advantages of mHealth Technology in Postdischarge Self-Management of Patients With HNC

mHealth technology can accelerate patient communication, facilitate home monitoring and self-management, and improve overall patient health. The use of mHealth in the postdischarge self-management of patients with HNC addresses many issues and offers several advantages, including the following:

- Real-time monitoring and tracking: Although patients are the best recorders of their daily health experiences, clinicians and others are unlikely to have full access to patients in their living environments after hospital discharge. Consequently, capturing timely information about patients' actual experiences, health status, and outcomes can be challenging. mHealth apps can help patients monitor and track their health in real time. Using sensors and devices, patients can measure physiological metrics (eg, heart rate, respiration, blood pressure, and weight) and record them. These data can be shared with doctors to better understand their condition and make necessary adjustments to the treatment plans. Meanwhile, because PROs have become the focus of health care research, scholars have begun to explore whether mHealth technology can be an effective tool for PROs. Findings have shown that the use of mHealth technology can help patients with daily symptom recording and outcome reporting. Furthermore, PROs have demonstrated strong concordance with CROs [20].
- Personalized health management: mHealth technology can provide a platform for patients and their families to obtain different health services in the face of different needs, as well as better intervention and management of patients after they leave the hospital. mHealth apps can provide personalized health management solutions tailored to individual patient differences and needs. The apps can provide customized dietary advice, exercise plans, medication reminders, follow-up reminders, and other functions according to the patient's condition and treatment plan, thereby helping them effectively manage their health.
- Education and information resources: While patients are receiving treatment within a hospital, doctors and nurses can provide a steady stream of health care support, including daily treatment, health information, and guidance. In contrast, once patients are discharged, although doctors and nurses may initially explain the key points of posthospitalization rehabilitation and care, the passage of time and various factors, both subjective and objective, such as patient compliance, can make it challenging for patients to effectively self-manage their physical and psychological well-being. mHealth apps can provide patients with health education and the necessary information resources they require anytime, anywhere. Patients can access professional medical knowledge about diseases, treatment options, medication, and more through these apps while they are at home, helping them better understand their condition, treatment options, recovery methods, etc.
- Social and psychological support: Due to factors related to the disease and treatment modalities, patients with HNC usually face issues concerning an altered body image, which may also affect their ability to resume work and engage in social activities. Therefore, social and psychological support for these patients requires careful consideration. However, this is a long-term process that cannot be fully addressed and promptly implemented during hospitalization. mHealth apps connect patients with other patients or health care professionals to provide social and psychological support. Patients can use these apps to share their experiences and offer advice to other patients. Moreover, they can have online consultations and communications with health care professionals, thus reducing anxiety and the feeling of isolation.
- Improving adherence: Undoubtedly, for doctors and nurses, managing and providing follow-up care for discharged patients pose a significant challenge. The brief, 1-time health education at the time of discharge makes it difficult to address the various medical, health, and psychosocial difficulties encountered by patients during home rehabilitation and to help patients establish good self-management ability and compliance with home treatment. Patients who leave the hospital often experience problems during the follow-up phase, including difficulty in communication, difficulty in management, poor adherence, suboptimal recovery, and a poor QoL. mHealth apps, however, can provide medication reminders, treatment plan tracking, and other functions to help patients better follow doctors' advice and treatment plans.

These findings suggest that novel mHealth technology is more likely to be welcomed by the patient population than traditional forms of follow-up and that patients who use mHealth technology show better adherence [13,14,17-19]. This suggests that mHealth technology may be an effective tool to provide follow-up and enhance the self-management of postdischarge patients in the future.

Overall, the advent of mobile technology has transformed the health ecosystem by changing the way individuals communicate and providing patients and health care providers with a wide range of supportive tools to monitor and manage health information, thereby facilitating better health care delivery. The advantages of mHealth in the home self-management of patients with HNC after discharge from the hospital include real-time monitoring, personalized health management, access to educational and information resources, social and psychological support, and improved treatment adherence. These benefits can help patients more effectively manage their health, improve their QoL, and work closely with their health care teams.

Limitations of mHealth Technology in Postdischarge Self-Management of Patients With HNC

Although most studies have demonstrated the positive impact of mHealth technology on the postdischarge self-management of patients with HNC, some have shown no significant improvement in patient self-management with respect to other outcome indicators [16]. This may be because mHealth technology is just beginning to be applied to the population with HNC and the amount of available literature on its application is limited. Although the number of RCTs among the available studies is relatively high, the research questions and outcome indicators vary significantly across these studies. Therefore, there is a scarcity of directly comparable studies in terms of outcome evaluation.

In addition to physical and psychological outcome indicators, few studies have examined standard technical usability evaluations and economic efficiency indicators. Wall et al [26] reported associated health service costs (service time, consumables, treatment resources), patient-related costs (travel and wages), and patient-related HRQoL statistics. Their study resulted in a comparison of the total costs of different forms of mHealth interventions, revealing significant cost savings for health care services and consumers with HNC, while achieving comparable HRQoL outcomes at a lower cost [26]. However, apart from this study, there is limited existing literature that reports on the social and economic benefits of using mHealth technology for both patients and hospitals. Various factors, such as software development and operational costs, equipment use and maintenance costs, actual changes in patient financial stress, and the overall social and economic benefits to hospitals, are areas where further data and research are needed. Such research is essential to demonstrate the technological advantages of mHealth and the associated individual societal benefits generated.

As the number and scale of mobile apps increase with the increasing use of mHealth technology, more attention should be paid to usability evaluation throughout the software development process. Timely usability evaluation before, during, and after development is an important measure for understanding user needs, improving the design of software features, and improving user experience and satisfaction. However, this is not well described in existing research.

At the same time, we must also consider the impact of mHealth technology on doctors and nurses, even as it brings convenience to patients. Questions regarding whether the use of mHealth technology consumes additional time for health care professionals, whether it increases their workload beyond their regular work, whether health care professionals are satisfied with their mHealth technology experience, and how hospital managers reconcile conflicts and contradictions arising from this technology are all aspects that have received less attention in existing research. Therefore, future research should explore these areas to achieve a comprehensive understanding.

Finally, the issues of data privacy and security should be given sufficient attention. mHealth apps handle large amounts of personal health data; therefore, data privacy and security are important concerns. Ensuring the security and privacy of patient data is challenging, and appropriate security measures must be adopted to protect such data.

Implications for Future Research

By addressing the limitations and research weaknesses of this study, future research can advance to explore the effectiveness of mHealth technology in the postdischarge self-management of patients with HNC. This can be achieved through rigorous methods, such as conducting an RCT with a robust study design, expanding the scope of the study and the number of users, and adopting scientific theoretical guidance for the selection and evaluation of outcome indicators.

Hotspots for future research will also focus on the application of artificial intelligence in mHealth, the integration of virtual reality and augmented reality technologies in oncology treatment, and the development of remote monitoring and diagnosis based on mobile devices. Future research will also need to overcome various challenges, including issues related to data privacy and security, technology usability and ease of use, data analysis and use, and evaluation of the effectiveness and efficacy of mHealth apps. Addressing these challenges requires interdisciplinary collaboration and sustained research efforts.

Limitations

This study was a scoping review conducted to broadly include and summarize the existing literature. In addition, this study only included studies published in English, and studies published in other languages were excluded from the discussion.

Conclusion

mHealth technology has been applied to the postdischarge self-management of patients with HNC. The main interventions in mHealth technology for improving postdischarge self-management include symptom monitoring and reporting, functional rehabilitation training, access to health care information, telehealth service support, peer-to-peer communication, and follow-up/review reminders. Outcome evaluations of the use of mHealth technology were discussed in terms of technology usability and patient satisfaction, indicators related to self-management of symptoms, PROs, adherence, the HRQoL, and the impact on somatic/psychological aspects, with most studies showing a positive impact.

Therefore, based on the limited research data available to date, mHealth technology can effectively help patients with HNC in self-management and postdischarge interventions. This plays an important role in meeting patients' health information needs, reducing their somatic and psychological burdens, and enhancing their QoL. Future research should aim to conduct higher-quality RCTs for usability evaluation and cost–economic benefit analysis.

Acknowledgments

This study was supported by the Shanghai Association for Science and Technology (21002411300), the Innovation Research Team of High-level Local Universities in Shanghai (SHSMU-ZDCX20212802), and the establishment of a clinical multicenter oncology integrated Chinese and Western medicine nursing database (Hlgy1901sjk).

Authors' Contributions

LYF designed the study, developed the search strategy, performed the literature search, analyzed the data, and was the lead author of the manuscript. CWH oversaw the process, codeveloped the search strategy, provided guidance on the development of inclusion and exclusion criteria, discussed the findings, and contributed to the editing of the manuscript. LYJ and YL participated in the codevelopment of the search strategy and analyzed and discussed the findings. HLL provided guidance on reporting the study results; contributed to the analysis, methods, and discussion; and edited the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and search statement.
[DOCX File , 21 KB - mhealth_v11i1e49051_app1.docx]

Multimedia Appendix 2
Overview of studies.
[DOCX File , 28 KB - mhealth_v11i1e49051_app2.docx]

Multimedia Appendix 3
Quality assessment of included studies.
[DOCX File , 24 KB - mhealth_v11i1e49051_app3.docx]

Multimedia Appendix 4
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
[PDF File (Adobe PDF File), 102 KB - mhealth_v11i1e49051_app4.pdf]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May 04;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. Mayo Clin Proc 2016 Mar;91(3):386-396. [doi: 10.1016/j.mayocp.2015.12.017] [Medline: 26944243]
3. Parke SC, Langelier DM, Cheng JT, Kline-Quiroz C, Stubblefield MD. State of rehabilitation research in the head and neck cancer population. functional impact vs. impairment-focused outcomes. Curr Oncol Rep 2022 Apr;24(4):517-532. [doi: 10.1007/s11912-022-01227-x] [Medline: 35182293]
4. Jacot W, Colinet B, Bertrand D, Lacombe S, Bozonnat M, Daurès JP, OncoLR Health Network. Quality of life and comorbidity score as prognostic determinants in non-small-cell lung cancer patients. Ann Oncol 2008 Aug;19(8):1458-1464 [FREE Full text] [doi: 10.1093/annonc/mdn064] [Medline: 18356134]
5. Blazeby J, Brookes S, Alderson D. Prognostic value of quality of life scores in patients with oesophageal cancer. Br J Surg 2000 Mar;87(3):362-373. [doi: 10.1046/j.1365-2168.2000.01383-8.x] [Medline: 10718971]
6. Thompson TL, Pagedar NA, Karnell LH, Funk GF. Factors associated with mortality in 2-year survivors of head and neck cancer. Arch Otolaryngol Head Neck Surg 2011 Nov;137(11):1100-1105. [doi: 10.1001/archoto.2011.179] [Medline: 22106233]
7. Ganz PA, Lee JJ, Siau J. Quality of life assessment. An independent prognostic variable for survival in lung cancer. Cancer 1991 Jun 15;67(12):3131-3135. [doi: 10.1002/1097-0142(19910615)67:12<3131::aid-cncr2820671232>3.0.co;2-4] [Medline: 1710541]
8. Ma Y, Zhao C, Zhao Y, Lu J, Jiang H, Cao Y, et al. Telemedicine application in patients with chronic disease: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2022 Apr 19;22(1):105 [FREE Full text] [doi: 10.1186/s12911-022-01845-2] [Medline: 35440082]

9. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 09;19(1):e11 [FREE Full text] [doi: 10.2196/jmir.6575] [Medline: 28069561]
10. World Health Organization. Global Diffusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016.
11. Caputo MP, Rodriguez CS, Padhya TA, Mifsud MJ. Telehealth interventions in head and neck cancer patients: a systematic review. Cancer Nurs 2023 Dec 01;46(5):E320-E327. [doi: 10.1097/NCC.0000000000001130] [Medline: 37607382]
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
13. Wall LR, Ward EC, Cartmill B, Hill AJ, Isenring E, Byrnes J, et al. Prophylactic swallowing therapy for patients with head and neck cancer: a three-arm randomized parallel-group trial. Head Neck 2020 May;42(5):873-885. [doi: 10.1002/hed.26060] [Medline: 31903689]
14. Di R, Li G. Use of a smartphone medical app improves complications and quality of life in patients with nasopharyngeal carcinoma who underwent radiotherapy and chemotherapy. Med Sci Monit 2018 Sep 04;24:6151-6156 [FREE Full text] [doi: 10.12659/MSM.908146] [Medline: 30178780]
15. Pfeifer MP, Keeney C, Bumpous J, Schapmire TJ, Studts JL, Myers J, et al. Impact of a telehealth intervention on quality of life and symptom distress in patients with head and neck cancer. J Community Support Oncol 2015 Jan;13(1):14-21 [FREE Full text] [doi: 10.12788/jcso.0101] [Medline: 25839061]
16. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GAP, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
17. Starmer HM, Klein D, Montgomery A, Goldsmith T, McCarroll L, Richmon J, et al. Head and neck virtual coach: a randomized control trial of mobile health as an adjunct to swallowing therapy during head and neck radiation. Dysphagia 2023 Jun;38(3):847-855. [doi: 10.1007/s00455-022-10506-5] [Medline: 35960394]
18. Wang T, Su J, Leung K, Liang S, Wu S, Wang H. Effects of a mouth-opening intervention with remote support on adherence, the maximum interincisal opening, and mandibular function of postoperative oral cancer patients: a randomized clinical trial. Eur J Oncol Nurs 2019 Jun;40:111-119. [doi: 10.1016/j.ejon.2019.04.001] [Medline: 31229200]
19. Van Cleave JH, Fu MR, Bennett AV, Concert C, Riccobene A, Tran A, et al. The usefulness of the Electronic Patient Visit Assessment (ePVA) as a clinical support tool for real-time interventions in head and neck cancer. Mhealth 2021;7:7 [FREE Full text] [doi: 10.21037/mhealth-19-250] [Medline: 33634190]
20. Ma D, Orner D, Ghaly MM, Parashar B, Ames JW, Chen WC, et al. Automated health chats for symptom management of head and neck cancer patients undergoing radiation therapy. Oral Oncol 2021 Nov;122:105551. [doi: 10.1016/j.oraloncology.2021.105551] [Medline: 34700280]
21. Graboyes EM, Maurer S, Park Y, Marsh CH, McElligott JT, Day TA, et al. Evaluation of a novel telemedicine-based intervention to manage body image disturbance in head and neck cancer survivors. Psychooncology 2020 Dec;29(12):1988-1994 [FREE Full text] [doi: 10.1002/pon.5399] [Medline: 32350999]
22. Wang T, Huang R, Yang S, Chou C, Chen L. Evaluating the effects of a mobile health app on reducing patient care needs and improving quality of life after oral cancer surgery: quasiexperimental study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18132 [FREE Full text] [doi: 10.2196/18132] [Medline: 32716303]
23. van den Brink JL, Moorman PW, de Boer MF, Hop WCJ, Pruyn JFA, Verwoerd CDA, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007;14(2):198-205 [FREE Full text] [doi: 10.1197/jamia.M2199] [Medline: 17213498]
24. Lin T, Wang Y, Huang C. Effects of a mobile oral care app on oral mucositis, pain, nutritional status, and quality of life in patients with head and neck cancer: a quasi-experimental study. Int J Nurs Pract 2022 Aug;28(4):e13042. [doi: 10.1111/ijn.13042] [Medline: 35224811]
25. Shah M, Douglas J, Carey R, Daftari M, Smink T, Paisley A, et al. Reducing ER visits and readmissions after head and neck surgery through a phone-based quality improvement program. Ann Otol Rhinol Laryngol 2021 Jan;130(1):24-31. [doi: 10.1177/0003489420937044] [Medline: 32567391]
26. Wall LR, Kularatna S, Ward EC, Cartmill B, Hill AJ, Isenring E, et al. Economic analysis of a three-arm RCT exploring the delivery of intensive, prophylactic swallowing therapy to patients with head and neck cancer during (chemo)radiotherapy. Dysphagia 2019 Oct;34(5):627-639. [doi: 10.1007/s00455-018-9960-1] [Medline: 30515560]

Abbreviations

CRO: clinician-reported outcome
EORTC-QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30

EORTC-QLQ-H&N35: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 35 and Head and Neck Module
ePVA: Electronic Patient Visit Assessment
HNC: head and neck cancer
HNSCC: head and neck squamous cell carcinoma
HRQoL: health-related quality of life
mHealth: mobile health
PRO: patient-reported outcome
QoL: quality of life
RCT: randomized controlled trial

Edited by L Buis; submitted 16.05.23; peer-reviewed by E Ward, R Di; comments to author 10.07.23; revised version received 14.08.23; accepted 03.10.23; published 23.10.23.

Please cite as:
Li Y, Chen W, Liang Y, Yang L, Hou L
Evaluation of Mobile Health Technology Interventions for the Postdischarge Management of Patients With Head and Neck Cancer: Scoping Review
JMIR Mhealth Uhealth 2023;11:e49051
URL: https://mhealth.jmir.org/2023/1/e49051
doi:10.2196/49051
PMID:37870887

Review

Factors Influencing the Acceptance and Adoption of Mobile Health Apps by Physicians During the COVID-19 Pandemic: Systematic Review

Sultan Alsahli[1,2], MSc; Su-yin Hor[1], PhD; Mary Lam[3], PhD

[1]School of Public Health, Faculty of Health, University of Technology Sydney, Sydney, Australia

[2]Department of Health Information Technology and Management, College of Public Health and Health Informatics, Umm Al-Qura University, Makkah, Saudi Arabia

[3]Department of Health and Biomedical Sciences, STEM College, RMIT University, Melbourne, Australia

Corresponding Author:
Sultan Alsahli, MSc
School of Public Health
Faculty of Health
University of Technology Sydney
235 Jones St.
Sydney, 2007
Australia
Phone: 61 (02) 9514 2000
Email: Sultan.alsahli@student.uts.edu.au

Abstract

Background: During the COVID-19 pandemic, the provision of and access to health care have been uniquely challenging, particularly during lockdowns or when dealing with COVID-19 cases. Health care professionals have had to provide patients with the necessary health care. However, delivering health care services while reducing face-to-face interaction puts an immense strain on health systems that are already overburdened. Against this backdrop, it is now more critical than ever to ensure the accessibility of health care services. Such access has been made increasingly available through mobile health (mHealth) apps. These apps have the potential to significantly improve health care outcomes and expectations and address some of the challenges confronting health care systems worldwide. Despite the advantages of mHealth, its acceptance and adoption remain low. Hence, health care organizations must consider the perceptions and opinions of physicians if the technology is to be successfully implemented.

Objective: The objective of this systematic review was to explore and synthesize the scientific literature on the factors influencing the acceptance and adoption of mHealth among physicians during the COVID-19 pandemic.

Methods: A systematic review of the studies published between March 2020 and December 2022 was conducted using the MEDLINE, Scopus, Embase, and ProQuest databases. The database search yielded an initial sample of 455 potential publications for analysis, of which 9 (2%) met the inclusion criteria. The methodology of this review was based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

Results: The factors influencing mHealth acceptance and adoption by physicians were divided into perceived barriers and perceived facilitators, which were further grouped into the following 3 major thematic categories: technological, individual, and organizational barriers and facilitators, respectively. The technological barriers were accessibility, technical issues, usefulness, and data management; individual barriers were perceived patient barriers, time and workload pressure, technical literacy, knowledge of mHealth, and peer support; and organizational barriers were financial factors, management support and engagement, data security, telemonitoring policy, and collaboration. The technological facilitators of uptake were technical factors, clinical usefulness, and data management; individual facilitators were patient-related care, intrinsic motivation, collaboration, and data sharing (individual); and organizational facilitators were workflow-related determinants, organizational financial support, recommendation of mHealth services, and evidence-based guidelines.

Conclusions: This review summarized the evidence on the factors influencing mHealth acceptance and adoption by physicians during the COVID-19 pandemic. The main findings highlighted the importance of addressing organizational readiness to support physicians with adequate resources, shifting the focus from technological to patient-centered factors, and the seamless integration of mHealth into routine practice during and beyond the pandemic.

Trial Registration: PROSPERO CRD42022356125; https://tinyurl.com/2mmhn5yu

(JMIR Mhealth Uhealth 2023;11:e50419) doi:10.2196/50419

KEYWORDS

mobile health; mHealth; mobile app; adoption; acceptance; barrier; attitude; physician; doctor; practitioner; mobile phone

Introduction

Background

On March 11, 2020, the World Health Organization (WHO) declared the outbreak of the COVID-19 pandemic [1], a crisis that has put pressure on health care systems around the world [2,3], with multiple waves of infections and deaths [4,5]. A recent report by the WHO stated that there have been 757,264,511 confirmed cases of COVID-19, of which 6,850,594 (0.9%) have been fatalities [6].

During this period, the provision of and access to health care have been uniquely challenging [3,7,8], particularly during lockdowns or when dealing with COVID-19 cases. Health care professionals have had to provide patients with the necessary health care. However, delivering health care services while reducing face-to-face interaction puts an immense strain on health systems that are already overburdened [9]. Against this backdrop, it is now more critical than ever to ensure the accessibility of health care services. Such access has been made increasingly available through mobile health (mHealth) apps, given the advancements in information and communication technology. These apps have the potential to significantly improve health care outcomes and expectations and address some of the challenges confronting health care systems worldwide [10-14].

mHealth falls under the broader umbrella of eHealth, which encompasses the use of electronic technologies and digital communication to enhance health care delivery [15-17]. However, mHealth technologies differ from conventional eHealth technologies in that they are specifically designed for use on mobile devices, and as such, mHealth apps do not rely solely on computers and wired internet connections, which makes them more accessible [18]. In addition, mHealth extends beyond medical consultations (more commonly known as telemedicine), offering features such as symptom tracking, mental health support, fitness tracking, medication reminders, personalized support, and access to health-related information [18-21]. Using mHealth is a popular strategy because it is user driven, readily available, and often reasonably priced [22].

The WHO [23] acknowledged that there is no widely accepted definition of mHealth, but it could be understood as the practice of using mobile devices for health care. More specifically, it refers to the capability to use mobile devices to collect health care–relevant data from patients in real time and use such information to monitor, diagnose, and treat patients [24]. It has the potential to benefit both health care professionals and patients during the COVID-19 pandemic [14,25,26]. For instance, it can improve the delivery of health care services, reduce health care professional and patient exposure to infectious diseases, and minimize patient demand for facilities [27,28]. In addition, mHealth apps use location data and proximity alerts to notify users if they were in close contact with someone who later tested positive for COVID-19 [29,30]. These timely alerts empower people to self-isolate, get tested, and inform their health care providers, helping break the chain of transmission [29,31]. It also offers opportunities for health care professionals to remotely consult and share data with their colleagues [32,33]. Furthermore, mHealth not only enables patients to receive remote consultation but also improves their adherence to medication and delivers disease education [20,25,33,34].

Despite the above-mentioned advantages, the acceptance and adoption of mHealth remain low [35-38]. The factors that influence technology acceptance and adoption are likely to vary across target users [39,40]. Physicians, for example, can stimulate changes in the health care sector and play a critical role in mHealth acceptance and adoption, depending on whether they themselves embrace this new technology. As explained by Cajita et al [41], patients are willing to accept and adopt mHealth when their physicians recommend it. Hence, health care organizations must consider the perceptions and opinions of physicians if the technology is to be successfully implemented [42].

Objectives

Before the COVID-19 pandemic, the acceptance and adoption of technology for work duties were a matter of personal or organizational preference [43]. This orientation was changed by the crisis, which compelled technology use in work environments, thereby accelerating the process of digitization in all sectors, including health care. As previously stated, physicians have been forced to provide health care services remotely [44], and they have accepted and adopted mHealth because of physical distancing restrictions. This situation may affect their continued use of the technology, which is one of the success factors for acceptance and adoption [38]. However, Keuper et al [44] found in their study that only a few physicians intend to continue offering remote health care services in the future. A possibility is that the COVID-19 pandemic has changed the behavioral intentions and perceptions of people regarding digital transformation [45,46]. Thus, the factors influencing technology acceptance and adoption have also likely changed [47], or new factors might have emerged. Shedding light on these factors can facilitate the acceptance and adoption of mHealth and help health care professionals provide services during the COVID-19 pandemic and other similar crises in the future.

Although previous reviews have analyzed mHealth acceptance and adoption by physicians [42,48,49], to the best of our knowledge, none of these reviews have focused on this topic in the context of the COVID-19 pandemic. This systematic review intended to fill this void. This review can benefit policy makers and mHealth providers by presenting an updated and thorough assessment of important issues that affect mHealth acceptance and adoption among physicians. This review can also help them design a strategy for promoting mHealth acceptance and adoption and derive potential benefits from this technology. Finally, this review provides opportunities for follow-up research by identifying potential gaps in mHealth acceptance and adoption.

Methods

Overview

The methodology of this review was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [50], which provides guidelines for a reliable and rigorous literature review (Multimedia Appendix 1). The review protocol was registered and published in advance with PROSPERO (CRD42022356125). The review focused on quantitative, qualitative, and mixed method studies to identify the factors that influenced the acceptance and adoption of mHealth among physicians during the COVID-19 pandemic.

Search Strategy

MEDLINE (Ovid), Scopus (Elsevier), Embase (Ovid), and ProQuest databases were searched for studies published in the English language. As the aim of this review was to explore mHealth acceptance and adoption factors during the pandemic, the time frame selected was from 2020 to 2022. The search strategy was established based on the population, intervention, comparator, and outcome (PICO) framework [51]. Specifically, we searched for studies revolving around physicians (population); the use of mHealth apps, including smartphones, portable digital devices, and tablets (intervention); and mHealth acceptance and adoption (outcome). Comparators were not relevant to this review.

Initially, combinations of Medical Subject Headings (MeSH) terms, keywords, and terminologies were used with reference to the following 3 categories: "mHealth," "acceptance or adoption," and "physician" (Multimedia Appendix 2). The more specific search terms used were as follows: "mobile health," "mHealth," "mHealth," or "mobile app"; "adoption," "acceptance," "barrier," or "attitude"; and "physician," "doctor," or "practitioner."

Study Selection

We used Covidence (Veritas Health Innovation, Ltd), a web-based collaboration software platform, to support the screening of the identified studies, all of which were uploaded onto the platform. A 2-step screening procedure was conducted to evaluate the relevance of the studies. In the first step, the titles and abstracts of the studies were screened independently by 2 reviewers (SA and SH). Any disagreements between the reviewers at the first step were discussed until a consensus was reached, or a third reviewer assisted in resolving the disagreement. In the second step, the studies that met the inclusion criteria were subjected to full-text screening carried out independently by 2 reviewers (SA and ML). Any disagreements at this point were resolved through discussion, or a third reviewer (SH) aided in the resolution.

Inclusion and Exclusion Criteria

The inclusion criteria were studies that (1) focused on the acceptance and adoption of mHealth primarily by physicians, (2) addressed factors influencing acceptance and adoption, (3) were peer reviewed, and (4) were published in English. The exclusion criteria were studies that (1) examined other health care technologies, such as electronic health records and electronic medical records; (2) focused solely on participants other than physicians (ie, patients, nurses, and midwives); and (3) collected data before the COVID-19 pandemic.

Quality Assessment

The studies included in the final data synthesis were assessed for methodological quality using the Quality Assessment with Diverse Studies (QuADS) criteria [52]. The QuADS is a 13-criteria tool developed to evaluate the quality of different designs, including quantitative, qualitative, and mixed methods research. For each criterion, a study can derive a score ranging from 0 (no mention at all) to 3 (full details), with the maximum possible score being 39. A QuADS score was calculated for each study, after which the item scores were summed and divided by the maximum possible score to obtain an overall quality assessment for each study. Studies with scores lower than 50%, ranging from 50% to 70%, and greater than 70% were classified as being of low, moderate, and high methodological quality, respectively [53]. Two authors (SA and SH) independently assessed the studies, and disagreements were resolved through discussion (Multimedia Appendix 3 [28,54-61]).

Data Extraction and Synthesis

Given the heterogeneous factors identified in the included studies, conducting a meta-analysis synthesis was not possible. Instead, the results on factors influencing the acceptance and adoption of mHealth among physicians were narratively synthesized. The selected studies were subjected to data extraction, with their titles, abstracts, and full texts screened, after which the required information was obtained using a predefined data extraction form. This form included the following details: authors, year of publication, location, study design, sample size, targeted population, theoretical framework, and influencing factors. To ensure the validity of these details, 2 reviewers (SA and ML) independently recorded them. Differences or disagreements were resolved through discussion.

Results

Overview

The database search yielded an initial sample of 455 potential publications for analysis. Of these 455 publications, 117 (25.7%) duplicates were eliminated. The titles and abstracts of the remaining 338 (85.3%) publications were reviewed, resulting in 314 (92.9%) publications being discarded at this stage for failing to meet the inclusion criteria. This remaining 24 (7.1%) publications underwent full-text review, of which 15 (62%) were eliminated because they did not meet the inclusion criteria (Figure 1). The final sample consisted of 9 published papers, whose key features are highlighted in Table 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). mHealth: mobile health.

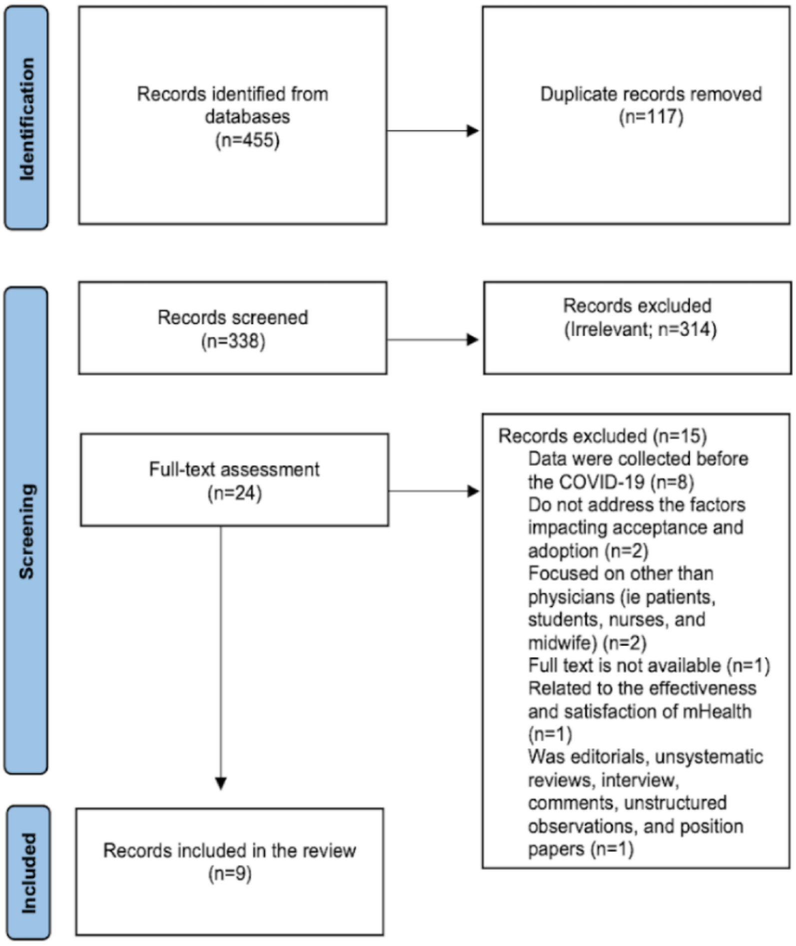


Table 1. Characteristics of the included studies (N=9).

Study	Country	Study Design	Participants (physicians), n	Targeted population	Specialty	Theoretical framework	Assessment tool	QuADS[a] score (%)
Aquino et al [54]	Canada	Qualitative	5	Clinicians and patients	Obstetricians	NR[b]	Interviews	59
Artanian et al [28]	Canada	Qualitatibe	5	Clinicians and patients	Cardiologists	Chaudoir multilevel framework	Interviews	87
Bhatt and Chakraborty [55]	India	Quantitative	316	Physicians	Multiple specialties	UTAUT[c]	Questionnaire	51
Dahlhausen et al [56]	Germany	Mixed methods	1295	Physicians	Multiple specialties	NR	Interviews and questionnaire	79
Fleddermann et al [57]	United States	Qualitative	13	Physicians	Multiple specialties	NR	Interviews	69
Jackson et al [58]	United States	Qualitative	29	Clinicians (physicians, nurses, diabetes educators, and dietitians)	Obstetricians	Stakeholder co-design framework	Focus groups and interviews	77
Li et al [59]	Australia	Qualitative	13	Clinicians and patients	Obstetricians	NR	Interviews	59
Mansour [60]	Egypt	Quantitative	203	Physicians	Multiple specialties	NR	Questionnaire	51
Wu et al [61]	China	Quantitative	393	Physicians	Multiple specialties	UTAUT	Questionnaire	72

[a]QuADS: Quality Assessment with Diverse Studies.

[b]NR: not reported.

[c]UTAUT: Unified Theory of Acceptance and Use of Technology.

Characteristics of the Included Studies

As shown in Table 1, of the 9 included studies, 2 (22%) each were conducted in the United States [57,58] and Canada [28,54], whereas 1 (11%) each was conducted in India [55], Australia [59], China [61], Egypt [60], and Germany [56]. A total of 5 (56%) studies focused on physicians [55-57,60,61], and 2 (22%) studies included patients as well [28,59]. Moreover, 1 (11%) study included practicing nurses in addition to physicians and patients [54], whereas another (11%) involved physicians, nurses, diabetes educators, dietitians, and lactation counselors [58]. From the perspective of specialization, most studies (5/9, 56%) involved physicians with multiple specialties [55-57,60,61], whereas other studies (4/9, 44%) involved cardiologists and obstetricians [28,54,58,59]. More than half (5/9, 56%) of the studies did not mention the use of a theoretical framework. A total of 2 (22%) studies used the Unified Theory of Acceptance and Use of Technology [55,61], 1 (11%) adopted a stakeholder co-design framework [58], and another used the Chaudoir multilevel framework [28]. Most studies (5/9, 56%) followed a qualitative approach that entailed conducting semistructured interviews and focus group discussions [28,54,57-59]. Overall, 3 (33%) studies adopted a quantitative approach entailing questionnaire administration [55,60,61], and only 1 (11%) used a mixed methods approach, in which questionnaires were administered and semistructured interviews were conducted [56].

Quality Assessment

As mentioned earlier, the studies were assessed using the QuADS tool to evaluate quality and risk of bias [52]. The methodological quality of the examined studies ranged from 51% to 87%. Overall, 4 (44%) studies had high-quality methodologies (scores of 72% to 87%), 5 (56%) studies had moderate-quality methodologies (scores ranging from 51% to 69%), and no study had low scores.

Factors Affecting Physicians' Acceptance and Adoption of mHealth Technologies

Perceived Barriers

Overview

All but 1 (11%) [61] of the 9 reviewed papers reported on perceived barriers to the acceptance and adoption of mHealth technologies by physicians. These barriers are summarized in Table 2. The literature is characterized by inconsistency in the use of theoretical frameworks to categorize barriers, and no single framework captures all relevant factors without some form of extension. Therefore, in this review, perceived barriers were grouped based on common themes and mapped into the following 3 major thematic categories: technological, individual, and organizational barriers (Figure 2).

Table 2. Barriers to the acceptance and adoption of mobile health (mHealth) technologies among physicians.

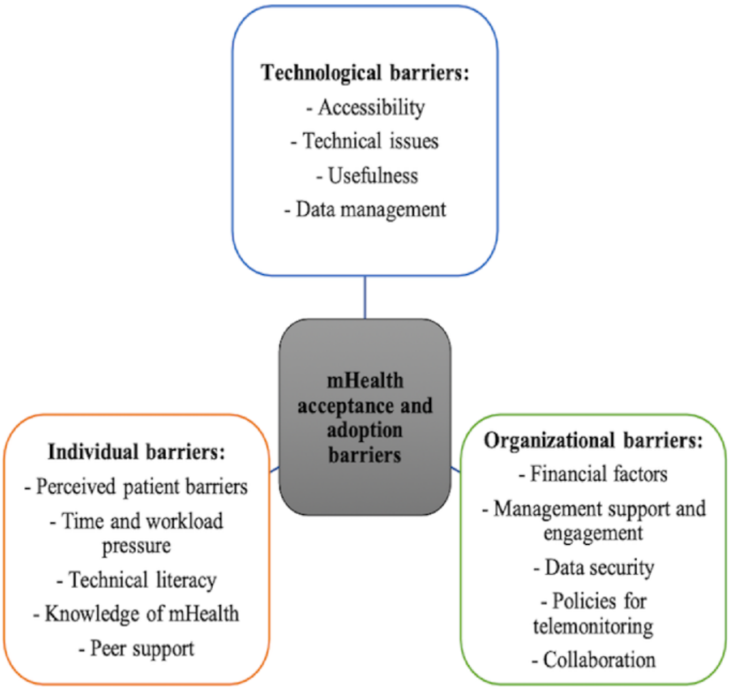
Study	Technological barriers	Individual barriers	Organizational barriers
Aquino et al [54]	• Lack of availability of telemonitoring systems for patients at a high risk for preeclampsia • Clinical utility: additional value in care management	• Increased clinician workload	• Lack of health system policies: limited guidelines for the telemonitoring of patients at a high risk for preeclampsia • Lack of access to appropriate resources (eg, validated BP[a] cuffs) • Financial cost (eg, cost of home BP monitor for patients)
Artanian et al [28]	• Lack of preparedness to implement telemonitoring: uncertainty regarding the functionality, operationalization, and integration of technology	• Patient preference for face-to-face contact • Patient acceptance of long-term technology use	• Lack of resources for supporting telemonitoring intervention: in the absence of a dedicated coordinator, time consuming for clinicians • Financial and economic factors: costs associated with resources for sustaining telemonitoring (eg, additional staff) • Physician remuneration: lack of compensation for services
Bhatt and Chakraborty [55]	• NR[b]	• Limited confidence (technology anxiety) • Lack of skill set for using mHealth services	• NR
Dahlhausen et al [56]	• Technical concerns: training needs, technical integration issues, and lack of technical support • Clinical utility: uncertainties about benefits and insufficient medical evidence • Low availability of technology	• Increased workload • Lack of awareness • Perceived low competence due to insufficient knowledge about differentiating mHealth platforms • Medicolegal concerns about potential liabilities for mistreatment	• Data protection and security • Financial factor: lack of reimbursement for mHealth-related medical services • Limitations of infrastructures: workflow-related issues (eg, workflow adjustments and training needs)
Fleddermann et al [57]	• Lack of adequate access to technology (among patients) • Challenges in navigating the technology • Competition from other similar apps • Lack of relatable content	• Lack of time • Competing priorities • Perceived lack of patient motivation (resistance to change) • Lack of peer support during internet-based treatment • Lack of in-person interaction for guiding patient use of mHealth	• Uncertainty regarding privacy and confidentiality • Limited organizational support and engagement • Limitations of infrastructures and workflows • Pandemic impact: disruption to the provision of services and challenges in shifting to hybrid care delivery and retaining patients
Jackson et al [58]	• Lack of evidence-based mHealth resources • Reliability of internet resources • Concern over ease of use and operationalization • Lack of credibility	• Limited familiarity, awareness, and knowledge of mHealth availability and utility • Low patient engagement in the long term	• Formal organizational structure: reliance on provider knowledge networks
Li et al [59]	• Accuracy of devices and uncertainty about technology reliability • Challenges related to integration with other health record systems • Clinical utility or usefulness: lack of evidence on the effectiveness of mHealth monitoring in pregnancy	• Pregnant women needing training to measure BP correctly • Difficulty with the sustainability of and compliance with the collection of data on pregnant women, especially due to cultural and linguistic barriers • Extra workload due to the review of monitoring data • Skill set required to accurately analyze the data	• Limited communication among clinicians from multiple disciplines: multidisciplinary approach or communication needed to consider pregnancy symptoms, risk factors, test findings, and data about babies • Concerns about patient data privacy • Limitation of resources for supporting mHealth • Financial cost of technology (especially among patients from low socioeconomic backgrounds)

Study	Technological barriers	Individual barriers	Organizational barriers
Mansour [60]	• Lack of training on using mHealth technologies • Lack of appropriate and relevant content • Failure of mobile network connection • Potential for the misuse of collected information	• Lack of time for using technology • Lack of technical skills • Lack of interest in, knowledge about, or awareness of the benefits of mHealth technologies • Lack of language skills • Communication barriers: demographic characteristics of patients (age, education, and gender)	• Financial cost of technology implementation • Concerns about personal data privacy and security
Wu et al [61]	• NR	• NR	• NR

[a]BP: blood pressure.

[b]NR: not reported.

Figure 2. Themes of barriers to mobile health (mHealth) acceptance and adoption by physicians.



Technological Barriers

The technological barriers to acceptance and adoption were further classified into the following 4 key subthemes identified from 8 (89%) of the 9 examined studies: accessibility, technical issues, usefulness, and data management. Technical issues were the most frequently reported barriers, including functionality (eg, concern over ease of use and operationalization) [28,57,58] and technical support (eg, technical issues in daily operations) [56]. Features related to usefulness, such as the clinical utility, added value, and evidence-based effectiveness of mHealth in care management (eg, lack of or insufficient evidence of benefit for patients), were other significant impediments to the use of mHealth technologies [54,56,58,59]. Concerns related to data management, including integration issues (eg, challenges with integration into clinical health records and poor integration or compatibility with existing practice software and tools) [28,56,59], were also raised. Lack of access [54,56,57], reliability [58,59], and limited connectivity (eg, concern about weak or failure of mobile network connectivity) [60] were cited by the rest of the studies.

Individual Barriers

Individual intrinsic (eg, confidence) and extrinsic (eg, technical competence) barriers emerged from the 8 (89%) of the 9 explored studies and were categorized into the following 5 key subthemes: perceived patient barriers, time and workload pressure, technical literacy, knowledge of mHealth, and peer support. Patient-related factors were the most prominently cited individual barriers, with patient acceptance or motivation (eg, perceived lack of patient motivation due to resistance to change) and sustained compliance with long-term technology use (eg, difficulty with the sustainability of and compliance with the collection of data on patients, especially due to cultural and linguistic barriers) being central concerns [28,57-60]. Time pressure and extra workload (eg, the additional work required for physicians to monitor patient data) [54,56,57,59,60] were reported as impediments to mHealth use by health care professionals. Other barriers mentioned were limited technical skills and confidence (eg, lack of language skills and technology anxiety) [55,57-60], the lack of knowledge about differentiating between mHealth platforms and awareness of mHealth benefits [56,58,60], and the lack of peer support [57].

Organizational Barriers

Organizational barriers were divided into 5 central subthemes: financial factors, management support and engagement, data security, technology policy, and collaboration. The most commonly reported barrier at the organizational level was financial factors, including the cost of mHealth apps and reimbursement issues. These issues involved costs associated with mHealth implementation (eg, the cost of devices) for both physicians [60] and patients [54,59], especially for those with low socioeconomic status [59], and the lack of or insufficient reimbursement for mHealth-related medical services (eg, responding to follow-up questions from patients) [56]. Other central barriers included the need for organizational engagement, lack of human resource support (eg, hiring a dedicated mHealth coordinator to reduce the workload of clinicians), lack of infrastructure [28,54,56-59], and lack of training [60]. The rest of the hindrances to mHealth uptake were the lack of policies related to data security (eg, uncertainty about the privacy and security of personal health data) [56,57,59,60], lack of evidence-based telemonitoring guidelines [54], and lack of communication among health care providers [59].

Perceived Facilitators

Overview

All the included studies discussed the perceived facilitators of mHealth acceptance and adoption by health care providers (Table 3). Similar to the barriers, the facilitators were categorized into technological, individual, and organizational facilitators (Figure 3).

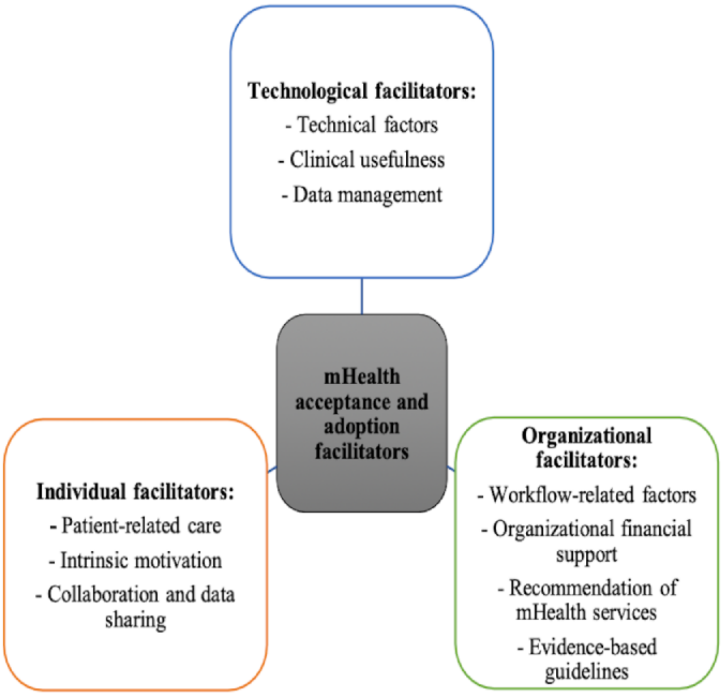
Table 3. Facilitators of the acceptance and adoption of mobile health (mHealth) technologies among physicians.

Study	Technological facilitators	Individual facilitators	Organizational facilitators
Aquino et al [54]	• Evidence-based action prompts generated from patient data based on guidelines for patients at a high risk for preeclampsia • Functionality: automatic data entry into telemonitoring systems	• Perceived benefits: self-management tool for patients • Effective display of patient data to facilitate trend detection and the visualization of patient health status	• Facilitation of decision-making for clinicians by integrating evidence-based protocols and standards for patients at a high risk for preeclampsia
Artanian et al [28]	• Functionality: ease of use of telemonitoring systems and their seamless integration into clinical practice and patient's daily routine • Clinical utility: access to daily data for providing accurate information about patient well-being	• Engagement of eligible patients in telemonitoring	• Availability of organizational resources: dedicated staff • Advantageous over standard care owing to overcoming limitations in clinic space and the optimization of clinical resources • Establishment of reimbursement models • Adequate information on how to implement telemonitoring
Bhatt and Chakraborty [55]	• Streamlined data handling for patient care management	• Self-confidence or self-efficacy of physicians in handling technology requirements • Performance expectancy • Personal innovativeness	• NR[a]
Dahlhausen et al [56]	• Clinical utility: data and more accessible medical evidence • Functionality: opportunities to navigate or test mHealth apps • Additional information about mHealth platforms • Compatibility of mHealth with existing infrastructures and workflows	• Patient motivation or patient request to use mHealth tools	• Recommendations by peers or medical associations • Provision of provider reimbursement for mHealth-related medical services • Extensive training with incentives (eg, certification for continuing medical education)
Fleddermann et al [57]	• Integration of technology use into routine workflows • Technological support for facilitating engagement	• Significant levels of clinician engagement for supporting patient use of mHealth platforms, especially for supporting the management of challenges encountered by patients unable to access typical in-person treatment during isolation • Collaboration with other staff using mHealth technologies	• Recommendation by physicians for potential mHealth benefits • Ongoing training
Jackson et al [58]	• Functionality: patient-centered participatory design of customized functions and educational features, including data-tracking, motivational feedback, and bidirectional communication capabilities • Clinical utility: potential to streamline clinical activities and resources • Clinical integration (into routine prenatal care)	• Provision for continued practical patient education to promote self-care management • Clinician engagement with patient education	• Integration of activities related to behavioral health changes into the patient's daily routine

Study	Technological facilitators	Individual facilitators	Organizational facilitators
Li et al [59]	• Functionality alert: function for the early detection of issues and timely interventions • User-friendly and comes with an automatic data capture feature • Access to data from multiple sources and integration of data with health records • Demonstration of impact and evidence-based evaluation studies before implementation • Compatibility with current practices of risk assessment and care for pregnant women with potential for multidisciplinary approach	• Integrated tailored educational content and feedback for pregnant women based on conditions and risks • Additional education and monitoring for pregnant women at high risk to improve data collection compliance and engagement	• Recommendations by clinicians with indications for potential benefits • Provision of ongoing education and training on using mHealth technologies
Mansour [60]	• Simplicity, user-friendliness, and convenience of mHealth apps (eg, detection of COVID-19 symptoms, pulse oximeter, and COVID-19 health-monitoring apps) • Access to COVID-19–related services and updated information • Clinical utility: support for frequent health monitoring and preventive health care	• Self-confidence in using technology • Increased patient knowledge, improved patient engagement and medication adherence, and faster access to providers • Communication and consultation with peers or colleagues and data sharing with other providers	• Recommendation of mHealth use by physician
Wu et al [61]	• Effort expectancy (ease and simplicity of mHealth)	• Behavioral intention of physicians to use mHealth was significantly affected by intrinsic motivations (altruism and cognitive trust) • High internet-based ratings affect sense of self-worth and contribute to positive participation in web-based health services	• Integration of mHealth into the national health system • Facilitating conditions, such as technical and human resource support, have a positive effect on mHealth adoption

[a]NR: not reported.

Figure 3. Themes of facilitators of mobile health (mHealth) acceptance and adoption by physicians.



Technological Facilitators

The 3 main subthemes related to technological facilitators were technical factors, clinical usefulness, and data management. Technical factors were subdivided into access, functionality, and technical support domains, which were discussed in most of the reviewed studies (8/9, 89%) [28,54,56-61]. Specifically, of the 9 reviewed studies, 7 (78%) highlighted functionality and ease of use as important features for engaging providers [28,54,56,58-61]. For instance, the clinicians participating in these studies applauded the ease with which patients with diabetes can use mHealth systems to track blood sugar levels in real time and the advantage of direct feeds to providers [54]. Technological support was a critical facilitator of mHealth use [57]. Of the 9 studies, 2 (22%) identified access to mHealth services, such as data collection from multiple sources [59] and updated information [60], as facilitators of successful uptake by health care providers.

Among the 9 included studies, 5 (56%) discussed clinical utility and usefulness as factors that favor adoption [28,56,58-60]. Providers are more likely to use mHealth services when they perceive mHealth technologies as potentially streamlining patient management care and clinical resources; some examples are technologies that allow the monitoring of prescription changes and updating of medical charts or clinical notes [58]. Usefulness pertained primarily to the availability of accurate real-time information about patient well-being [28], additional support for current prenatal care practices [59], and frequent health monitoring [60]. Evidence-based evaluation studies and accessible evidence of the usefulness of mHealth platforms also potentially facilitate the adoption of mHealth technologies by health care providers [56,59]. Finally, facilitators that support adoption and sustained use were data management, including the integration of mHealth technologies into routine clinical practice and health records [28,56-59] and streamlined data handling for patient management [55].

Individual Facilitators

Individual facilitators were divided into the following 3 central subthemes: patient-related care, intrinsic motivation, and collaboration and data sharing. Facilitators related to patients were central in most of the reviewed studies (7/9, 78%). These included the perceptions (of physicians) that mHealth technologies have the potential to support self-managed care and provide real-time feedback [54,58], allow faster access to health care providers [60], integrate mHealth into patient routines with tailored content [59], improve patient engagement [58-60], and provide support to patients who are unable to access typical in-person clinical treatment given the isolation prompted by the COVID-19 pandemic [57]. In particular, physicians are predisposed to use mHealth services when their integration increases the efficiency of daily patient flow, data management, patient diagnosis, and other clinical activities [55,58]. During the pandemic, especially when clinic access was largely restricted, the promotion of mHealth as a patient self-care management tool was one of the key factors in physicians' decision to adopt this innovation as a critical supportive tool in clinical care [54,57,58]. This decision is further supported by the effectiveness of mHealth in advancing multidisciplinary communication, as is the case, for example, with pregnancy care, for which access to data from multiple disciplines or sources is needed [59]. In addition, health care professionals with self-confidence, self-efficacy [55,60], altruism, and cognitive trust [61] in the reliability of technology are inclined to engage with and use mHealth platforms. These factors were rounded up through collaboration with peers or other users to share experiences and knowledge as well as data sharing with other providers [56,57,60].

Organizational Facilitators

Organizational facilitators were divided into the following 4 key subthemes: workflow-related factors, organizational financial support, recommendation of mHealth services, and evidence-based guidelines. Among the 9 included studies, 3 (33%) pinpointed workflow-related factors, such as the availability of support for streamlining clinical resources and activities and improvement of infrastructure for seamless workflow, as key facilitators [28,58,61]. In particular, organizational human resource support, such as the hiring of a dedicated coordinator to reduce physician workload [28] and address training needs [56,57,59,61], was highly advocated as a facilitator of mHealth uptake by physicians. Moreover, widespread adoption was found to be motivated by organizational financial support deployed via the establishment of reimbursement models [28] and the provision of financial incentives or reimbursement for mHealth services [56]. Effective implementation was also regarded as facilitated by the recommendation of mHealth services by trusted leaders, such as medical associations [56] or other physicians [57,59]. Other important facilitators of successful uptake included the integration of evidence-based standards and guidelines for telemonitoring into practice to facilitate clinical decision-making [54] and the integration of mHealth into the national health system [61]. None of the included studies reported specific facilitators regarding legal issues related to the security and privacy of patient data.

Discussion

Summary of the Main Findings

The COVID-19 pandemic has clearly been a catalyst of the wider acceptance and adoption of mHealth interventions worldwide, with studies frequently reporting benefits such as minimized risk of transmission, increased patient involvement, and reduced burden on hospitals and health care expenditure [9,62,63]. Nevertheless, the move toward mHealth apps as a model of care delivery during the pandemic has revealed several shortcomings in stimulating physicians' uptake of such technologies. This review explored the factors influencing mHealth acceptance and adoption by physicians as the COVID-19 pandemic evolves. Factors related to the technological, individual, and organizational domains were identified.

Critical Barriers to mHealth Acceptance and Adoption

Evidence suggests that a number of barriers have persisted since the prepandemic period [42,48,49]. This finding corresponds to the work of Zakerabasali et al [42], who reviewed evidence from 18 articles and identified 18 technical, individual, and health care system barriers. Similar to the findings in this review, the authors identified the lack of technical infrastructure, concerns about privacy issues, and the lack of workflow compatibility as barriers to mHealth adoption. Other principal barriers were limited technical literacy, preference for face-to-face interaction, financial factors, and health system policies [42]. Another prepandemic systematic review conducted in 2020 identified 55 barriers, including the lack of clinical training, the lack of technical support, the lack of compatibility with the existing workflow, and patient-related factors [48]. Consistent with the aforementioned studies, a systematic review conducted in 2016 identified 81 barriers, with emphasis placed on cost and time issues as well as difficulties in patient-professional interaction [49].

Although some of the perceived barriers that we found were similar to those identified in explorations carried out before the pandemic, we were able to identify other factors that are specific to acceptance and adoption during the pandemic. Examples include challenges accompanying the shift to hybrid care delivery to retain patients affected by the implementation of mHealth tools by physicians. The transition to internet-based treatment during the COVID-19 pandemic has disrupted services by dramatically reducing clinical caseloads, an issue that highlights patients' preference for face-to-face appointments. Clinicians also lamented the considerable difficulty involved in assisting and guiding patients in downloading and signing up to an mHealth app [57]. As can be seen, the pandemic has highlighted the need to improve organizational readiness by making workflow adjustments to allow time for the introduction of mHealth tools to patients and the effective implementation of such innovations in practice. Another novel finding of this systematic review is that physicians perceive low competence in dealing with mHealth technologies as a result of insufficient knowledge and information regarding differentiating between mHealth platforms [56]. Collectively, these findings point to the importance of organizational support during *business as usual* periods to provide physicians with adequate education and training on the use of emerging mHealth tools.

Systematic reviews conducted before the pandemic differently emphasized barriers to mHealth adoption. Whereas cost issues and patient-professional interaction were reported as the most common barriers in an early systematic review [49], technical difficulties, particularly the lack of technical support, the lack of compatibility with the existing workflow, and patient-related challenges, were underscored as principal impediments in a more recent analysis [48]. In addition to technical and cost factors, privacy concerns were one of the most cited barriers in the examined studies [42]. To these lists, our study added limited financial support and technical and privacy issues as common barriers to uptake. However, in contrast to prepandemic reviews, this review identified patient-related factors, such as patient preference, engagement, and compliance, as the most frequently reported determinants of uptake during the pandemic. On these bases, we can conclude that the pandemic has shifted the focus from a technological perspective to a more patient-centered perspective in recognizing the main challenges to mHealth adoption and integration into practice.

Leading Facilitators of mHealth Acceptance and Adoption

Some of the common facilitators of mHealth uptake evaluated in this study were consistent with those reported before the pandemic. These include perceived usefulness and ease of use, perceived patient-related benefits (eg, improved patient care, interprofessional collaborations, and data sharing), ongoing technical support and training, and financial support for technology implementation and integration with practice systems [48,49]. However, this review is distinct from prior research in terms of facilitators that are specific to the context of the pandemic.

The most prominent facilitators before the pandemic were those related to organizational workflow, such as infrastructure, training, resource allocation, perceived efficiency, improved reimbursement, and compatibility with workflow [48,49]. Against the backdrop of the pandemic, the central facilitators were the individual factors associated with the intrinsic motivation of physicians and patient-related matters. For instance, the behavioral intention of physicians to use mHealth apps was significantly influenced by self-efficacy [55], and intrinsic motivation was potentially strengthened by altruism and cognitive trust (perceived reliability) linked to competence in using mHealth platforms [61]. Recent studies confirmed that cognitive trust strongly influences the use of digital technologies, suggesting that it is essential to cultivate physicians' trust in mHealth adoption through their sense of altruism [64] while their self-efficacy in the sustained intention to use mHealth platforms is elevated [65].

In our review, individual factors related to patient acceptance for greater engagement in and long-term commitment to using mHealth services were demonstrated to be critical to sustained uptake by physicians. High levels of physician engagement in promoting the benefits of mHealth apps for treatment [57] and clinician involvement with patient education [58] were also regarded as necessary for supporting patient access and the use of mHealth tools. This was especially important during periods of enforced isolation, as mHealth use fostered connections and supported the management of patients unable to access face-to-face treatment [57].

Furthermore, although addressing legal issues was one of the organizational factors that facilitated mHealth adoption before the pandemic [48], none of the reviewed studies discussed security and data protection. This deficiency can be attributed to the changes to regulations made by some countries during the global outbreak to provide further security guidance and support the more extensive use of telehealth [66]. In this situation, the attention of physicians could have been diverted from legal issues to concerns about their patients. Altogether, the available evidence highlights the importance of physicians' intrinsic self-motivation in supporting a patient-centered approach. The focus should be directed to patient benefits as critical facilitators of successful acceptance and adoption in the context of the COVID-19 pandemic.

It is worth noting that there are varying factors influencing the acceptance and adoption of mHealth across limited-resource and high-resource countries. For instance, in limited-resource countries, Mansour [60] and Bhatt and Chakraborty [55] highlighted barriers, including the lack of language, technical skills, and training. By contrast, some studies in high-resource countries emphasized that mHealth apps were easy to use and integrated well into clinicians' routines [28,58]. This variation can be attributed to the fact that health care systems in high-resource countries commonly have well-established training programs that integrate the latest medical advancements for health care professionals. By contrast, limited-resource countries may face challenges in providing sufficient training and education programs for health care professionals because of limited resources and funding [67-70]. Consequently, health care professionals in limited-resource countries may have limited opportunities for training and may not have the same skills and knowledge as their peers in high-resource countries.

Although our findings indicate that health care professionals have a generally positive attitude toward mHealth, there are variations in attitudes across various medical specialties [56,60]. For example, Dahlhausen et al [56] highlighted that neurologists have a mostly favorable perspective toward mHealth apps, whereas orthopedists and trauma surgeons hold somewhat less positive attitudes toward these apps. In line with our findings, a survey conducted by Zaslavsky et al [39] revealed differences in attitudes toward implementing mHealth apps across different medical specializations. Understanding these differences is crucial for customizing strategies to promote the adoption of mHealth among various medical specialties.

Limitations and Recommendations for Future Research

Although this review contributes to the understanding of the factors influencing the acceptance and adoption of mHealth technologies among physicians, some limitations must be acknowledged. Most studies (6/9, 67%) were conducted in developed countries (eg, the United States, Canada, and Germany) [28,54,56,57], which means that our understanding of the factors influencing the acceptance and adoption of mHealth among physicians in developing countries is limited. Moreover, more than half (5/9, 56%) of the studies [28,54,57-59] used qualitative methods, such as semistructured interviews and focus group discussions, to gather data. Therefore, generalizing the results of this review may be challenging. In addition, this review might not have incorporated relevant papers that were not listed in the databases that were searched and that were published in a language other than English, which would have helped identify more factors that influence the acceptance and adoption of mHealth among physicians.

We provide several recommendations for future research. Identifying the factors that affect the acceptance and adoption of technologies such as mHealth is an ongoing process [57]. Hence, there is a need for more extensive research on these behaviors of physicians, especially in limited-resource countries. Research in limited-resource countries is necessary to understand whether there are different opportunities and constraints. In addition, robust methodologies, such as mixed methods approaches, are required to uncover the factors influencing acceptance and adoption. Mixed methods research can overcome the disadvantages associated with quantitative or qualitative approaches, thereby enriching the findings. For example, some researchers claim that quantitative exploration insufficiently advances the understanding of contexts or areas in which people live, as the voices of participants are not directly heard [71]. Qualitative studies might be considered deficient because of a researcher's subjective interpretations, the bias that results from these, and the difficulty in generalizing findings [71]. Finally, the identified factors could help policy makers make decisions aimed at implementing mHealth successfully. These factors may facilitate physicians' acceptance and adoption of mHealth technologies.

Conclusions

The pandemic has highlighted and expanded the avenues in which mHealth can aid clinical decision-making and improve the quality of care. This review summarized the evidence on the factors influencing mHealth acceptance and adoption by physicians during the COVID-19 pandemic. The main findings of this review highlighted the importance of addressing organizational readiness to support physicians with adequate resources, shifting the focus from technological to patient-centered factors, and the seamless integration of mHealth into routine practice during and beyond the pandemic.

Authors' Contributions

SA, the first author, wrote the manuscript, and ML and SH provided insightful feedback on the manuscript. All the authors were involved in screening the studies, extracting data, and synthesizing the findings. SA was funded by a Doctor of Philosophy (PhD) scholarship from the Saudi Arabian Cultural Mission in Australia. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.

[DOCX File , 21 KB - mhealth_v11i1e50419_app1.docx]

Multimedia Appendix 2

Search strategy.

[DOCX File , 16 KB - mhealth_v11i1e50419_app2.docx]

Multimedia Appendix 3
Risk-of-bias assessment of the included studies.
[DOCX File , 17 KB - mhealth_v11i1e50419_app3.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2023-01-15]
2. Labrague LJ, de Los Santos JA, Fronda DC. Factors associated with missed nursing care and nurse-assessed quality of care during the COVID-19 pandemic. J Nurs Manag 2022 Jan;30(1):62-70 [FREE Full text] [doi: 10.1111/jonm.13483] [Medline: 34590383]
3. Mira JJ, Cobos A, Martínez García O, Bueno Domínguez MJ, Astier-Peña MP, Pérez Pérez P, SARS-CoV-2 Second Victims Working Group. An acute stress scale for health care professionals caring for patients with COVID-19: validation study. JMIR Form Res 2021 Mar 09;5(3):e27107 [FREE Full text] [doi: 10.2196/27107] [Medline: 33687343]
4. Nguyen M. The psychological benefits of COVID-19 vaccination. Adv Public Health 2021 Nov 3;2021:1-9 [FREE Full text] [doi: 10.1155/2021/1718800]
5. Woods JA, Hutchinson NT, Powers SK, Roberts WO, Gomez-Cabrera MC, Radak Z, et al. The COVID-19 pandemic and physical activity. Sports Med Health Sci 2020 Jun;2(2):55-64 [FREE Full text] [doi: 10.1016/j.smhs.2020.05.006] [Medline: 34189484]
6. Who coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/?mapFilter=cases [accessed 2023-05-18]
7. Mitra M, Basu M. A study on challenges to health care delivery faced by cancer patients in India during the COVID-19 pandemic. J Prim Care Community Health 2020 Jan;11:2150132720942705 [FREE Full text] [doi: 10.1177/2150132720942705] [Medline: 32830619]
8. Roy CM, Bollman EB, Carson LM, Northrop AJ, Jackson EF, Moresky RT. Assessing the indirect effects of COVID-19 on healthcare delivery, utilization and health outcomes: a scoping review. Eur J Public Health 2021 Jul 13;31(3):634-640 [FREE Full text] [doi: 10.1093/eurpub/ckab047] [Medline: 33755130]
9. Houlding E, Mate KK, Engler K, Ortiz-Paredes D, Pomey MP, Cox J, et al. Barriers to use of remote monitoring technologies used to support patients with COVID-19: rapid review. JMIR Mhealth Uhealth 2021 Apr 20;9(4):e24743 [FREE Full text] [doi: 10.2196/24743] [Medline: 33769943]
10. Nicholas J, Bell IH, Thompson A, Valentine L, Simsir P, Sheppard H, et al. Implementation lessons from the transition to telehealth during COVID-19: a survey of clinicians and young people from youth mental health services. Psychiatry Res 2021 May;299:113848 [FREE Full text] [doi: 10.1016/j.psychres.2021.113848] [Medline: 33725578]
11. Nguyen HL, Tran K, Doan PL, Nguyen T. Demand for mobile health in developing countries during COVID-19: Vietnamese’s perspectives from different age groups and health conditions. Patient Prefer Adherence 2022 Feb 02;16:265-284 [FREE Full text] [doi: 10.2147/PPA.S348790] [Medline: 35140459]
12. Lin Y, Lou M. Effects of mHealth-based interventions on health literacy and related factors: a systematic review. J Nurs Manag 2021 Apr;29(3):385-394. [doi: 10.1111/jonm.13175] [Medline: 33040429]
13. Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y. A questionnaire for assessing user satisfaction with mobile health apps: development using Rasch measurement theory. JMIR Mhealth Uhealth 2020 May 26;8(5):e15909 [FREE Full text] [doi: 10.2196/15909] [Medline: 32452817]
14. John Leon Singh H, Couch D, Yap K. Mobile health apps that help with COVID-19 management: scoping review. JMIR Nurs 2020 Aug 6;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
15. Alenoghena CO, Onumanyi AJ, Ohize HO, Adejo AO, Oligbi M, Ali SI, et al. eHealth: a survey of architectures, developments in mHealth, security concerns and solutions. Int J Environ Res Public Health 2022 Oct 11;19(20):13071 [FREE Full text] [doi: 10.3390/ijerph192013071] [Medline: 36293656]
16. Kernebeck S, Busse TS, Böttcher MD, Weitz J, Ehlers J, Bork U. Impact of mobile health and medical applications on clinical practice in gastroenterology. World J Gastroenterol 2020 Aug 07;26(29):4182-4197 [FREE Full text] [doi: 10.3748/wjg.v26.i29.4182] [Medline: 32848328]
17. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2020. URL: https://iris.who.int/handle/10665/274134?&locale-attribute=de [accessed 2023-04-07]
18. Alkhalifah T. A structural equation modelling of governing factors influencing patient acceptance of mobile health in Saudi Arabia: a modified UTAUT model. Int J E-Serv Mob Appl 2022;14(1):1-17 [FREE Full text] [doi: 10.4018/ijesma.295963]
19. Chen Y, Xu Q. The willingness to use mobile health and its influencing factors among elderly patients with chronic heart failure in Shanghai, China. Int J Med Inform 2021 Dec 16;158:104656 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104656] [Medline: 34933173]

20. Ni Z, Wu B, Yang Q, Yan LL, Liu C, Shaw RJ. An mHealth intervention to improve medication adherence and health outcomes among patients with coronary heart disease: randomized controlled trial. J Med Internet Res 2022 Mar 09;24(3):e27202 [FREE Full text] [doi: 10.2196/27202] [Medline: 35262490]
21. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
22. Gomez Quiñonez S, Walthouwer MJ, Schulz DN, de Vries H. mHealth or eHealth? Efficacy, use, and appreciation of a web-based computer-tailored physical activity intervention for Dutch adults: a randomized controlled trial. J Med Internet Res 2016 Nov 09;18(11):e278 [FREE Full text] [doi: 10.2196/jmir.6171] [Medline: 27829576]
23. mHealth: new horizons for health through mobile technologie. World Health Organization. 2011. URL: https://iris.who.int/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1 [accessed 2023-04-20]
24. Almotiri SH, Khan MA, Alghamdi MA. Mobile health (m-Health) system in the context of IoT. In: Proceedings of the 4th International Conference on Future Internet of Things and Cloud Workshops. 2016 Presented at: FiCloudW '16; August 22-24, 2016; Vienna, Austria p. 39-42 URL: https://ieeexplore.ieee.org/document/7592698 [doi: 10.1109/w-ficloud.2016.24]
25. Yang S, Chen Y, Zhou L, Huang Y, Dai J. Willingness to adopt mHealth among Chinese parents during the COVID-19 outbreak: cross-sectional questionnaire study. JMIR Mhealth Uhealth 2021 Jan 27;9(1):e23155 [FREE Full text] [doi: 10.2196/23155] [Medline: 33406052]
26. Schmeelk S, Davis A, Li Q, Shippey C, Utah M, Myers A, et al. Monitoring symptoms of COVID-19: review of mobile apps. JMIR Mhealth Uhealth 2022 Jun 01;10(6):e36065 [FREE Full text] [doi: 10.2196/36065] [Medline: 35609313]
27. El-Sherif DM, Abouzid M, Elzarif MT, Ahmed AA, Albakri A, Alshehri MM. Telehealth and artificial intelligence insights into healthcare during the COVID-19 pandemic. Healthcare (Basel) 2022 Feb 18;10(2):385 [FREE Full text] [doi: 10.3390/healthcare10020385] [Medline: 35206998]
28. Artanian V, Ware P, Rac VE, Ross HJ, Seto E. Experiences and perceptions of patients and providers participating in remote titration of heart failure medication facilitated by telemonitoring: qualitative study. JMIR Cardio 2021 Nov 25;5(2):e28259 [FREE Full text] [doi: 10.2196/28259] [Medline: 34842546]
29. Almalki M, Giannicchi A. Health apps for combating COVID-19: descriptive review and taxonomy. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
30. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
31. Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, et al. Health observation app for COVID-19 symptom tracking integrated with personal health records: proof of concept and practical use study. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e19902 [FREE Full text] [doi: 10.2196/19902] [Medline: 32568728]
32. Khan NU, Rasheed S, Sharmin T, Ahmed T, Mahmood SS, Khatun F, et al. Experience of using mHealth to link village doctors with physicians: lessons from Chakaria, Bangladesh. BMC Med Inform Decis Mak 2015 Aug 05;15(1):62 [FREE Full text] [doi: 10.1186/s12911-015-0188-9] [Medline: 26242574]
33. Smith R, Menon J, Rajeev JG, Feinberg L, Kumar RK, Banerjee A. Potential for the use of mHealth in the management of cardiovascular disease in Kerala: a qualitative study. BMJ Open 2015 Nov 17;5(11):e009367 [FREE Full text] [doi: 10.1136/bmjopen-2015-009367] [Medline: 26576813]
34. Sleurs K, Seys SF, Bousquet J, Fokkens WJ, Gorris S, Pugin B, et al. Mobile health tools for the management of chronic respiratory diseases. Allergy 2019 Jul;74(7):1292-1306. [doi: 10.1111/all.13720] [Medline: 30644567]
35. Alam MZ, Hu W, Kaium MA, Hoque MR, Alam MM. Understanding the determinants of mHealth apps adoption in Bangladesh: a SEM-Neural network approach. Technol Soc 2020 May;61:101255 [FREE Full text] [doi: 10.1016/j.techsoc.2020.101255]
36. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019 [FREE Full text] [doi: 10.1111/isj.12179]
37. Kaium MA, Bao Y, Alam MZ, Hasan N, Hoque MR. Understanding the insight of factors affecting mHealth adoption. Int J Res Bus Soc Sci 2019 Oct 26;8(6):181-200 [FREE Full text] [doi: 10.20525/ijrbs.v8i6.522]
38. Hoque MR, Rahman MS, Nipa NJ, Hasan MR. Mobile health interventions in developing countries: a systematic review. Health Informatics J 2020 Dec 21;26(4):2792-2810 [FREE Full text] [doi: 10.1177/1460458220937102] [Medline: 32691659]
39. Zaslavsky O, Chu F, Renn BN. Patient digital health technologies to support primary care across clinical contexts: survey of primary care providers, behavioral health consultants, and nurses. JMIR Form Res 2022 Feb 25;6(2):e32664 [FREE Full text] [doi: 10.2196/32664] [Medline: 35212642]
40. Moon JW, Kim YG. Extending the TAM for a World-Wide-Web context. Inf Manag 2001 Feb;38(4):217-230 [FREE Full text] [doi: 10.1016/S0378-7206(00)00061-6]
41. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]

42. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021 Oct;27(4):267-278 [FREE Full text] [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
43. Ishaq E, Bashir S, Zakariya R, Sarwar A. Technology acceptance behavior and feedback loop: exploring reverse causality of TAM in post-COVID-19 scenario. Front Psychol 2021 Sep 13;12:682507 [FREE Full text] [doi: 10.3389/fpsyg.2021.682507] [Medline: 34589017]
44. Keuper J, Batenburg R, Verheij R, van Tuyl L. Use of E-health in Dutch general practice during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Nov 26;18(23):12479 [FREE Full text] [doi: 10.3390/ijerph182312479] [Medline: 34886204]
45. Zhong L, Coca-Stefaniak JA, Morrison AM, Yang L, Deng B. Technology acceptance before and after COVID-19: no-touch service from hotel robots. Tour Rev 2022;77(4):1062-1080 [FREE Full text] [doi: 10.1108/tr-06-2021-0276]
46. Srisathan WA, Naruetharadhol P. A COVID-19 disruption: the great acceleration of digitally planned and transformed behaviors in Thailand. Technol Soc 2022 Feb;68:101912 [FREE Full text] [doi: 10.1016/j.techsoc.2022.101912] [Medline: 35125560]
47. An MH, You SC, Park RW, Lee S. Using an extended technology acceptance model to understand the factors influencing telehealth utilization after flattening the COVID-19 curve in South Korea: cross-sectional survey study. JMIR Med Inform 2021 Jan 08;9(1):e25435 [FREE Full text] [doi: 10.2196/25435] [Medline: 33395397]
48. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
49. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
50. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
51. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7(1):16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
52. Harrison R, Jones B, Gardner P, Lawton R. Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. BMC Health Serv Res 2021 Feb 15;21(1):144 [FREE Full text] [doi: 10.1186/s12913-021-06122-y] [Medline: 33588842]
53. Medlinskiene K, Tomlinson J, Marques I, Richardson S, Stirling K, Petty D. Barriers and facilitators to the uptake of new medicines into clinical practice: a systematic review. BMC Health Serv Res 2021 Nov 05;21(1):1198 [FREE Full text] [doi: 10.1186/s12913-021-07196-4] [Medline: 34740338]
54. Aquino M, Griffith J, Vattaparambil T, Munce S, Hladunewich M, Seto E. Patients' and providers' perspectives on and needs of telemonitoring to support clinical management and self-care of people at high risk for preeclampsia: qualitative study. JMIR Hum Factors 2022 Feb 07;9(1):e32545 [FREE Full text] [doi: 10.2196/32545] [Medline: 35129445]
55. Bhatt V, Chakraborty S. Intrinsic antecedents to mHealth adoption intention: an SEM-ANN approach. Int J Electron Gov Res 2022 Mar;18(2):1-17 [FREE Full text] [doi: 10.4018/ijegr.298139]
56. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
57. Fleddermann K, Molfenter T, Jacobson N, Horst J, Roosa MR, Boss D, et al. Clinician perspectives on barriers and facilitators to implementing e-health technology in substance use disorder (SUD) treatment facilities. Subst Abuse 2021 Oct 26;15:11782218211053360 [FREE Full text] [doi: 10.1177/11782218211053360] [Medline: 34720585]
58. Jackson J, Leziak K, Niznik CM, Yee LM. Health care providers' utilization of and perspectives on mobile health technology for diabetes and pregnancy support. Diabetes Spectr 2021 Aug;34(3):257-267 [FREE Full text] [doi: 10.2337/ds20-0040] [Medline: 34511852]
59. Li J, Silvera-Tawil D, Varnfield M, Hussain MS, Math V. Users' perceptions toward mHealth technologies for health and well-being monitoring in pregnancy care: qualitative interview study. JMIR Form Res 2021 Dec;5(12):e28628 [FREE Full text] [doi: 10.2196/28628] [Medline: 34860665]
60. Mansour E. A survey of Egyptian physicians' awareness and use of coronavirus-related mHealth applications. Inf Dev 2021 Oct;39(2):271-286 [FREE Full text] [doi: 10.1177/02666669211049494]
61. Wu P, Zhang R, Luan J, Zhu M. Factors affecting physicians using mobile health applications: an empirical study. BMC Health Serv Res 2022 Jan 04;22(1):24 [FREE Full text] [doi: 10.1186/s12913-021-07339-7] [Medline: 34983501]
62. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]

63. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
64. Jiang Y, Lau AK. Understanding post-adoption behavioral intentions of mobile health service users: an empirical study during COVID-19. Int J Environ Res Public Health 2023 Feb 22;20(5):3907 [FREE Full text] [doi: 10.3390/ijerph20053907] [Medline: 36900918]
65. Luo Y, Wang G, Li Y, Ye Q. Examining protection motivation and network externality perspective regarding the continued intention to use M-Health apps. Int J Environ Res Public Health 2021 May 26;18(11):5684 [FREE Full text] [doi: 10.3390/ijerph18115684] [Medline: 34073247]
66. de Oliveira Andrade A, Soares AB, de Andrade Palis A, Cabral AM, Barreto CG, de Souza DB, et al. On the use of telemedicine in the context of COVID-19: legal aspects and a systematic review of technology. Res Biomed Eng 2021 Mar 13;38(1):209-227 [FREE Full text] [doi: 10.1007/s42600-021-00133-8]
67. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. Health Equity 2018 Aug;2(1):174-181 [FREE Full text] [doi: 10.1089/heq.2018.0037] [Medline: 30283865]
68. Mossialos E, Courtin E, Naci H, Benrimoj S, Bouvy M, Farris K, et al. From "retailers" to health care providers: transforming the role of community pharmacists in chronic disease management. Health Policy 2015 May;119(5):628-639. [doi: 10.1016/j.healthpol.2015.02.007] [Medline: 25747809]
69. Elhadi M, Msherghi A, Alkeelani M, Alsuyihili A, Khaled A, Buzreg A, et al. Concerns for low-resource countries, with under-prepared intensive care units, facing the COVID-19 pandemic. Infect Dis Health 2020 Nov;25(4):227-232 [FREE Full text] [doi: 10.1016/j.idh.2020.05.008] [Medline: 32631682]
70. Getachew E, Woldeamanuel Y, Manyazewal T. Capacity and readiness assessment of healthcare facilities for digital health interventions against tuberculosis and HIV in Addis Ababa, Ethiopia. Front Digit Health 2022 Feb 28;4:821390 [FREE Full text] [doi: 10.3389/fdgth.2022.821390] [Medline: 35295619]
71. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. 3rd edition. Thousand Oaks, CA: Sage Publications; 2018.

Abbreviations

MeSH: Medical Subject Headings
mHealth: mobile health
PICO: population, intervention, comparator, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QuADS: Quality Assessment with Diverse Studies
WHO: World Health Organization

Edited by L Buis; submitted 30.06.23; peer-reviewed by W Wei, K Masters; comments to author 15.08.23; revised version received 13.09.23; accepted 04.10.23; published 08.11.23.

Please cite as:
Alsahli S, Hor SY, Lam M
Factors Influencing the Acceptance and Adoption of Mobile Health Apps by Physicians During the COVID-19 Pandemic: Systematic Review
JMIR Mhealth Uhealth 2023;11:e50419
URL: https://mhealth.jmir.org/2023/1/e50419
doi:10.2196/50419
PMID:37938873

Review

Collection and Analysis of Adherence Information for Software as a Medical Device Clinical Trials: Systematic Review

Emily Grayek[1], PhD; Tamar Krishnamurti[2], PhD; Lydia Hu[1], MSc; Olivia Babich[3], MSc; Katherine Warren[4], BSc; Baruch Fischhoff[5], PhD

[1]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, United States
[2]Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[3]University of Pittsburgh School of Medicine, Pittsburgh, PA, United States
[4]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, United States
[5]Department of Engineering and Public Policy, Institute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, PA, United States

Corresponding Author:
Emily Grayek, PhD
Department of Engineering and Public Policy
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213
United States
Phone: 1 (412) 268 2000
Email: egrayek@andrew.cmu.edu

Abstract

Background: The rapid growth of digital health apps has necessitated new regulatory approaches to ensure compliance with safety and effectiveness standards. Nonadherence and heterogeneous user engagement with digital health apps can lead to trial estimates that overestimate or underestimate an app's effectiveness. However, there are no current standards for how researchers should measure adherence or address the risk of bias imposed by nonadherence through efficacy analyses.

Objective: This systematic review aims to address 2 critical questions regarding clinical trials of software as a medical device (SaMD) apps: How well do researchers report adherence and engagement metrics for studies of effectiveness and efficacy? and What efficacy analyses do researchers use to account for nonadherence and how appropriate are their methods?

Methods: We searched the Food and Drug Administration's registration database for registrations of repeated-use, patient-facing SaMD therapeutics. For each such registration, we searched ClinicalTrials.gov, company websites, and MEDLINE for the corresponding clinical trial and study articles through March 2022. Adherence and engagement data were summarized for each of the 24 identified articles, corresponding to 10 SaMD therapeutics. Each article was analyzed with a framework developed using the Cochrane risk-of-bias questions to estimate the potential effects of imperfect adherence on SaMD effectiveness. This review, funded by the Richard King Mellon Foundation, is registered on the Open Science Framework.

Results: We found that although most articles (23/24, 96%) reported collecting information about SaMD therapeutic engagement, of the 20 articles for apps with prescribed use, only 9 (45%) reported adherence information across all aspects of prescribed use: 15 (75%) reported metrics for the *initiation* of therapeutic use, 16 (80%) reported metrics reporting adherence between the initiation and discontinuation of the therapeutic (*implementation*), and 4 (20%) reported the discontinuation of the therapeutic (*persistence*). The articles varied in the reported metrics. For trials that reported adherence or engagement, there were 4 definitions of initiation, 8 definitions of implementation, and 4 definitions of persistence. All articles studying a therapeutic with a prescribed use reported effectiveness estimates that might have been affected by nonadherence; only a few (2/20, 10%) used methods appropriate to evaluate efficacy.

Conclusions: This review identifies 5 areas for improving future SaMD trials and studies: use consistent metrics for reporting adherence, use reliable adherence metrics, preregister analyses for observational studies, use less biased efficacy analysis methods, and fully report statistical methods and assumptions.

(JMIR Mhealth Uhealth 2023;11:e46237) doi:10.2196/46237

KEYWORDS

mobile health; mHealth; adherence; evaluation; usability; efficacy; systematic review; application; compliance; safety; effectiveness; engagement; risk; medical device; clinical trials

Introduction

Background

There are over 350,000 health-related apps on the market, each claiming to improve certain aspects of physical or mental health [1]. A small fraction of these apps is subject to Food and Drug Administration (FDA) regulations. Regulators, health care providers, and patients need to understand how these apps compare with alternatives (eg, pharmaceuticals) that undergo rigorous evaluation. As with pharmaceuticals, the risks and benefits of apps depend on how well people use them. Incorrect assumptions about adherence in clinical trials can lead to incorrect regulatory and treatment decisions. With pharmaceuticals, these risks are reduced by the gold standard practice of intent-to-treat analysis, which estimates effectiveness based on actual, typically imperfect, use. This standard is not the norm in trials of digital health apps, leading to an unknown risk of bias (ROB) in the estimated effects. Here, we provide a systematic review of current practices in FDA-regulated apps, leading to recommendations for reducing the risks of bias revealed by the review.

The FDA focuses on the regulation of software as a medical device (SaMD) therapeutics intended to prevent, diagnose, or treat diseases [2]. If a predicate therapeutic exists, applicants may use the FDA's 510k pathway to prove that their therapeutic is substantially equivalent to the predicate therapeutic (ie, with the same intended use, technological characteristics, and benefits and risks of an approved or cleared therapeutic [3]). In the absence of a predicate therapeutic, SaMD therapeutics follow the FDA's De Novo pathway, which requires evidence that the therapeutic is safe and effective. The FDA established the Digital Health Center of Excellence to create innovative ways to regulate SaMDs [4], which, for example, are easier to update than pharmaceuticals. One such innovation, reviewed under the FDA's precertification pilot program, conducted *excellence appraisals* of software companies. This program tested a streamlined approach to approving and updating therapeutics for companies that have demonstrated quality practices [5,6]. Other innovations have been applied across all FDA departments, such as allowing clearance, approval, and marketing claims based on "real-world evidence" [7]. There are also proposals, created outside FDA, specifying standard processes (eg, performance reporting standards) for clinical trials of low-risk digital health apps not subject to regulatory oversight [8]. Given the novelty of SaMDs and the associated regulatory environment, the FDA has the need and opportunity to create guidance and requirements for addressing adherence in future trials. We hope to inform that process.

A systematic review by Milne-Ives et al [9] found that approximately three-fourths of digital health app trials collected and reported basic adherence information, such as the number of dropouts. These trials reported a variety of app engagement metrics, with only one-third reporting >60% use. Prior systematic reviews of digital health apps reported similar simple summary statistics (eg, average adherence and dropout rates), with few details on how adherence data were collected and analyzed [9-14]. This systematic review extends that work by examining, in detail, how adherence and engagement information is collected, analyzed, and reported. It considers how those practices affect the estimates of *effectiveness* and *efficacy*, defined as the app's effect in the entire sample, regardless of adherence, and the app's effect in the adherent subgroup, reflecting the moderating effect of adherence. This review focuses on digital health apps with a reasonably well-defined evidentiary base, namely, those that followed the FDA's De Novo or 510k pathways.

Criteria for Evaluation

ROB Framework

Imperfect adherence can cause underestimation or overestimation of the safety and efficacy of a SaMD. For example, a therapeutic's efficacy and side effects may be underestimated, if trial participants use it sparingly, but consistent use is assumed. Conversely, efficacy may be overestimated if adherence reflects neglected confounding variables (eg, income and lifestyle factors). As a hypothetical example, researchers evaluating an app to reduce the risk of preeclampsia may observe a reduced rate not because of participant adherence but because participants adhering to the app were recipients of commercial health insurance. To evaluate the ROB owing to imperfect adherence, we used the adherence components of the Cochrane ROB Assessment (version 2.0) [15], a well-documented tool for systematic reviews and meta-analyses. To determine the ROB from nonadherence, the ROB tool first asks, "Was there nonadherence to the assigned intervention regimen that could have affected participants' outcomes?" If outcomes could have been affected, the ROB tool then asks, "Was an appropriate analysis used to estimate the effect of adhering to the intervention?" We developed criteria to answer each question based on research regarding adherence metrics and common methods of analyzing efficacy.

Adherence and Engagement Metrics

Adherence refers to how well participants use an intervention, as defined by a protocol or recommendation for use. Engagement refers to how participants use an intervention, irrespective of the intended use of the app. Engagement data can be used to measure adherence for a digital health app. As both adherence and engagement can affect the outcomes of a trial, we have reported both. When collecting and reporting adherence and engagement statistics, researchers must consider 3 facets of use [16]: *initiation*, when a person starts using an intervention; *implementation*, how a person uses the intervention between initiation and discontinuation; and *persistence*, how long a person uses the intervention before discontinuation.

Which metrics are collected and how they are collected can also affect the ability to conduct efficacy analyses and the analyses' potential bias. For instance, adherence with recommendations from the therapeutic (eg, using backup contraception when an app detects fertility) could also affect effectiveness estimates. Without collecting this information, researchers would be unable to analyze efficacy in terms of adherence to behavioral recommendations. Therefore, we report adherence and engagement with both the therapeutic and its recommendations. The mechanism of collecting adherence and engagement information can act as a potential confounder if it prompts additional engagement with the therapeutic compared with real-world engagement. Reminders used to increase adherence (eg, email messages) can also be confounders if they are not part of the therapeutic design. To account for these potential confounders, we recorded whether reminders and mechanisms for measuring adherence and engagement were internal to the app or external (ie, an additional component not found in the marketed app). We found few prior studies or analysis plans that determined the level of adherence or engagement required to have a clinical effect. This level of adherence can vary depending on the therapeutic being used. Without a study or trial analysis plan defining low adherence or evidence of the level of adherence needed to produce a clinical effect, we cannot conclusively assess whether adherence is low or not because of insufficient information.

Analysis of Efficacy

In evaluating efficacy analyses, we ask how well a trial or study fulfills the assumptions required by its efficacy analysis method. There are 3 commonly used estimates of efficacy: the *average treatment effect* (ATE), *per-protocol effect*, and *dose-response effect*. Table 1 describes each estimate, the common analysis methods for calculating estimates, and the assumptions required for unbiased estimates. Multimedia Appendix 1 [17-22] includes definitions of the following assumptions: consistency, positivity, ignorability, exclusion restriction, strong monotonicity, and the stable unit treatment value assumption (SUTVA). In addition to the requirements in Table 1, researchers should preregister their analyses of effectiveness and efficacy to reduce the risk of capitalization on chance [23].

Table 1. Methods of analysis commonly used to account for imperfect adherence and the assumptions required for unbiased estimates.

Estimate of efficacy and common analysis methods	Assumptions for unbiased estimates
ATE[a]: estimates the average effect of treatment	

• ATE analysis
  • Evaluates groups according to their treatment group regardless of adherence.
  • Estimates efficacy if adherence is modified with regular reminders to participants.
	• SUTVA[b]
	• Consistency[c]
	• Positivity
	• Ignorability

• ITT[d] analysis
  • Evaluates groups according to their assigned treatment regardless of adherence.
  • Estimates efficacy if adherence is modified with regular reminders to participants.
	• SUTVA
	• Consistency[c]
	• Randomization (fulfills positivity, exclusion restriction, and ignorability)

Per-protocol effect: estimates the average effect of adhering to the treatment assignment	

• Complier average causal effect or local average treatment effect
  • Evaluates the per-protocol effect for the adherent subpopulation.
  • Evaluates groups based on an adherence threshold. Nonadherent participants in the treatment group are labeled as never-takers. It is assumed that the effect of the never-takers is equal in both groups.
	• SUTVA
	• Consistency[c,e]
	• Randomization (fulfills positivity, ignorability, exclusion restriction, and strong monotonicity)

• Generalized estimation • Evaluates groups based on an adherence threshold. Groups are evaluated based on adherence over time such as never-takers, early-takers, late-takers, and always-takers.	• SUTVA • Consistency[c,e] • Positivity • Ignorability (sequential exchangeability)
• As-treated analysis • Evaluates groups based on an adherence threshold. Nonadherent participants in the treatment group are considered part of the control group.	• SUTVA • Consistency[c,e] • Positivity • Ignorability (conditional independence of adherence and outcomes)
• Per-protocol analysis • Evaluates groups based on an adherence threshold. Excludes nonadherent participants in the treatment group.	• SUTVA • Consistency[c,e] • Positivity • Ignorability (conditional independence of adherence and outcomes)
Dose-response effect: estimates the effect of adherence on the treatment	
• Dose-response analysis (IV[f] method) • Evaluates adherence as a mediator for all participants using an IV to fulfill the mechanism ignorability assumption.	• SUTVA • Consistency[c,e] • Randomization (fulfills positivity, ignorability, exclusion restriction, and strong monotonicity)
• Dose-response analysis (confounder adjustment) • Evaluates adherence as a mediator for all participants using confounder adjustment to fulfill the mechanism ignorability assumption.	• SUTVA • Consistency[c,e] • Positivity • Ignorability (conditional independence of adherence and outcomes)

[a]ATE: average treatment effect.

[b]SUTVA: stable unit treatment value assumption.

[c]Consistent definition of treatment.

[d]ITT: intent-to-treat.

[e]Consistent definition of adherence.

[f]IV instrumental variable.

We applied our framework, which was developed based on the Cochrane ROB, to evaluate how well existing trials and studies meet our standards, with the goal of improving future trials. We examined the completeness of their reporting and the appropriateness of the procedures reported. By focusing on SaMD therapeutics, the most rigorously evaluated digital health apps, we sought to identify improvements for future studies on all digital health apps.

Methods

Screening

A 2-stage search strategy was used to identify all product codes and registrations for patient-facing SaMDs, with intended repeated use for at least 2 weeks, that the FDA had approved or cleared before March 2022. In the first stage, 2 reviewers independently searched the FDA product code database for product codes related to SaMDs. We searched the device name, definition, physical state, and technical method attributes for the keywords "software," "mobile," "digital," and "application." In the second stage, we searched the FDA registration database for these product codes. We examined each registration's supporting documents, De Novo decision summaries, and 510k decision summaries to determine whether the product met our inclusion criteria.

We then searched ClinicalTrials.gov, product websites, and MEDLINE for peer-reviewed publications corresponding to each included product. For the ClinicalTrials.gov search, we used the product and company names as keywords, individually and in combination, to identify clinical trials. We included all publications that evaluated the effectiveness or efficacy of the included products, including both randomized controlled trials (RCTs) and observational studies. We reviewed all publications listed at the end of the ClinicalTrials.gov registration for potential inclusion. For the MEDLINE search, product and company names were used as keywords. For the product website search, publications listed as clinical evidence on company websites were included. Two reviewers independently screened each publication, examining the title and abstract as well as the full text, where appropriate. Reviewer disagreements were reconciled by discussion. We screened and included only those articles published before March 2022. We did not include pilot or feasibility studies.

For example, the first stage of the search identified the PYT product code when the "device name" field was searched for "software." All registrations coded as PYT (ie, "Device, Fertility Diagnostic, Contraceptive, Software Application") were then evaluated for inclusion based on corresponding supporting documents, 510k decision summaries, and De Novo decision summaries. One included 510k for this product code was for the Clue app, K193330. In the second stage, we searched ClinicalTrials.gov using the keywords "Clue," "Clue Birth Control," "Biowink," "Dynamic Optimal Timing," and "Cycle Technologies." We searched MEDLINE using the keywords "Dynamic Optimal Timing," "Biowink," and "NCT02833922." Finally, we searched the product website [24] for clinical trial documents.

Data Extraction

For each publication, one reviewer extracted data and the other reviewer checked the accuracy of the data. Differences were reconciled by discussion between the reviewers. The Cochrane Data Collection Form for Intervention Reviews [25] was completed with clinical trial characteristics, including the design, number of participants, sampling method, interventions, and outcomes.

The remainder of the data extraction form was created using the criteria for reporting adherence metrics described in the *Adherence and Engagement Metrics* section and the assumptions for the associated efficacy analysis method described in the *Analysis of Efficacy* section. Given the diversity of the apps and outcomes, we reported each metric that a clinical trial or study reported separately, without averaging across different metrics. When evaluating efficacy analyses, we categorized trials or studies as fulfilling the positivity condition if they had a control group. We categorized trials as fulfilling the consistency condition if they had definitions of treatment and adherence that avoided hidden variations of treatment that might affect participants differently.

Some assumptions, referenced in Table 1 and described in Multimedia Appendix 1, could not be fully evaluated. One such assumption is SUTVA, which requires no interaction between units of observation that could affect a result. Although it is impossible to prove that this assumption holds, some trial designs afford greater confidence than others. For example, if a trial has no central clinical team and treatment is administered only through an app, it would be difficult for participants to interact with the clinical research staff. By contrast, if clinical research staff interact with both the control and treatment groups, they might treat participants in the 2 groups in ways that affect their independence. We categorized a trial as fulfilling SUTVA if it had no central clinical team or if it had mechanisms for reducing the risk of interaction between participants or between participants and staff.

Similarly, it is impossible to fully evaluate the assumption that there are no unmeasured confounders. Instead, we asked whether the researchers demonstrated awareness of confounders by listing potential confounders explicitly and reporting their rationale for selecting them.

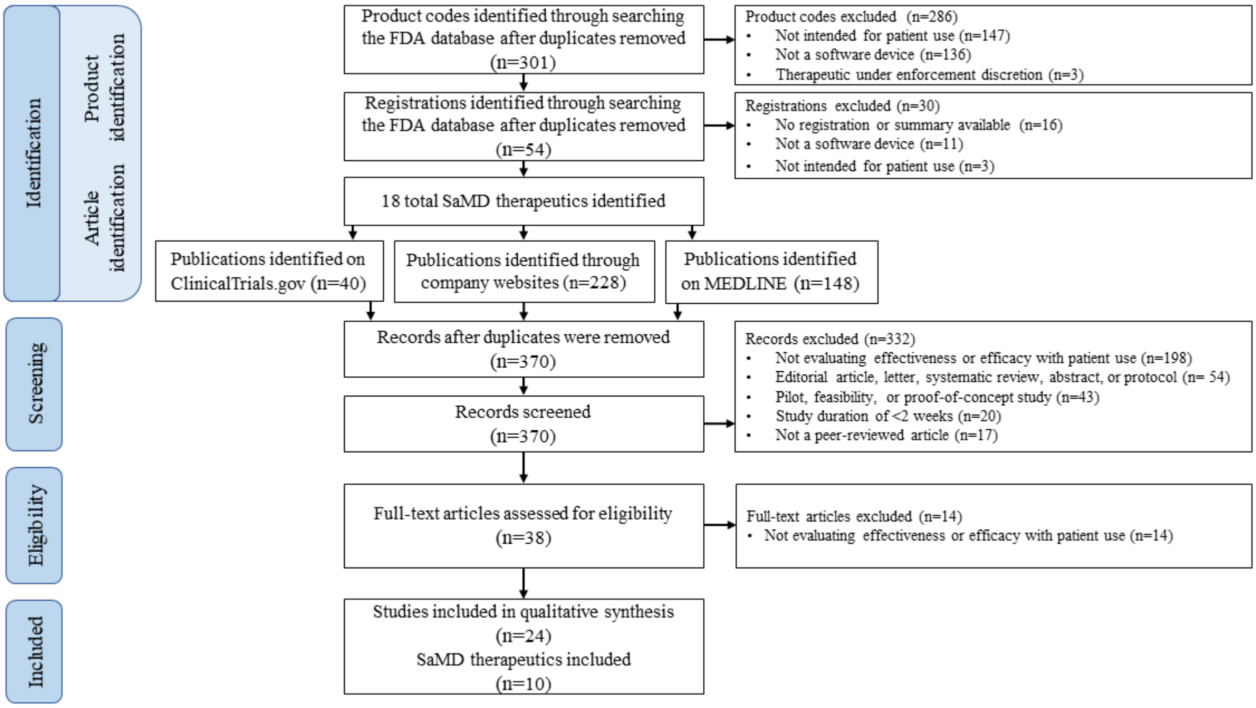
The results in the *Adherence Metrics* section and *Analysis of Efficacy* section below summarize practices for the included trials using means or counts as appropriate. Given the heterogeneity of the therapeutics and outcomes, we did not estimate the overall impact of all biases. The protocols and preregistrations referenced in the included articles were used as supporting documents. The protocol for this review was registered on the Open Science Framework [26], which includes the data extraction forms and extracted data. Article screening data, extracted data, and summarized extracted data are also available in Multimedia Appendices 2-4 [27-50].

Results

Included Trials

Figure 1 shows the completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. The 2-stage search for SaMD therapeutics identified 5% (15/301) of product codes and 44% (24/54) of registrations as potential SaMDs. These registrations included 18 unique SaMD therapeutics. Our search of ClinicalTrials.gov, company websites, and MEDLINE identified 40, 228, and 148 articles, respectively. After screening and removal of duplicate articles, 24 articles, involving 10 products, met all the inclusion criteria. A total of 8 products were excluded because clinical trials or observational studies evaluating efficacy for at least 2 weeks were not found in our literature search.

Figure 1. The 2-stage strategy used to identify trials and studies of software as a medical device (SaMD) therapeutics. The Food and Drug Administration (FDA) databases were first searched for SaMD therapeutics that would be used by patients for at least 2 weeks. In the second stage, ClinicalTrials.gov, MEDLINE, and company websites were then searched for articles evaluating effectiveness or efficacy for these products when used by patients for at least 2 weeks.



As seen in Tables 2 and 3, the 24 included articles (22 total trials) studied a variety of SaMD therapeutics, including those intended to treat irritable bowel syndrome, insomnia, substance use disorder, and attention-deficit/hyperactivity disorder. All the SaMD therapeutics were mobile apps and will be referred to as apps for the remainder of the article. Table 3 shows an even mix of apps intended for continual use or module-based apps. Most trials (18/22, 82%) specified a recommended dose for their app, such as the frequency of use or the number of modules to complete. Overall, 11 (50%) trials or studies studied apps used a module-based design with a recommended dose for the app [27,34-39,45-50], whereas 7 (32%) trials or studies studied apps used a continual use design with a recommended dose for the app [31,32,40-44]. Apps without a recommended dose only used the continual use design (4/22, 18%) [28-30,33].

Table 2. Included articles and associated products.

Product and condition treated	Study, year	Title
Apple Irregular Arrhythmia Notification		
Irregular arrhythmia notification	Perez et al [27], 2019	Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation
BlueStar		
Diabetes management	Quinn et al [28], 2011	Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control
Diabetes management	Agarwal et al [29], 2019	Mobile App for Improved Self-Management of Type 2 Diabetes: Multicenter Pragmatic Randomized Controlled Trial
Diabetes management	Dugas et al [30], 2020	Engagement and Outcomes Associated with Contextual Annotation Features of a Digital Health Solution
Clue		
Contraceptive	Jennings et al [31], 2018	Estimating six-cycle efficacy of the Dot app for pregnancy prevention
Contraceptive	Jennings et al [32], 2019	Perfect- and typical-use effectiveness of the Dot fertility app over 13 cycles: results from a prospective contraceptive effectiveness trial
DexCom G6		
Diabetes management	Akturk et al [33], 2021	Real-World Evidence and Glycemic Improvement Using Dexcom G6 Features
EndeavorRx		
Videogame treatment for ADHD[a]	Kollins et al [34], 2020	A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial
Videogame treatment for ADHD	Kollins et al [35], 2021	Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD
Videogame treatment for ADHD	Gallen et al [36], 2022	Enhancing neural markers of attention in children with ADHD using a digital therapeutic
Mahana		
CBT[b] for IBS[c]	Everitt et al [37][d], 2019	Assessing telephone-delivered cognitive–behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial
CBT for IBS	Everitt et al [38][d], 2019	Therapist telephone-delivered CBT and web-based CBT compared with treatment as usual in refractory irritable bowel syndrome: the ACTIB three-arm RCT
CBT for IBS	Everitt et al [39], 2019	Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial
Natural Cycles		
Contraceptive	Berglund Scherwitzl et al [40], 2016	Fertility awareness-based mobile application for contraception
Contraceptive	Berglund Scherwitzl et al [41], 2017	Perfect-use and typical-use Pearl Index of a contraceptive mobile app
Contraceptive	Bull et al [42], 2019	Typical use effectiveness of Natural Cycles: postmarket surveillance study investigating the impact of previous contraceptive choice on the risk of unintended pregnancy
Contraceptive	Pearson et al [43], 2021	Natural Cycles app: contraceptive outcomes and demographic analysis of UK users
Contraceptive	Pearson et al [44], 2021	Contraceptive Effectiveness of an FDA-Cleared Birth Control App: Results from the Natural Cycles U.S. Cohort
ReSet		
CBT for SUD[e]	Campbell et al [45], 2014	Internet-delivered treatment for substance abuse: a multisite randomized controlled trial
ReSet-O		
CBT for OUD[f]	Christensen et al [46][g], 2014	Adding an Internet-delivered treatment to an efficacious treatment package for opioid dependence

Product and condition treated	Study, year	Title
CBT for OUD	Maricich et al [47], 2021	Real-world evidence for a prescription digital therapeutic to treat opioid use disorder
CBT for OUD	Maricich et al [48], 2021	Real-world use and clinical outcomes after 24 weeks of treatment with a prescription digital therapeutic for opioid use disorder
CBT for OUD	Maricich et al [49][g], 2021	Safety and efficacy of a prescription digital therapeutic as an adjunct to buprenorphine for treatment of opioid use disorder
Somryst		
CBT for Insomnia	Ritterband et al [50], 2017	Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up A Randomized Clinical Trial

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]CBT: cognitive behavioral therapy.

[c]IBS: irritable bowel syndrome.

[d]Everitt et al [37] and Everitt et al [38] were based on the same trial.

[e]SUD: substance use disorder.

[f]OUD: opioid use disorder.

[g]Christensen et al [46] and Maricich et al [49] were based on the same trial.

Table 3. Summary of devices and trials included in the study (n=22).

Characteristics	Values
Therapeutic indication for use, n (%)	
Contraceptive	7 (32)
Videogame treatment for ADHD[a]	3 (14)
Irregular arrhythmia notification	1 (5)
Diabetes management	4 (18)
Cognitive behavioral therapy	7 (32)
IBS[b]	2 (9)
Insomnia	1 (5)
Substance use disorder	4 (18)
Type of therapeutic, n (%)	
Recommended use with module design	11 (50)
Recommended use with continual use design	7 (32)
No recommended use with module design	0 (0)
No recommended use with continual use design	4 (18)
Trial design	
RCT[c] [28,29,34,37-39,45,46,50], n (%)	8 (36)
Participants (in comparison groups), mean (SD)	290 (120)
Trial length (d), mean (SD)	300 (270)
Observational [27,30-33,35,36,40-44,47-49], n (%)	14 (64)
Participants (in comparison groups), mean (SD)	5100 (7000)
Trial length (d), mean (SD)	230 (140)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]IBS: irritable bowel syndrome.

[c]RCT: randomized controlled trial.

Most trials (14/22, 64%) were observational, with the remainder being RCTs (8/22, 36%). On average, the RCTs recruited 290 (SD 120) participants and lasted 300 (SD 270) days. On average, the observational trials recruited 5100 (SD 7000) participants and lasted 230 (SD 140) days.

Adherence Metrics

Table 4 summarizes how the articles measured and reported each of the 3 aspects of adherence. As each article could report different adherence metrics for the same trial or study and report separate analyses, duplicate trials and studies were counted twice. Of the 24 articles, 23 (96%) collected information about app engagement. All apps that provided recommendations (8/8, 100%) also collected information about adherence to their recommendations [27,31,32,40-44]. Of the 23 articles that collected adherence information, 2 (9%) reported that adherence information was collected externally from the marketed app [31,32]. Three articles reported that researchers attempted to increase adherence by notifying inactive patients [34-36]. One reported the use of in-app notifications and 2 reported using email notifications.

Table 4. Summary of adherence metrics (N=24)[a].

Adherence metrics	Values, n (%)	Each reported metric (%), mean (SD)
Trial collected information about app engagement	23 (96)	N/A[b]
Trial collected information about adherence to recommendations (n=8 articles for apps that gave recommendations)	8 (100)	N/A
Adherence information collected outside of the marketed app (n=23 articles for apps that collected adherence information)	2 (9)	N/A
Adherence notification sent outside of app (n=3 articles reported sending adherence notifications)	2 (67)	N/A
Engagement metrics (metric is not measuring prescribed use)		
Initiation	2 (8)	N/A
Initial app use, core completion, or activity use [30,33]	2 (8)	52 (35)
Implementation	2 (8)	N/A
Completed sessions, modules, or activities [29,30]	2 (8)	20 (22)
Log-in days [29]	1 (4)	23[c]
Persistence	7 (29)	N/A
Percentage of participants continuing use at 1 y [31,32,40,41,43,44]	6 (25)	52 (12)
Number of days participants used the app [30]	1 (4)	153[c]
Adherence metrics (metric is measuring prescribed use)		
Initiation	15 (63)	N/A
Provided at least 20 d of data [40-44]	5 (21)	100 (0)
Initial app use, core completion, or activity use [35,36,38,47-50]	7 (29)	98 (4)
Entered at least 2 period start dates [31,32]	2 (8)	100 (0)
Initiation of video in response to app alert [27]	1 (4)	44[c]
Implementation	16 (67)	N/A
Completed sessions, modules, or activities [34-36,45,49]	5 (21)	88 (16)
Completed at least 4 sessions and 1 call [37-39]	3 (13)	64 (5)
Completed half of the modules [47,48]	2 (8)	76 (13)
Completed ≥8 core modules [47,48]	2 (8)	87 (9)
Percentage of logged intercourse on red days [43,44]	2 (8)	23 (0)
Percentage of total days intercourse logged on red days (ie, days where the user did not follow app recommendations) [42]	1 (4)	2[c]
Percentage of perfect use cycles (ie, menstruation cycles where the user followed all trial recommendations) [32,41]	2 (8)	17 (10)
Log-in days [40,43,44]	3 (13)	47 (19)
Persistence	4 (17)	N/A
Participants using the app at week 12 [47,48]	2 (8)	4 (17)
Completed all core modules [38,47,48,50]	2 (8)	49 (19)
Study reported all prescribed facets of adherence (n=20 studies that prescribed a recommended use of the app)	9 (45)	N/A

[a]The left-hand columns report what percentage of articles reported adherence or engagement information and what metrics were used by each article. The right-hand columns report the mean and SD for all the articles that reported that metric.

[b]N/A: not applicable for summary of facets of adherence.

[c]SD values are not applicable as only 1 article was included.

A total of 4 articles studied a product without prescribing how often to use the app. Engagement was reported in 3 articles on these products. Of the 24 articles, engagement was reported for 2 (8%) in terms of initiation, 2 (8%) in terms of implementation, and 1 (4%) in terms of persistence. Two continual use therapeutics prescribed app use in terms of initiation and implementation but not persistence. As such, 25% (6/24) of the articles studying these apps reported engagement persistence metrics.

Of the 24 articles, 15 (63%) reported initiation in 4 different ways (eg, the number of users who finished the first app module and the number of users who entered 20 data points into the app). Seven articles excluded participants who did not initiate app use, leading to a high adherence for their adherence metrics. Of the 24 articles, 16 (67%) reported implementation, with 9 different definitions (eg, proportion of days between starting and stopping the use of an app that users logged their temperature and the number of perfect use cycles reported by women [ie, abstaining or using contraception on all high-risk days]). Of the 24 articles, 4 (17%) reported persistence, with 2 different definitions (participants using the app over the prescribed period and participants completing the prescribed number of modules). Table 4 reports the percentage of studies and the average adherence across trials and studies that used each metric. Of the 20 articles that prescribed use of the app, only 9 (45%) reported all prescribed facets of adherence [32,39-44,47,48].

ROB: "Nonadherence to the Assigned Intervention Regimen"

Of the 24 articles, 4 (17%) only reported engagement information, as there was no prescribed amount of app use. We found that the outcomes of the remaining articles could have been affected by nonadherence. Of the 83% (20/24) of articles for apps with prescribed use, 25% (5/20) reported adherence at or below their definition of low adherence for at least 1 facet of adherence. Of the remaining 15 articles, 12 (80%) reported that there was some nonadherence with the app for any prescribed facet of adherence or the app's behavior recommendations but did not provide a definition of low adherence. These articles provided insufficient information to determine whether adherence was sufficient for each app. The remaining 3 articles did not report sufficient information about each prescribed facet of adherence to judge adherence.

Analysis of Efficacy

Table 5 summarizes the effectiveness and efficacy estimates from each article. Of the 24 articles, 20 (83%) estimated the app's effectiveness as the ATE for all participants. Of these 20 articles, 11 (55%) preregistered their analysis of effectiveness. A higher percentage of RCTs preregistered their effectiveness analysis (7/9, 78%) compared with observational studies (4/11, 36%). Of the 24 articles, 15 (63%) estimated efficacy in terms of the ATE, per-protocol effect, or dose-response effect. Of these 15 articles, only 5 (33%) preregistered an efficacy analysis. Preregistration was more common for RCTs (3/6, 50%) than for observational trials (2/9, 22%).

Table 5. Summary of efficacy estimates (N=24).

Efficacy estimates	Values, n (%)	References
Effectiveness estimate	20 (83)	—[a]
None	4 (17)	[30,33,34,36]
Average treatment effect	20 (83)	[27-29,31,32,35,37-50]
Preregistered effectiveness analysis (n=20)	11 (55)	—
RCT[b] (n=9)	7 (78)	[28,29,37,39,45,49,50]
Observational (n=11)	4 (36)	[27,31,32,35]
Efficacy estimate	15 (63)	—
None	9 (38)	[27,28,31,35,40,42,45,49]
Average treatment effect	2 (8)	[34,36]
Per-protocol effect	10 (42)	[30,32,33,37-39,41,43,44,50]
Dose-response effect	3 (13)	[29,47,48]
Preregistered efficacy analysis (n=15)	5 (33)	—
RCT (n=6)	3 (50)	[34,37,39]
Observational (n=9)	2 (22)	[32,36]

[a]References not listed for summary rows.

[b]RCT: randomized controlled trial.

Table 6 characterizes the articles in terms of how well they meet the assumptions for their method of analysis. Of the 24 articles, 2 (8%) estimated efficacy in terms of ATE [34,36]. One of them used intent-to-treat analysis and met the relevant reporting requirement [34], and the other article calculated the ATE for an observational trial [36]. It met the criteria for SUTVA and had a clear definition of the treatment condition. However, it did not meet the positivity condition and lacked a control condition. The study adjusted for 1 confounder without saying how it was chosen.

Table 6. Fulfillment of required assumptions for efficacy analyses (n=14).

Estimate category, analysis method, and article	SUTVA[a], n (%)	Positivity, n (%)	Consistency, n (%)		Exclusion restriction, n (%)	Strong monotonicity, n (%)	Assignment mechanism ignorability			
			Clear treatment definition	Clear adherence definition			Randomization, n (%)	Conditional independence of treatment and outcomes	Sequential exchangeability	Conditional independence of adherence and outcomes
								Control variables	Control variables	Control variables
Average treatment effect (n=2)										
Intent-to-treat analysis (n=1)	1 (100)	1 (100)	1 (100)	NR[b]	1 (100)	1 (100)	1 (100)	NR	NR	NR
Kollins et al [34] (n=1)	1 (100)	1 (100)	1 (100)	NR	1 (100)	1 (100)	1 (100)	NR	NR	NR
Average treatment effect analysis (n=1)	1 (100)	0 (0)	1 (100)	NR	NR	NR	NR	NR	NR	NR
Gallen et al [36] (n=1)	1 (100)	0 (0)	1 (100)	NR	NR	NR	NR	Basic response time	NR	NR
Per-protocol effect (n=9)										
Complier average causal effect analysis (n=1)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	NR	NR	NR
Everitt et al [37,38] (n=1)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	NR	NR	NR
Generalized estimation (n=0)	—[c]	—	—	—	NR	NR	NR	—	NR	—
As-treated analysis (n=3)	2 (67)	1 (33)	3 (100)	3 (100)	NR	NR	NR	NR	NR	N/A[d]
Ritterband et al [50] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	Baseline ISI[e]
Dugas et al [30] (n=1)	1 (100)	1 (100)	1 (100)	1 (100)	NR	NR	NR	NR	NR	Time and demographic characteristics
Akturk et al [33] (n=1)	1 (100)	0 (0)	0 (0)	0 (0)	NR	NR	NR	NR	NR	None
Per-protocol analysis (n=5)	5 (100)	1 (20)	5 (100)	5 (100)	NR	NR	NR	NR	NR	N/A
Everitt [39] (n=1)	1 (100)	1 (100)	1 (100)	1 (100)	NR	NR	NR	NR	NR	Known baseline predictors of missingness at 12 months (IMD[f] and IBS-SSS[g])
Berglund Scherwitzl et al [41] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None

Estimate category, analysis method, and article	SUTVA[a], n (%)	Positivity, n (%)	Consistency, n (%)		Exclusion restriction, n (%)	Strong monotonicity, n (%)	Assignment mechanism ignorability			
			Clear treatment definition	Clear adherence definition			Randomization, n (%)	Conditional independence of treatment and outcomes	Sequential exchangeability	Conditional independence of adherence and outcomes
								Control variables	Control variables	Control variables
Jennings et al [32] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None
Pearson et al [43] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None
Pearson et al [44] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None
Dose-response effect (n=3)										
Dose-response analysis (IV[h]method; n=0)	—	—	—	—	—	—	—	—	—	—
Dose-response analysis (confounder adjustment method; n=3)	3 (100)	1 (33)	3 (100)	3 (100)	NR	NR	NR	NR	NR	N/A
Agarwal et al [29] (n=1)	1 (100)	1 (100)	1 (100)	1 (100)	NR	NR	NR	NR	NR	Baseline hemoglobin A_{1c}
Maricich et al [47] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None
Maricich et al [48] (n=1)	1 (100)	0 (0)	1 (100)	1 (100)	NR	NR	NR	NR	NR	None

[a]SUTVA: stable unit treatment value assumption.

[b]NR: not required (for the analysis method).

[c]No included articles used the analysis method.

[d]N/A: not applicable (count is not applicable for listed control variables).

[e]ISI: insomnia severity index.

[f]IMD: index of multiple deprivation.

[g]IBS-SSS: irritable bowel syndrome symptom severity score.

[h]IV: instrumental variable.

Of the 14 articles that estimated efficacy, 9 (64%) estimated efficacy in per-protocol effect terms (ie, treatment effect for adherent participants). One trial (2 articles) calculated the complier average causal effect (CACE), or local ATE (LATE), and provided evidence of meeting its assumptions [37,38]. Three articles used as-treated analysis [30,33,50]. Three of these articles had strong support for the SUTVA assumption. Two articles met the requirements for consistency, whereas the third article did not, as it defined treatment loosely. Two articles accounted for confounders but did not mention how they were chosen. Five articles used per-protocol analysis [32,39,41,43,44]. All articles had strong support for the SUTVA assumption and clear definitions of treatment and adherence. One article used an RCT design, provided evidence of positivity, and accounted for the baseline predictors of missingness. Four articles had no control cohort and did not account for any potential confounders of adherence.

Three articles estimated dose-response effects [29,47,48], treating adherence as a moderator. All 3 articles had strong support for the SUTVA assumption and provided clear definitions of treatment and adherence. In total, 33% (1/3) of the articles used an RCT design, providing evidence of positivity. This paper corrected for 1 confounder without saying how it was chosen.

ROB: "Analysis Used to Estimate the Effect of Adhering to the Intervention"

Of the 20 articles with a recommended dose, only 2 (10%) used an appropriate method of analysis to estimate the impact of nonadherence. Both reported on a trial that calculated CACE or LATE based on a preregistered plan, demonstrating compliance with its assumptions [37,38].

Discussion

Recommendations for Future Trials

Our systematic review of the SaMD literature found 24 articles evaluating the clinical evidence for 10 unique apps. These apps covered a breadth of treatment areas, risk levels, and prescribed uses.

Adherence Metrics

On the basis of our evaluation of adherence metrics, we identified the following key issues and opportunities to address them in future SaMD trials and studies:

1. Trial and study reporting was inconsistent. Many trials did not report all 3 facets of adherence. Trials used many definitions for each facet of adherence, limiting comparisons.
2. Some trials measure, analyze, and report adherence in ways likely to produce estimates inconsistent with those experienced with actual use.

As mentioned in the *Adherence Metrics* section of the results, most trials (23/24, 96%) collected some engagement information, but only a minority (9/20, 45%) reported all the prescribed facets of adherence. Most trials reported metrics for initiation (17/24, 71%) and implementation (18/24, 75%), and fewer trials (11/24, 46%) reported metrics for persistence. Persistence may have been reported less often because studies often reported persistence solely in terms of study dropout (adherence to trial or study protocols) but not discontinued app use. For example, 1 common outcome of trials evaluating an app treating substance use disorder was the number of days until the last face-to-face therapy session. This metric addresses 1 aspect of persistence for the treatment but neglects persistence for use of the app.

When an app offered behavior recommendations, adherence was often reported only for adherence to app recommendations or app use. For example, many contraceptive studies had complete reports on sexual activity but no reports on how often the temperature or cycle start information was logged. Such missing information could help physicians reviewing the literature to provide recommendations or warnings to patients regarding products with low adherence or engagement, better informing their patients' consumer choices.

Within just these few articles, there were many definitions of adherence, even for apps with similar treatment mechanisms or application areas. This variety limits the possibilities for meta-analysis or app comparisons (eg, is engagement higher when 75% of users complete half of the modules or when users complete 75% of the modules on average?). In an ideal world, patients or their care providers would be able to compare adherence and engagement metrics across similar apps to choose the app that has the best outcomes and highest levels of engagement.

We recommend that the FDA's guidance or voluntary standards determine which metrics should be collected and reported. Both guidance and standards could recognize that the most important metrics would vary across treatment areas and app design. The FDA's guidance could provide broad recommendations for researchers to collect and report adherence information for all prescribed facets of adherence. Voluntary standards for each treatment could benefit from further studies of engagement, which would identify which metrics are most important for each treatment area. Standards would enable developers, providers, and consumers to compare the usability and efficacy of similar apps and researchers to conduct meta-analyses for apps in a treatment area. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) and CONSORT (Consolidated Standards of Reporting Trials) provide examples of reporting protocols that could be adapted. The required information could be reported in the text, as supplemental material, or on an external site such as the Open Science Framework.

Several trials included external notifications intended to increase engagement that nontrial users would not have, such as manual notifications prompted by nonadherence. For example, most trials of an app treating attention-deficit/hyperactivity disorder sent email notifications to participants in response to nonactivity. This raises the question of how the app would perform without these notifications. The collection of trial-related information can itself introduce statistical bias if the information is solicited differently than would be the case with normal use. For example, studies on a contraceptive app collected information about sexual behavior daily within the study's app version, whereas the marketed version did not collect that information at the time of approval. Users in the trial might have been more engaged with the app; therefore, they were more likely to follow its recommendations compared with typical users. The marketed version of the app has since been updated to allow users to report sexual activity.

We recommend either regulatory guidance or voluntary standards that require trials to collect adherence information in ways similar to actual use. We recognize that, for some therapeutics, in-app notifications would not be practical or as effective. In such cases, regulators could demand postmarket evidence that measured effectiveness in real-world settings without manual notifications or external mechanisms for collecting adherence used in trials or studied are absent, aligns with measured effectiveness.

Analysis of Efficacy

On the basis of our evaluation of efficacy analysis, we identified the following key issues and opportunities to address them in future SaMD trials and studies:

1. Preregistration is common for analyses of effectiveness, especially for RCTs. Preregistration is uncommon for efficacy analyses of RCTs or observational studies.

2. Many articles analyzed efficacy with statistically biased methods.
3. Few articles reported evidence of meeting the assumptions for their efficacy analyses.

Preregistration is the accepted practice for protecting studies from *p-hacking* (ie, capitalization on chance), such as trying different analyses until a desired or expected result is obtained. Efficacy analyses for app trials are particularly susceptible to such practices, given their many variables and alternative definitions of metrics. For example, estimates of per-protocol effects depend on how researchers dichotomize participants into adherent and nonadherent groups. Different thresholds can produce different estimates.

Currently, regulators recommend that observational studies used to generate real-world evidence create and follow a protocol and analysis plan [7]. They also recommend that manufacturers follow a presubmission process to receive feedback on the plan. This consultation process is different than preregistration, where researchers publicly state their planned analyses and outcomes. Although both processes can reduce the risk of p-hacking, preregistration has the advantage of allowing the public to access study information in a standard, time-stamped manner.

In addition, we found that preregistration was more common for effectiveness analyses than for efficacy analyses. Effectiveness is likely to be the primary outcome of interest for a study and would be the primary concern for regulators reviewing an analysis plan. To protect against statistical bias, it is also important to specify in advance which analyses will be conducted in the case of low adherence and how low adherence will be defined. The analysis plan for 2 articles studying an app for the treatment of irritable bowel syndrome provides an excellent reference for defining adherence metrics and specifying analysis for low-adherence cases [37,38].

We recommend that regulators require preregistration of all app trials, in a standard format that specifies the planned metrics and analyses. Prespecified plans should include the analytical method used, any threshold for dichotomizing adherence, and the plan to account for confounders. Preregistrations should also address the conditions and methods for analyzing efficacy, such as the threshold for low adherence that would trigger efficacy analyses. Voluntary standards could mandate preregistration before updated regulatory requirements are implemented. Scientific journals might impose standards more quickly than regulatory bodies, as they have for preregistration of interventional trials [51,52].

Most trials studied efficacy and effectiveness using various methods. Most studies estimated a per-protocol effect using as-treated or per-protocol analysis, methods that studies have found to produce statistically biased results owing to insufficient adjustments for selection bias and confounders [53,54]. Our review also found insufficient accounting for confounders.

Only 1 trial (2 articles) used the preferred CACE, or LATE, analysis [37,38]. However, many other trials could have used this method, given their RCT designs. CACE, or LATE, analysis accounts for confounders without perfect knowledge of the relationships between outcomes and confounders using randomization as an instrumental variable.

We recommend that regulatory guidance or voluntary standards require less biased methods of estimating the per-protocol effects when the trial or study design allows. For example, an RCT using per-protocol analysis should use CACE, or LATE, analysis instead, treating access to treatment assignment as an instrumental variable. With observational studies, less biased instrumental variables approach methods are often not possible, given the lack of a control condition. In such cases, confounder adjustment could be used, with explicit acknowledgment of its limitations. Most articles reported satisfying the requirements for SUTVA, positivity, and consistency. However, the validity of all analytical methods also depends on satisfying ignorability, namely, accounting for confounders related to treatment and adherence, using an instrumental variable or confounder adjustment. Confounder adjustment is needed for ATE analysis with observational studies, as-treated analysis, per-protocol analysis, and dose-response analysis with confounder adjustment. Table 6 shows that confounders were often not even considered for these efficacy analyses. Even when confounders were considered, the rationale for choosing them was often not stated.

We recommend regulatory guidance or voluntary standards that clearly specify how researchers should choose and report confounders for efficacy analyses. Given that confounders will vary by treatment area, regulatory guidance should focus on general best practices, such as including transparent, preregistered methods for confounder selection. Whenever possible, confounders should be selected based on prior knowledge of causal relationships [55,56]. Voluntary standards could identify confounders for common treatment areas when such research exists. Without such research, empirical methods of confounder selection could be used, with the disclosure of potential bias in the selection method.

Conclusions

Most of the trials included in our systematic review report data suggesting nonadherence that could have affected the effectiveness of the app, without sufficiently evaluating efficacy in these circumstances. Appropriate use of SaMDs requires an understanding of how adherence could function as a moderator of the outcomes. Realistic, unbiased efficacy estimates are needed by regulators evaluating apps, health care providers potentially prescribing them, consumers deciding whether to use them (or seek other treatments), and vendors trying to improve their products.

The challenge of producing unbiased estimates will grow if real-world evidence studies are used more often to estimate the effectiveness of SaMD. Together, our findings illustrate the range (and inconsistencies) of the approaches used to measure and account for adherence. Without clear regulatory guidance or voluntary standards that specify how researchers should choose adherence metrics, perform efficacy analyses, and report their methods, it is unreasonable to expect that researchers will provide the information necessary to evaluate the potential effect of adherence on trial outcomes. More rigorous and consistent reporting and analyses are needed to facilitate decisions about individual products and to aggregate knowledge across products. Future SaMD clinical trials and studies may be improved by producing consensus standards on the definitions of adherence for similar products and studying the role of confounders for product areas. Without accurate efficacy estimates, SaMDs will not fulfill their potential to improve health outcomes with minimal risk.

Limitations and Future Work

Our review excluded qualitative and exploratory studies, thus potentially missing insights found in them. For example, exploratory studies might reveal how prescribed dosages were determined, filling a gap in this study. Our review may have also missed proprietary studies that identified confounders of adherence or developed ways to improve adherence, filling other gaps. Although our search method was thorough, following the protocol described in the *Screening* section, studies that would have been found with other protocols may have been missed. A complementary strategy for future reviews would be to use the Digital Therapeutics Alliance product page to identify additional products as a starting point for looking for related evaluation studies. As few digital health apps qualify as SaMDs, our review reflects only a small portion of the clinical trials studying digital health apps. However, as these apps are subject to the most stringent regulatory requirements, they might be expected to have the highest quality evaluations. If so, future trials and studies on all digital health apps could benefit from implementing the recommendations of this study.

Acknowledgments

The authors would like to thank the Richard King Mellon Foundation for the financial support for the research, authorship, and publication of this paper (grant 220510 AWD00000732).

Authors' Contributions

EG, TK, and BF conceived of and planned the systematic review. EG wrote the manuscript and received critical feedback from TK and BF. LH, EG, and KW screened the articles for review. LH, OB, and KW reviewed the extracted data and provided feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of terms and assumptions.
[DOCX File , 44 KB - mhealth_v11i1e46237_app1.docx]

Multimedia Appendix 2
Article screening.
[PDF File (Adobe PDF File), 908 KB - mhealth_v11i1e46237_app2.pdf]

Multimedia Appendix 3
Extracted data.
[PDF File (Adobe PDF File), 1760 KB - mhealth_v11i1e46237_app3.pdf]

Multimedia Appendix 4
Summarized data.
[PDF File (Adobe PDF File), 479 KB - mhealth_v11i1e46237_app4.pdf]

Multimedia Appendix 5
PRISMA Checklist.
[PDF File (Adobe PDF File), 66 KB - mhealth_v11i1e46237_app5.pdf]

Multimedia Appendix 6
PRISMA Abstract Checklist.
[PDF File (Adobe PDF File), 45 KB - mhealth_v11i1e46237_app6.pdf]

References

1. Digital health trends 2021. IQVIA. 2021 Jul 22. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2022-06-14]

2. Digital health innovation action plan. US Food and Drug Administration. 2020. URL: https://www.fda.gov/media/106331/download [accessed 2021-07-11]
3. The 510(k) program: evaluating substantial equivalence in premarket notifications [510(k)]. U.S. Department of Health and Human Services. 2014. URL: https://www.fda.gov/media/82395/download [accessed 2023-07-27]
4. Statement from FDA Commissioner Scott Gottlieb, M.D., and Center for Devices and Radiological Health Director Jeff Shuren, M.D., J.D., on agency efforts to work with tech industry to spur innovation in digital health. U.S. Food and Drug Administration. URL: https://tinyurl.com/bdde8vb2 [accessed 2022-08-26]
5. Digital health software precertification (pre-cert) pilot program. U.S. Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program [accessed 2022-09-26]
6. The software precertification (Pre-Cert) pilot program: tailored total product lifecycle approaches and key findings. U.S. Food and Drug Administration. 2022. URL: https://www.fda.gov/media/161815/download [accessed 2023-07-18]
7. Use of real-world evidence to support regulatory decision-making for medical devices. U.S. Food and Drug Administration. 2017. URL: https://www.fda.gov/media/99447/download [accessed 2023-07-27]
8. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 Oct 17;2(1):38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
9. Milne-Ives M, Lam C, de Cock C, van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
10. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR Mhealth Uhealth 2017 Jul 20;5(7):e94 [FREE Full text] [doi: 10.2196/mhealth.7088] [Medline: 28729235]
11. Shek AC, Biondi A, Ballard D, Wykes T, Simblett SK. Technology-based interventions for mental health support after stroke: a systematic review of their acceptability and feasibility. Neuropsychol Rehabil 2021 Apr 13;31(3):432-452. [doi: 10.1080/09602011.2019.1701501] [Medline: 31833819]
12. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
13. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419 [FREE Full text] [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
14. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile health for pediatric weight management: systematic scoping review. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
15. Risk of Bias 2 (RoB 2) tool. The Cochrane Collaboration. 2022. URL: https://methods.cochrane.org/risk-bias-2 [accessed 2022-02-28]
16. de Geest S, Zullig LL, Dunbar-Jacob J, Helmy R, Hughes DA, Wilson IB, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). Ann Intern Med 2018 Jul 03;169(1):30-35 [FREE Full text] [doi: 10.7326/M18-0543] [Medline: 29946690]
17. Rubin DB. Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). Psychol Methods 2010 Mar;15(1):38-46 [FREE Full text] [doi: 10.1037/a0018537] [Medline: 20230101]
18. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2023.
19. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983 Apr;70(1):41-55 [FREE Full text] [doi: 10.1093/biomet/70.1.41]
20. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol 2017 Apr 01;46(2):756-762 [FREE Full text] [doi: 10.1093/ije/dyw323] [Medline: 28039382]
21. Sagarin BJ, West SG, Ratnikov A, Homan WK, Ritchie TD, Hansen EJ. Treatment noncompliance in randomized experiments: statistical approaches and design issues. Psychol Methods 2014 Sep;19(3):317-333. [doi: 10.1037/met0000013] [Medline: 24773358]
22. Mazza GL. Examining dose-response effects in randomized experiments with partial adherence. Arizona State University. 2018. URL: https://core.ac.uk/download/pdf/157755537.pdf [accessed 2021-01-20]
23. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci USA 2018 Mar 13;115(11):2600-2606 [FREE Full text] [doi: 10.1073/pnas.1708274114] [Medline: 29531091]
24. Home Page. Clue. URL: https://helloclue.com/ [accessed 2023-10-18]
25. Data extraction forms. The Cochrane Collaboration. 2022. URL: https://dplp.cochrane.org/data-extraction-forms [accessed 2022-02-28]
26. Grayek E. A systematic review of the collection and analysis of adherence information for software as a medical device clinical trials. Open Science Framework. 2021. URL: https://osf.io/z46un/?view_only=e06d49f44852454c80a086347a18f28e [accessed 2022-09-21]

27. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
28. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep 25;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
29. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
30. Dugas M, Wang W, Crowley K, Iyer AK, Peeples M, Shomali M, et al. Engagement and outcomes associated with contextual annotation features of a digital health solution. J Diabetes Sci Technol 2022 Jul 23;16(4):804-811 [FREE Full text] [doi: 10.1177/1932296820976409] [Medline: 33355003]
31. Jennings VH, Haile LT, Simmons RG, Fultz HM, Shattuck D. Estimating six-cycle efficacy of the Dot app for pregnancy prevention. Contraception 2019 Jan;99(1):52-55 [FREE Full text] [doi: 10.1016/j.contraception.2018.10.002] [Medline: 30316782]
32. Jennings V, Haile LT, Simmons RG, Spieler J, Shattuck D. Perfect- and typical-use effectiveness of the dot fertility app over 13 cycles: results from a prospective contraceptive effectiveness trial. Eur J Contracept Reprod Health Care 2019 Apr;24(2):148-153. [doi: 10.1080/13625187.2019.1581164] [Medline: 30880509]
33. Akturk HK, Dowd R, Shankar K, Derdzinski M. Real-world evidence and glycemic improvement using Dexcom G6 features. Diabetes Technol Ther 2021 Mar;23(S1):S21-S26 [FREE Full text] [doi: 10.1089/dia.2020.0654] [Medline: 33434439]
34. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RS, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
35. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med 2021 Mar 26;4(1):58 [FREE Full text] [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]
36. Gallen CL, Anguera JA, Gerdes MR, Simon AJ, Cañadas E, Marco EJ. Enhancing neural markers of attention in children with ADHD using a digital therapeutic. PLoS One 2021;16(12):e0261981 [FREE Full text] [doi: 10.1371/journal.pone.0261981] [Medline: 34972140]
37. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, et al. Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. Gut 2019 Sep;68(9):1613-1623 [FREE Full text] [doi: 10.1136/gutjnl-2018-317805] [Medline: 30971419]
38. Everitt H, Landau S, Little P, Bishop FL, O'Reilly G, Sibelli A, et al. Therapist telephone-delivered CBT and web-based CBT compared with treatment as usual in refractory irritable bowel syndrome: the ACTIB three-arm RCT. Health Technol Assess 2019 Apr;23(17):1-154 [FREE Full text] [doi: 10.3310/hta23170] [Medline: 31042143]
39. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, et al. Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial. Lancet Gastroenterol Hepatol 2019 Nov;4(11):863-872 [FREE Full text] [doi: 10.1016/S2468-1253(19)30243-2] [Medline: 31492643]
40. Berglund Scherwitzl E, Gemzell Danielsson K, Sellberg JA, Scherwitzl R. Fertility awareness-based mobile application for contraception. Eur J Contracept Reprod Health Care 2016 Jun 22;21(3):234-241 [FREE Full text] [doi: 10.3109/13625187.2016.1154143] [Medline: 27003381]
41. Berglund Scherwitzl E, Lundberg O, Kopp Kallner H, Gemzell Danielsson K, Trussell J, Scherwitzl R. Perfect-use and typical-use Pearl Index of a contraceptive mobile app. Contraception 2017 Dec;96(6):420-425 [FREE Full text] [doi: 10.1016/j.contraception.2017.08.014] [Medline: 28882680]
42. Bull J, Rowland S, Lundberg O, Berglund-Scherwitzl E, Gemzell-Danielsson K, Trussell J, et al. Typical use effectiveness of Natural Cycles: postmarket surveillance study investigating the impact of previous contraceptive choice on the risk of unintended pregnancy. BMJ Open 2019 Mar 23;9(3):e026474 [FREE Full text] [doi: 10.1136/bmjopen-2018-026474] [Medline: 30904873]
43. Pearson JT, Chelstowska M, Rowland S, Mcilwaine E, Benhar E, Berglund Scherwitzl E, et al. Natural Cycles app: contraceptive outcomes and demographic analysis of UK users. Eur J Contracept Reprod Health Care 2021 Apr;26(2):105-110. [doi: 10.1080/13625187.2020.1867844] [Medline: 33539252]
44. Pearson JT, Chelstowska M, Rowland SP, Benhar E, Kopp-Kallner H, Berglund Scherwitzl E, et al. Contraceptive effectiveness of an FDA-cleared birth control app: results from the natural cycles U.S. cohort. J Womens Health (Larchmt) 2021 Jun;30(6):782-788. [doi: 10.1089/jwh.2020.8547] [Medline: 33370220]
45. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. Am J Psychiatry 2014 Jun;171(6):683-690 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13081055] [Medline: 24700332]

46. Christensen DR, Landes RD, Jackson L, Marsch LA, Mancino MJ, Chopra MP, et al. Adding an Internet-delivered treatment to an efficacious treatment package for opioid dependence. J Consult Clin Psychol 2014 Dec;82(6):964-972 [FREE Full text] [doi: 10.1037/a0037496] [Medline: 25090043]
47. Maricich YA, Xiong X, Gerwien R, Kuo A, Velez F, Imbert B, et al. Real-world evidence for a prescription digital therapeutic to treat opioid use disorder. Curr Med Res Opin 2021 Feb;37(2):175-183. [doi: 10.1080/03007995.2020.1846023] [Medline: 33140981]
48. Maricich YA, Gerwien R, Kuo A, Malone DC, Velez FF. Real-world use and clinical outcomes after 24 weeks of treatment with a prescription digital therapeutic for opioid use disorder. Hosp Pract (1995) 2021 Dec 03;49(5):348-355. [doi: 10.1080/21548331.2021.1974243] [Medline: 34461801]
49. Maricich YA, Bickel WK, Marsch LA, Gatchalian K, Botbyl J, Luderer HF. Safety and efficacy of a prescription digital therapeutic as an adjunct to buprenorphine for treatment of opioid use disorder. Curr Med Res Opin 2021 Feb 07;37(2):167-173 [FREE Full text] [doi: 10.1080/03007995.2020.1846022] [Medline: 33140994]
50. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. JAMA Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
51. de Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004 Sep;364(9438):911-912. [doi: 10.1016/S0140-6736(04)17034-7] [Medline: 15364170]
52. Gopal AD, Wallach JD, Aminawung JA, Gonsalves G, Dal-Ré R, Miller JE, et al. Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals. Trials 2018 Aug 23;19(1):448 [FREE Full text] [doi: 10.1186/s13063-018-2825-y] [Medline: 30134950]
53. Merrill PD, McClure LA. Dichotomizing partial compliance and increased participant burden in factorial designs: the performance of four noncompliance methods. Trials 2015 Nov 17;16:523 [FREE Full text] [doi: 10.1186/s13063-015-1044-z] [Medline: 26573840]
54. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. Biometrics 2009 Jun;65(2):640-649 [FREE Full text] [doi: 10.1111/j.1541-0420.2008.01066.x] [Medline: 18510650]
55. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol 2019 Mar;34(3):211-219 [FREE Full text] [doi: 10.1007/s10654-019-00494-6] [Medline: 30840181]
56. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002 Jan 15;155(2):176-184. [doi: 10.1093/aje/155.2.176] [Medline: 11790682]

Abbreviations

ATE: average treatment effect
CACE: complier average causal effect
CONSORT: Consolidated Standards of Reporting Trials
FDA: Food and Drug Administration
LATE: local average treatment effect
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROB: risk of bias
SaMD: software as a medical device
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SUTVA: stable unit treatment value assumption

Edited by L Buis; submitted 07.02.23; peer-reviewed by H Luderer, B Imbert; comments to author 20.04.23; revised version received 31.07.23; accepted 25.08.23; published 15.11.23.

Please cite as:
Grayek E, Krishnamurti T, Hu L, Babich O, Warren K, Fischhoff B
Collection and Analysis of Adherence Information for Software as a Medical Device Clinical Trials: Systematic Review
JMIR Mhealth Uhealth 2023;11:e46237
URL: https://mhealth.jmir.org/2023/1/e46237
doi:10.2196/46237
PMID:37966871

Review

App Characteristics and Accuracy Metrics of Available Digital Biomarkers for Autism: Scoping Review

Sonia Ponzo[1,2], MSc, PhD; Merle May[1], BSc, MSc; Miren Tamayo-Elizalde[1], MSc, PhD; Kerri Bailey[1], MSc, PhD; Alanna J Shand[1,3], BSc, MSc; Ryan Bamford[1], BSc, MSc; Jan Multmeier[1], MSc, PhD; Ivan Griessel[1], BA; Benedek Szulyovszky[1], BSc, MSc; William Blakey[1,4], BSc; Sophie Valentine[1], BSc, MRES; David Plans[1,5], MSc, PhD

[1]Healios Limited, London, United Kingdom

[2]Institute of Health Informatics, University College London, London, United Kingdom

[3]Institute of Epidemiology and Healthcare, University College London, London, United Kingdom

[4]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, United Kingdom

[5]Department of Psychology, Royal Holloway, University of London, London, United Kingdom

Corresponding Author:
Sonia Ponzo, MSc, PhD
Healios Limited
4A Tileyard, Tileyard Rd
London, N7 9AH
United Kingdom
Phone: 44 0330 124 4222
Email: sonia.ponzo@healios.org.uk

Abstract

Background: Diagnostic delays in autism are common, with the time to diagnosis being up to 3 years from the onset of symptoms. Such delays have a proven detrimental effect on individuals and families going through the process. Digital health products, such as mobile apps, can help close this gap due to their scalability and ease of access. Further, mobile apps offer the opportunity to make the diagnostic process faster and more accurate by providing additional and timely information to clinicians undergoing autism assessments.

Objective: The aim of this scoping review was to synthesize the available evidence about digital biomarker tools to aid clinicians, researchers in the autism field, and end users in making decisions as to their adoption within clinical and research settings.

Methods: We conducted a structured literature search on databases and search engines to identify peer-reviewed studies and regulatory submissions that describe app characteristics, validation study details, and accuracy and validity metrics of commercial and research digital biomarker apps aimed at aiding the diagnosis of autism.

Results: We identified 4 studies evaluating 4 products: 1 commercial and 3 research apps. The accuracy of the identified apps varied between 28% and 80.6%. Sensitivity and specificity also varied, ranging from 51.6% to 81.6% and 18.5% to 80.5%, respectively. Positive predictive value ranged from 20.3% to 76.6%, and negative predictive value fluctuated between 48.7% and 97.4%. Further, we found a lack of details around participants' demographics and, where these were reported, important imbalances in sex and ethnicity in the studies evaluating such products. Finally, evaluation methods as well as accuracy and validity metrics of available tools were not clearly reported in some cases and varied greatly across studies. Different comparators were also used, with some studies validating their tools against the Diagnostic and Statistical Manual of Mental Disorders criteria and others through self-reported measures. Further, while in most cases, 2 classes were used for algorithm validation purposes, 1 of the studies reported a third category (indeterminate). These discrepancies substantially impact the comparability and generalizability of the results, thus highlighting the need for standardized validation processes and the reporting of findings.

Conclusions: Despite their popularity, systematic evaluations and syntheses of the current state of the art of digital health products are lacking. Standardized and transparent evaluations of digital health tools in diverse populations are needed to assess their real-world usability and validity, as well as help researchers, clinicians, and end users safely adopt novel tools within clinical and research practices.

(JMIR Mhealth Uhealth 2023;11:e52377) doi:10.2196/52377

KEYWORDS

autism; diagnostics; digital biomarkers; digital health; mobile apps; neurodevelopmental conditions

Introduction

Autism is a common form of neurodivergence estimated to affect 1% of the population worldwide, with prevalence rates reported at 1% in the United Kingdom [1] and 1.85% in the United States [2]. Despite this high prevalence rate, diagnostic delays are common, with difficulties receiving an initial service referral [1] and families in the United Kingdom [3] and United States [4] reporting between 2 and 3 years to receive a diagnosis from the onset of symptoms, respectively. On top of severely impacting the quality of life of those awaiting assessments and their families [5], diagnostic delays may increase the likelihood and severity of comorbidities [6]. Further, current diagnostic processes for autism rely solely on subjective clinician interpretations derived from standardized assessment tools, leading to potential misdiagnosis [7] and accentuation of phenomena such as masking [8].

Digital health products, such as mobile apps, have the potential to aid the diagnostic process due to their scalability and ease of access. Apps also offer the possibility of providing additional ecological information collected in users' home environments to clinicians during the assessment phase. Specifically, digital biomarkers, that is, digital tools that collect information about the behavioral characteristics and physiological processes of individuals affected by a condition, have shown promise in identifying the presence of a disorder in several diagnostic domains (eg, cognitive impairment and dementia [9], depression [10], and learning disabilities [11]). Recommendations for developers and the scientific community at large outline the importance of transparent communication of the algorithms underlying digital biomarkers as well as plans for iterative evaluations of such products [12]. Further, multimodal approaches that prioritize cognitive and behavioral assessments have been recognized as promising in aiding precision diagnostics and personalized therapeutics [13]. While research on digital biomarkers in autism is still in its infancy and specific recommendations for the development and validation of these products are lacking, there are digital health tools available to researchers, clinicians, and end users.

Nevertheless, not many resources synthesizing the characteristics and evaluation outcomes of these products are widely available. The aim of the current scoping review is to summarize the evidence about existing digital biomarker tools so that researchers in the field of autism, clinicians, and end users are provided with up-to-date information to make informed decisions regarding their usefulness and adoption.

Methods

Search Strategy

A structured search, following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guideline [14], was conducted on August 21, 2023 (with further manual searches conducted ad hoc), in the databases MEDLINE through PubMed and Elsevier's Scopus. The search terms included "autism digital biomarker"; "autism app"; and related synonyms, truncations, and Medical Subject Headings. Additional searches were conducted through the US Food and Drug Administration (FDA) website [15] and Google to find regulatory submissions and additional materials, respectively. If data reported as part of a regulatory submission had been published in a peer-reviewed journal, the peer-reviewed journal was used. A review protocol was not published; however, the full search strategy can be found in Table S1in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Only full-text primary research studies published in English and in peer-reviewed journals, as well as regulatory submissions published on the regulatory body website, were included (Table S2 in Multimedia Appendix 1). Further, studies were included if they reported accuracy and validity metrics from available digital biomarkers. Apps using digital versions of existing standardized autism assessments or telehealth adaptations of in-person assessments were also excluded.

Data Extraction and Analysis

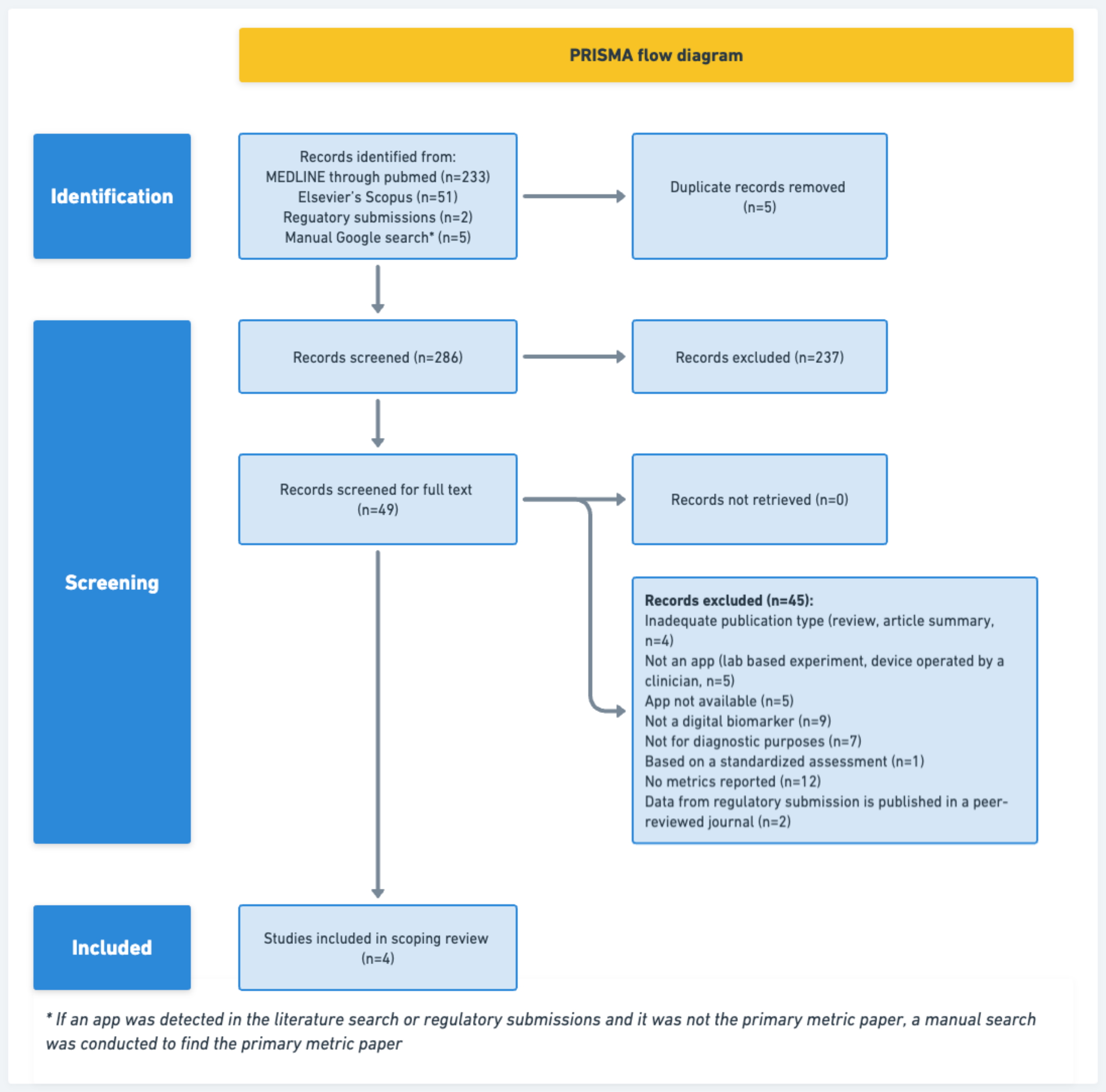
Data were extracted and analyzed by 3 reviewers. Titles and abstracts were screened once with a reason for exclusion provided by 1 reviewer, which was then inspected by another reviewer. Full texts of eligible papers were reviewed independently by at least 2 reviewers, and the inclusion and exclusion of studies were discussed as a team. The full data extraction form can be found in Multimedia Appendix 1; Table S3 in Multimedia Appendix 1 provides further study details and metrics about the included studies.

Results

Overview

We found 286 studies and regulatory submissions, of which 49 were eligible for full-text screening. A total of 4 studies met our criteria and were included in the review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the literature search and selection process.

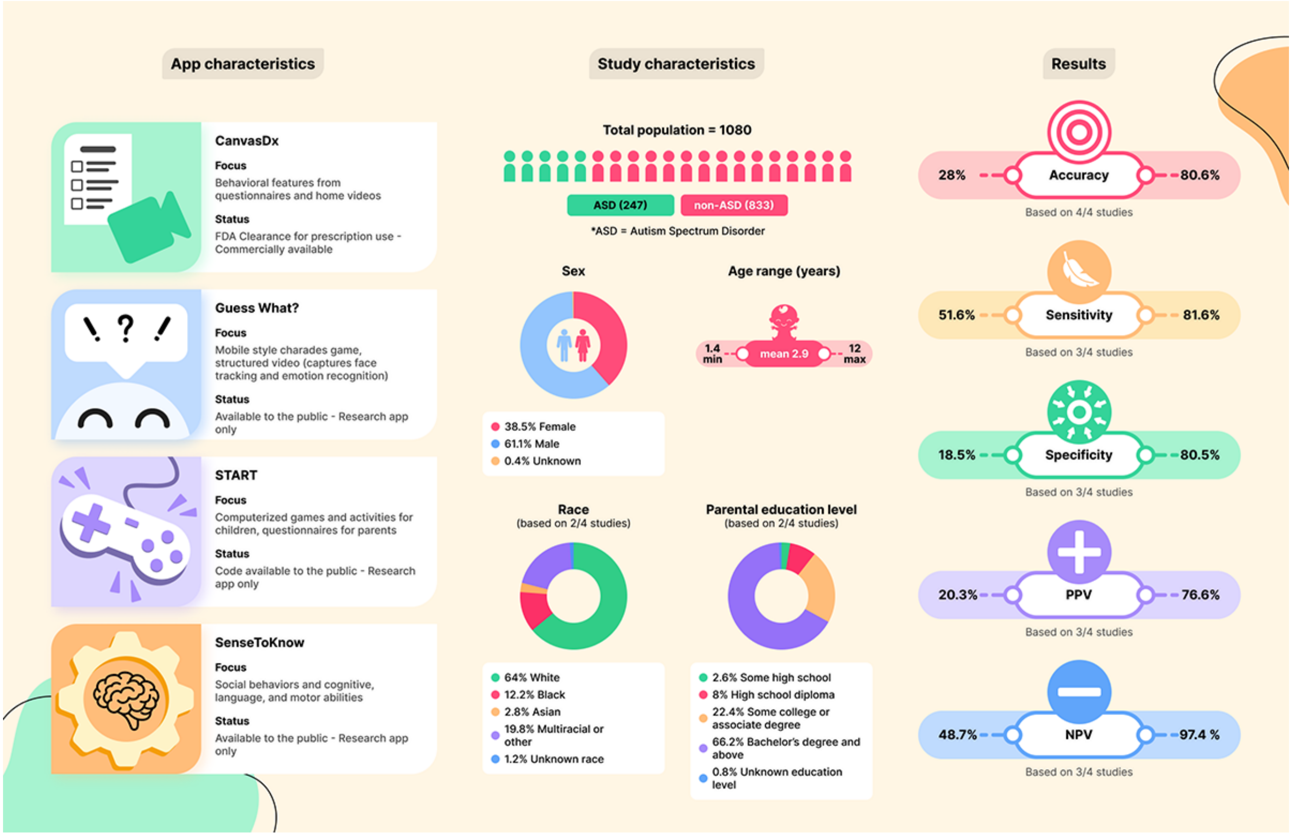


App Characteristics and Regulatory Aspects

A total of 4 studies involving 4 unique apps were included in the review (Figure 2), of which 1 (Canvas Dx [16]) was an FDA-cleared commercial product and 3 (Guess What? [17], START [18], and SenseToKnow [19]) were research apps.

The apps from the included studies target children between 17 and 144 months (1.4-12 years) of age who do not exhibit significant sensory or motor impairments (due to the nature of the tasks). Canvas Dx uses questionnaires from caregivers and health practitioners and videos to provide a diagnostic indication. Guess What? acquires structured videos of the interaction between child and parent during a charades-style game and applies face tracking and emotion recognition techniques. START measures social, sensory, and motor skills through games and activities for children and a questionnaire for parents. SenseToKnow displays specifically designed movies aimed at eliciting autism-related attention and motor behaviors while recording the child's responses through the front camera of the device.

Figure 2. Summary of app characteristics, study details, and accuracy metrics for all included studies. ASD: autism spectrum disorder; FDA: US Food and Drug Administration; NPV: negative predictive value; PPV: positive predictive value.



As for their technical approaches, 2 studies were administered through smartphones and 2 through tablets. Most apps (3/4, 75%) classified individuals as either autistic or nonautistic, although CanvasDx included an "undetermined" category. Further, all the studies used tree-based classification methods, particularly gradient-boosted decision trees. In terms of real-world usability, most studies (3/4, 75%) reported results from usability testing, interviews with clinicians and families, or app quality scores, showing good acceptability and feasibility results and high (93.9%) quality scores. All tools emphasize that their intended use is screening or diagnostic aid rather than stand-alone diagnostics. CanvasDx also warns that results may be potentially unreliable in individuals with specific medical conditions, such as epilepsy or genetic disorders (Table S4 in Multimedia Appendix 1).

A total of 3 out of the 4 apps from the included studies are available for download in the United States (links are included in Table S3 in Multimedia Appendix 1), with 1 (Guess What?) also being available in the United Kingdom. As for START, only the app code has been published. All apps are free to download. Regarding device compatibility, Canvas Dx and Guess What? can be downloaded on either Android or iOS; SenseToKnow is only available for iOS; and the code for START is for Android implementation only.

Studies Characteristics

Studies included to validate these 4 apps involved 1080 individuals. Participants were children aged between 17 and 144 months (1.4-12 years; mean 2.9 SD 1.0 years), with a mean autism prevalence of 22.9% (range 10.3%-57.1%). The pooled sex split was 38.5% (416/1080; range 30.5%-43.4%) female, 61.1% (660/1080; range 56.6%-69.5%) male, and 0.4% (4/1080; 1 study only: 8.2%) unknown. Race was reported for 2 studies: 2.8% (25/900; range 1.5%-4.2%) Asian, 12.2% (110/900; range 11.4%-13.2%) Black, 64% (576/900; range 53.9%-73.2%) White, 19.8% (178/900; range 13.9%-28.7%) multiracial or other, and 1.2% (11/900; range 0.2%-2.4%) unknown, and ethnicity only reported independently in 1 study: 10.5% (50/475) Hispanic or Latino and 89.5% (425/475) non-Hispanic or non-Latino. A total of 2 studies also reported parental education level: 2.6% (23/900; range 2.1%-3.1%) some high school, 8% (72/900; range 5.9%-10.4%) high school diploma, 22.4% (202/900; range 11.2%-35.1%) some college or associate degree, 66.2% (596/900; range 50.4%-80.4%) bachelor's degree and above, and 0.8% (7/900; range 0.4%-1.2%) unknown. Samples varied in size and ranged from 49 to 475.

Accuracy and Validity Metrics

When metrics were not reported by the authors but sensitivity, specificity, and prevalence were indicated, they were calculated using the formulas in Table 1. Overall, accuracy fluctuated between 28% and 80.6%. Sensitivity and specificity also varied, ranging from 51.6% to 81.6% and 18.5% to 80.5%, respectively. Similarly, positive predictive value ranged from 20.3% to 76.6%, and negative predictive value fluctuated between 48.7% and 97.4%.

Table 1. Accuracy and validity metrics of the included digital biomarker apps for autism classification.

App name	Sample size, n	Autism prevalence (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	Comparator
Canvas Dx	425	28.7	28[c]	51.6[d]	18.5[d]	20.3[c]	48.7[c]	ECD[e]
Guess What?[f]	49	57.1	73	76	69[c]	76.6[c]	68.3[c]	Parent-reported diagnosis
START	131	36.6	61.6	Unable to calculate	Unable to calculate	Unable to calculate	Unable to calculate	ECD
SenseToKnow	475	10.3	80.6[c]	81.6	80.5	32.5[c]	97.4[c]	ECD

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]Calculated using the following formulas: accuracy = (sensitivity) × (prevalence) + (specificity) × (1 – prevalence); PPV = (sensitivity × prevalence) / {(sensitivity × prevalence) + ([1 – specificity] × [1 – prevalence])}; and NPV = (specificity × [1 – prevalence]) / {(specificity × [1 – prevalence]) + ([1 – sensitivity] × prevalence)}.

[d]For 3 classes (autism, uncertain, and nonautism) classification. If participants who received an uncertain or indeterminate class decided by the classifier were removed, for the remaining 31.8% (135/425) of participants who received a determinate output (autism or nonautism), sensitivity and specificity increased to 98.4% and 78.9%, respectively.

[e]Expert clinician diagnosis.

[f]Metrics from the feasibility study. Ongoing study for the validation of the GuessWhat? app (registered trial NCT04739982 [20])

Discussion

Overview

This study investigated the available digital biomarker tools for autism diagnosis. We found 4 products targeted at children, exploring a variety of domains, ranging from attention and looking behaviors to analysis of social, sensory, and motor skills. Of the examined research studies, 1 included commercial apps with medical device classification, and 3 involved unregulated research apps.

All studies use digital biomarkers of known domains that have been shown to be indicative of autism [21]. Most products use video analysis (often accompanied by questionnaires) to extract features of interest, whereas 1 of the included tools collects behavioral measures from interactive tasks. Both parent questionnaires and batteries assessing children during interactive tasks are part of well-known diagnostic assessments for autism, for example, Autism Diagnostic Observation Schedule (ADOS-2) [22,23], Autism Diagnostic Interview-Revised [24], and TELE-ASD-PEDS [25]. While assessment of different domains ensures higher coverage of potentially relevant behaviors, the variety in the evaluated domains limits the comparability between tools. Nevertheless, each of these digital biomarkers has strengths and weaknesses. CanvasDx obtained FDA approval for commercialization and is therefore undergoing robust validation processes. Further, its classification algorithm is very sensitive to extreme cases (very high or very low risk of autism); however, the inclusion of an indeterminate class makes it more difficult to identify less severe cases. Another limitation of CanvasDx is that it is not clinician independent. The strengths of Guess What? include its blended diagnostic and therapeutic approach, which offers an end-to-end solution for users; its wider age range of use; and its availability in multiple geographies. Its diagnostic and therapeutic nature is also its biggest limitation, as this tool was not originally designed as a diagnostic tool, and evaluation data are limited. The START app combines multiple assessment domains (social functioning and motor and sensory behaviors) and has been validated with diverse communities in mind. Nevertheless, the validation study does not report information about misdiagnoses, which limits the interpretability of its results. Finally, SenseToKnow offers combined biomarkers for social behavior as well as cognitive, language, and motor abilities within an assessment shorter than 10 minutes. The major limitation in its validation data pertains to the lower prevalence of autism-positive cases, which highlights the need for further evaluations of the tool.

Availability to download the apps was restricted to specific geographies, with unclear documentation about accessibility. Further, 1 of the tools (START) has only been published within a code repository, thus limiting accessibility to the wider population. Given that one of the main advantages of mobile apps is their scalability and ease of access [26], restrictions based on geography may negatively impact the adoption of such products in both research and clinical settings.

In terms of study details, only 2 studies reported full demographic information and socioeconomic variables, despite research showing differences in the adoption of health apps highly correlate with higher education and income [27]. The fact that most parents reported having a bachelor's degree or above further impacts the generalizability of the results to individuals from different socioeconomic backgrounds and poses questions around overall real-world usability. Both sex and race and ethnicity data were unbalanced, with most study participants being male and White. Evidence suggests that phenomena such as masking are more common in female individuals [28,29], thus highlighting the need for digital biomarkers to be validated in balanced sex populations. Similarly, findings outline longer diagnostic delays among minorities and underserved communities, with an associated lack of early interventions during pivotal developmental years [30,31]. Thus, if app developers fail to assess the usability, acceptability, and validity of these tools in underserved communities and diverse populations, the potential of digital health products to address barriers to entry to diagnostic and therapeutic pathways may be significantly reduced. As such, research evaluating digital biomarkers for autism should aim to assess their performance in these demographics to successfully increase health equity and help tackle diagnostic complexities.

Contrary to the Standards for Reporting of Diagnostic Accuracy Studies [32], half (2/4, 50%) of the resources did not provide the full array of accuracy and validity metrics nor the confusion matrix (Table S3 in Multimedia Appendix 1), with 1 of the studies (START) only reporting accuracy (but no sensitivity or specificity nor confusion matrix), which substantially limits the interpretability of their findings. The observed high variability in the presented app metrics may be justified by discrepancies in the evaluation methods. First, differences in the prevalence of autism substantially impact the calculation of classic accuracy metrics, therefore not rendering a reliable picture of the true classifiers' performance and limiting generalizability [33]. Similarly, most studies (3/4, 75%) used 2 classes (autism vs nonautism) for their algorithm validation, whereas 1 added an "indeterminate" class. Additionally, the studies evaluated their apps using different comparators, with most (3/4, 75%) measuring the performance of their algorithms against expert clinician diagnosis based on the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) criteria and another 1 referring to parent-reported diagnosis. Notably, clinicians may reach a diagnostic outcome using different diagnostic criteria (eg, DSM-5 or International Classification of Diseases-11) or supplement their clinical judgment with the use of assessment tools (eg, ADOS-2) [34,35]. Additionally, geographic differences exist in parent-reported symptomology when using common assessment tools [36], of which accuracy and validity metrics have also been shown to vary by country [37]. The abovementioned factors significantly impact any attempt to draw direct comparisons between studies. As such, app developers and researchers should aim to validate their products within the context of standardized evaluation frameworks that advocate for validation in diverse populations and with multiple comparators. Further, findings should be reported in a uniform fashion and should include the prevalence of the disorder, full metrics, confusion matrices, binary classification, and comparator of choice to facilitate the assessment of the potential for digital biomarkers to represent effective screening tools for autism.

This review highlighted important gaps in the literature surrounding the development and testing of digital biomarkers for autism. First, it remains unclear whether and to what extent these tools are typically developed with input from both end users and clinicians [38]. Frameworks outlining guidelines for the design and development of digital biomarkers would help unify approaches across companies and research entities and ensure standards of quality and safety [39]. Our work also outlined exciting trends toward the development of mobile-first digital technologies. It has been shown that while individuals in underserved communities often do not have access to desktop computers or laptops, the availability of mobile phones is typically higher [18]. Therefore, developing mobile-first tools could help ensure health equity in communities where conditions like autism suffer from greater diagnostic delays [40]. Finally, digital biomarker tools typically provide ecologically valid information that may not otherwise be available to clinicians. As such, including these tools as part of the waiting list for autism assessments could provide clinicians with valuable information ahead of the assessment, which in turn could help prioritize more severe cases.

Conclusions

Digital health products are increasingly gaining popularity, yet systematic syntheses of the current state of the art are lacking. Our work highlighted how diversity in the development and evaluation of digital biomarkers aiding in the detection of autism may impact their real-world usability and adoption in communities where these tools may have the most positive impact. As such, standardized and transparent development and evaluation frameworks, recommending assessing the validity of digital biomarkers in diverse populations, are needed to guide researchers, clinicians, and end users in making informed decisions about whether to consider their adoption within research settings and clinical pathways.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Information on the included app's metrics and on the search strategy.

[XLSX File (Microsoft Excel File), 578 KB - mhealth_v11i1e52377_app1.xlsx]

Multimedia Appendix 2

PRISMA-ScR checklist.

[PDF File (Adobe PDF File), 643 KB - mhealth_v11i1e52377_app2.pdf]

References

1. Locke J, Ibanez LV, Posner E, Frederick L, Carpentier P, Stone WL. Parent perceptions about communicating with providers regarding early autism concerns. Pediatrics 2020;145(Suppl 1):S72-S80 [FREE Full text] [doi: 10.1542/peds.2019-1895J] [Medline: 32238533]
2. Bougeard C, Picarel-Blanchot F, Schmid R, Campbell R, Buitelaar J. Prevalence of autism spectrum disorder and co-morbidities in children and adolescents: a systematic literature review. Front Psychiatry 2021;12:744709 [FREE Full text] [doi: 10.3389/fpsyt.2021.744709] [Medline: 34777048]
3. The All Party Parliamentary Group on Autism. The Autism Act, 10 years on: a report from the All Party Parliamentary Group on Autism on understanding, services and support for autistic people and their families in England. National Autistic Society. 2018. URL: https://pearsfoundation.org.uk/wp-content/uploads/2019/09/APPGA-Autism-Act-Inquiry-Report.pdf [accessed 2023-11-07]
4. Gordon-Lipkin E, Foster J, Peacock G. Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder. Pediatr Clin North Am 2016;63(5):851-859 [FREE Full text] [doi: 10.1016/j.pcl.2016.06.007] [Medline: 27565363]
5. Rabba AS, Dissanayake C, Barbaro J. Parents' experiences of an early autism diagnosis: insights into their needs. Res Autism Spectr Disord 2019;66:101415. [doi: 10.1016/j.rasd.2019.101415]
6. Jadav N, Bal VH. Associations between co-occurring conditions and age of autism diagnosis: implications for mental health training and adult autism research. Autism Res 2022;15(11):2112-2125 [FREE Full text] [doi: 10.1002/aur.2808] [Medline: 36054777]
7. Hus Y, Segal O. Challenges surrounding the diagnosis of autism in children. Neuropsychiatr Dis Treat 2021;17:3509-3529 [FREE Full text] [doi: 10.2147/NDT.S282569] [Medline: 34898983]
8. Adamou M, Jones SL, Wetherhill S. Predicting diagnostic outcome in adult autism spectrum disorder using the autism diagnostic observation schedule, second edition. BMC Psychiatry 2021;21(1):24 [FREE Full text] [doi: 10.1186/s12888-020-03028-7] [Medline: 33423664]
9. Ding Z, Lee TL, Chan AS. Digital cognitive biomarker for mild cognitive impairments and dementia: a systematic review. J Clin Med 2022;11(14):4191 [FREE Full text] [doi: 10.3390/jcm11144191] [Medline: 35887956]
10. Asare KO, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. Pervasive Mob Comput 2022;83:101621 [FREE Full text] [doi: 10.1016/j.pmcj.2022.101621]
11. Toki EI, Tatsis G, Tatsis VA, Plachouras K, Pange J, Tsoulos IG. Employing classification techniques on SmartSpeech biometric data towards identification of neurodevelopmental disorders. Signals 2023;4(2):401-420 [FREE Full text] [doi: 10.3390/signals4020021]
12. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
13. Koehler JC, Falter-Wagner CM. Digitally assisted diagnostics of autism spectrum disorder. Front Psychiatry 2023;14:1066284 [FREE Full text] [doi: 10.3389/fpsyt.2023.1066284] [Medline: 36816410]
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
15. US Food and Drug Administration. URL: https://www.fda.gov/ [accessed 2023-11-07]
16. Megerian JT, Dey S, Melmed RD, Coury DL, Lerner M, Nicholls CJ, et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. NPJ Digit Med 2022;5(1):57 [FREE Full text] [doi: 10.1038/s41746-022-00598-6] [Medline: 35513550]
17. Deveau N, Washington P, Leblanc E, Husic A, Dunlap K, Penev Y, et al. Machine learning models using mobile game play accurately classify children with autism. Intell Based Med 2022;6:100057 [FREE Full text] [doi: 10.1016/j.ibmed.2022.100057] [Medline: 36035501]
18. Dubey I, Bishain R, Dasgupta J, Bhavnani S, Belmonte MK, Gliga T, et al. Using mobile health technology to assess childhood autism in low-resource community settings in India: an innovation to address the detection gap. Autism 2023:13623613231182801 [FREE Full text] [doi: 10.1177/13623613231182801] [Medline: 37458273]
19. Perochon S, Di Martino JM, Carpenter KLH, Compton S, Davis N, Eichner B, et al. Early detection of autism using digital behavioral phenotyping. Nat Med 2023;29(10):2489-2497 [FREE Full text] [doi: 10.1038/s41591-023-02574-3] [Medline: 37783967]
20. Wall DP. Evaluating mobile game based therapy among children with autism, a randomized clinical trial. Report no.: NCT04739982. ClinicalTrials.gov. 2022. URL: https://clinicaltrials.gov/study/NCT04739982 [accessed 2023-08-31]
21. Frye RE, Vassall S, Kaur G, Lewis C, Karim M, Rossignol D. Emerging biomarkers in autism spectrum disorder: a systematic review. Ann Transl Med 2019;7(23):792 [FREE Full text] [doi: 10.21037/atm.2019.11.53] [Medline: 32042808]
22. Lord C, Rutter M, DiLavore P, Risi S, Gotham K, Bishop S. Autism Diagnostic Observation Schedule, Second Edition. Torrance, CA: Western Psychological Services; 2006.
23. Lord C, Luyster RJ, Gotham K, Guthrie W. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module. Western Psychological Services: Torrance, CA; 2012.

24. Lord C, Rutter M, le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994;24(5):659-685. [doi: 10.1007/BF02172145] [Medline: 7814313]
25. Corona L, Hine J, Nicholson A, Stone C, Swanson A, Wade J, et al. TELE-ASD-PEDS: a telemedicine-based ASD evaluation tool for toddlers and young children. Vanderbilt University Medical Center. 2020. URL: https://vkc.vumc.org/vkc/triad/tele-asd-peds [accessed 2023-11-07]
26. Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 2020;5(8):759-769 [FREE Full text] [doi: 10.1016/j.bpsc.2019.11.015] [Medline: 32085921]
27. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? a secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
28. Hull L, Petrides KV, Mandy W. The female autism phenotype and camouflaging: a narrative review. Rev J Autism Dev Disord 2020;7(4):306-317 [FREE Full text] [doi: 10.1007/s40489-020-00197-9]
29. Tierney S, Burns J, Kilbey E. Looking behind the mask: social coping strategies of girls on the autistic spectrum. Res Autism Spectr Disord 2016;23:73-83. [doi: 10.1016/j.rasd.2015.11.013]
30. Keehn RM, Ciccarelli M, Szczepaniak D, Tomlin A, Lock T, Swigonski N. A statewide tiered system for screening and diagnosis of autism spectrum disorder. Pediatrics 2020;146(2):e20193876 [FREE Full text] [doi: 10.1542/peds.2019-3876] [Medline: 32632023]
31. Aylward BS, Gal-Szabo DE, Taraman S. Racial, ethnic, and sociodemographic disparities in diagnosis of children with autism spectrum disorder. J Dev Behav Pediatr 2021;42(8):682-689 [FREE Full text] [doi: 10.1097/DBP.0000000000000996] [Medline: 34510108]
32. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6(11):e012799 [FREE Full text] [doi: 10.1136/bmjopen-2016-012799] [Medline: 28137831]
33. Bone D, Goodwin MS, Black MP, Lee CC, Audhkhasi K, Narayanan S. Applying machine learning to facilitate autism diagnostics: pitfalls and promises. J Autism Dev Disord 2015;45(5):1121-1136 [FREE Full text] [doi: 10.1007/s10803-014-2268-6] [Medline: 25294649]
34. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. Eur Child Adolesc Psychiatry 2013;22(6):329-340. [doi: 10.1007/s00787-013-0375-0] [Medline: 23322184]
35. Charman T, Gotham K. Measurement issues: screening and diagnostic instruments for autism spectrum disorders—lessons from research and practise. Child Adolesc Ment Health 2013;18(1):52-63 [FREE Full text] [doi: 10.1111/j.1475-3588.2012.00664.x] [Medline: 23539140]
36. Matson JL, Matheis M, Burns CO, Esposito G, Venuti P, Pisula E, et al. Examining cross-cultural differences in autism spectrum disorder: a multinational comparison from Greece, Italy, Japan, Poland, and the United States. Eur Psychiatry 2017;42:70-76. [doi: 10.1016/j.eurpsy.2016.10.007] [Medline: 28212508]
37. Medda JE, Cholemkery H, Freitag CM. Sensitivity and specificity of the ADOS-2 algorithm in a large German sample. J Autism Dev Disord 2019;49(2):750-761 [FREE Full text] [doi: 10.1007/s10803-018-3750-3] [Medline: 30238180]
38. Wingfield B, Miller S, Yogarajah P, Kerr D, Gardiner B, Seneviratne S, et al. A predictive model for paediatric autism screening. Health Informatics J 2020;26(4):2538-2553 [FREE Full text] [doi: 10.1177/1460458219887823] [Medline: 32191164]
39. Dunlap JJ. Autism spectrum disorder screening and early action. J Nurse Pract 2019;15(7):496-501. [doi: 10.1016/j.nurpra.2019.04.001]
40. Divan G, Bhavnani S, Leadbitter K, Ellis C, Dasgupta J, Abubakar A, et al. Annual research review: achieving universal health coverage for young children with autism spectrum disorder in low- and middle-income countries: a review of reviews. J Child Psychol Psychiatry 2021;62(5):514-535 [FREE Full text] [doi: 10.1111/jcpp.13404] [Medline: 33905120]

Abbreviations

ADOS-2: Autism Diagnostic Observation Schedule
DSM-5: Diagnostic and Statistical Manual of Mental Disorders-5
FDA: US Food and Drug Administration
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by L Buis, G Eysenbach; submitted 05.09.23; peer-reviewed by W Zhao, Q Qi, M Mamat, T Valentine; comments to author 03.10.23; revised version received 30.10.23; accepted 02.11.23; published 17.11.23.

Please cite as:
Ponzo S, May M, Tamayo-Elizalde M, Bailey K, Shand AJ, Bamford R, Multmeier J, Griessel I, Szulyovszky B, Blakey W, Valentine S, Plans D
App Characteristics and Accuracy Metrics of Available Digital Biomarkers for Autism: Scoping Review
JMIR Mhealth Uhealth 2023;11:e52377
URL: https://mhealth.jmir.org/2023/1/e52377
doi:10.2196/52377
PMID:37976084

Review

Exploring Adherence to Pelvic Floor Muscle Training in Women Using Mobile Apps: Scoping Review

Rosie C Harper[1,2,3], BSc; Sally Sheppard[1,3], Grad Dip Phys; Carly Stewart[4], PhD, Prof Dr; Carol J Clark[1], PhD, Prof Dr

[1]Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, United Kingdom
[2]National Institute of Health Research ARC Wessex, Southampton, United Kingdom
[3]University Hospitals Dorset NHS Foundation Trust, Poole, United Kingdom
[4]Faculty of Science and Technology, Bournemouth University, Bournemouth, United Kingdom

Corresponding Author:
Rosie C Harper, BSc
Faculty of Health and Social Sciences
Bournemouth University
Bournemouth Gateway Building
12 St Paul's Lane
Bournemouth, BH8 8GP
United Kingdom
Phone: 44 01202 524111
Email: rharper1@bournemouth.ac.uk

Abstract

Background: Pelvic floor dysfunction is a public health issue, with 1 in 3 women experiencing symptoms at some point in their lifetime. The gold standard of treatment for pelvic floor dysfunction is supervised pelvic floor muscle training (PFMT); however, adherence to PFMT in women is poor. Mobile apps are increasingly being used in the National Health Service to enable equity in the distribution of health care and increase accessibility to services. However, it is unclear how PFMT mobile apps influence PFMT adherence in women.

Objective: We aimed to identify which behavior change techniques (BCTs) have been used in PFMT mobile apps, to distinguish the core "capability, opportunity, and motivation" (COM) behaviors targeted by the BCTs used in PFMT mobile apps, and to compare the levels of PFMT adherence in women between those using PFMT mobile apps and those receiving usual care.

Methods: We conducted a scoping review of the literature. Published quantitative literature that compared the use of a PFMT mobile app to a control group was included to address the objectives of the study. The electronic bibliographic databases searched included MEDLINE, CINAHL, Scopus, Web of Science, and PEDro, along with CENTRAL. Studies were also identified from reference searching of systematic reviews. Original articles written in English from 2006 onward were included. Nonexperimental quantitative studies, qualitative studies, studies that use male participants, case studies, web-based interventions, and interventions that use vaginal probes were excluded. Narrative synthesis was conducted on eligible articles based on the aims of the study.

Results: Of the 114 records retrieved from the search, a total of 6 articles met the eligibility and inclusion criteria. The total number of participants in the studies was 471. All PFMT mobile apps used the BCT "prompts and cues." Opportunity was the core COM behavior targeted by the PFMT mobile apps. Higher levels of adherence to PFMT were observed among women using PFMT mobile apps.

Conclusions: Digital "prompts and cues" are a BCT commonly used in PFMT mobile apps, and further research is required to practically assess whether a future randomized controlled trial that investigates the effectiveness of digital "prompts and cues" on PFMT adherence in women can be conducted.

(JMIR Mhealth Uhealth 2023;11:e45947) doi:10.2196/45947

KEYWORDS

adherence; behavior change; mHealth; mobile apps; pelvic floor muscle training; women

Introduction

Pelvic floor dysfunction (PFD) is a public health issue, with 1 in 3 women experiencing symptoms at some point in their lifetime [1,2]. PFD is a long-term condition that can worsen or reoccur in women at predictable points in their lives, including the childbearing years, menopause, and older age [2].

Symptoms of PFD include urinary incontinence (including mixed, urge, and stress urinary incontinence), pelvic organ prolapse, fecal incontinence, and sexual dysfunction, with stress urinary incontinence (SUI) as the most prevalent symptom [3,4]. SUI is defined by the International Continence Society as "the complaint of any involuntary loss of urine on effort or physical exertion or on sneezing or coughing" [5]. The gold standard of prevention and treatment for SUI is supervised pelvic floor muscle training (PFMT) [2,6,7]. Adherence to PFMT is key to its effectiveness in the treatment of symptoms of PFD [7,8]. The evidence suggests PFMT for a minimum of 3 months reduces symptoms of PFD and increases the likelihood of long-term adherence [2,9].

Adherence in relation to exercise prescription is the commitment with which a person sticks to a prescribed regime, which may be registered in terms of their compliance with the regime [7,10]. Exercise adherence can be measured in multiple ways, including exercise diaries [11,12] and exercise session attendance [13]. Self-efficacy, an individual's belief in their own capability to successfully execute specific actions to attain certain outcomes, has been suggested to be a key determinant in women's short- and long-term adherence to PFMT [12]. Adherence to PFMT in women is key to preventing and treating symptoms of PFD [12]. However, women's adherence to PFMT is poor and is associated with forgetting to complete PFMT [14].

Self-efficacy is suggested to be a key factor influencing women's short- and long-term adherence to PFMT [12]. Pelvic health physiotherapists use behavior change techniques (BCT) to increase women's self-efficacy to facilitate engagement with PFMT [2]. However, the effectiveness of these techniques is hindered by limited access to services and delayed continuity of care [15]. Furthermore, the impact of COVID-19 on health care services has exacerbated the issue, as the redirection of resources through the suspension of outpatient services has led to a backlog of patients awaiting specialist care [16].

Digital health technology is being progressively integrated into the National Health Service to enable equitable distribution of health care amidst ever-growing pressures [17]. Among these technologies, mobile apps have gained prominence as valuable tools for self-management, particularly among individuals with long-term conditions [18]. Mobile apps offer convenient and timely health information delivery, comparing favorably with web-based interventions [18-21]. In the current health care landscape, marked by challenges to service delivery, mobile apps offer a promising pathway for facilitating PFMT interventions. Apps incorporating effective BCTs should actively guide and support women in their PFMT and maximize adherence.

By identifying the BCTs used within PFMT mobile apps and the core behaviors they address, it is expected that future digital interventions aimed at improving PFMT adherence in women could be designed to focus on the most effective core behavior. The Behavior Change Wheel (BCW) is a theoretically underpinned framework using 19 different behavioral science frameworks and provides a standardized taxonomy for the characteristics of behavioral interventions [22]. The Capability, Opportunity, Motivation, Behavior (COM-B) model sits at the core of the BCW and names the 3 key elements of behavior change: capability, opportunity, and motivation. The Capability, Opportunity, Motivation, Behavior Change Wheel (COM-BCW) links the theoretical domains framework with the theoretical models of behavior. This means behavioral interventions can be characterized and evaluated using the COM-BCW through the identification of BCTs [23,24]. Experts in PFMT adherence have recommended using the COM-B to map the behavioral theories underpinning interventions to improve PFMT adherence in women [7].

Some systematic review findings suggest that PFMT mobile apps increase PFMT adherence in women [8,21,25,26] and their effectiveness in the treatment of PFD [27]. However, it is unclear how they work. The aim of one study [28] was to determine the encouraging features of PFMT mobile apps using persuasive system design to suggest features for new PFMT mobile apps that incorporate the COM-B model. However, to date, no review has mapped the core behaviors targeted by PFMT mobile apps through the identification of BCTs using the COM-B model. The findings of this scoping review will add to the body of literature around the theoretical underpinnings of existing interventions used in PFMT mobile apps and their influence on women's adherence to PFMT compared with usual care.

Therefore, the aim of the scoping review is to provide a comprehensive overview of the existing literature of studies using PFMT mobile apps. This review seeks to address the following objectives: (1) identify the BCTs used in PFMT mobile apps, (2) identify the core behaviors of the COM-B targeted by the BCTs in PFMT mobile apps, and (3) compare and explore the levels of PFMT adherence among women using PFMT apps and those receiving usual care.

Through these aims and objectives, this scoping review seeks to contribute to the understanding of the existing interventions within PFMT apps to illuminate their potential role in supporting and promoting adherence to PFMT among women.

Methods

Rationale

The scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [29,30]. To develop clearly defined research questions, the process created by the Joanna Briggs Institute was followed [31].

Search Strategy and Inclusion

The following electronic bibliographic databases were searched: MEDLINE, CINAHL, Scopus, Web of Science, and PEDro, along with CENTRAL. Studies were also identified from reference searching of systematic reviews. The search dates ranged from 2006 to January 2023. Studies before the year 2006 were excluded because mobile technology was unable to support mobile apps and the concept of self-management had not been widely used before this date [32].

Search terms followed the “patient or population, intervention, comparison, and outcomes” model and included the following terms: “women,” “mobile apps,” “mHealth,” “pelvic floor muscle,” “kegel,” “compliance,” and “adherence.” Multimedia Appendix 1 contains a list of all the search terms. A title and abstract search was conducted in January 2023. The full search strategy is presented in Multimedia Appendix 2. The a priori protocol was developed and published on PROSPERO (International Prospective Register of Systematic Reviews) [33]. The “patient or population, intervention, comparison, and outcomes” eligibility criteria for this study are presented in Textbox 1.

The exclusion criteria were as follows: nonexperimental quantitative studies, qualitative studies, studies involving men, case studies, web-based interventions, and adolescent women aged 18 years or younger. Web-based interventions were excluded because of the technological advances in phones to keep up with user activity. Only studies written in English were included. Unpublished studies were not sought. For the review, adherence outcomes were self-reported exercise adherence, in-app adherence measures, and self-efficacy. There was no restriction on what measurement tools could be used to measure adherence. Studies were excluded if they did not measure these concepts.

Textbox 1. The “patient or population, intervention, comparison, and outcomes” eligibility criteria.

Eligibility criteria

- Population: women aged 18 years or older, either asymptomatic or symptomatic of stress urinary incontinence, urge urinary incontinence, or mixed urinary incontinence.
- Intervention: mobile apps for urinary incontinence that support pelvic floor muscle training.
- Comparison: standard care, supervised physiotherapy or home-based training, education delivered through apps, waitlist control, and no treatment.
- Outcome: primary outcome (adherence to pelvic floor muscle training) and secondary outcome (self-efficacy).

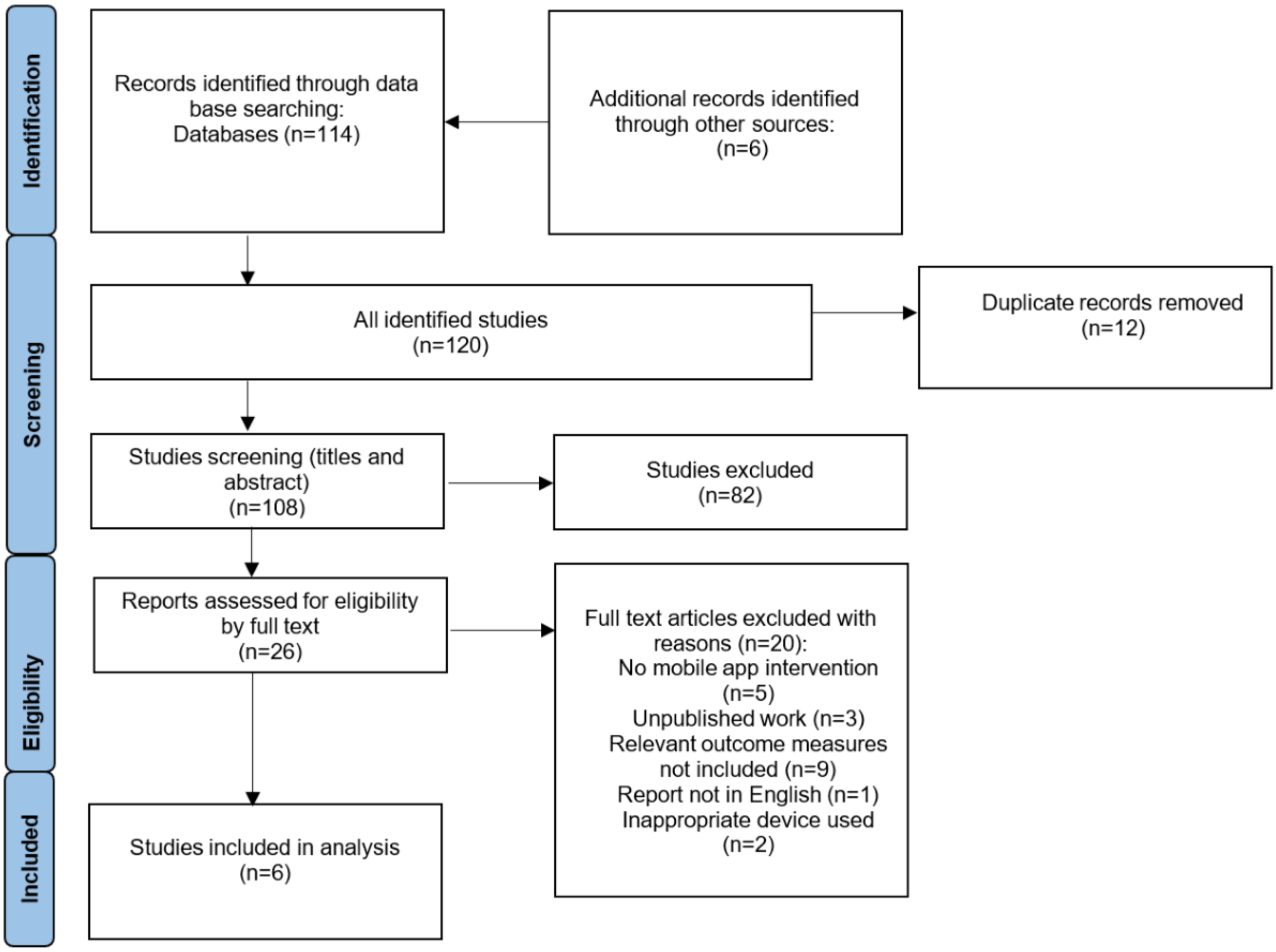
Data Extraction

The integrity of the review was demonstrated using the PRISMA-ScR (Figure 1). Studies were retrieved using a search strategy, and those from additional sources were screened based on their titles and abstracts by the first author (RCH) to determine if they met the inclusion criteria mentioned above. Studies that fulfilled the inclusion criteria were retrieved and independently assessed by the authors (RCH and CJC). The researchers were not blinded to each other’s decisions.

A standardized data extraction form was used by 2 authors to independently extract data on the name and location of the study, study participants, study design, PFMT program, description of PFMT mobile app features, outcome measures, and results in the eligible studies.

Discrepancies between authors RCH and CJC were identified and resolved through discussion. The data were recorded in an Excel (Microsoft Corp) spreadsheet (Multimedia Appendix 3 [34-39]).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart of study selection.



Data Synthesis

Adherence data were analyzed descriptively. A meta-analysis was not possible because of the clinical and methodological diversity in the included studies [40]. The aim of the analysis was to explore differences in adherence levels between the intervention and comparator groups.

BCTs identified in PFMT mobile apps were mapped using the COM-B model [22]. To determine which BCTs used in the intervention group target which core behavior of the COM-B model, the step-by-step mapping guidance to design behavioral interventions using the COM-BCW was used in reverse [22,24]. Rather than designing an intervention, the reverse allowed the identification of each core behavior from each BCT using the theoretical domains framework linked with the COM-B [22]. The 5-step process went as follows:

1. Descriptions of the intervention group were extracted.
2. BCTs were identified based on the above descriptions using BCT Taxonomy Version 1 [22].
3. BCTs were linked to their associated theoretical domains according to the theoretical domains framework [22].
4. Each of the theoretical domains was then linked to the COM-B model.
5. The core element of "capability, opportunity, and motivation" was then identified.

The characteristics of the intervention group were extracted into an Excel spreadsheet (Multimedia Appendix 3 [34-39]).

The description of the PFMT mobile app features and their purpose in the included papers allowed the authors (RCH and CJC) to identify which BCTs they used. The individual PFMT mobile apps were not downloaded. It was assumed that the authors of the included studies followed the correct way to report complex digital interventions and that their description was accurate. The descriptions of the digital interventions were included in the data extraction table (Table 1). All recognized BCTs were included. The BCT identification was completed by authors RCH and CJC. Any discrepancies between the identified BCTs were resolved through discussion. All PFMT mobile apps include one or more BCTs. No BCTs in the taxonomy were excluded based on perceived appropriateness in health and care settings, for example, future punishment.

Using the description of the PFMT mobile app features and their purpose in the included papers, the associated BCT was identified; for example, "daily reminder notification" is associated with "prompts and cues." Descriptions of PFMT mobile app features with no associated BCT were excluded. The identified BCTs were then grouped and mapped in a Word (Microsoft Corp) table (Table 2).

Table 1. Data extraction table of included studies.

Paper	Country	Study participants	Study design	PFMT[a] program	Description of PFMT mobile app features	Outcome measures
Jaffar et al [38]	Malaysia	26 pregnant women aged between 21 and 39 years	Pilot RCT[b]; app: n=16; control: n=10	A total of 11 contractions a day (10 fast and 1 slow). Length of a sustained contraction is 2-10 seconds. Repeated 3 times a day for 8 weeks.	Educational video; training timer; symptom calendar; daily reminder notification; progress chart; frequently asked questions	Exercise Adherence Rating Scale; Broome Self-efficacy Scale for Practicing Pelvic Floor Exercises Questionnaire
Araujo et al [35]	Brazil	33 women aged between 36 and 67 years	RCT; app: n=17; control: n=16	A total of 4 contractions a day (3 fast and 1 slow). Length of a sustained contraction is 8 seconds. Repeated 2 times a day for 12 weeks.	Visual feedback; reminders; audio feedback with PFMT; self-reported perception of improvement; exercise log	In-app adherence measures; self-reported adherence measure
Kinouchi and Ohashi [39]	Japan	58 postnatal women aged between 31 and 37 years	HCT[c]; app: n=29; control: n=29	A total of 6 contractions a day (fast unspecified and 6 slow). Length of a sustained contraction is 3-6 seconds. Repeated 3 times a day for 12 weeks.	Reminder system	Exercise implementation rate; median training intensity; exercise frequency
Wang et al [36]	China	108 pregnant women aged between 23 and 34 years	RCT; app: n=54; control: n=54	Participants were given the option of either (1) an unspecified number of slow contractions holding for >2 seconds twice a day for 12 weeks, or (2) a total of 150 pelvic floor contractions per day for 12 weeks.	Systematic audio guidance PFMT program; audio reminders; a dynamic graph to support PFMT	Broome Self-efficacy Scale for Practicing Pelvic Floor Exercises Questionnaire
Asklund et al [34]	Sweden	123 women aged between 27 and 72 years	RCT; app: n=62; control: n=61	12 “levels” of PFMT programs (basic and advanced). Repeated 3 times a day for 12 weeks.	Information (on the pelvic floor, SUI[d], and lifestyle factors); statistics function that calculates the number of exercises performed; reminders; graphic to support PFMT	Self-reported adherence measure
Wadensten et al [37]	Sweden	123 women aged between 31 and 77 years	RCT; app: n=60; control: n=63	12 “levels” of PFMT programs (basic and advanced). Repeated 3 times a day for 12 weeks.	Information (on PFMT, the bladder, psychological topics, and lifestyle advice); reinforcement messages and daily reminders; exercise log; bladder diary	Perceived ability to correctly perform PFMT

[a]PFMT: pelvic floor muscle training.

[b]RCT: randomized controlled trial.

[c]HCT: historical controlled trial.

[d]SUI: stress urinary incontinence.

Table 2. The features of pelvic floor muscle training (PFMT) mobile apps and their associated behavior change techniques (BCTs).

Study authors	Intervention features	BCTs used in the intervention group			
		Self-monitoring of behavior	Self-monitoring of outcomes of behavior	Instruction on how to perform behavior	Prompts and cues
Jaffar et al [38]	Educational video, training timer, symptom calendar, daily reminder notification, progress chart, and frequently asked questions		✓	✓	✓
Asklund et al [34]	Information (on the pelvic floor, SUI[a], and lifestyle factors), statistics function that calculates the number of exercises performed, reminders, and graphic to support PFMT	✓		✓	✓
Kinouchi and Ohashi [39]	Reminder system				✓
Wang et al [36]	Systematic audio guidance PFMT program, audio reminders, and a dynamic graph to support PFMT			✓	✓
Araujo et al [35]	Visual feedback, reminders, audio feedback with PFMT, self-reported perception of improvement, and exercise log	✓	✓	✓	✓
Wadensten et al [37]	Information (on PFMT, the bladder, psychological topics, and lifestyle advice), reinforcement messages and daily reminders, exercise log, and bladder diary	✓	✓	✓	✓

[a]SUI: stress urinary incontinence.

The BCTs in the PFMT mobile apps and usual care groups were then directly linked to their associated theoretical domains [22]. For example, "prompts and cues" are associated with the "environmental context and resources." The final step involved linking the theoretical domains with the COM-B components; for example, "environmental context and resources" is linked with "physical opportunity." This process allowed the core behavior (the BCT was targeting) to be identified according to the process described in the BCW, a guide to designing interventions [22], which is displayed in Table 3.

Identified BCTs were narratively synthesized following the guidance for conducting systematic scoping reviews [41].

Table 3. The identified behavior change techniques (BCTs) and how they can be linked to a core behavior of the Capability, Opportunity, Motivation, Behavior (COM-B) using the theoretical domains framework (TDF).

Behavior change techniques used in different studies	Associated TDF domain	COM-B component identified in behavior analysis	COM core behavior
Prompts and cues [34-39]	Environmental context and resources	Physical opportunity	Opportunity
Self-monitoring of behavior [34,35,37]	Knowledge	Psychological capability	Capability
Self-monitoring of outcomes of behavior [35,37,38]	Knowledge	Psychological capability	Capability
Instruction on performing behavior [34-38]	Knowledge	Psychological capability	Capability

Results

Study Inclusion

Figure 1 details the process of selecting studies according to PRISMA-ScR standards. A total of 114 studies were retrieved and screened against the inclusion and exclusion criteria. A total of 6 studies met the inclusion criteria, including 4 randomized controlled trials (RCTs) [34-37], 1 pilot RCT [38], and 1 historical controlled trial [39].

Characteristics of Selected Studies

The studies included were published between 2017 and 2022. A total of 2 studies were published in Sweden [34,37], and the rest were published in Brazil [35], China [36], Japan [39], and Malaysia [38]. Sample sizes ranged from 26 to 129, for a total of 471 participants. All studies were treatment studies [34-39], and no prevention studies were found. A detailed summary of all the characteristics of the studies selected can be found in Table 1.

Sample Population

Women aged between 21 and 72 years were recruited in the included studies, with half of the studies including women aged 40 years or younger [36,38,39]. Women included in all 6 studies had access to a smartphone [37-39], and 3 studies required internet access [34,37,38]. A total of 2 studies involved pregnant women between 26 and 32 weeks of gestation [36,38], and 2 studies recruited women following recent vaginal delivery [36,39]. A total of 4 studies recruited women who had symptoms of SUI [34-36,38]. A total of 3 studies recruited women with mixed urinary incontinence [35,37,38], and 1 study recruited women with urge urinary incontinence [38].

Intervention and Control

The study length ranged from 2 to 6 months. In 4 studies [34,37-39] PFMT was recommended 3 times per day, and in 1 study, PFMT was recommended twice a day [35]. Alternatively, in 1 study [36], participants were given the option of completing PFMT twice a day or completing 150 pelvic floor muscle contractions a day [36]. The number of fast contractions varied between 0 and 10, the number of slow contractions between 1 and 6, or training time was recommended up to 15 minutes [36]. A total of 2 studies [34,37], stated their intervention group had different "combinations" of PFMT programs participants could use, but this was not expanded. The variety of PFMT training programs and modes of delivery used in the included studies highlights the lack of standardization of PFMT programs and could potentially influence the outcomes. In addition, a total of 3 studies [36,38,39] used a higher PFMT frequency and duration than recommended [9].

A total of 2 studies [35,39] used health care professionals to provide PFMT information to both the intervention and control groups. A specialist physiotherapist taught PFMT using surface electromyography and pelvic floor examination in 1 study [35]. Midwives taught participants through verbal instruction and written information in the other [39]. In comparison, 1 study used trained members of the research team to deliver PFMT [36]. In this study, all participants were given 2 sessions of a 45-minute presentation on pelvic floor education and one-on-one PFMT practice at the start and just before the end of the study [36]. The first presentation focused on the anatomy of the pelvic floor and the PFMT technique. The second focused on the importance of sustaining PFMT [36].

A total of 1 study offered 3 follow-ups with a continuous health care professional that included a pelvic floor examination for all participants [35]. One study offered follow-up telephone contact from a member of the research team once a month to encourage all participants to complete PFMT [36]. In 1 study [38], participants in the controlled group received usual antenatal care from a midwife.

A total of 2 studies provided written PFMT information to the control group [34,35], 1 study [38] postponed the provision of the intervention app until the study was complete, and 2 studies [36,39] provided no additional information. In 1 study [37], the control group received a restricted version of the PFMT mobile app used in the intervention group. A total of 4 studies [34,36,37,39] provided digital support to the intervention group using a PFMT mobile app. In summary, usual care varied between the studies.

It is evident that the variation in PFMT programs, modes of delivery, and levels of professional involvement underscores the lack of uniformity in PFMT across the studies. This makes it difficult to compare the results between studies.

Outcome Measures

Adherence was the primary outcome measure for 4 studies [35,36,38,39], with the remaining studies reporting it as a secondary outcome [34,37]. The primary outcomes of these studies were the International Consultation on Incontinence Questionnaire for Urinary Incontinence Short Form [42] and the Lower Urinary Tract Symptoms Short Form [42], used in 1 study [34].

The outcome measures used to measure PFMT adherence included the Exercise Adherence Rating Scale [38,43], the Broome Self-efficacy Scale for Practicing Pelvic Floor Exercises Questionnaire [36,44], self-reported adherence using questionnaires [39], or the Visual Analog Scale scores [35]. In one study [35], in-app adherence measures were also used that recorded the number of PFMT protocol repetitions the participants in the intervention groups completed. The findings suggest there was no agreement on how to measure adherence.

The BCTs Used in PFMT Mobile Apps

The BCTs used in the intervention groups were "self-monitoring of behavior," "self-monitoring of outcomes of behavior," "instruction on how to perform behavior," and "prompts and cues" (Table 2). The findings suggest that a combination of BCTs were used in the intervention group.

The BCT "prompts and cues" were used in all intervention groups [34-39]. In 5 of 6 intervention groups, the BCT "instruction on how to perform behavior" was used [34-38]. A total of 2 intervention groups used all the BCTs mentioned above [35,37], whereas 1 used only "prompts and cues" [39]. In summary, "prompts and cues" were the most frequently used BCT in the intervention groups.

Core Behaviors of the COM-B Framework Targeted by the BCTs Used in PFMT Mobile Apps

The COM-B core behaviors targeted by the BCTs used in the studies were opportunity and capability (Table 3). Opportunity was targeted using "prompts and cues" in all the intervention groups [34-39]. Capability was the core behavior targeted using the BCTs "self-monitoring of behavior," "self-monitoring of outcomes of behavior," and "instruction on how to perform behavior" in the intervention group [34-38]. The findings suggest that the core behavior opportunity was most frequently targeted using "prompts and cues" in the intervention groups.

PFMT Adherence

The levels of adherence to PFMT were greater in the intervention group compared with the control group, according to 2 studies [34,39]. In 1 study [34], daily PFMT adherence in the intervention group was 40% compared with 3% in the control group, with 60% of the control group reporting "sporadic" PFMT. Conversely, in 1 study [39], daily PFMT adherence was reported at 15 repetitions a day over 7 days in the intervention group compared with once a day over 3 days in the control group. In summary, higher levels of daily adherence to PFMT were observed in the intervention group.

In the control group, PFMT adherence decreased steadily over time [35,36]. In comparison, PFMT adherence increased over time in the control group of another study [38]. Participants in the control group, where adherence increased over time, received a PFMT mobile app at the end of the study period, which may have influenced the results.

Discussion

Principal Results

To achieve objective 1, we conducted an analysis of the BCTs used within studies with PFMT mobile app features. Our analysis identified a combination of BCTs with a range of strategies to facilitate adherence to PFMT, including "self-monitoring of behavior," "self-monitoring of outcomes of behavior," "instruction on how to perform behavior," and "prompts and cues." "Prompts and cues" were the most widely used BCT. To achieve objective 2, we mapped the BCTs used in the studies onto the core behaviors of the COM-B model. This highlighted the central emphasis on the core behavior of opportunity, evident through the prevalence of "prompts and cues" across all intervention groups. Additionally, "self-monitoring of outcomes of behavior" and "instruction on how to perform behavior" were used to address the core behavior capability. The widespread use of "prompts and cues" highlights their role in fostering PFMT engagement by making the most of opportune moments. Furthermore, the incorporation of "self-monitoring of behavior" and "instruction on performing behavior" empowers participants with the necessary knowledge and confidence for effectively executing PFMT. To address objective 3, data on PFMT adherence were extracted from the included studies. This included adherence data from self-efficacy measures and patients' self-reported adherence measures. The findings suggest higher levels of PFMT adherence were observed in women using PFMT mobile apps; however, there was heterogeneity in the outcome measures used in the included studies, and meta-analysis was not possible.

Digital "prompts and cues" were used by all PFMT mobile apps. Digital "prompts and cues" are a BCT that target the core behavior opportunity according to the COM-B model. Digital reminders are a "persuasive" behavioral intervention that may be valuable when it comes to changing women's behavior around PFMT, since women commonly associate poor adherence with forgetting to complete PFMT [14,28]. Persuasive behavioral interventions can help women overcome known barriers to PFMT by creating more positive attitudes toward the exercises [28]. These findings are supported by a systematic review that found daily PFMT reminders "stimulate" daily adherence to PFMT in women [21]. Similarly, a systematic review that aimed to determine the persuasive features used in PFMT mobile apps recognized that reminder features were a way of assisting users to remember to perform PFMT, which the authors mapped under the core behavior motivation rather than opportunity [28], which demonstrates how elements of BCT are disputed. In the wider literature, digital reminders in populations with long-term conditions, including diabetes, suggest that digital reminders increase adherence to medication and physical activity [45-47], although none of these papers discuss how behaviors might be influenced. In this review, the results suggest digital PFMT prompts increased daily PFMT adherence in women, and it might be argued that leveraging either or both of the core behaviors of opportunity or motivation. However, further rigorous research is required to explore the effects of digital PFMT prompts on women's PFMT adherence while controlling other behavioral variables.

Some studies have observed higher levels of PFMT adherence among women using PFMT mobile apps [34,39]. Although the intervention groups reported higher levels of daily adherence, women did not meet the daily PFMT recommendations of 3 times a day [2]. In 1 study [35], women completed PFMT 1.6 times a day on average, and in another [37], only 10% of women completed PFMT 3 times a day. One reason for the lower levels of PFMT adherence may have been the heterogeneity of the outcome measures of adherence that were inconsistently monitored and poorly described [12]. Alternatively, the heterogeneity in PFMT program prescription may have influenced adherence to PFMT, with 3 studies prescribing a higher dosage of PFMT [36,38,39] than recommended [9]. The high frequency and intensity of PFMT recommendations are at odds with the "easy, accessible, social, timely" principles of behavior change [48]. If the exercises and the intensity of the regime appear difficult, it is likely that women are less likely to engage in and adhere to PFMT. This is particularly important to consider given the fact that no prevention studies were identified in the scoping review, and women may find adhering to PFMT even harder if they are asymptomatic.

Women's adherence to PFMT appeared to decrease around 3 months [35,36,38]. These findings are supported by other studies that have explored adherence rates in people with chronic disease and national guidelines [2,48], suggesting digital prompts may only be valuable in the short term. It is recognized that short-term PFMT adherence is greater than long-term adherence [7]. Previous research indicated that behavior change can take up to a year [49]. In the included studies of this review, there were none that followed up a year postintervention, and therefore we are unable to make a comment on the long-term nature of behavioral change. The decrease in PFMT adherence after 3 months may reflect a normal drop in adherence levels [49] because women notice an improvement in their symptoms and no longer consider the exercises to be beneficial. It has been suggested that "social opportunity" rather than "physical opportunity" (referenced in the COM-B model) is important when it comes to behavioral maintenance, since actions are more likely to be sustained when in line with group norms [49]. Further behavioral research on PFMT adherence in women should increase the length of study to 1 year and explore the role of “social opportunity” on adherence.

Conversely, in the control group of 1 study, adherence increased at 2 months [38]. A potential reason for this is that the control group received “usual antenatal care” that involved contact with health care professionals. The influence of health care experts on exercise adherence is “pivotal,” according to other studies [50]. Arguably, contact and follow-up with a health care professional as a “credible source” may be a valuable BCT in relation to PFMT adherence in women because it holds women accountable to complete PFMT [22]. The use of health care professionals to teach PFMT and follow-up participants in some of the included studies may have led to confounding behavioral factors that influenced individuals’ adherence to PFMT. Future research should consider the influence health care professionals have on participants PFMT adherence when conducting future research on PFMT mobile apps.

Strengths and Limitations

Using a scoping review methodology, we have been able to identify a breadth of experimental studies investigating the effect of PFMT apps on adherence to PFMT in women as an intervention. To the authors’ knowledge, this study was the first scoping review to identify and discuss specific BCTs embedded within existing features of PFMT mobile apps and to map the core behaviors targeted using the COM-BCW framework. By doing so, this research contributes to a deeper comprehension of the factors influencing PFMT adherence among women and highlights the potential use of PFMT mobile apps, particularly those with digital prompts, to increase adherence to PFMT. Additionally, the study uses the COM-B model to provide a holistic understanding of how capability, opportunity, and motivation are used in the context of PFMT adherence, bridging the gap between theoretical concepts and application that can inform the development and enhancement of interventions targeting PFMT adherence.

A limitation of the study was the way BCTs were assigned to the features of PFMT mobile apps based on descriptions of the apps and app features in the included papers. One BCT was assigned per app feature, and features that did not specifically relate to any of the 93 BCTs were not included. Inaccurate descriptions of mobile app features by the authors of the included studies may have resulted from misattribution or underestimation of the BCTs and therefore misrepresentation of the core behaviors targeted by PFMT mobile apps. The misrepresentation of core behaviors may have made the core behaviors that encourage women to adhere to PFMT unclear. Additional BCTs may have been identified if the PFMT mobile apps were downloaded or original papers around the app development were sought and appraised.

Another limitation of the study was the exclusion of studies without a control group. The exclusion of additional studies, such as observational studies, meant that studies with potentially fuller descriptions of PFMT mobile apps and their intended behavioral mechanisms were not included in the analysis. Arguably, this limits the depth of the analysis in the study by limiting the comparison and corroboration between a greater number of PFMT mobile apps that would have better addressed the first 2 objectives of this study. Alternatively, the inclusion of studies with a control group increased our confidence in the findings that suggest levels of adherence were higher in women using a PFMT mobile app compared to an alternative variable and addressed objective 3.

This review identified the use of mobile apps for the treatment of PFMT and the behaviors being leveraged; however, due to the heterogeneity of the PFMT regimes across the studies, it was not possible to compare. In the future, it will be important to identify the minimum effective training dosage women engage with and why.

Conclusion

The findings suggest there was a trend in higher levels of PFMT adherence in women using PFMT mobile apps over 3 months. However, there was no consistency across the studies in the PFMT regimes and outcome measures. Digital “prompts and cues” were used to encourage adherence and may be the BCT that leveraged improved adherence. Opportunity was the most targeted core behavior in the intervention groups of the included studies. Further research is required to practically assess whether a future RCT that investigates the effectiveness of digital “prompts and cues” on PFMT adherence in women can be conducted.

Acknowledgments

This work was supported by the National Institute for Health Research (Applied Research Collaboration Wessex Doctoral Award, in conjunction with Bournemouth University and University Hospitals Dorset). The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms.

[DOCX File , 13 KB - mhealth_v11i1e45947_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - mhealth_v11i1e45947_app2.docx]

Multimedia Appendix 3
Raw data extraction table of the included studies.
[XLSX File (Microsoft Excel File), 22 KB - mhealth_v11i1e45947_app3.xlsx]

Multimedia Appendix 4
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 552 KB - mhealth_v11i1e45947_app4.pdf]

References

1. Burkhart R, Couchman K, Crowell K, Jeffries S, Monvillers S, Vilensky J. Pelvic floor dysfunction after childbirth: Occupational impact and awareness of available treatment. OTJR (Thorofare N J) 2021 Apr;41(2):108-115. [doi: 10.1177/1539449220970881] [Medline: 33176560]
2. Pelvic floor dysfunction: prevention and non-surgical management. The National Institute of Health and Care Excellence. 2021. URL: https://tinyurl.com/cu5k7jhb [accessed 2023-11-17]
3. Wood LN, Anger JT. Urinary incontinence in women. BMJ 2014 Sep 15;349:g4531. [doi: 10.1136/bmj.g4531] [Medline: 25225003]
4. Hagen S, Stark D, Glazener C, Dickson S, Barry S, Elders A, et al. Individualised pelvic floor muscle training in women with pelvic organ prolapse (POPPY): a multicentre randomised controlled trial. Lancet 2014;383(9919):796-806 [FREE Full text] [doi: 10.1016/S0140-6736(13)61977-7] [Medline: 24290404]
5. Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, Lee J, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Neurourol Urodyn 2010;29(1):4-20 [FREE Full text] [doi: 10.1002/nau.20798] [Medline: 19941278]
6. Hay-Smith EJC, Herderschee R, Dumoulin C, Herbison GP. Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women. Cochrane Database Syst Rev 2011(12):CD009508. [doi: 10.1002/14651858.CD009508] [Medline: 22161451]
7. Dumoulin C, Hay-Smith J, Frawley H, McClurg D, Alewijnse D, Bo K, et al. 2014 consensus statement on improving pelvic floor muscle training adherence: International Continence Society 2011 state-of-the-science seminar. Neurourol Urodyn 2015 Sep;34(7):600-605. [doi: 10.1002/nau.22796] [Medline: 25998603]
8. Widdison R, Rashidi A, Whitehead L. Effectiveness of mobile apps to improve urinary incontinence: a systematic review of randomised controlled trials. BMC Nurs 2022;21(1):32 [FREE Full text] [doi: 10.1186/s12912-022-00812-6] [Medline: 35090464]
9. Woodley SJ, Lawrenson P, Boyle R, Cody JD, Mørkved S, Kernohan A, et al. Pelvic floor muscle training for preventing and treating urinary and faecal incontinence in antenatal and postnatal women. Cochrane Database Syst Rev 2020;5(5):CD007471 [FREE Full text] [doi: 10.1002/14651858.CD007471.pub4] [Medline: 32378735]
10. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;2014(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
11. Odell J, Clark C, Hunnisett A, Ahmed OH, Branney J. Manual therapy for chronic migraine: a pragmatic randomised controlled trial study protocol. Chiropr Man Therap 2019;27:11 [FREE Full text] [doi: 10.1186/s12998-019-0232-4] [Medline: 30962877]
12. Dumoulin C, Alewijnse D, Bo K, Hagen S, Stark D, Van Kampen M, et al. Pelvic-Floor-Muscle Training adherence: tools, measurements and strategies-2011 ICS state-of-the-science seminar research paper II of IV. Neurourol Urodyn 2015;34(7):615-621. [doi: 10.1002/nau.22794] [Medline: 25998493]
13. Farrance C, Tsofliou F, Clark C. Adherence to community based group exercise interventions for older people: a mixed-methods systematic review. Prev Med 2016;87:155-166 [FREE Full text] [doi: 10.1016/j.ypmed.2016.02.037] [Medline: 26921655]
14. Borello-France D, Burgio KL, Goode PS, Ye W, Weidner AC, Lukacz ES, et al. Adherence to behavioral interventions for stress incontinence: rates, barriers, and predictors. Phys Ther 2013;93(6):757-773 [FREE Full text] [doi: 10.2522/ptj.20120072] [Medline: 23431210]
15. Salmon VE, Hay-Smith EJC, Jarvie R, Dean S, Terry R, Frawley H, et al. Implementing pelvic floor muscle training in women's childbearing years: a critical interpretive synthesis of individual, professional, and service issues. Neurourol Urodyn 2020;39(2):863-870 [FREE Full text] [doi: 10.1002/nau.24256] [Medline: 31845393]
16. Sairally BZF, Clark TJ. Prioritisation of outpatient appointments and elective surgery in gynaecology. Best Pract Res Clin Obstet Gynaecol 2021;73:2-11 [FREE Full text] [doi: 10.1016/j.bpobgyn.2021.03.002] [Medline: 33883091]

17. Health technology assesssment. World Health Organization. 2022. URL: https://www.who.int/health-topics/health-technology-assessment#tab=tab_1 [accessed 2022-01-06]
18. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
19. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
20. Chatzigeorgiou IM, Andreou GT. A systematic review on feedback research for residential energy behavior change through mobile and web interfaces. Renew Sust Energ Rev 2021;135:110187. [doi: 10.1016/j.rser.2020.110187]
21. Nagib ABL, Riccetto C, Martinho NM, Pennisi PRC, Blumenberg C, Paranhos LR, et al. Use of mobile apps for controlling of the urinary incontinence: a systematic review. Neurourol Urodyn 2020;39(4):1036-1048. [doi: 10.1002/nau.24335] [Medline: 32187704]
22. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback Publishing; 2014.
23. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
24. Murphy K, Berk J, Muhwava-Mbabala L, Booley S, Harbron J, Ware L, et al. Using the COM-B model and behaviour change wheel to develop a theory and evidence-based intervention for women with gestational diabetes (IINDIAGO). BMC Public Health 2023;23(1):894 [FREE Full text] [doi: 10.1186/s12889-023-15586-y] [Medline: 37189143]
25. Bernard S, Boucher S, McLean L, Moffet H. Mobile technologies for the conservative self-management of urinary incontinence: a systematic scoping review. Int Urogynecol J 2020;31(6):1163-1174. [doi: 10.1007/s00192-019-04012-w] [Medline: 31267139]
26. Hou Y, Feng S, Tong B, Lu S, Jin Y. Effect of pelvic floor muscle training using mobile health applications for stress urinary incontinence in women: a systematic review. BMC Womens Health 2022;22(1):400 [FREE Full text] [doi: 10.1186/s12905-022-01985-7] [Medline: 36192744]
27. Xu P, Wang X, Guo P, Zhang W, Mao M, Feng S. The effectiveness of eHealth interventions on female pelvic floor dysfunction: a systematic review and meta-analysis. Int Urogynecol J 2022;33(12):3325-3354 [FREE Full text] [doi: 10.1007/s00192-022-05222-5] [Medline: 35616695]
28. Jaffar A, Tan CE, Mohd-Sidik S, Admodisastro N, Goodyear-Smith F. Persuasive technology in an mHealth app designed for pelvic floor muscle training among women: systematic review. JMIR Mhealth Uhealth 2022;10(3):e28751 [FREE Full text] [doi: 10.2196/28751] [Medline: 35315777]
29. Peters MDJ, Marnie C, Colquhoun H, Garritty CM, Hempel S, Horsley T, et al. Scoping reviews: reinforcing and advancing the methodology and application. Syst Rev 2021;10(1):263 [FREE Full text] [doi: 10.1186/s13643-021-01821-3] [Medline: 34625095]
30. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
31. Pollock D, Davies EL, Peters MDJ, Tricco AC, Alexander L, McInerney P, et al. Undertaking a scoping review: a practical guide for nursing and midwifery students, clinicians, researchers, and academics. J Adv Nurs 2021;77(4):2102-2113 [FREE Full text] [doi: 10.1111/jan.14743] [Medline: 33543511]
32. Belisario JSM, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013;2013(11):CD010013 [FREE Full text] [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
33. Harper RC, Clark CJ, Stewart C. The role of pelvic floor muscle training mobile apps on pelvic floor muscle exercise adherence in women: a systematic review. PROSPERO. International Prospective Register of Systematic Reviews. 2022. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022323789 [accessed 2023-11-17]
34. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017;36(5):1369-1376 [FREE Full text] [doi: 10.1002/nau.23116] [Medline: 27611958]
35. Araujo CC, de A Marques A, Juliato CRT. The adherence of home pelvic floor muscles training using a mobile device application for women with urinary incontinence: a randomized controlled trial. Female Pelvic Med Reconstr Surg 2020;26(11):697-703. [doi: 10.1097/SPV.0000000000000670] [Medline: 30624250]
36. Wang X, Xu X, Luo J, Chen Z, Feng S. Effect of app-based audio guidance pelvic floor muscle training on treatment of stress urinary incontinence in primiparas: a randomized controlled trial. Int J Nurs Stud 2020;104:103527. [doi: 10.1016/j.ijnurstu.2020.103527] [Medline: 32058140]
37. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A mobile app for self-management of urgency and mixed urinary incontinence in women: randomized controlled trial. J Med Internet Res 2021;23(4):e19439 [FREE Full text] [doi: 10.2196/19439] [Medline: 33818395]

38. Jaffar A, Mohd Sidik S, Foo CN, Muhammad NA, Abdul Manaf R, Suhaili N. Preliminary effectiveness of mHealth app-based pelvic floor muscle training among pregnant women to improve their exercise adherence: a pilot randomised control trial. Int J Environ Res Public Health 2022;19(4):2332 [FREE Full text] [doi: 10.3390/ijerph19042332] [Medline: 35206520]
39. Kinouchi K, Ohashi K. Smartphone-based reminder system to promote pelvic floor muscle training for the management of postnatal urinary incontinence: historical control study with propensity score-matched analysis. PeerJ 2018;6:e4372 [FREE Full text] [doi: 10.7717/peerj.4372] [Medline: 29441244]
40. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions, 2nd Edition. Chichester, UK: John Wiley & Sons; 2019.
41. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015;13(3):141-146 [FREE Full text] [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
42. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. Neurourol Urodyn 2004;23(4):322-330. [doi: 10.1002/nau.20041] [Medline: 15227649]
43. Newman-Beinart NA, Norton S, Dowling D, Gavriloff D, Vari C, Weinman JA, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy 2017;103(2):180-185. [doi: 10.1016/j.physio.2016.11.001] [Medline: 27913064]
44. Broome BA. Development and testing of a scale to measure self-efficacy for pelvic muscle exercises in women with urinary incontinence. Urol Nurs 1999;19(4):258-268. [Medline: 10889770]
45. Louch G, Dalkin S, Bodansky J, Conner M. An exploratory randomised controlled trial using short messaging service to facilitate insulin administration in young adults with type 1 diabetes. Psychol Health Med 2013;18(2):166-174. [doi: 10.1080/13548506.2012.689841] [Medline: 22646659]
46. Schwebel FJ, Larimer ME. Using text message reminders in health care services: a narrative literature review. Internet Interv 2018;13:82-104 [FREE Full text] [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]
47. Yasmin F, Banu B, Zakir SM, Sauerborn R, Ali L, Souares A. Positive influence of short message service and voice call interventions on adherence and health outcomes in case of chronic disease care: a systematic review. BMC Med Inform Decis Mak 2016;16:46 [FREE Full text] [doi: 10.1186/s12911-016-0286-3] [Medline: 27106263]
48. Hallsworth M, Snijders V, Burd H, Prestt J, Judah G, Huf S, et al. Applying behavioral insights: simple ways to improve health outcomes. World Innovation Summit for Health. Doha, Qatar; 2016. URL: https://www.bi.team/publications/applying-behavioural-insights-simple-ways-to-improve-health-outcomes/
49. Behaviour change: individual approaches. The National Institute of Health and Care Excellence. 2014. URL: https://www.nice.org.uk/guidance/ph49/resources/behaviour-change-individual-approaches-pdf-1996366337989 [accessed 2022-01-06]
50. Kwok BC, Wong WP, Remedios L. Improving centre-based group exercise participation of older adults using the behaviour change wheel. BMJ Open Qual 2021;10(1):e001078 [FREE Full text] [doi: 10.1136/bmjoq-2020-001078] [Medline: 33589505]

Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
COM-B: Capability Opportunity Motivation Behavior
COM-BCW: Capability Opportunity Motivation Behavior change wheel
PFD: pelvic floor dysfunction
PFMT: pelvic floor muscle training
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SUI: stress urinary incontinence
VAS: Visual Analog Scale

Edited by L Buis; submitted 23.01.23; peer-reviewed by J Hay-Smith, W Perveen, R May; comments to author 15.05.23; revised version received 01.07.23; accepted 02.11.23; published 30.11.23.

Please cite as:
Harper RC, Sheppard S, Stewart C, Clark CJ
Exploring Adherence to Pelvic Floor Muscle Training in Women Using Mobile Apps: Scoping Review
JMIR Mhealth Uhealth 2023;11:e45947
URL: https://mhealth.jmir.org/2023/1/e45947
doi:10.2196/45947
PMID:38032694

Review

Mobile Technology–Based Interventions for Stroke Self-Management Support: Scoping Review

Alexandra N Thompson[1,2], MScOT; Deirdre R Dawson[1,2,3], PhD; Jean Michelle Legasto-Mulvale[1,4], MScPT; Nivetha Chandran[1], HBSc; Chelsea Tanchip[1], MSc; Veronika Niemczyk[5], MScOT; Jillian Rashkovan[3]; Saisa Jeyakumar[3], HBSc; Rosalie H Wang[1,3,6], PhD; Jill I Cameron[1,3,6], PhD; Emily Nalder[1,3,6], PhD

[1]Rehabilitation Sciences Institute, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Rotman Research Institute, Baycrest Health Sciences, Toronto, ON, Canada

[3]Department of Occupational Science and Occupational Therapy, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[4]Department of Physical Therapy, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[5]School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[6]KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

Corresponding Author:
Emily Nalder, PhD
Rehabilitation Sciences Institute
Temerty Faculty of Medicine
University of Toronto
Suite 160
500 University Avenue
Toronto, ON, M5G 1V7
Canada
Phone: 1 416 978 5937
Email: emily.nalder@utoronto.ca

Abstract

Background: There is growing interest in enhancing stroke self-management support using mobile health (mHealth) technology (eg, smartphones and apps). Despite this growing interest, "self-management support" is inconsistently defined and applied in the poststroke mHealth intervention literature, which limits efforts to synthesize and compare evidence. To address this gap in conceptual clarity, a scoping review was conducted.

Objective: The objectives were to (1) identify and describe the types of poststroke mHealth interventions evaluated using a randomized controlled trial design, (2) determine whether (and how) such interventions align with well-accepted conceptualizations of self-management support (the theory by Lorig and Holman and the Practical Reviews in Self-Management Support [PRISMS] taxonomy by Pearce and colleagues), and (3) identify the mHealth functions that facilitate self-management.

Methods: A scoping review was conducted according to the methodology by Arksey and O'Malley and Levac et al. In total, 7 databases were searched. Article screening and data extraction were performed by 2 reviewers. The data were analyzed using descriptive statistics and content analysis.

Results: A total of 29 studies (26 interventions) were included. The interventions addressed 7 focal areas (physical exercise, risk factor management, linguistic exercise, activities of daily living training, medication adherence, stroke education, and weight management), 5 types of mobile devices (mobile phones or smartphones, tablets, wearable sensors, wireless monitoring devices, and laptops), and 7 mHealth functions (educating, communicating, goal setting, monitoring, providing feedback, reminding, and motivating). Collectively, the interventions aligned well with the concept of self-management support. However, on an individual basis (per intervention), the alignment was less strong.

Conclusions: On the basis of the results, it is recommended that future research on poststroke mHealth interventions be more theoretically driven, more multidisciplinary, and larger in scale.

(JMIR Mhealth Uhealth 2023;11:e46558) doi:10.2196/46558

KEYWORDS

stroke; chronic disease; self-management; rehabilitation; mobile health; mHealth; eHealth; telehealth; telemedicine; digital health; mobile phone

Introduction

Background

Managing the chronic effects of stroke (eg, mobility problems, cognitive impairment, and depression) has become a global health priority because of its enormous burden on health care systems [1,2]. In Canada, >400,000 people live with the effects of stroke, and by 2038, this number is expected to increase to nearly 700,000 [3]. To meet the needs of this growing population and address international priorities, self-management support interventions for stroke are of growing interest to researchers and health care professionals. Broadly defined, self-management support is a complex intervention that provides people with knowledge, confidence, and skills to manage their chronic condition [4]. Self-management support interventions have been shown to improve a variety of health outcomes after stroke, including risk factor control [5], functional ability [6], participation [6], and quality of life [7]. They have also been recommended in recent clinical practice guidelines [8]. Unfortunately, however, because of limited health care budgets and unequal access to rehabilitation, few Canadians have the opportunity to participate in self-management support interventions following stroke [8,9]. Increased access to timely, effective, and low-cost stroke self-management support could be provided through mobile health (mHealth) technology–based (eg, smartphone app–based) interventions.

Despite the growing potential, need, and interest in enhancing stroke self-management support interventions with mHealth, the evidence for its effectiveness remains unclear. In previous reviews of poststroke mHealth interventions, connections were drawn to "self-management support"; however, the concept was never explicitly defined or operationalized [10-14]. In these previous reviews, self-management support was discussed in a way that suggests that it is a newly emerging concept in the literature on poststroke mHealth interventions. Specifically, in the abstract and introduction of 3 reviews, self-management was framed as a key concept in the rationale for the review [10,12,14]. For example, in 1 review, mHealth for self-management was described as a "new strategy for stroke rehabilitation" [10]. In the discussion of 2 reviews, improved self-management was highlighted as an important outcome of mHealth use [11,13]. In the conclusion of 1 review, identifying literature on mHealth interventions to support self-management was stated as the purpose of the study [11]. Although clearly emphasizing an interest in the concept, without explicit definitions or operationalizations, the literature remains challenging to synthesize and compare, which may lead future reviews to draw incorrect conclusions about intervention effectiveness [15]. To our knowledge, no review has addressed this lack of conceptual clarity; that is, no review has aimed to map the literature on poststroke mHealth interventions according to well-accepted conceptualizations of self-management support.

Objectives

To address this gap in the literature, we conducted a scoping review. This method was selected for its utility in clarifying key concepts in the literature, identifying key characteristics related to a concept, and identifying and analyzing knowledge gaps in an emerging field [16]. The objectives were to (1) identify and describe the types of poststroke mHealth interventions evaluated using a randomized controlled trial (RCT) design, (2) determine whether (and how) such interventions align with well-accepted conceptualizations (theory [17] and taxonomy [18]) of self-management support, and (3) identify the mHealth functions that facilitate self-management. The purpose of this study was to identify gaps in the literature and recommendations for future research related to mHealth-enhanced stroke self-management support.

Methods

Design

Using well-established methods [19,20], a scoping review was conducted. The protocol was not registered. A critical appraisal of the included studies was not conducted as the aim of this review was to map the breadth and depth of conceptualizations, not to draw conclusions about intervention effectiveness [16]. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist is provided in Multimedia Appendix 1 [21].

Identifying Relevant Studies

In consultation with 2 librarians, ANT searched MEDLINE, Embase, PsycINFO, CINAHL, AMED, Scopus, and ProQuest Dissertations and Theses Global 3 times (October 2-3, 2020, February 28, 2022, and July 10, 2023). The second and third searches were conducted to identify new literature published between 2020 and 2022 and between 2022 and 2023. The search terms captured 2 search concepts: stroke and mHealth (see Multimedia Appendix 2 for the full Ovid search strategy).

Selecting Studies

ANT, JML-M, NC, CT, VN, JR, and SJ conducted level-1 (title and abstract) and level-2 (full-text) screening in duplicate using Covidence (Veritas Health Innovation). Disagreements were resolved through consensus-based discussion. Studies were included if the article reported original research, the study included human participants with stroke or transient ischemic attack, the study evaluated an mHealth intervention (mHealth defined using 2 definitions: those of the World Health Organization—"[the] medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants [PDAs], and other wireless devices" [22]—and Akter et al [23]—"focusing on any wireless technologies [e.g., Bluetooth, GSM, GPRS/3G, Wi-Fi, WiMAX] to transmit various health-related data content and services through mobile devices, including mobile phones, smartphones, PDAs, laptops and Tablet PCs"), and the study was an RCT. The search was limited to RCTs as a preliminary search identified a large number of studies using an RCT design. In addition, as RCTs are typically regarded as the highest in quality and presumably are the farthest along in the technology development process, their influence on research and practice was thought to be the most significant. Studies were excluded if the sample was mixed (eg, acquired brain injury), the intervention included client (person with stroke)–facing technology or equipment that was not clearly mobile and wireless, the article did not report any outcome measures related to intervention effectiveness, and the article was not written in English.

Charting the Data

ANT developed the data-charting form in collaboration with DRD and EN. ANT charted the data verbatim and then JML-M, NC, CT, VN, JR, and SJ verified the data. Data were charted from the included articles as well as from supplementary materials and protocol papers when referenced. The data-charting form included study characteristics (eg, study aims and outcome measures), participant characteristics (eg, time since stroke and sex or gender), and intervention characteristics (based on the Template for Intervention Description and Replication checklist [24]). Visual information related to the intervention characteristics was also charted (eg, screenshots of apps).

Collating, Summarizing, and Reporting the Results

ANT completed the data analysis in collaboration with DRD, EN, RHW, and JIC. Quantitative data were analyzed using descriptive statistics, and qualitative data were analyzed using conventional content analysis (objective 1) and directed content analysis (objectives 2 and 3) [25]. Directed content analysis for objective 2 was guided by the theory by Lorig and Holman [17] and the Practical Reviews in Self-Management Support (PRISMS) taxonomy by Pearce et al [18] as they are widely cited, slightly different conceptualizations of self-management support (see Multimedia Appendix 3 [17,18] for the operational definitions of codes). Directed content analysis for objective 3 was guided by the definition of mHealth functions by Cameron et al [26] ("the verbs describing the behavior of the system"), examples from previous research on mHealth functions [27-31], and dictionary definitions [32-38] (see Multimedia Appendix 4 [29-38] for the operational definitions of codes).

Results

Study Characteristics

A total of 29 studies describing 26 interventions were included (see Figure 1 for the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flow diagram [39]). The studies were published between 2007 and 2023 and were from Asia (13/29, 45%), Europe (8/29, 28%), North America (4/29, 14%), Africa (2/29, 7%), and Australia (2/29, 7%). Of the 29 studies, 1 (3%) was a doctoral dissertation [40] and the remaining 28 (97%) were peer-reviewed journal articles. A total of 34% (10/29) of the studies were considered pilot, proof-of-concept, or feasibility studies. The sample sizes ranged from 11 to 4298. Table 1 presents the study and participant characteristics.

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

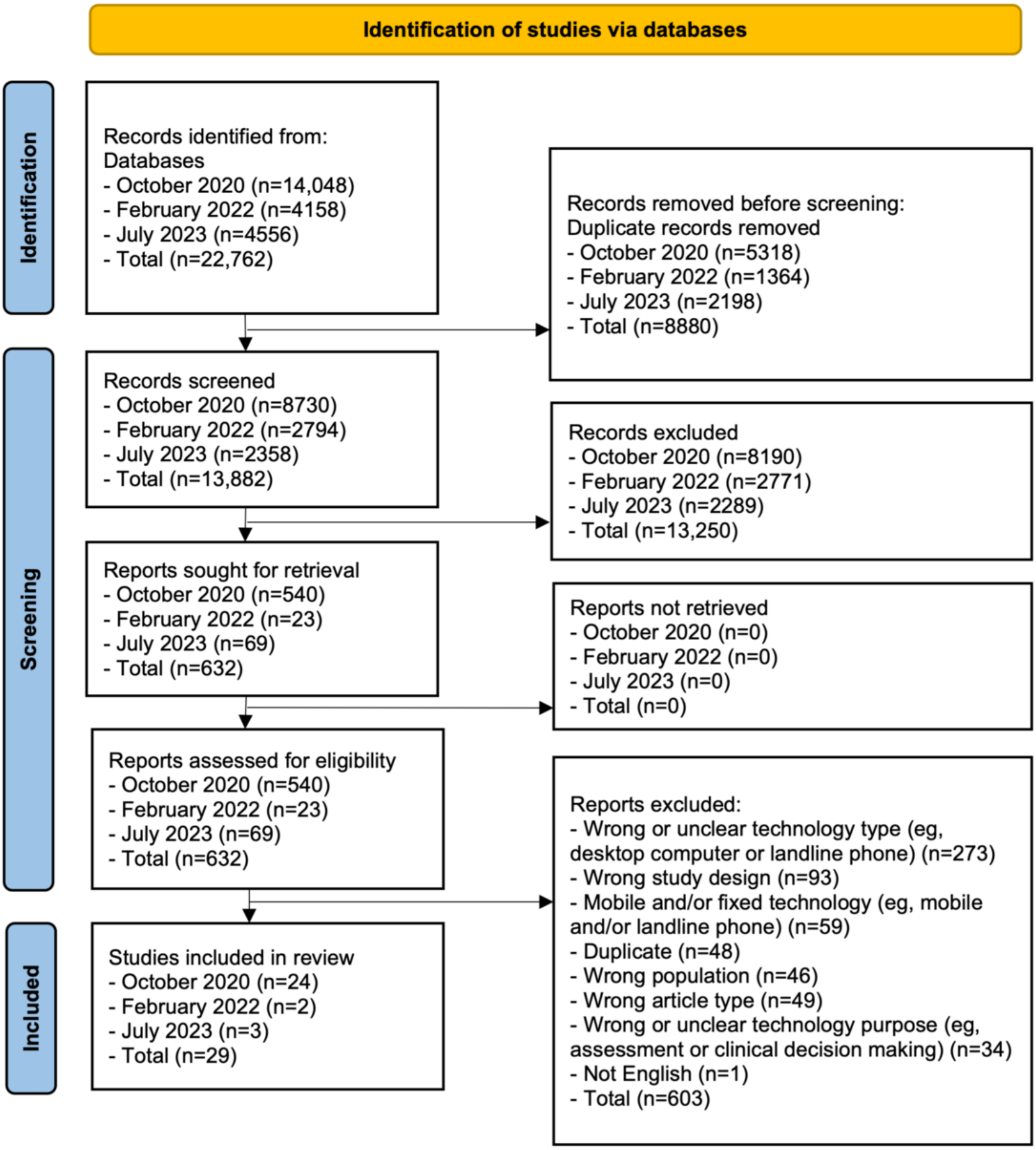


Table 1. Study and participant characteristics.

Study	Country	Study aims	Outcome measures	Sample size	Stroke type	Time since stroke or stroke stage	Age (y)	Sex or gender (male or men; %)
Asano et al [41]	Singapore	Effectiveness	Late-Life Function and Disability Instrument; timed 5-Meter Walk Test; 2-minute walk test; Modified Barthel Index; Activities-Specific Balance Confidence Scale; EQ-5D	98 (IG[a]: 50; CG[b]: 48)	78% ischemic; 22% hemorrhagic	Within 4 wk	Mean 64.1 (range 40.5-89.6)	52
Emmerson et al [42]	Australia	Effectiveness	Self-report logbook; Wolf Motor Function Test; customized questionnaire	58 (IG: 28; CG: 30)	86% ischemic; 15% hemorrhagic	Median 120 (range 58-226) d	Mean 66 (SD 16)	63
Ferrete Ruíz et al [43]	Spain	Effectiveness	Mississippi Aphasia Screening Test; minutes of device use	30 (IG: 23; CG: 7)	100% ischemic	Within 7 d	IG: mean 73.20 (SD 9.53); CG: mean 72.40 (SD 2.79)	50
Grau-Pellicer et al [44]	Spain	Effectiveness	Participant reports of community ambulation and sedentary behavior; 10-Meter Walk Test; 6-Minute Walk Test; Timed Up and Go test; Barthel Index; EQ-5D; satisfaction questionnaire	34 (IG: 21; CG: 13)	Not reported	IG: mean 18.92 (SD 27.6; range 1-96) mo; CG: mean 20.85 (SD 59.74; range 1-252) mo	IG: mean 62.96 (SD 11.87; range 33-89); CG: mean 68.53 (SD 11.53; range 41-83)	51
Hankinson et al [45]	Australia	Pilot study; feasibility; effectiveness	Adherence to intervention; Fugl-Meyer Assessment	15 (IG: 6; CG: 9)	Not reported	0-180 d	Not reported	59
Ifejika et al [46]	United States	Pilot; feasibility; preliminary effectiveness	Reduction in total body weight; compliance with the weight loss intervention; Patient Health Questionnaire–9; systolic blood pressure; serum low-density lipoprotein value; proportion of total hemoglobin; proportion of serum coagulation factor VIII	25 (IG: 13; CG: 12)	Not reported	Acutely hospitalized	Mean 54.1 (SD 9.4)	56
Jang and Jang [47]	South Korea	Effectiveness	Manual Muscle Test; Manual Function Test; Purdue Pegboard Test	21 (IG: 10; CG: 11)	19% ischemic; 81% hemorrhagic	Mean 963 (SD 799) d	Mean 44.5 (SD 16.5)	71
Kamal et al [48]	Pakistan	Effectiveness	Morisky Medication Adherence Scale; systolic and diastolic blood pressure; satisfaction questionnaires	162 (IG: 83; CG: 79)	84% ischemic; 17% hemorrhagic	IG: median 2 (range 1-5) mo; CG: median 2 (range 1-4) mo	IG: mean 56.07 (SD 1.5); CG: mean 57.62 (SD 1.3)	68
Kamal et al [49]	Pakistan	Effectiveness; safety	Systolic and diastolic blood pressure; blood sugar (HbA_{1c}[c]); blood cholesterol (low-density lipoprotein level); mortality; modified Rankin Scale; National Institutes of Health Stroke Scale; Barthel Index	277 (IG: 141; CG: 136)	Not reported	Not reported	IG: mean 60.6 (SD 12); CG: mean 59.7 (SD 14.3)	67

Study	Country	Study aims	Outcome measures	Sample size	Stroke type	Time since stroke or stroke stage	Age (y)	Sex or gender (male or men; %)
Kamwesiga et al [50]	Sweden	Feasibility study; preliminary effectiveness	Logbook to record adherence; Canadian Occupational Performance Measure; self-efficacy scale; Stroke Impact Scale 3.0 Uganda version; Barthel Index; Occupational Gaps Questionnaire, Ugandan version	28 (IG: 13; CG: 15)	75% ischemic; 21% hemorrhagic; 4% unspecified	3-6 mo—IG: 10 (76.9%) and CG: 10 (66.7%); 7-11 mo—IG: 3 (23.1%) and CG: 3 (20%); 1-2 y—IG: 0 and CG: 2 (13.3%)	IG: mean 61.2 (SD 15); CG: mean 58.5 (SD 14)	25
Kang et al [51]	South Korea	Effectiveness	Regional House-Brackmann Facial Nerve Grading System; length between the corner of the mouth and the earlobe	21 (IG: 10; CG: 11)	95% ischemic; 5% hemorrhagic	Within 12 wk	IG: mean 63.1 (SD 10.3); CG: mean 55.6 (SD 16)	62
Kang et al [52]	Taiwan	Effectiveness	Stroke knowledge questionnaire; EQ-5D	63 (IG: 30; CG: 33)	43% ischemic; 57% hemorrhagic	Not reported	IG: mean 50.47 (SD 10.82); CG: mean 52.33 (SD 11.03)	68
Kenny et al [53]	United Kingdom	Feasibility; acceptability; preliminary effectiveness	Motor Status Scale; Leeds Movement Performance Index; General Self-Efficacy Scale; diary to record time spent exercising	11 (IG: 5; CG: 6)	77% ischemic; 8% hemorrhagic; 15% unspecified	Not reported	Mean 73.46 (range 41-88)	46
Kim et al [54]	South Korea	Effectiveness	Activities-Specific Balance Confidence Scale; Dynamic Gait Index; Four Square Step Test; Functional Ambulation Categories; Timed Up and Go test; up-stair and down-stair times; spatiotemporal parameters of gait (velocity and cadence)	18 (IG: 9; CG: 9)	40% ischemic; 60% hemorrhagic	IG: mean 5.68 (SD 1.04) mo; CG: mean 4.76 (SD 2.65) mo	IG: mean 58.3 (SD 11.8); CG: mean 51.8 (SD 13.7)	65
Labovitz et al [55]	United States	Effectiveness	Pill count; plasma sampling; data from artificial intelligence platform	27 (IG: 15; CG: 12)	100% ischemic	Not reported	Mean 57 (SD 13.17); median 59 (range 30-79)	46
Lakshminarayan et al [56]	United States	Pilot, proof-of-concept study; feasibility; usability; acceptability; preliminary effectiveness	Number of days blood pressure data were transmitted; systolic blood pressure; Morisky Medication Adherence Scale	50 (IG: 28; CG: 22)	Not reported	Acute	IG: mean 63.1 (SD 9.7; range 42-81); CG: mean 68.3 (SD 10.0; range 46-85); withdrawn: mean 60.33 (SD 13.7; range 47-84)	68
Maresca et al [57]	Italy	Pilot study; effectiveness	Token Test; Esame Neuropsicologico Per l'Afasia; Aphasic Depression Rating Scale; EQ-5D; Psychosocial Impact of Assistive Devices Scale	30 (IG: 15; CG: 15)	63% ischemic; 37% hemorrhagic	Not reported	Mean 51.2 (SD 11.3)	47
Moon et al [58]	South Korea	Effectiveness	Functional Dysphagia Scale; penetration-aspiration scale; visual analog satisfaction scale	16 (IG: 8; CG: 8)	88% ischemic; 13% hemorrhagic	IG: mean 22.75 (SD 9.21) d; CG: mean 21 (SD 9.02) d	IG: mean 54.13 (SD 5.41); CG: mean 55.38 (SD 14.88)	56

Study	Country	Study aims	Outcome measures	Sample size	Stroke type	Time since stroke or stroke stage	Age (y)	Sex or gender (male or men; %)
Øra et al [59]	Norway	Pilot study; preliminary effectiveness	Norwegian Basic Aphasia Assessment; Verb and Sentence Test; Communicative Effectiveness Index	62 (IG: 32; CG: 30)	69% ischemic; 18% hemorrhagic; 13% both	≤3 mo—IG: 16 (50%) and CG: 12 (40%); 3-12 mo—IG: 5 (15.6%) and CG: 4 (13.3%); >12 mo—IG: 11 (34.4%) and CG: 14 (46.7%)	IG: mean 64.7 (SD 11.7); CG: mean 65 (SD 12.2)	66
Pandian et al [60]	India	Effectiveness	Composite end point of recurrent stroke, high-risk transient ischemic attack, acute coronary syndrome, and all-cause mortality; change in BMI; physical activity total metabolic equivalent (min/wk); current smoking; current alcohol intake; modified Rankin Scale; medication noncompliance; systolic and diastolic blood pressure (mm Hg); fasting blood sugar (mg/dL); low-density lipoprotein cholesterol (mg/dL); triglycerides (mg/dL)	4298 (IG: 2148; CG: 2150)	83% ischemic; 17% hemorrhagic	2 d to 3 mo	IG: median 56 (range 18-88); CG: median 56 (range 18-89)	73
Radomski [40]	United States	Effectiveness	Everyday habit questionnaire; self-reported adherence to self-care checklist; Functional Independence Measure; Frenchay Activities Index; Caregiver's Burden Scale; performance time for self-care task (seconds)	15 (IG: 5; CG 1: 5; CG 2: 5)	Not reported	Not reported	Mean 59 (SD 14)	80
Sarfo et al [61]	Ghana	Pilot; feasibility; preliminary effectiveness	Systolic and diastolic blood pressure; medication possession ratio; perceived confidence scale; Treatment Self-Regulation Questionnaire	56 (IG: 29; CG: 27)	77% ischemic; 23% hemorrhagic	<1 mo	IG: mean 54.3 (SD 11.9); CG: mean 55.9 (SD 13.7)	65
Sarfo et al [62]	Ghana	Pilot study; preliminary effectiveness	Systolic and diastolic blood pressure; medication possession ratio score; Morisky Medication Adherence Scale; perceived confidence scale; Treatment Self-Regulation Questionnaire; Telemedicine Satisfaction and Usefulness Questionnaire; hypertension and stroke knowledge 14-item questionnaire	55 (IG: 28; CG: 27)	77% ischemic; 23% hemorrhagic	<1 mo	IG: mean 54.3 (SD 11.9); CG: mean 55.9 (SD 13.7)	65
Tomori et al [63]	Japan	Pilot study; feasibility; preliminary effectiveness	36-Item Short Form Health Survey; Brunnstrom recovery stages; Functional Independence Measure; Client Satisfaction Questionnaire; duration of stay	37 (IG: 16; CG: 21)	Not reported	≥30 d; subacute	Mean 66.22 (SD 10.64)	67

Study	Country	Study aims	Outcome measures	Sample size	Stroke type	Time since stroke or stroke stage	Age (y)	Sex or gender (male or men; %)
Vahlberg et al [64]	Sweden	Effectiveness	6-Minute Walk Test (m); chair stand test (s); 10-Meter Walk Test (m/s); Short Physical Performance Battery	79 (IG: 40; CG: 39)	72% ischemic; 11% hemorrhagic; 17% transient ischemic attack	Mean 6 (SD 4.4) d	IG: mean 63.9 (SD 10.1); CG: mean 63.9 (SD 10.8)	63
Vahlberg et al [65]	Sweden	Effectiveness	Fat-free mass (kg); fat mass (kg); BMI; body weight (kg); HbA_{1c}; serum insulin-like growth factor; low- and high-density lipoprotein cholesterol; self-reported health; mortality	71 (IG: 36; CG: 35)	72% ischemic; 11% hemorrhagic; 17% transient ischemic attack	Median 5 d	IG: mean 63.9 (SD 10); CG: mean 63.9 (SD 10)	63
Wan et al [66]	China	Effectiveness	Systolic and diastolic blood pressure; Health-Promoting Lifestyle Profile II	158 (IG: 80; CG: 78)	100% ischemic	Within 1 mo	Median 63.81	65
Wang et al [67]	China	Effectiveness	Health-Promoting Lifestyle Profile II; systolic and diastolic blood pressure; modified Rankin Scale; stroke recurrence	151 (IG: 76; CG: 75)	100% ischemic	Within 1 mo	Median 63.80	66
Wang et al [68]	China	Effectiveness	Systolic blood pressure; Self-Management Ability Scale; Morisky Medication Adherence Scale; BMI; blood low-density lipoprotein	193 (IG: 98; CG: 95)	67% ischemic; 33% hemorrhagic	Not reported	IG: mean 42.75 (SD 0.16); CG: mean 41.32 (SD 2.16)	61

[a]IG: intervention group.
[b]CG: control group.
[c]HbA_{1c}: glycated hemoglobin.

Participant Characteristics

In total, 62% (18/29) of the studies included participants with both ischemic and hemorrhagic stroke, and 7% (2/29) also included transient ischemic attack. Of the 23 studies that reported participants' stroke stage or time since stroke, 16 (70%) focused on the subacute stage (7 d to 6 mo after stroke). The average age of the participants ranged from 42 to 74 years (weighted average 57, weighted SD 4.46). No studies differentiated between sex and gender. A total of 83% (24/29) of the studies included more male participants or men than female participants or women, ranging from 25% to 80% of male participants or men. Some studies reported on participants' education (15/29, 52%), marital status (8/29, 28%), employment status (6/29, 21%), and geographic location (5/29, 17%), and fewer studies reported on race (3/29, 10%), ethnicity (1/29, 3%), and income (2/29, 7%).

Objective 1: Types of Poststroke mHealth Interventions

Multimedia Appendix 5 [40-68] summarizes the interventions individually and the following sections summarize the interventions collectively, according to selected items from the Template for Intervention Description and Replication checklist [24].

Why: Describe Any Rationale, Theory, or Goal of the Elements Essential to the Intervention

mHealth technology was rationalized as a strategy to improve intervention effectiveness (18/26, 69%), access (13/26, 50%), convenience (6/26, 23%), and cost-effectiveness (5/26, 19%). A total of 5 interventions were explicitly based on a theory, model, framework, or taxonomy: self-determination theory (n=2, 40%); Health Belief Model (n=2, 40%); social cognitive theory (n=1, 20%); the International Classification of Functioning, Disability, and Health framework (n=1, 20%); the Coventry, Aberdeen, and London–Refined taxonomy of behavior change techniques (n=1, 20%); and a proposed ecological model of adherence to rehabilitation treatment recommendations (n=1, 20%). Common goals of the interventions were to improve outcomes related to treatment or medication adherence (10/26, 38%), motor or physical activity (8/26, 31%), functional ability or independence (5/26, 19%), speech, language, or swallowing (5/26, 19%), hypertension or blood pressure control (5/26, 19%), risk factor control (5/26, 19%), and quality of life (3/26, 12%).

What: Describe Any Physical or Informational Materials Used and Each of the Procedures, Activities, or Processes Used in the Intervention

A total of 7 focal areas were identified: physical exercise (10/26, 38%), risk factor management (5/26, 19%), linguistic exercise (3/26, 12%), activities of daily living (ADLs) training (3/26, 12%), medication adherence (2/26, 8%), stroke education (2/26, 8%), and weight management (1/26, 4%). In total, 5 types of mobile devices were used: mobile phones or smartphones (17/26, 65%), tablets (9/26, 35%), wearable sensors (5/26, 19%; eg, pedometers or wearable bracelets), wireless monitoring devices (4/26, 15%; eg, Bluetooth sphygmomanometers or Bluetooth blood glucose meters), and laptops (1/26, 4%). Within devices, the features used included: apps (15/26, 58%), messaging (12/26, 46%; eg, via an app or SMS text messaging), phone calling (7/26, 27%), videos (6/26, 23%), videoconferencing (3/26, 12%), and email (2/26, 8%). All but 4 interventions (22/26, 85%) were self-directed, and 8% (2/26) were gamified.

Who Provided: For Each Category of Intervention Provider, Describe Their Background

The interventions were provided by researchers (9/26, 35%), occupational therapists (7/26, 27%), physical therapists (4/26, 15%), nurses (4/26, 15%), speech-language pathologists (2/26, 8%), physicians (2/26, 8%), pharmacists (1/26, 4%), neuropsychologists (1/26, 4%), brain and heart health managers (1/26, 4%), allied health professionals (1/26, 4%), clinicians (1/26, 4%), and clinic staff (1/26, 4%). In total, 12% (3/26) were provided by a multidisciplinary team of health care professionals.

How: Describe the Modes of Delivery of the Intervention and Whether It Was Provided Individually or in a Group

A total of 85% (22/26) of the interventions were delivered both virtually (eg, via videoconferencing or SMS text messaging) and in-person (eg, in-person orientation or clinic visits). In total, 77% (20/26) were individual based (delivered to the individual with stroke), 38% (10/26) were dyad based (delivered to the individual with stroke and their caregiver or family member), and 8% (2/26) were group based (delivered to groups of people with stroke).

Where: Describe the Types of Locations Where the Intervention Occurred

In total, 58% (15/26) of the interventions occurred both at the hospital or clinic (in-person component) and the participants' home (virtual component).

When and How Much: Describe the Number of Times the Intervention Was Delivered and Over What Period

Intervention delivery time ranged from 14 days to 1 year, with the most common being 4 weeks (5/26, 19%) and 6 months (5/26, 19%). Session frequency varied (twice/d to once every 2-3 mo), as did session length (5 min to 1 h). This variability reflects a wide range of session types (eg, exercise sessions, education sessions, blood pressure self-monitoring, and clinic visits). There was also variability in the dosage of technology used, such as the schedule for sending and receiving messages (twice/d to once/wk) and the amount of time connected to the devices (eg, 1 intervention required participants to wear a pedometer at all times except when sleeping, bathing, or swimming).

Tailoring: If the Intervention Was Planned to be Personalized, Titrated, or Adapted, Describe What, Why, When, and How

A total of 69% (18/26) of the interventions involved tailoring to the person with stroke (eg, abilities, goals, or preferred music). In total, 8% (2/26) of the interventions involved tailoring to the caregiver or family member (eg, preferred ADLs) [40,50]. A total of 12% (3/26) of the interventions involved self-tailoring by the person with stroke (eg, education topics [52] or exercises [53,64,65]).

Objective 2: Alignment With Self-Management Support Theory and Taxonomy

Of the 29 conceptual variables, 26 (90%) were coded at least once. The number of interventions coded per variable ranged from 0 to 25 (mean 8.55). The number of variables coded per intervention ranged from 2 to 15 (mean 9.54). Figure 2 [17,18,40-68] presents the extent and range of alignment, and Table 2 presents the nature of alignment.

Figure 2. Extent and range of alignment between poststroke mobile technology–based interventions and self-management support theory and taxonomy. Please see Multimedia Appendix 6 for a larger version.

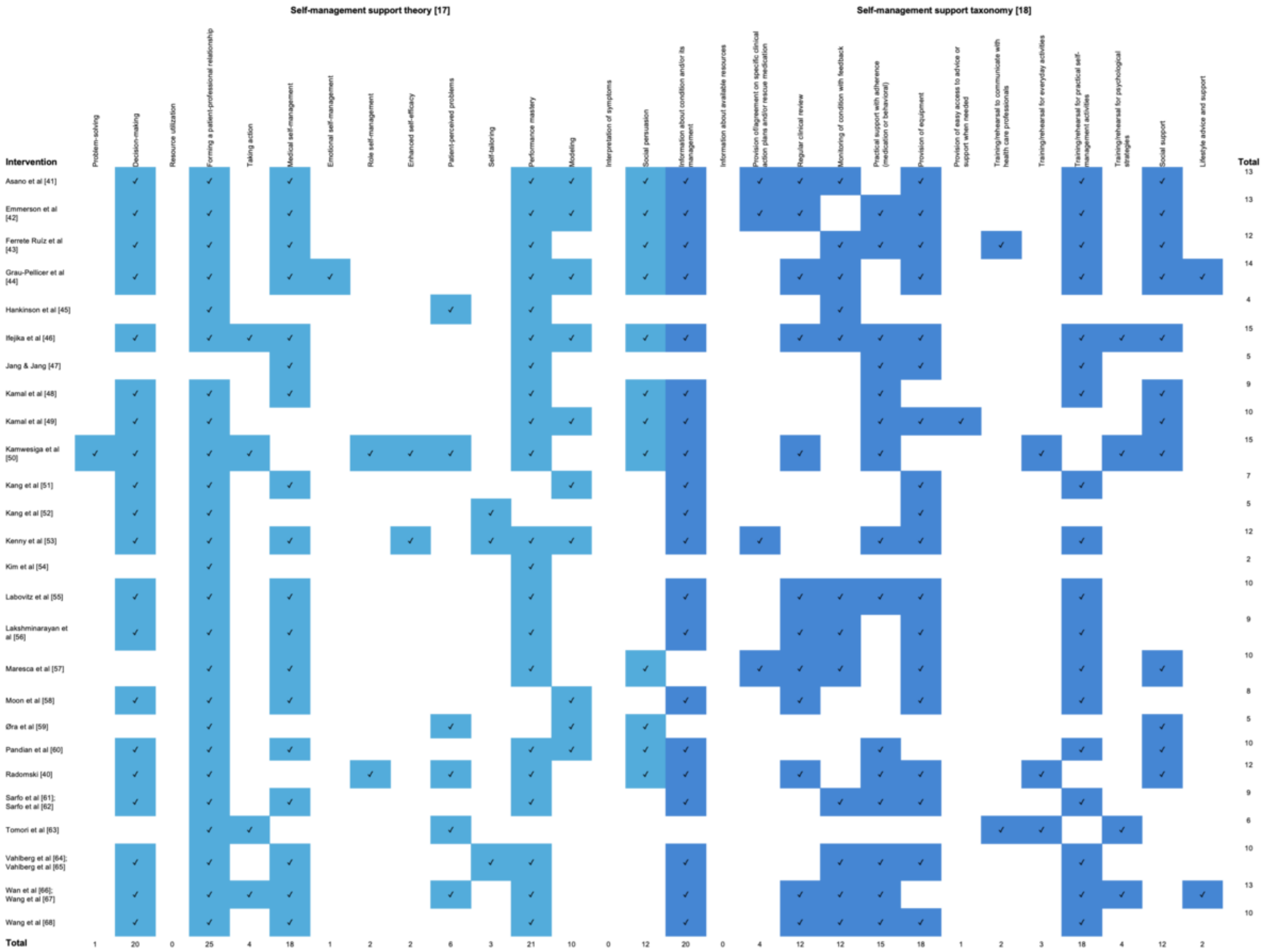


Table 2. Nature of alignment between poststroke mobile technology–based interventions and self-management support theory and taxonomy.

Conceptual variable	Intervention examples
Self-management support theory [17]	
Core self-management skills	
Problem-solving	Clients developed skills in problem-solving through training focused on the Target-Plan-Perform-Prove strategy [50].
Decision-making	Clients developed skills in decision-making through learning information about stroke, rehabilitation, medications, lifestyle, and risk factors (eg, stroke history, heart disease, atrial fibrillation, obesity, age, sleep patterns, diet, exercise, and blood pressure [40,46,48-50,52,60-62,66-68]), as well as by learning how to perform exercises [41-44,51,53,58,60,64,65], take medications [55], measure blood pressure [56,61,62,66-68], and measure blood glucose [68].
Resource utilization	Not reported
Forming a patient-professional relationship	Clients developed skills in forming a relationship with a professional through participating in virtual (eg, videoconference or phone call) and in-person sessions [40-46,48-60,63,66-68], as well as by sending and receiving messages (eg, via an app or SMS text messaging [40,44,48,50,60-62,64-68]).
Taking action	Clients developed skills in taking action through goal-setting training focused on weight loss [46], ADLs[a] [50,63], and self-management (eg, medication adherence [66,67]).
Self-management tasks or behaviors	
Medical self-management	Clients practiced the medical tasks of performing self-directed physical or linguistic exercises [41-44,47,51,53,57,58,60,64,65], taking medications [48,55], and self-monitoring physical health data (weight, diet, and exercise [46,68]; sleep [68]; fatigue [44]; blood pressure [56,61,62,66-68]; and blood glucose [68]).
Emotional self-management	Clients practiced the emotional task of self-monitoring mood through app-based questionnaires [44].
Role self-management	Clients practiced maintaining or changing behaviors related to ADLs or life roles (eg, dressing or cooking) in the context of their home environment [40,50].
Mechanism of change	
Enhanced self-efficacy	Self-efficacy was used as an outcome measure [50,53].
Characteristics of self-management support	
Patient-perceived problems	Interventions were based on client-identified problems, concerns, or goals [40,45,50,59,63,66,67].
Self-tailoring	Self-tailoring was encouraged by instructing clients to select their own educational content based on their time and needs [52], practice the exercises as often as they wished [53], and gradually modify the exercises based on preference and perceived strenuous intensity [64,65].
Efficacy enhancement	
Performance mastery	Performance mastery was promoted through feedback (eg, via an app or wearable sensor, through SMS text messaging, or from a therapist [40-47,49,54-57,61,62,64,65,68]) and self-reflection (eg, through checklists or diaries or from reminder messages prompting a response [40,43,47-50,53,60,64-67]).
Modeling	Modeling was offered through group-based formats [44,59]; demonstration videos (eg, of the therapist or client performing the exercises or of animated characters taking medications [41,42,49,53,58]); fictional and nonfictional stories about people with stroke that are age appropriate, country specific, and culturally relevant (eg, Mahatma Gandhi was a character in a story created for an Indian audience [60]); distorted mirror reflections via an app that allowed clients to watch the reflection of the unaffected half of their face as if it were the affected half [51]; and culturally competent counseling by clinicians of clients' ethnic groups [46].
Interpretation of symptoms	Not reported
Social persuasion	Social persuasion was promoted through group-based (eg, a WhatsApp group was created for clients to motivate each other to maintain an active lifestyle [44,59]) and dyad (client-caregiver or family member)-based (eg, caregivers were instructed to support the client at home in using the technology and adhering to the program [40-43,46,48-50,57,60]) formats.
Self-management support taxonomy [18]	
Information about condition and/or its management	Clients were provided with general information about stroke, rehabilitation, medications, lifestyle, and risk factors (eg, stroke history, heart disease, atrial fibrillation, obesity, age, sleep patterns, diet, exercise, and blood pressure [40,46,48-50,52,60-62,66-68]) as well as general instruction on how to perform exercises [41-44,51,53,58,60,64,65], take medications [55], measure blood pressure [56,61,62,66-68], and measure blood glucose [68].

Conceptual variable	Intervention examples
Information about available resources	Not reported
Provision of or agreement on specific clinical action plans and/or rescue medication	Clients were provided with specific, tailored instruction on how to perform physical or linguistic exercises [41,42,53,57].
Regular clinical review	Clients connected with health care professionals on a regular basis for a scheduled review of their condition and self-management (eg, via videoconference, phone call, messaging, or in-person sessions [40-42,44,46,50,55-58,66-68]).
Monitoring of condition with feedback	Clients monitored symptoms, behaviors, or objective measures related to their condition (eg, exercise adherence data from wearable sensors, weight management data through tracking calories consumed, medication adherence data via app-based artificial intelligence, and blood pressure data from wireless monitoring devices [41,43-46,55-57,61,62,64-68]). Professionals also reviewed the monitored data and provided clients with feedback [41,44,55-57,61,62,68].
Practical support with adherence (medication or behavioral)	Practical support with adherence was provided to clients in the form of reminder alarms [42], reminder messages (eg, via push notifications within an app or through SMS text messaging [40,46,48-50,55,61,62,64-68]), reminder phone calls [48,50,60], and reminder sheets (eg, checklists, diaries, and calendars [40,43,47,50,53,60,64-67]).
Provision of equipment	Self-monitoring or self-management was enabled, assisted, or promoted through the provision of tablets [41-43,47,49,51,53,57], smartphones [40,56,58,61,62], wearable sensors (eg, pedometers or wearable bracelets [41,44,64,65,68]), wireless monitoring devices (eg, Bluetooth sphygmomanometers or Bluetooth blood glucose meters [41,56,61,62,68]), apps [43,44,46,47,49,51,52,55,56,61,62,68], and measuring cups [46].
Provision of easy access to advice or support when needed	Clients were provided with a 24/7 stroke helpline number [49].
Training or rehearsal to communicate with health care professionals	Clients developed communication skills by using apps to communicate their needs with nursing staff [43] and participate in shared decision-making with occupational therapists for goal setting [63].
Training or rehearsal for everyday activities	Clients developed skills to support ADLs (eg, dressing, cooking, or knitting [40,50,63]).
Training or rehearsal for practical self-management activities	Clients developed specific, practical skills in performing self-directed physical or linguistic exercises [41-44,47,51,53,57,58,60,64,65], taking medications [48,55], and self-monitoring health data (weight, diet, and exercise [46,68]; sleep [68]; fatigue [44]; blood pressure [56,61,62,66-68]; blood glucose [68]; and mood [44]).
Training or rehearsal for psychological strategies	Clients developed psychological skills in goal setting [46,50,63,66,67] and problem-solving [50].
Social support	Social support was facilitated through group-based (eg, a WhatsApp group was created for clients to facilitate peer support [44,59]) and dyad (client-caregiver or family member)-based (eg, caregivers were encouraged to watch educational videos with the client and engage in a follow-up discussion with a professional afterward or instructed to support the client at home in using the technology and adhering to the program [40-43,46,48-50,57,60]) formats.
Lifestyle advice and support	Lifestyle advice and support were provided to clients through a peer-based WhatsApp group [44] and nurse-led education and goal-setting sessions [66,67].

[a]ADL: activity of daily living.

Objective 3: mHealth Functions That Facilitate Self-Management

Across all conceptual variables and interventions, 7 mHealth functions were identified as facilitating self-management: educating, communicating, goal setting, monitoring, providing feedback, reminding, and motivating.

Discussion

Principal Findings

Overview

To our knowledge, this is the first scoping review to map the literature on poststroke mHealth interventions according to a self-management support theory and taxonomy. A total of 29 studies describing 26 interventions were included. Overall, we found that the interventions addressed 7 focal areas, 5 types of mobile devices, and 7 mHealth functions. Collectively, the interventions aligned well with the concept of self-management support. However, on an individual basis (per intervention), the alignment was less strong. The following sections further explain how this review extends previous reviews on poststroke mHealth [10-14,69-72] and telehealth [73] interventions in relation to the study objectives and interventions included.

Objective 1: Types of Poststroke mHealth Interventions

Focal Areas: Current Trends and Gaps

Our first objective was to identify and describe the types of poststroke mHealth interventions evaluated using an RCT design. Speaking to such types, 7 focal areas were identified: physical exercise, risk factor management, linguistic exercise, ADLs training, medication adherence, stroke education, and weight management. These 7 focal areas have been identified in previous reviews on poststroke mHealth interventions [10-14,69-72]; however, the included interventions varied. Compared with previous reviews, 45% (13/29) of the studies included in our review (12 interventions) had not been previously identified. Similar to previous reviews, this review found the most common focal area to be physical exercise, likely reflecting the rising trend within the general population of using mobile technology to promote physical fitness in everyday life [74,75]. Hence, the literature clearly supports continued research on poststroke mHealth interventions for physical exercise. Also consistent with previous reviews, this review did not identify any interventions focused on mood or fatigue. Considering the high prevalence of poststroke depression, anxiety, and fatigue, this is a serious gap that should be addressed in future research [76]. Surprisingly, unlike 5 previous reviews [10,11,13,70,71], this review did not identify any interventions focused on cognition. This difference was due to the varying eligibility criteria (eg, study design). Given this difference across reviews as well as the high prevalence of poststroke cognitive impairment [76], future research on poststroke mHealth interventions for cognition is encouraged to progress toward the level of RCTs.

mHealth Technology: Positioning on the Spectrum of Definitions

Regarding the types of technology used in the interventions, our review identified 5 types of mobile devices (mobile phones or smartphones, tablets, wearable sensors, wireless monitoring devices, and laptops) and 6 features within these devices (apps, messaging, phone calling, videos, videoconferencing, and email). This wide range of technologies resulted from our novel approach to defining mHealth. Previous reviews on poststroke mHealth interventions have defined mHealth either very narrowly, focusing on a few specific mobile devices or features (eg, mobile phones [72], wearable activity monitors [69], or mobile apps for phones [10,11,14,69,71] and tablets [10,11,14,70,71]), or very broadly, focusing on mHealth in general and including devices and features that may not be entirely mobile and wireless (eg, computer programs [10,12,71], telephone calls [12], and web-based applications [13]). Our review was interested in the literature between these 2 ends of the narrow-broad spectrum of mHealth definitions. We followed the recommendation of Cameron et al [26] to define mHealth in a way that captures the "combinatorial complexity" of the mobile system and used 2 open-ended definitions of mHealth [22,23]. Thus, we captured additional literature on poststroke mHealth interventions by focusing on entirely mobile *systems* (technology and equipment) of any type (devices and features). As the field of mHealth continues to grow, we suggest that future reviews explicitly position themselves on this narrow-broad spectrum of mHealth definitions so that the literature can be more readily interpreted and applied. In addition, future work should build on that by Cameron et al [26] to further deepen our understanding of the mobile system.

Objective 2: Alignment With Self-Management Support Theory and Taxonomy

Our second objective was to determine whether (and how) the included interventions aligned with well-accepted conceptualizations (theory [17] and taxonomy [18]) of self-management support. Collectively, the interventions addressed 90% (26/29) of the conceptual variables, whereas individual interventions only addressed an average of 33% (9.54/29) of the conceptual variables. This discrepancy speaks to the potential for improvements in the alignment between poststroke mHealth interventions and the concept of self-management support. The results also revealed key conceptual variables missing from the literature, such as "emotional self-management" and "information about available resources." Hence, the results suggest that future research should be more closely aligned with the theory and taxonomy of self-management support. Previous reviews on poststroke mHealth interventions [10-14,69-72] have not mapped the literature in this way. However, a review of poststroke telehealth interventions [73] used the PRISMS taxonomy [18] in a similar way, further validating the relevance of this approach.

Objective 3: mHealth Functions That Facilitate Self-Management

Our third objective was to identify the mHealth functions that facilitate self-management. A total of 7 mHealth functions were identified: educating, communicating, goal setting, monitoring, providing feedback, reminding, and motivating. These 7 functions, although together framed as facilitating self-management, are not inherently specific to self-management support interventions as they speak generally to what the intervention *does*, not specifically to what the intervention is *about*. Viewing mHealth functions in this way, as generic "verbs describing the behavior of the system" [26] or as action words that link technology capabilities with intervention components, has not been done in past reviews on poststroke mHealth interventions [10-14,69-72]. However, this approach to conceptualizing mHealth functions does align with other work in the broader field of mHealth [26,30]. Future research is encouraged to build on this approach and use the identified functions to describe how specific technology capabilities are linked to specific intervention components. Specifically linking technology capabilities with intervention components is important as it would allow for more systematic examinations as to what it is about delivery through mHealth that may be superior or not to other intervention delivery modalities (eg, is educating on sensitive topics via mHealth better than via in-person groups?).

Recommendations for Future Research

The purpose of this study was to identify gaps in the literature and recommendations for future research related to mHealth-enhanced stroke self-management support. In total, 3 overarching recommendations for future research were identified. First, future research should be more explicit about the theories their interventions are based on as well as their conceptualizations of self-management support. Using theory and other conceptualizations in this way would help promote a common language of self-management support and ensure that all conceptual variables are considered, which could ultimately improve intervention adherence, effectiveness, replicability, and uptake in clinical practice. Second, future research should be more multidisciplinary so that a wider range of conceptual variables can be addressed per intervention. This multidisciplinary approach to improving alignment would likely lead to more comprehensive, holistic, and effective interventions. Third, future research should use larger sample sizes and consider using pragmatic trial designs to establish real-world effectiveness.

Limitations

The search was limited to the English language, so the findings may be biased toward English-speaking countries, although 15 countries were represented. Directed content analysis, as with any qualitative approach, involves subjectivity; to address this, operational definitions for codes were used and reported. Finally, this review focused on RCTs, so the findings may be biased toward more traditional or RCT-suited interventions. Given the challenges associated with conducting RCTs on technology-based interventions [77], future reviews should consider including other study designs.

Conclusions

This scoping review clarified the concept of self-management support in the literature on poststroke mHealth interventions by mapping studies according to well-accepted conceptualizations of self-management support. On the basis of the results, it is recommended that future research on poststroke mHealth interventions be more theoretically driven, more multidisciplinary, and larger in scale.

Acknowledgments

This research is part of the doctoral work of ANT, which was funded by the Queen Elizabeth II Graduate Scholarship in Science and Technology, Ontario Graduate Scholarship, Ydessa Hendeles Graduate Scholarship at Baycrest Award, and Toronto Rehabilitation Institute Student Scholarship. EN is supported by a Canada Research Chair (Tier 2) in Resiliency and Rehabilitation funded by the Canada Research Chairs Program. The authors gratefully acknowledge Erica Nekolaichuk and Jessica Shiers-Hanley, librarians at the University of Toronto, for their assistance in developing the search strategy.

Authors' Contributions

ANT conceptualized and designed the study under the guidance of DRD and EN. ANT searched the databases. ANT, JML-M, NC, CT, VN, JR, and SJ screened the articles. ANT extracted the data. JML-M, NC, CT, VN, JR, and SJ verified the data. ANT analyzed the data and wrote the manuscript in collaboration with DRD, EN, RHW, and JIC. All authors have reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 25 KB - mhealth_v11i1e46558_app1.docx]

Multimedia Appendix 2
Search strategy for the Ovid databases (MEDLINE, Embase, PsycINFO, and AMED).
[DOCX File , 15 KB - mhealth_v11i1e46558_app2.docx]

Multimedia Appendix 3
Operational definitions used in the directed content analysis related to self-management support theory and taxonomy (objective 2).
[DOCX File , 28 KB - mhealth_v11i1e46558_app3.docx]

Multimedia Appendix 4
Operational definitions used in the directed content analysis related to mobile health functions (objective 3).
[DOCX File , 17 KB - mhealth_v11i1e46558_app4.docx]

Multimedia Appendix 5
Intervention characteristics according to selected items from the Template for Intervention Description and Replication checklist.
[DOCX File , 51 KB - mhealth_v11i1e46558_app5.docx]

Multimedia Appendix 6

Extent and range of alignment between poststroke mobile technology–based interventions and self-management support theory and taxonomy.
[PNG File , 540 KB - mhealth_v11i1e46558_app6.png]

References

1. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Feb 07;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7] [Medline: 25468153]
2. McKevitt C, Fudge N, Redfern J, Sheldenkar A, Crichton S, Rudd AR, et al. Self-reported long-term needs after stroke. Stroke 2011 May;42(5):1398-1403. [doi: 10.1161/STROKEAHA.110.598839] [Medline: 21441153]
3. Krueger H, Koot J, Hall RE, O'Callaghan C, Bayley M, Corbett D. Prevalence of individuals experiencing the effects of stroke in Canada: trends and projections. Stroke 2015 Aug;46(8):2226-2231. [doi: 10.1161/STROKEAHA.115.009616] [Medline: 26205371]
4. Adams K, Greiner AC, Corrigan JM. The 1st Annual Crossing the Quality Chasm Summit: A Focus on Communities. Washington, DC: National Academies Press; 2004.
5. Sakakibara BM, Kim AJ, Eng JJ. A systematic review and meta-analysis on self-management for improving risk factor control in stroke patients. Int J Behav Med 2017 Feb;24(1):42-53 [FREE Full text] [doi: 10.1007/s12529-016-9582-7] [Medline: 27469998]
6. Warner G, Packer T, Villeneuve M, Audulv A, Versnel J. A systematic review of the effectiveness of stroke self-management programs for improving function and participation outcomes: self-management programs for stroke survivors. Disabil Rehabil 2015 Jan 12;37(23):2141-2163. [doi: 10.3109/09638288.2014.996674] [Medline: 25579669]
7. Fryer C, Luker J, McDonnell M, Hillier S. Self management programmes for quality of life in people with stroke. Cochrane Database Syst Rev 2016 Aug 22;2016(8):CD010442 [FREE Full text] [doi: 10.1002/14651858.CD010442.pub2] [Medline: 27545611]
8. Mountain A, Patrice Lindsay M, Teasell R, Salbach NM, de Jong A, Foley N, et al. Canadian stroke best practice recommendations: rehabilitation, recovery, and community participation following stroke. Part two: transitions and community participation following stroke. Int J Stroke 2020 Oct 27;15(7):789-806. [doi: 10.1177/1747493019897847] [Medline: 31983292]
9. Blacquiere D, Lindsay MP, Foley N, Taralson C, Alcock S, Balg C, Heart and Stroke Foundation Canadian Stroke Best Practice Committees. Canadian stroke best practice recommendations: Telestroke best practice guidelines update 2017. Int J Stroke 2017 Oct 26;12(8):886-895. [doi: 10.1177/1747493017706239] [Medline: 28441928]
10. Zhou X, Du M, Zhou L. Use of mobile applications in post-stroke rehabilitation: a systematic review. Top Stroke Rehabil 2018 Sep 13;25(7):1-11. [doi: 10.1080/10749357.2018.1482446] [Medline: 30209991]
11. Burns SP, Terblanche M, Perea J, Lillard H, DeLaPena C, Grinage N, et al. mHealth intervention applications for adults living with the effects of stroke: a scoping review. Arch Rehabil Res Clin Transl 2021 Mar;3(1):100095 [FREE Full text] [doi: 10.1016/j.arrct.2020.100095] [Medline: 33778470]
12. Adcock AK, Haggerty T, Crawford A, Espinosa C. mHealth impact on secondary stroke prevention: a scoping review of randomized controlled trials among stroke survivors between 2010-2020. Mhealth 2022;8:19 [FREE Full text] [doi: 10.21037/mhealth-21-27] [Medline: 35449509]
13. Marwaa MN, Guidetti S, Ytterberg C, Kristensen HK. Use of mobile/tablet and web-based applications to support rehabilitation after stroke: a scoping review. J Rehabil Med 2022 Apr 07;54:jrm00269 [FREE Full text] [doi: 10.2340/jrm.v54.452] [Medline: 35174871]
14. Rintala A, Kossi O, Bonnechère B, Evers L, Printemps E, Feys P. Mobile health applications for improving physical function, physical activity, and quality of life in stroke survivors: a systematic review. Disabil Rehabil 2022 Nov 02:1-15. [doi: 10.1080/09638288.2022.2140844] [Medline: 36325613]
15. Jonkman NH, Schuurmans MJ, Jaarsma T, Shortridge-Baggett LM, Hoes AW, Trappenburg JC. Self-management interventions: proposal and validation of a new operational definition. J Clin Epidemiol 2016 Dec;80:34-42 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.08.001] [Medline: 27531245]
16. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
17. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
18. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CL, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr 15;21(2):73-82 [FREE Full text] [doi: 10.1177/1355819615602725] [Medline: 26377727]
19. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

20. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2023-02-08]
23. Akter S, D'Ambra J, Ray P. Development and validation of an instrument to measure user perceived service quality of mHealth. Inf Manag 2013 Jun;50(4):181-195. [doi: 10.1016/j.im.2013.03.001]
24. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348(mar07 3):g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
25. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Cameron JD, Ramaprasad A, Syn T. An ontology of and roadmap for mHealth research. Int J Med Inform 2017 Apr;100:16-25. [doi: 10.1016/j.ijmedinf.2017.01.007] [Medline: 28241934]
27. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
28. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
29. Bhargava H, Tanghetti J. Mobile health technologies. In: Mangal V, Mandelli A, Karmarkar U, La Rocca A, editors. The UCLA Anderson Business and Information Technologies (BIT) Project. Singapore, Singapore: World Scientific; 2016:167-185.
30. Smahel D, Elavsky S, Machackova H. Functions of mHealth applications: a user's perspective. Health Informatics J 2019 Sep 10;25(3):1065-1075 [FREE Full text] [doi: 10.1177/1460458217740725] [Medline: 29121831]
31. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: a scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
32. Definition of educate. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/educate [accessed 2023-02-08]
33. Definition of communicate. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/communicate [accessed 2023-02-08]
34. Definition of goal setting. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/goal-setting [accessed 2023-02-08]
35. Definition of monitor. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/monitor [accessed 2023-02-08]
36. Definition of feedback. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/feedback [accessed 2023-02-08]
37. Definition of remind. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/remind [accessed 2023-02-08]
38. Definition of motivate. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/motivate [accessed 2023-02-08]
39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
40. Radomski MV. Impact of post-discharge habit training of self care skills on independence, caregiver burden, and development of automaticity for survivors of recent stroke. University of Minnesota. 2007. URL: https://www.proquest.com/openview/3bc519083a10f08639f8d53a3ad21732/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2023-11-18]
41. Asano M, Tai BC, Yeo FY, Yen SC, Tay A, Ng YS, et al. Home-based tele-rehabilitation presents comparable positive impact on self-reported functional outcomes as usual care: the Singapore Tele-technology Aided Rehabilitation in Stroke (STARS) randomised controlled trial. J Telemed Telecare 2021;27(4):231-238. [doi: 10.1177/1357633x19868905]
42. Emmerson KB, Harding KE, Taylor NF. Home exercise programmes supported by video and automated reminders compared with standard paper-based home exercise programmes in patients with stroke: a randomized controlled trial. Clin Rehabil 2017 Aug 05;31(8):1068-1077. [doi: 10.1177/0269215516680856] [Medline: 27920262]
43. Ferrete Ruíz EJ, de Torres Chacón MD, Perez Sánchez S, Domínguez Mayoral A, Gamero García MA, Barragan Prieto A, et al. Breaking the communication barrier using new technologies in aphasic patients post-stroke. Rev Cient Soc Esp Enferm Neurol 2023 Jul;58:14-20 [FREE Full text] [doi: 10.1016/j.sedeng.2022.06.002]
44. Grau-Pellicer M, Lalanza J, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020 Jul 02;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]

45. Hankinson K, Shaykevich A, Vallence A, Rodger J, Rosenberg M, Etherton-Beer C. A tailored music-motor therapy and real-time biofeedback mobile phone app ('GotRhythm') to promote rehabilitation following stroke: a pilot study. Neurosci Insights 2022 May 19;17:26331055221100587 [FREE Full text] [doi: 10.1177/26331055221100587] [Medline: 35615116]
46. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
47. Jang SH, Jang WH. The effect of a finger training application using a tablet PC in chronic hemiparetic stroke patients. Somatosens Mot Res 2016 Jun 29;33(2):124-129. [doi: 10.1080/08990220.2016.1197117] [Medline: 27357969]
48. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
49. Kamal A, Khoja A, Usmani B, Magsi S, Malani A, Peera Z, et al. Effect of 5-minute movies shown via a mobile phone app on risk factors and mortality after stroke in a low- to middle-income country: randomized controlled trial for the stroke caregiver dyad education intervention (Movies4Stroke). JMIR Mhealth Uhealth 2020 Jan 28;8(1):e12113 [FREE Full text] [doi: 10.2196/12113] [Medline: 32012080]
50. Kamwesiga JT, Eriksson GM, Tham K, Fors U, Ndiwalana A, von Koch L, et al. A feasibility study of a mobile phone supported family-centred ADL intervention, F@ce™, after stroke in Uganda. Global Health 2018 Aug 15;14(1):82 [FREE Full text] [doi: 10.1186/s12992-018-0400-7] [Medline: 30111333]
51. Kang J, Chun MH, Choi SJ, Chang MC, Yi YG. Effects of mirror therapy using a tablet PC on central facial paresis in stroke patients. Ann Rehabil Med 2017 Jun;41(3):347-353 [FREE Full text] [doi: 10.5535/arm.2017.41.3.347] [Medline: 28758071]
52. Kang Y, Shen H, Lin C, Elwyn G, Huang S, Wu T, et al. Does a mobile app improve patients' knowledge of stroke risk factors and health-related quality of life in patients with stroke? A randomized controlled trial. BMC Med Inform Decis Mak 2019 Dec 21;19(1):282 [FREE Full text] [doi: 10.1186/s12911-019-1000-z] [Medline: 31864348]
53. Kenny M, Gilmartin J, Thompson C. Video-guided exercise after stroke: a feasibility randomised controlled trial. Physiother Theory Pract 2022 May 20;38(5):609-620 [FREE Full text] [doi: 10.1080/09593985.2020.1790072] [Medline: 32684076]
54. Kim JH, Park SG, Lim HJ, Park GC, Kim MH, Lee BH. Effects of the combination of rhythmic auditory stimulation and task-oriented training on functional recovery of subacute stroke patients. J Phys Ther Sci 2012;24(12):1307-1313. [doi: 10.1589/jpts.24.1307]
55. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
56. Lakshminarayan K, Westberg S, Northuis C, Fuller CC, Ikramuddin F, Ezzeddine M, et al. A mHealth-based care model for improving hypertension control in stroke survivors: pilot RCT. Contemp Clin Trials 2018 Jul;70:24-34 [FREE Full text] [doi: 10.1016/j.cct.2018.05.005] [Medline: 29763657]
57. Maresca G, Maggio MG, Latella D, Cannavò A, De Cola MC, Portaro S, et al. Toward improving poststroke aphasia: a pilot study on the growing use of telerehabilitation for the continuity of care. J Stroke Cerebrovasc Dis 2019 Oct;28(10):104303. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104303] [Medline: 31371144]
58. Moon JH, Heo SJ, Jung JH. Effects of orofacial muscles exercise program on swallowing function and satisfaction in sub-acute stroke patients with dysphagia. Med-Leg Update 2019;19(1):623 [FREE Full text] [doi: 10.5958/0974-1283.2019.00111.7]
59. Øra HP, Kirmess M, Brady MC, Partee I, Hognestad RB, Johannessen BB, et al. The effect of augmented speech-language therapy delivered by telerehabilitation on poststroke aphasia-a pilot randomized controlled trial. Clin Rehabil 2020 Mar 05;34(3):369-381. [doi: 10.1177/0269215519896616] [Medline: 31903800]
60. SPRINT INDIA trial collaborators. Secondary prevention with a structured semi-interactive stroke prevention package in INDIA (SPRINT INDIA): a multicentre, randomised controlled trial. Lancet Glob Health 2023 Mar;11(3):e425-e435 [FREE Full text] [doi: 10.1016/S2214-109X(22)00544-7] [Medline: 36796986]
61. Sarfo F, Treiber F, Gebregziabher M, Adamu S, Patel S, Nichols M, et al. PINGS (phone-based intervention under nurse guidance after stroke): interim results of a pilot randomized controlled trial. Stroke 2018 Jan;49(1):236-239 [FREE Full text] [doi: 10.1161/STROKEAHA.117.019591] [Medline: 29222227]
62. Sarfo FS, Treiber F, Gebregziabher M, Adamu S, Nichols M, Singh A, PINGS Team. Phone-based intervention for blood pressure control among Ghanaian stroke survivors: a pilot randomized controlled trial. Int J Stroke 2019 Aug 22;14(6):630-638. [doi: 10.1177/1747493018816423] [Medline: 30465630]
63. Tomori K, Nagayama H, Ohno K, Nagatani R, Saito Y, Takahashi K, et al. Comparison of occupation-based and impairment-based occupational therapy for subacute stroke: a randomized controlled feasibility study. Clin Rehabil 2015 Aug 07;29(8):752-762. [doi: 10.1177/0269215514555876] [Medline: 25381345]
64. Vahlberg B, Lundström E, Eriksson S, Holmbäck U, Cederholm T. Effects on walking performance and lower body strength by short message service guided training after stroke or transient ischemic attack (The STROKEWALK Study): a randomized

controlled trial. Clin Rehabil 2021 Feb;35(2):276-287 [FREE Full text] [doi: 10.1177/0269215520954346] [Medline: 32942914]
65. Vahlberg BM, Lundström E, Eriksson S, Holmback U, Cederholm T. Potential effects on cardiometabolic risk factors and body composition by short message service (SMS)-guided training after recent minor stroke or transient ischaemic attack: post hoc analyses of the STROKEWALK randomised controlled trial. BMJ Open 2021 Oct 18;11(10):e054851 [FREE Full text] [doi: 10.1136/bmjopen-2021-054851] [Medline: 34663672]
66. Wan L, Zhang X, You L, Ruan H, Chen S. The efficacy of a comprehensive reminder system to improve health behaviors and blood pressure control in hypertensive ischemic stroke patients: a randomized controlled trial. J Cardiovasc Nurs 2018 Nov;33(6):509-517. [doi: 10.1097/JCN.0000000000000496] [Medline: 29901484]
67. Wang M, Shen M, Wan L, Mo M, Wu Z, Li L, et al. Effects of a comprehensive reminder system based on the health belief model for patients who have had a stroke on health behaviors, blood pressure, disability, and recurrence from baseline to 6 months: a randomized controlled trial. J Cardiovasc Nurs 2020 Mar;35(2):156-164. [doi: 10.1097/JCN.0000000000000631] [Medline: 31904693]
68. Wang S, Li Y, Tian J, Peng X, Yi L, Du C, et al. A randomized controlled trial of brain and heart health manager-led mHealth secondary stroke prevention. Cardiovasc Diagn Ther 2020 Oct;10(5):1192-1199 [FREE Full text] [doi: 10.21037/cdt-20-423] [Medline: 33224743]
69. Lynch EA, Jones TM, Simpson DB, Fini NA, Kuys SS, Borschmann K, ACTIOnS Collaboration. Activity monitors for increasing physical activity in adult stroke survivors. Cochrane Database Syst Rev 2018 Jul 27;7(7):CD012543 [FREE Full text] [doi: 10.1002/14651858.CD012543.pub2] [Medline: 30051462]
70. Pugliese M, Ramsay T, Johnson D, Dowlatshahi D. Mobile tablet-based therapies following stroke: a systematic scoping review of administrative methods and patient experiences. PLoS One 2018 Jan 23;13(1):e0191566 [FREE Full text] [doi: 10.1371/journal.pone.0191566] [Medline: 29360872]
71. Szeto SG, Wan H, Alavinia M, Dukelow S, MacNeill H. Effect of mobile application types on stroke rehabilitation: a systematic review. J Neuroeng Rehabil 2023 Jan 24;20(1):12 [FREE Full text] [doi: 10.1186/s12984-023-01124-9] [Medline: 36694257]
72. Avcı A, Gün M. The effect of mobile phone-based interventions on blood pressure in stroke patients: a systematic review of randomized controlled trials. Florence Nightingale J Nurs 2023 Jun;31(2):122-130 [FREE Full text] [doi: 10.5152/FNJN.2023.22218] [Medline: 37404215]
73. Hwang N, Park J, Chang M. Telehealth interventions to support self-management in stroke survivors: a systematic review. Healthcare (Basel) 2021 Apr 15;9(4):472 [FREE Full text] [doi: 10.3390/healthcare9040472] [Medline: 33921183]
74. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Critical Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
75. Lupton D. Self-tracking, health and medicine. Health Sociol Rev 2017;26(1):1-5. [doi: 10.1080/14461242.2016.1228149]
76. Lanctôt KL, Lindsay MP, Smith EE, Sahlas DJ, Foley N, Gubitz G, et al. Canadian stroke best practice recommendations: mood, cognition and fatigue following stroke, 6th edition update 2019. Int J Stroke 2020 Aug 21;15(6):668-688. [doi: 10.1177/1747493019847334] [Medline: 31221036]
77. Wang RH, Kenyon LK, McGilton KS, Miller WC, Hovanec N, Boger J, et al. The time is now: a FASTER approach to generate research evidence for technology-based interventions in the field of disability and rehabilitation. Arch Phys Med Rehabil 2021 Sep;102(9):1848-1859. [doi: 10.1016/j.apmr.2021.04.009] [Medline: 33992634]

Abbreviations

ADL: activity of daily living
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PRISMS: Practical Reviews in Self-Management Support
RCT: randomized controlled trial

Edited by L Buis; submitted 16.02.23; peer-reviewed by M Peeples, E Nabovati; comments to author 03.07.23; revised version received 03.10.23; accepted 18.10.23; published 06.12.23.

Please cite as:
Thompson AN, Dawson DR, Legasto-Mulvale JM, Chandran N, Tanchip C, Niemczyk V, Rashkovan J, Jeyakumar S, Wang RH, Cameron JI, Nalder E
Mobile Technology–Based Interventions for Stroke Self-Management Support: Scoping Review
JMIR Mhealth Uhealth 2023;11:e46558
URL: https://mhealth.jmir.org/2023/1/e46558
doi:10.2196/46558
PMID:38055318

Original Paper

Collecting Food and Drink Intake Data With Voice Input: Development, Usability, and Acceptability Study

Louise A C Millard[1,2], BSc, MSc, PhD; Laura Johnson[1,2,3,4], BSc, MSc, PhD; Samuel R Neaves[1,2], BSc, MSc, PhD; Peter A Flach[5], MSc, PhD; Kate Tilling[1,2], BSc, MSc, PhD; Deborah A Lawlor[1,2], MBChB, MSc, PhD

[1]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

[2]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]Centre for Exercise, Nutrition and Health Sciences, School for Policy Studies, University of Bristol, Bristol, United Kingdom

[4]Centre for Health, National Centre for Social Research (NatCen), London, United Kingdom

[5]Faculty of Engineering, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Louise A C Millard, BSc, MSc, PhD
Medical Research Council (MRC) Integrative Epidemiology Unit
University of Bristol
Oakfield House
Oakfield Grove
Bristol, BS8 2BN
United Kingdom
Phone: 44 0117 455 7676
Email: louise.millard@bristol.ac.uk

Abstract

Background: Voice-based systems such as Amazon Alexa may be useful for collecting self-reported information in real time from participants of epidemiology studies using verbal input. In epidemiological research studies, self-reported data tend to be collected using short, infrequent questionnaires, in which the items require participants to select from predefined options, which may lead to errors in the information collected and lack of coverage. Voice-based systems give the potential to collect self-reported information “continuously” over several days or weeks. At present, to the best of our knowledge, voice-based systems have not been used or evaluated for collecting epidemiological data.

Objective: We aimed to demonstrate the technical feasibility of using Alexa to collect information from participants, investigate participant acceptability, and provide an initial evaluation of the validity of the collected data. We used food and drink information as an exemplar.

Methods: We recruited 45 staff members and students at the University of Bristol (United Kingdom). Participants were asked to tell Alexa what they ate or drank for 7 days and to also submit this information using a web-based form. Questionnaires asked for basic demographic information, about their experience during the study, and the acceptability of using Alexa.

Results: Of the 37 participants with valid data, most (n=30, 81%) were aged 20 to 39 years and 23 (62%) were female. Across 29 participants with Alexa and web entries corresponding to the same intake event, 60.1% (357/588) of Alexa entries contained the same food and drink information as the corresponding web entry. Most participants reported that Alexa interjected, and this was worse when entering the food and drink information (17/35, 49% of participants said this happened often; 1/35, 3% said this happened always) than when entering the event date and time (6/35, 17% of participants said this happened often; 1/35, 3% said this happened always). Most (28/35, 80%) said they would be happy to use a voice-controlled system for future research.

Conclusions: Although there were some issues interacting with the Alexa skill, largely because of its conversational nature and because Alexa interjected if there was a pause in speech, participants were mostly willing to participate in future research studies using Alexa. More studies are needed, especially to trial less conversational interfaces.

(JMIR Mhealth Uhealth 2023;11:e41117) doi:10.2196/41117

KEYWORDS

digital health; data collection; voice-based approaches; Amazon Alexa; self-reported data; food and drink

Introduction

Epidemiological cohorts typically collect data at widely spaced time points (eg, every 1-5 years) [1,2]. Although some types of traits (eg, weight or height) are fairly stable or change gradually over time, others such as activity levels, blood glucose levels, mental well-being, and dietary intake can vary more acutely, for example, within days, hours, or even minutes. For these traits, prospectively capturing how they vary across time allows us to assess how this variability relates to other traits and disease. Some acutely varying traits can be collected continuously and objectively using wearable digital devices; for example, physical activity can be tracked using accelerometers or blood glucose can be measured using continuous glucose monitors [3]. For others, such as mental health traits and dietary intake, no objective approach to measuring within-day variation in these traits exists, and they need to be collected by self-report.

One possible approach to providing real-time self-reported information is verbal input, which could enable participants to conveniently enter free text. Over the last few years, several technology companies have released voice-controlled "smart" systems. These systems, such as Amazon Alexa, Google Assistant, and Samsung's Bixby, allow users to talk to a device rather than typing or pressing a button. They each have core functionality available by default (eg, saying the time when asked) and have developer platforms that allow anyone to produce and publish a custom voice-based app. This means that it is now technically possible to collect self-reported data continuously over a day or several days using verbal input.

Voice-based data collection may be most useful for collecting self-reported data that are both complex and variable across a day. One possible example is the food and drink consumed by a person and the time when they consume it. Traditionally, cohorts have collected dietary intake information using paper or web-based food frequency questionnaires or (less commonly) diaries. The limitations of these include retrospective recording, requiring conversion to an electronic form, potential for missing data because participants are not prompted for missing information, and the inconvenience of having to carry a diary. More recently, other approaches have been developed such as web-based dietary recall tools [4] and approaches using photographs [5-7]. Although these methods can collect detailed dietary information, they are burdensome, so they can only be used for short periods by highly motivated participants [3]. Approaches have been developed to detect eating events using wearable devices [8,9], for example, using wrist-worn accelerometers and gyroscopes [8]—these detect when an event occurs and not what was consumed. Wearable camera devices that capture images throughout the day have been trialed, but identifying and classifying food in images is challenging [10].

In this pilot study, we explored the potential of voice-based data collection in epidemiological research using food and drink diaries as an exemplar. Epidemiology studies are a challenging potential application of voice-based data collection because they are used to inform health policy and medical interventions; therefore, it is important to understand the biases in the collected data (eg, which food and drinks can be recorded correctly vs with error) to avoid incorrect conclusions being made. In addition, participation in epidemiological studies is predominantly altruistic, with participants usually receiving little direct benefit from participation, such that these studies aim to minimize participant burden to maximize participation. Our study has three key aims: (1) to demonstrate the technical feasibility of collecting data using Alexa, (2) to gain initial insight into participant acceptability, and (3) to provide an initial evaluation of the validity of the collected data. In general, we view the capture and processing of information as separate steps and, in this study, focused on demonstrating and evaluating the former.

Methods

Ethics Approval

Ethics approval was obtained from the University of Bristol Faculty of Health Sciences Research Ethics Committee (approval number 63861).

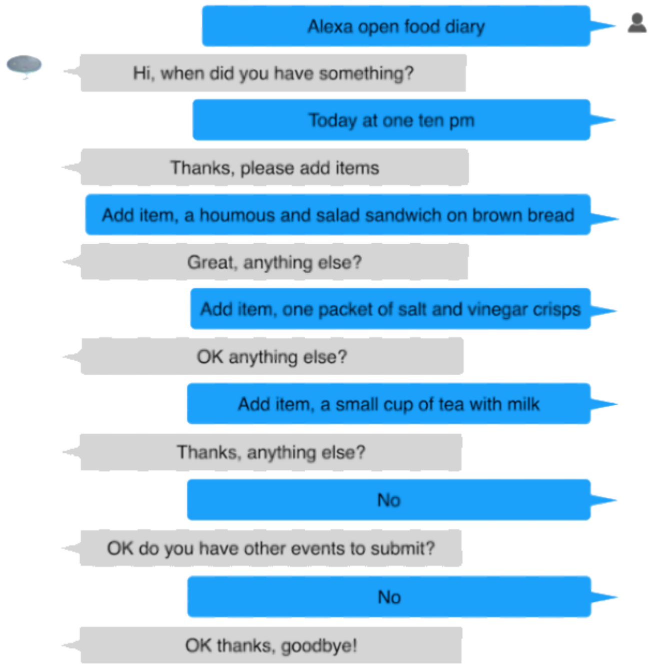
Study Participants

Power calculations based on 2 measures suggested that a sample size of at least 35 is needed (see details in Section S1 in Multimedia Appendix 1). We recruited volunteers from the University of Bristol staff and student email lists. Participants were compensated with a £30 (US $36) voucher after they submitted the postparticipation questionnaire (receiving the voucher was not dependent on them submitting any food diary entries).

Description of the System Architecture: a Voice-Based System Using Amazon Alexa

In this study, we used the Amazon Alexa voice system (a comparison with other voice-based systems such as Google Assistant and Samsung's Bixby is left for future work). The Alexa system enables the development of custom functionality, referred to as a custom skill. Alexa skills comprise intents that each define an interaction that a user can have with the skill. We developed a custom skill to collect food and drink intake events, with intents that allow participants to (1) add the date and time of an intake event, (2) add ≥1 items they ate or drank at this time, (3) cancel the event, (4) cancel the last item added to the event, and (5) submit the event. See example utterances in Table S1 in Multimedia Appendix 1 and an example conversation in Figure 1. Section S2 in Multimedia Appendix 1 provides further details on the system architecture.

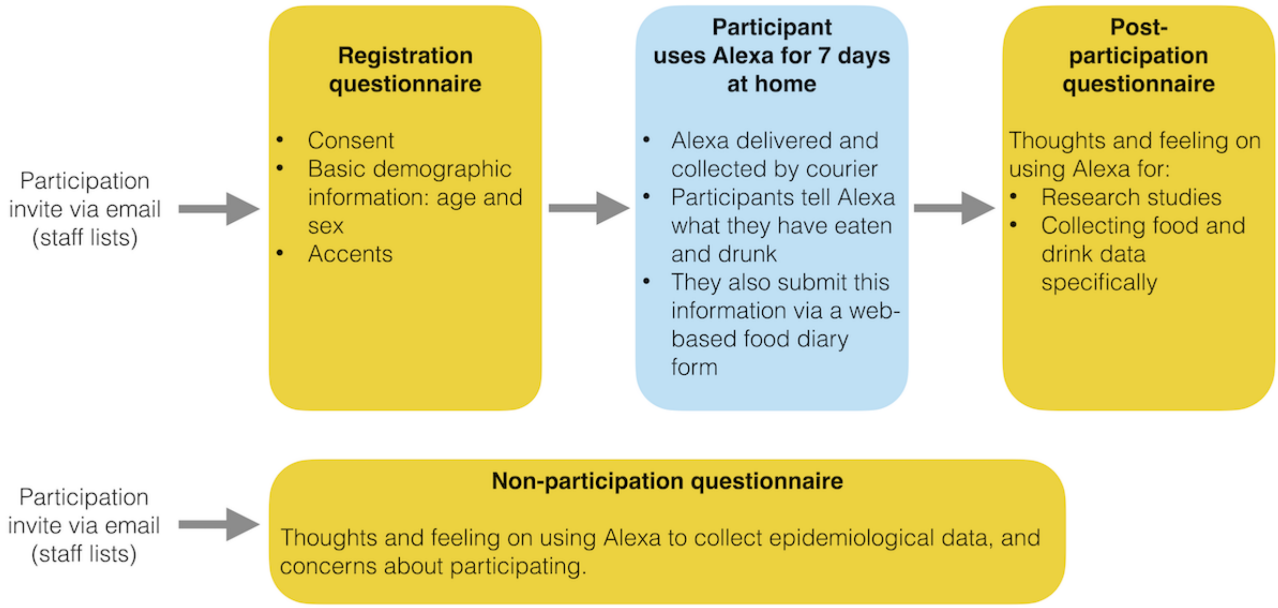
Figure 1. Example conversation of study participant with custom food diary skill.



Data Collection Protocol

An overview of the data collection protocol is shown in Figure 2. Owing to the COVID-19 pandemic, participants took part at home. We sent an initial email with an accompanying participation information sheet (Multimedia Appendix 2) inviting staff and students to participate in this study. Upon replying, participants were sent a preparticipation questionnaire asking for basic demographic information such as their age and sex (questionnaire 1 in Multimedia Appendix 3). On completion, participants were booked for a 7-day data collection period using Alexa.

Figure 2. Overview of data collection protocol.



The equipment was stored in the principal investigator's (LACM) home. On day 1 of the participants' data collection period using Alexa, the equipment was delivered to their home by courier, along with a participant guide (Multimedia Appendix 4). The participant was asked to set up the equipment and start using it as soon as possible. Participants were instructed with the following statement: "After you have had something to eat or drink, we would like you to submit your food and drink information to Alexa first, and then submit it on the web form." Entering the food and drink information using both Alexa and a web form (questionnaire 4 in Multimedia Appendix 3) allowed us to compare the data entered using these approaches (ie, relative validity [11]). As participants entered the date and time of the intake event, they were able to enter events consumed earlier on the same day or on a previous day (including those consumed outside the home). On day 7, the equipment was returned to the principal investigator's home via courier. Participants were then asked to complete a postparticipation questionnaire on their experiences during the study and the acceptability of using Alexa (questionnaire 2 in Multimedia Appendix 3).

To understand views on the acceptability of using voice-based interfaces more widely (beyond our participant group), we also sent a further invitation (to the same email lists) asking those who did not participate to complete a short questionnaire about their feelings on using voice-based devices and their reasons for not participating (questionnaire 3 in Multimedia Appendix 3).

The questionnaires were deployed via the University of Bristol REDCap (Research Electronic Data Capture; Vanderbilt University) secure web platform [12]. The content of the study emails is provided in Multimedia Appendix 5.

Analytical Sample

A participant flow diagram is shown in Figure S1 in Multimedia Appendix 1. Of the 45 participants who registered to participate, 1 (2%) withdrew and 7 (16%) were excluded owing to equipment issues (Section S3 in Multimedia Appendix 1). The remaining 82% (37/45) of participants comprised our analytical sample. Among these, 3% (1/37) of participants did not attempt to use Alexa. In addition, 19% (7/37) of participants had Alexa entries but no web diary entries completed within the 30 minutes directly following the Alexa submission. As Alexa and web entries must be submitted within 30 minutes to be identified as corresponding to the same intake event in our data processing approach (see below), the Alexa and web entries from these participants could not be compared. The entries from the remaining 78% (29/37) of participants were used to compare the information entered via the web form versus Alexa ("comparison" sample).

Data Preprocessing

Mapping Web Food Form Entries to Alexa Intake Events

The web and Alexa entries both included the following information: (1) intake timestamp—the date and time the participant (said they) ate or drank; (2) submission timestamp—the date and time the participant submitted the entry; and (3) intake items—≥1 food and drink intake items. To compare the content of the web and Alexa entries, we first undertook an automated process to identify Alexa and web entry pairs that correspond to the same intake event, referred to as counterpart entries. This was nontrivial because a participant might not have entered each entry with the web form immediately after entering it via Alexa or the intake timestamp entered via Alexa might have been recorded incorrectly (ie, Alexa might have heard the day or time stated by the participant incorrectly).

We identified counterpart entries using intake and submission timestamps. The process we used was as follows (illustrated in Figure S5 in Multimedia Appendix 1):

1. Identify counterparts as the set of entries in which the web and Alexa intake timestamps were within 5 minutes of each other, and the Alexa submission timestamp was up to 30 minutes before the web submission timestamp. The nonexact match of the intake time was because participants can tell Alexa this using a phrase such as "just now" or "ten minutes ago," which may not correspond exactly to the intake time entered using the web form.
2. Identify web counterpart entries of the Alexa submissions not matched in step 1 as the nearest subsequent web entry where one occurs within 30 minutes of the Alexa entry.

Comparing Food and Drink Descriptions in counterpart Web and Alexa Entries

We compared counterpart entries using 2 approaches, an automated approach and a systematic manual approach.

Automated Approach

We compared the text content of the counterpart entries by comparing the set of words contained in each. Entries were preprocessed to remove plurality of words (eg, "crisps" becomes "crisp") [13] and convert numbers to numeric values (eg, "one" and "a" both become 1). For each counterpart pair, we calculated the number of words in (1) the web word set but not the Alexa word set, (2) the Alexa word set but not the web word set, and (3) both word sets.

Systematic Manual Approach

Our systematic manual approach was conducted by LACM. As this approach has some degree of judgment, we also asked 5 researchers independent to the project (within the same unit but not involved in this study) to review 10 random entries (none repeated across researchers) so that we can evaluate the interresearcher variability of these manual evaluations.

We used a 2-step process to conduct this manual review. First, the intake items of each counterpart pair were compared to determine whether there was any similarity. If the set of items was completely different, then they were marked as most likely corresponding to different intake events (ie, the counterpart pairing did not work in this case, eg, "a cup of coffee with milk" vs "spaghetti bolognaise"). All other entries were performed in step 2.

Step 2 involved reviewing each counterpart entry and, for each, recording the number of food or drink items in a counterpart pair in the following categories:

1. Same item semantically (the 2 entries are equivalent with no additional or different information in each)
2. Same item but with different details (eg, "cup of tea" vs "mug of tea")
3. Same item, Alexa information has less detail (eg, "cheese and salad sandwich" vs "a sandwich")
4. Same item, Alexa item has more detail
5. Same item, misspelling in Alexa input, but still understandable, that is, there is no loss of information (eg, "to bagels" vs "two bagels")
6. Same item, misspelling in web form input, but still understandable
7. Same item, with Alexa entry issue, in which the consumed item is still identifiable (eg, "ball of yoghurt" rather than "bowl of yoghurt")
8. Item with major entry issue, such that it contains no food or drink information, or the main essence of the food or

drink is missing (eg, a "cough with milk" rather than "coffee with milk")

9. Extra Alexa item with major entry issue (which can happen if a participant makes a mistake or stops talking, then tries again so there is an extra item, eg, "two")
10. Extra Alexa item that is recognizable as a food or drink (ie, should not be assigned to category 9)
11. Extra web item

Table S2 in Multimedia Appendix 1 shows some example assignments using this approach.

The independent researchers who completed 10 entries were provided with an information sheet describing the task (Multimedia Appendix 6). We visually evaluated the agreement between the assignments of LACM and independent researchers using a stacked bar chart.

The automated and systematic manual approaches are complementary because the former is objective but is likely to be a more pessimistic assessment of agreement. This is because participants may not write an entry in the same way that they would speak it. For example, a participant might write "1 x apple. 1 bar of chocolate" but say "one apple and a chocolate bar," which has differences in the words used even though they are semantically the same.

Statistical Analyses

Use Summary

We summarized the participants' use of the web and Alexa approaches using the median and IQR of the number of submitted web and Alexa entries, respectively.

Comparison of Counterpart Diary Entries

We compared the intake timestamps in the counterpart pairs using a plot similar to a Bland-Altman plot but in which the x-axis is the intake time entered using the web form rather than the average. Assuming that the intake timestamp entered on the web form will be largely correct, this is to help show whether the intake time submitted via Alexa may be less accurate for particular times of the day. We summarized automated and systematic manual comparisons using stacked bar charts.

Summarizing the Number of Incomplete Attempts

The Alexa skill saves partial entries (ie, those that have not been submitted, perhaps because the internet connection was interrupted) in addition to completed entries. We estimated the median (IQR) number of unsuccessful attempts across participants.

Evaluating Participant Questionnaire Responses on Usability and Acceptability

We summarized the responses to the postparticipation questionnaire (questionnaire 2 in Multimedia Appendix 3) and the nonparticipation questionnaire (questionnaire 3 in Multimedia Appendix 3) by calculating the number of participants (and percentage) that responded to each questionnaire item option. Responses to free-text items were read and reread to identify the key themes.

The Alexa skill and web service code, and analysis code, are publicly available [14,15]. Git tag version 0.1 of the analysis code corresponds to the version of the analyses presented here.

Results

Participant Characteristics

The participant characteristics are summarized in Table 1. Most participants (30/37, 81%) were in their early adulthood (aged 20-39 years). Our sample included more female participants than male participants (23/37, 62% female). The majority (31/37, 84%) reported that they did not believe they had a strong regional UK accent, with 68% (25/37) reporting that they did not have an accent because English was a second language. In total, 43% (16/37) of participants have an Alexa device at home that they use.

Table 1. Participant demographics (n=37).

Participant characteristics[a]	Values, n (%)
Age range (years)	
<20	4 (11)
20-29	16 (43)
30-39	14 (38)
40-49	2 (5)
50-59	1 (3)
Sex	
Female	23 (62)
Male	14 (38)
Other	0 (0)
Has a regional UK accent	
No	31 (84)
A little	4 (11)
Yes	2 (5)
Has a non-English accent	
No	25 (68)
A little	9 (24)
Yes	3 (8)
Has a voice-controlled device	
No	16 (43)
Yes, but it is used by others, not me	5 (14)
Yes, and I use it	16 (43)

[a]The characteristics shown are those collected in this study.

Use Summary

On average, participants completed more web diary entries than Alexa entries (median number of entries was 17, IQR 13-27 compared with 11, IQR 7-21; paired 2-tailed *t* test *P* value <.001; comparison shown in Figure S6 in Multimedia Appendix 1). The median number of partial Alexa attempts across all participants was 6 (IQR 1-9).

Comparison of Counterpart Diary Entries

Intake Timestamp Comparison of Web Form Versus Alexa Entries

Across the 29 participants in the comparison subsample, there were 310 counterpart entries. Of these, 71.6% (222/310) had a matching timestamp (Figure S7 in Multimedia Appendix 1). The median proportion of completed counterpart entries with a matching timestamp across the participants was 0.67 (IQR 0.5-1).

Food and Drink Description Comparison of Web Form Versus Alexa Entries

The results comparing the submitted food and drink information using automated and manual comparison approaches are shown in Figures 3 and 4, respectively. Of the 310 counterpart entries manually reviewed, 21 (6.8%) were classified as corresponding to different intake events. The remaining 93.2% (289/310) of counterpart entries included 612 web form items and 588 Alexa items, with 33 extra web items (not identified in the counterpart Alexa entry) compared with 9 extra Alexa items (not found in the counterpart web entry). The majority (357/612, 58.3% and 357/588, 60.7% for the web and Alexa items, respectively) of the items entered via the web form and Alexa were the same, containing the same information. Of the 194 items that were identified as corresponding to the same intake item but containing different information, 64 (33%) had less detail from Alexa, 12 (6.2%) had more detail from Alexa, 15 (7.7%) had different detail, 3 (1.5%) had a web entry issue, 36 (18.6%) had an Alexa entry issue, 4 (2.1%) had spelling mistakes in the web version not the Alexa version, 59 (30.4%) had a misspelling in Alexa only, and 1 (0.5%) had a misspelling in both the Alexa and web input. Of the 59 items with an Alexa misspelling, 40 (68%) were owing to Alexa recording the word “to” rather than “two.” Of the 588 items entered via Alexa, 28 (5%) were classified as having a major entry issue.

We did not identify systematic differences in the assignments by LACM for the systematic manual approach compared with those of independent researchers (Figure S8 in Multimedia Appendix 1).

Figure 3. Summary of automated comparison of the food and drink information submitted using Alexa versus the web form. Results shown for 29 participants in the "comparison" sample. Each stacked bar shows the number of unique words in (1) the Alexa entry only, (2) both the web and Alexa entries, and (3) the web entry only. Each block of stacked bars shows the set of entries for a given participant.

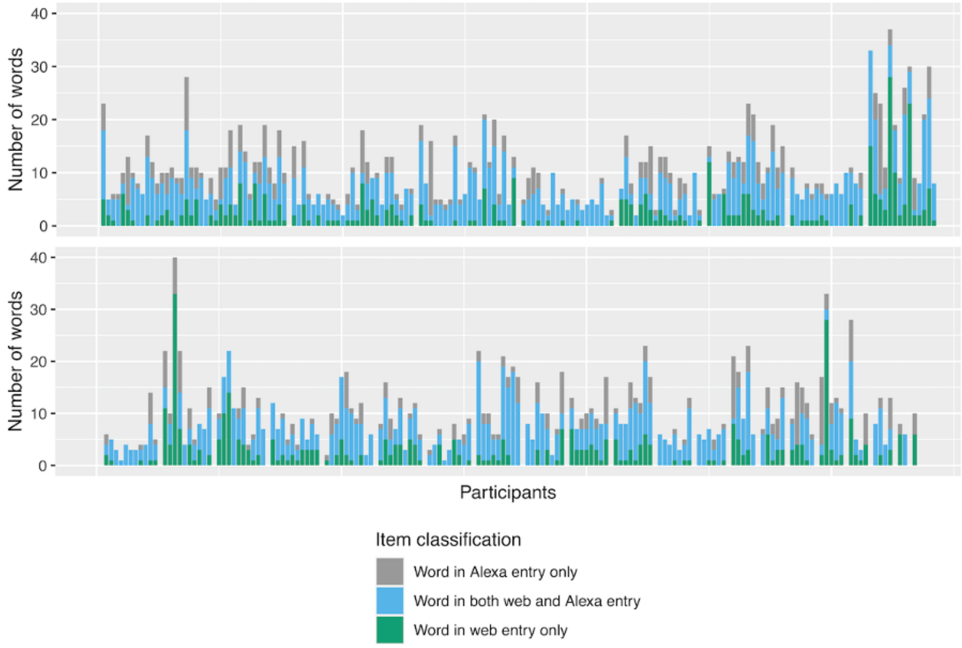
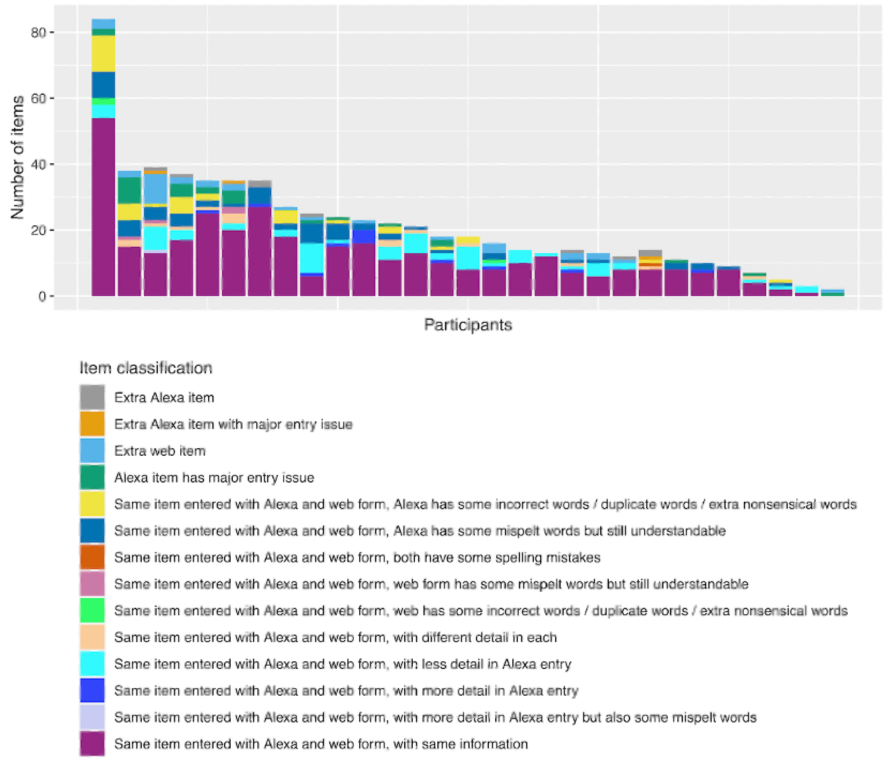


Figure 4. Summary of manual evaluation of the submitted food and drink information. Results shown for 29 participants in the "comparison" sample. Each stacked bar indicates the number of items in each category for each participant. Manual evaluation was conducted by LACM.



Evaluating Participant Questionnaire Responses on Usability and Acceptability

The summaries of the postparticipation questionnaire responses are provided in Table 2. Of the 35 participants who completed the postparticipation questionnaire, 26 (74%) said they would be happy to use a voice-controlled system at home for future research and 28 (80%) said they would be happy to use one on a wearable device (eg, a smart watch). Alexa sometimes interjected when participants were telling her *when* they ate or drank, with 20% (7/35) of participants saying that this happened often or always and 31% (11/35) saying that this happened occasionally. Alexa often interjected when participants were telling her *what* they ate or drank, with 51% (18/35) of participants saying that this happened often or always and 34% (12/35) saying that this happened occasionally. In terms of convenience, enjoyment, and efficiency, 51% (18/35), 60% (21/35), and 43% (15/35) of participants, respectively, said that they found using Alexa “OK” or better.

Of the 35 participants who completed the postparticipation questionnaire, 25 (71%) had previously used another approach to record their food and drink intake (Table S3 in Multimedia Appendix 1). Of the 13 participants who have previously used a traditional diary (on paper or a computer), 7 (54%) found using Alexa at least as convenient, 5 (56%) found using Alexa at least as enjoyable, and 2 (22%) found using Alexa at least as efficient. Of the 34% (12/35) of participants who have previously used MyFitnessPal, 75% (9/12) found Alexa at least as convenient, 64% (7/12) found Alexa at least as enjoyable, and 36% (4/12) found Alexa at least as efficient.

Table 2. Postparticipation questionnaire summary (n=35).

Questionnaire items[a]	Values, n (%)
Participant was able to accurately tell Alexa what they ate and drank (n=35)	
Completely agree	3 (9)
Somewhat agree	16 (46)
Neither agree not disagree	2 (6)
Somewhat disagree	12 (34)
Completely disagree	2 (6)
Participant was able to estimate accurate quantities describing how much they ate (n=34)	
Completely agree	3 (9)
Somewhat agree	11 (32)
Neither agree not disagree	10 (29)
Somewhat disagree	9 (26)
Completely disagree	1 (3)
Participant chose not to record particular snacks or meals (eg, because it was unhealthy; n=35)	
Completely agree	1 (3)
Somewhat agree	5 (14)
Neither agree not disagree	0 (0)
Somewhat disagree	7 (20)
Completely disagree	22 (63)
Participant sometimes chose to be selective with the truth (n=35)	
Completely agree	3 (9)
Somewhat agree	5 (14)
Neither agree not disagree	3 (9)
Somewhat disagree	6 (17)
Completely disagree	18 (51)
Participant felt they remembered to submit food and drink information (n=35)	
Completely agree	9 (26)
Somewhat agree	19 (54)
Neither agree not disagree	1 (3)
Somewhat disagree	6 (17)
Completely disagree	0 (0)
Alexa interjected when I had not finished telling her when I ate or drank (n=35)	
Never	4 (11)
Rarely	13 (37)
Occasionally	11 (31)
Often	6 (17)
Always	1 (3)
Alexa interjected when I had not finished telling her what I ate or drank (n=35)	
Never	2 (6)
Rarely	3 (9)
Occasionally	12 (34)
Often	17 (49)
Always	1 (3)

Questionnaire items[a]	Values, n (%)
How convenient or inconvenient did you find providing information using Alexa? (n=35)	
Very inconvenient	4 (11)
Somewhat inconvenient	13 (37)
OK	10 (29)
Somewhat convenient	7 (20)
Very convenient	1 (3)
How enjoyable or unenjoyable did you find providing information using Alexa? (n=35)	
Very unenjoyable	3 (9)
Somewhat unenjoyable	11 (31)
OK	15 (43)
Somewhat enjoyable	6 (17)
Very enjoyable	0 (0)
How efficient or inefficient did you find providing information using Alexa? (n=35)	
Very inefficient	5 (14)
Somewhat inefficient	15 (43)
OK	6 (17)
Somewhat efficient	8 (23)
Very efficient	1 (3)
How easy or hard did you find providing information using Alexa? (n=35)	
Could not use at all	0 (0)
Very hard	3 (9)
Somewhat hard	18 (51)
OK	7 (10)
Somewhat easy	7 (20)
Very easy	0 (0)
Happy to use a voice-controlled system (eg, Alexa) at home for research in the future (n=35)	
Yes	26 (74)
No	2 (6)
Not sure	7 (20)
Happy to use a voice-controlled system (eg, Alexa) on a wearable device such as a smart watch, for research (n=35)	
Yes	28 (80)
No	1 (3)
Not sure	6 (17)

[a]A summary of all items in the postparticipation questionnaire is provided in Table S3 in Multimedia Appendix 1.

Evaluating Nonparticipation Questionnaire Responses

Of the 69 participants who responded, 11 (16%) did not take part because of privacy concerns (with respect to Amazon, researchers collecting their diet information, or Alexa inadvertently listening to other conversations; Table 3). In total, 61% (42/69) stated that they would be happy to use Alexa at home for future research, whereas 57% (39/69) said that they would be happy to use Alexa on a wearable device for research purposes.

Table 3. Nonparticipation questionnaire summary (n=69).

Questionnaire items	Value, n (%)
Age range (years)	
<20	13 (19)
20-29	36 (52)
30-39	11 (16)
40-49	5 (7)
50-59	2 (3)
Prefer not to answer	2 (3)
Sex	
Female	51 (74)
Male	18 (26)
Other	0 (0)
Reasons did not take part	
Not available during the study session times	30 (43)
Data privacy concerns around Amazon collecting information on my diet	9 (13)
Data privacy concerns around researchers collecting information on my diet	3 (4)
Concerns that Alexa will inadvertently listen to other conversations	9 (13)
I do not eat or drink during my working hours	3 (4)
Picking up and returning the device was inconvenient	5 (7)
Other reason[a]	21 (30)
Has a voice-controlled device	
No	35 (51)
Yes, but it is used by others, not me	10 (14)
Yes, and I use it	24 (35)
Happy to use a voice-controlled system (eg, Alexa) at home for research in the future	
Yes	42 (61)
No	13 (19)
Not sure	14 (20)
Happy to use a voice-controlled system (eg, Alexa) on a wearable device such as a smart watch for research	
Yes	39 (57)
No	18 (26)
Not sure	12 (17)

[a]Participants who stated "other" were able to complete a free-text response; these are summarized in Section S4 in Multimedia Appendix 1. A summary of all items in the nonparticipation questionnaire is provided in Table S4 in Multimedia Appendix 1.

Discussion

In this study, we have demonstrated the technical feasibility of collecting data using Alexa for epidemiological research by successfully developing an Alexa skill to collect food and drink information and using it to collect data from 37 participants across a period of 7 days (5 full days). Our results provide useful initial insight into the participant acceptability of using this approach and validity of the collected data. On average, more entries were submitted via the web form than via Alexa. Our results suggest that intake date and time was largely entered accurately via Alexa. The majority of the Alexa entries (357/588, 60.7%) contained the same food and drink information as the corresponding web entry, according to our systematic manual approach. The most common differences were Alexa information having less detail or a homophone error (most often "to" rather than "two").

Overall, the usability of our Alexa skill was fairly poor. Most participants reported that Alexa interjected while they were trying to enter food and drink information (12/35, 34% of participants sometimes and 18/35, 51% often or always), with better results for the date or time of the intake event (11/35, 31% of participants sometimes and 7/35, 20% often or always). Several participants reported finding it difficult to avoid pausing while articulating what they ate or drank, which might cause Alexa to interject or cut out. Some reported reducing the information they provided, so that Alexa would be more likely to accept it. The participants also reported that Alexa sometimes did not understand or would exit the skill during use.

The voice interface we have trialed comprises our Alexa skill implementation and the Amazon back-end logic, and only the former is under our control. The implementation and deployment of the Alexa skill has several components, with many choices regarding the design of the voice interface, the technical infrastructure, and the study protocol (eg, location of data collection, which was home based in our study). Each of these factors may have affected the usability of the skill to collect food and drink information. Most notably, we conclude that the conversational interface of our skill (in which participants first tell Alexa the time, then each of the items consumed) was not successful, because when the skill inadvertently cut out (eg, because of multiple failed attempts to converse with Alexa or a poor internet connection), the participant would have to start that entry from the beginning. A less conversational interface in which the participant states the information without separate prompts would likely be more usable. Although our results suggest that Alexa may be more appropriate for entering short summaries of information, in the longer term, the integration of this approach with other approaches (such as a phone app) can be used to supplement voice-collected data. For example, using Alexa to log events directly after eating or drinking (eg, on a wearable device) and then entering more detail via a phone app when convenient. Therefore, while in this study, participants entered more events and provided more detail using the web form, there are some opportunities to improve the skill for future studies.

The strengths of our study include the collection of pilot data "in the wild" rather than in a controlled laboratory-based setting. We collected food and drink information via a web form, in addition to Alexa, to allow the comparison of the data collected using these approaches. Our study had several limitations. While asking participants to provide information via both Alexa and a web form was valuable, interactions with one of these approaches may have affected their interaction and perceived feelings toward the other. The Alexa and web entries in our data had no explicit link and identifying entries that corresponded to the same intake event was difficult. We could only assess the relative validity of the Alexa entries (relative to the web form entries), that is, we have no absolute ground truth. Although the intake date and time could be easily compared between the Alexa and web form entries in an automated manner, comparing the free-text food and drink information was nontrivial as differences in the way the participant conveyed this information would not necessarily amount to meaningful differences in the submitted information. Most of these limitations could be rectified by integrating this voice-based approach with a phone app in which the participant can review each entry and either correct it or mark it as correct, instead of requiring a web diary, so that validity can be assessed in an automated manner by evaluating the corrections made by the participant. This would likely increase the number of Alexa entries that could be evaluated and could also reduce the participant burden because entries would either need to be marked as correct or corrected, rather than inputting all the information on a web form. Additional strengths and limitations and details are provided in Section S5 in Multimedia Appendix 1.

Although other studies have used voice-based approaches in other health settings [16-19], to the best of our knowledge, this is the first study to assess collecting self-reported epidemiology data with a voice-based system (to the best of our knowledge, a previous grant that sought to create a voice-based interface did not achieve this objective [20]). Furthermore, although our focus was on using this technology for collecting epidemiological data, the results of our study are likely to be useful more broadly, for example, to inform the development of technologies for personalized health care or commercial systems (collecting self-reported data to track behavior).

Table 4 summarizes the main findings of this study. More studies are needed to understand the strengths and limitations of different approaches to collect epidemiological data using voice, for example, with different voice-based systems (eg, comparing Amazon Alexa vs Google Assistant), different types of devices (eg, wearables vs smartphones), different voice interface designs, particularly those that are less conversational, and to further evaluate biases in the collected data [21]. Although this study used an Amazon Echo Dot device situated in the participants' home, it is also possible to deploy an Alexa skill on other devices, for example, on smartphones and wearables. The acceptability of collecting epidemiological data with voice (including the length of time a participant may be willing to use such an approach), and the accuracy of the collected data, may differ depending on the device used (eg, because of differing levels of background noise when "on the go" vs in the home environment). Further studies are needed to investigate this. Voice-based approaches may be particularly useful in populations that might not be able to write (or write with ease), for example, those with learning difficulties, such as dyslexia, or certain diseases, such as motor neuron disease.

Table 4. Summary of the main results and implications for future research.

Results	Implications for future research
Voice-based data collection is technically feasible.	Future studies are needed to understand the strengths and limitations of different voice interfaces.
Conversational interface was a frustration for users because it could cut out (eg, owing to a poor internet connection) and the conversation would have to start from the beginning.	Design a less conversational voice interface.
Alexa more suited to entering short bits of information.	Integration with a phone app would allow supplementing information to be entered with voice entries.
The majority of the Alexa entries (357/588, 60.7%) contained the same food or drink information as the corresponding web entry, but a substantial proportion contained differences.	Trial and compare different voice-based systems such as the Google Assistant.
Matching voice entry with corresponding web form entry was difficult and many could not be matched (and therefore compared).	Use a phone app to evaluate the collected data by asking participant to validate the entry, either marking the entry as correct or providing a correction.

Acknowledgments

This work was supported by the University of Bristol and the UK Medical Research Council (grants MC_UU_00011/3 and MC_UU_00011/6). LACM received funding from a University of Bristol Vice-Chancellor's Fellowship. This study was also supported by the National Institute for Health and Care Research Biomedical Research Centre at the University Hospitals Bristol National Health Service Foundation Trust and the University of Bristol (specifically via a Biomedical Research Centre Directors Fund grant). DAL's contribution is supported by the British Heart Foundation (grants CH/F/20/90003 and AA/18/7/34219).

The authors are extremely grateful to the study participants. They are grateful to Mrs Shirley Jenkins for providing the administrative support for this study. The authors thank Professor Katrina Turner for helpful feedback on the qualitative review of the questionnaire responses. The authors also thank Mr Tom Clark, Dr Emily Kawabata, Dr Michael Lawton, Dr Nancy McBride, and Dr Tom Palmer, who conducted independent manual assessments of subsamples of food and drink entries.

Authors' Contributions

LACM and SRN conceptualized the use of Alexa for epidemiological data collection. LACM conceptualized the study, led the designing of the study, and wrote the first version of the manuscript. All authors contributed to the study design. All authors critically reviewed and revised the manuscript. LACM acts as the guarantor for this study.

Conflicts of Interest

DAL has received support from numerous national and international governments and charitable funders as well as Roche Diagnostics and Medtronic Ltd for work unrelated to this paper. KT received grant support including from the UK Government and US Government for research unrelated to this work.

Multimedia Appendix 1
Supplementary text, figures, and tables.
[PDF File (Adobe PDF File), 501 KB - mhealth_v11i1e41117_app1.pdf]

Multimedia Appendix 2
Study participant information sheet.
[PDF File (Adobe PDF File), 138 KB - mhealth_v11i1e41117_app2.pdf]

Multimedia Appendix 3
Study questionnaires.
[PDF File (Adobe PDF File), 198 KB - mhealth_v11i1e41117_app3.pdf]

Multimedia Appendix 4
Study participant guide.
[PDF File (Adobe PDF File), 412 KB - mhealth_v11i1e41117_app4.pdf]

Multimedia Appendix 5
Study emails.
[PDF File (Adobe PDF File), 118 KB - mhealth_v11i1e41117_app5.pdf]

Multimedia Appendix 6
Information sheet describing manual comparison task, of Alexa, and web entries.
[PDF File (Adobe PDF File), 140 KB - mhealth_v11i1e41117_app6.pdf]

References

1. Boyd A, Golding J, Macleod J, Lawlor D, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'--the index offspring of the Avon longitudinal study of parents and children. Int J Epidemiol 2013 Feb;42(1):111-127 [FREE Full text] [doi: 10.1093/ije/dys064] [Medline: 22507743]
2. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015 Mar;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
3. Lawlor DA, Lewcock M, Rena-Jones L, Rollings C, Yip V, Smith D, ALSPAC Executive. The second generation of The Avon Longitudinal Study of Parents and Children (ALSPAC-G2): a cohort profile. Wellcome Open Res 2019 Dec 16;4:36 [FREE Full text] [doi: 10.12688/wellcomeopenres.15087.2] [Medline: 31984238]
4. Simpson E, Bradley J, Poliakov I, Jackson D, Olivier P, Adamson AJ, et al. Iterative development of an online dietary recall tool: INTAKE24. Nutrients 2017 Feb 09;9(2):118 [FREE Full text] [doi: 10.3390/nu9020118] [Medline: 28208763]
5. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. J Hum Nutr Diet 2014 Jan;27 Suppl 1(0 1):72-81 [FREE Full text] [doi: 10.1111/jhn.12014] [Medline: 23848588]
6. Whitton C, Healy JD, Collins CE, Mullan B, Rollo ME, Dhaliwal SS, et al. Accuracy and cost-effectiveness of technology-assisted dietary assessment comparing the automated self-administered dietary assessment tool, intake24, and an image-assisted mobile food record 24-hour recall relative to observed intake: protocol for a randomized crossover feeding study. JMIR Res Protoc 2021 Dec 16;10(12):e32891 [FREE Full text] [doi: 10.2196/32891] [Medline: 34924357]
7. Healy J, Boushey C, Delp E, Zhu F, Collins C, Rollo M, et al. Mobile food record 24 hour recall (mFR24) was “easy” and an acceptable mobile health dietary assessment method. Curr Dev Nutr 2022 Jun;6(Supplement_1):767. [doi: 10.1093/cdn/nzac063.009]
8. Sharma S, Hoover A. Top-down detection of eating episodes by analyzing large windows of wrist motion using a convolutional neural network. Bioengineering (Basel) 2022 Feb 11;9(2):70 [FREE Full text] [doi: 10.3390/bioengineering9020070] [Medline: 35200423]
9. Zhang S, Zhao Y, Nguyen DT, Xu R, Sen S, Hester J, et al. NeckSense: a multi-sensor necklace for detecting eating activities in free-living conditions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Jun 15;4(2):1-26 [FREE Full text] [doi: 10.1145/3397313] [Medline: 34222759]
10. Skinner A, Toumpakari Z, Stone C, Johnson L. Future directions for integrative objective assessment of eating using wearable sensing technology. Front Nutr 2020 Jul 2;7:80 [FREE Full text] [doi: 10.3389/fnut.2020.00080] [Medline: 32714939]
11. Cade JE, Warthon-Medina M, Albar S, Alwan NA, Ness A, Roe M, DIET@NET consortium. DIET@NET: best practice guidelines for dietary assessment in health research. BMC Med 2017 Nov 15;15(1):202 [FREE Full text] [doi: 10.1186/s12916-017-0962-x] [Medline: 29137630]
12. University of Bristol clinical data server. REDCap. URL: https://sscmredcap.bris.ac.uk/redcap/ [accessed 2021-12-09]
13. Rudis B, Embrey B. Pluralize: pluralize and 'singularize' any (English) word. CRAN R Project. 2020. URL: https://cran.r-project.org/package=pluralize [accessed 2021-12-14]
14. Millard LAC. Food diary app with Alexa. GitHub. URL: https://github.com/MRCIEU/food-diary-alexa-pilot-aws [accessed 2022-12-14]
15. Millard LAC. Alexa food diary pilot analysis code. GitHub. URL: https://github.com/MRCIEU/food-diary-alexa-pilot-analysis [accessed 2022-12-14]
16. Jian W, Wang J, Rahmanti AR, Chien S, Hsu C, Chien C, et al. Voice-based control system for smart hospital wards: a pilot study of patient acceptance. BMC Health Serv Res 2022 Mar 03;22(1):287 [FREE Full text] [doi: 10.1186/s12913-022-07668-1] [Medline: 35236341]
17. Sharma A, Oulousian E, Ni J, Lopes R, Cheng M, Label J, et al. Voice-based screening for SARS-CoV-2 exposure in cardiovascular clinics. Eur Heart J Digit Health 2021 Sep;2(3):521-527 [FREE Full text] [doi: 10.1093/ehjdh/ztab055] [Medline: 36713601]
18. Yang S, Lee J, Sezgin E, Bridge J, Lin S. Clinical advice by voice assistants on postpartum depression: cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e24045 [FREE Full text] [doi: 10.2196/24045] [Medline: 33427680]

19. Kocaballi AB, Sezgin E, Clark L, Carroll JM, Huang Y, Huh-Yoo J, et al. Design and evaluation challenges of conversational agents in health care and well-being: selective review study. J Med Internet Res 2022 Nov 15;24(11):e38525 [FREE Full text] [doi: 10.2196/38525] [Medline: 36378515]
20. Weiss R. Mobile food intake visualization and voice recognize (FIVR). Grantome. URL: https://grantome.com/grant/NIH/U01-HL091738-04S2 [accessed 2022-12-14]
21. Preparing for the future: tackling the key challenges facing the UK's longitudinal population studies. CLOSER longitudinal research. 2019. URL: https://www.closer.ac.uk/wp-content/uploads/20190401-Preparing-for-the-future-Conference-report.pdf [accessed 2022-06-10]

Abbreviations

REDCap: Research Electronic Data Capture

Edited by L Buis, G Eysenbach; submitted 15.07.22; peer-reviewed by S Wang; comments to author 17.10.22; revised version received 06.01.23; accepted 03.02.23; published 31.03.23.

Please cite as:
Millard LAC, Johnson L, Neaves SR, Flach PA, Tilling K, Lawlor DA
Collecting Food and Drink Intake Data With Voice Input: Development, Usability, and Acceptability Study
JMIR Mhealth Uhealth 2023;11:e41117
URL: https://mhealth.jmir.org/2023/1/e41117
doi:10.2196/41117
PMID:37000476

Original Paper

G Tolerance Prediction Model Using Mobile Device–Measured Cardiac Force Index for Military Aircrew: Observational Study

Ming-Hao Kuo[1], MSc; You-Jin Lin[2], MSc; Wun-Wei Huang[3], BSc; Kwo-Tsao Chiang[4,5,6], MSc, MD; Min-Yu Tu[7,8,9], MD, PhD; Chi-Ming Chu[1,5,10,11,12,13], PhD; Chung-Yu Lai[6], PhD

[1]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei City, Taiwan
[2]Thoracic Department, China Medical University Beigang Hospital, Yunlin County, Taiwan
[3]Aviation Physiology Research Laboratory, Kaohsiung Armed Forces General Hospital Gangshan Branch, Kaohsiung City, Taiwan
[4]Superintendent Office, Taipei Veterans General Hospital Fonglin Branch, Hualien County, Taiwan
[5]School of Public Health, National Defense Medical Center, Taipei City, Taiwan
[6]Graduate Institute of Aerospace and Undersea Medicine, National Defense Medical Center, Taipei City, Taiwan
[7]Orthopedics Division, Taichung Armed Forces General Hospital, Taichung City, Taiwan
[8]Department of Health Business Administration, Meiho University, Pingtung County, Taiwan
[9]Department of Life Sciences, National Chung Hsing University, Taichung City, Taiwan
[10]Graduate Institute of Life Sciences, National Defense Medical Center, Taipei City, Taiwan
[11]Big Data Research Center, College of Medicine, Fu-Jen Catholic University, New Taipei City, Taiwan
[12]Department of Public Health, Kaohsiung Medical University, Kaohsiung City, Taiwan
[13]Department of Public Health, China Medical University, Taichung City, Taiwan

Corresponding Author:
Chung-Yu Lai, PhD
Graduate Institute of Aerospace and Undersea Medicine
National Defense Medical Center
Rm 8118, No 161, Sec 6, Minquan E Rd
Neihu Dist
Taipei City, 11490
Taiwan
Phone: 886 287923100 ext 19066
Email: multi0912@gmail.com

Abstract

Background: During flight, G force compels blood to stay in leg muscles and reduces blood flow to the heart. Cardiovascular responses activated by the autonomic nerve system and strengthened by anti-G straining maneuvers can alleviate the challenges faced during G loading. To our knowledge, no definite cardiac information measured using a mobile health device exists for analyzing G tolerance. However, our previous study developed the cardiac force index (CFI) for analyzing the G tolerance of military aircrew.

Objective: This study used the CFI to verify participants' cardiac performance when walking and obtained a formula for predicting an individual's G tolerance during centrifuge training.

Methods: Participants from an air force aircrew undertook high-G training from January 2020 to December 2022. Their heart rate (HR) in beats per minute and activity level per second were recorded using the wearable BioHarness 3.0 device. The CFI was computed using the following formula: *weight × activity / HR during resting or walking*. Relaxed G tolerance (RGT) and straining G tolerance (SGT) were assessed at a slowly increasing rate of G loading (0.1 G/s) during training. Other demographic factors were included in the multivariate regression to generate a model for predicting G tolerance from the CFI.

Results: A total of 213 eligible trainees from a military aircrew were recruited. The average age was 25.61 (SD 3.66) years, and 13.1% (28/213) of the participants were women. The mean resting CFI and walking CFI (WCFI) were 0.016 (SD 0.001) and 0.141 (SD 0.037) kg × G/beats per minute, respectively. The models for predicting RGT and SGT were as follows: *RGT = 0.066 × age + 0.043 × (WCFI × 100) – 0.037 × height + 0.015 × systolic blood pressure – 0.010 × HR + 7.724* and *SGT = 0.103 × (WCFI × 100) − 0.069 × height + 0.018 × systolic blood pressure + 15.899*. Thus, the WCFI is a positive factor for predicting the RGT and SGT before centrifuge training.

Conclusions: The WCFI is a vital component of the formula for estimating G tolerance prior to training. The WCFI can be used to monitor physiological conditions against G stress.

(JMIR Mhealth Uhealth 2023;11:e48812) doi:10.2196/48812

KEYWORDS

G force; baroreflex; anti-G straining maneuver; G tolerance; cardiac force index; anti-G suit; relaxed G tolerance; straining G tolerance; cardiac force ratio

Introduction

Background

On the Earth's surface, humans are exposed to gravitational forces. The applied acceleration of gravity, 9.8 m/s^2, is defined as 1 Gz (1 G) in the direction from the head to the feet when a person is standing vertically. During flight, changes in speed or direction result in acceleration. The same magnitude of inertial force is generated in the opposite direction of the acceleration. Because fighter aircrafts are highly agile, military pilots experience high levels of G force, which decrease blood pressure and cause massive shifts in and redistributions of bodily fluid, especially during acrobatic combat maneuvers. The cardiovascular system is highly sensitive to G force, and its ability to maintain sufficient cerebral perfusion can be impaired under high-G force. Military pilots commonly experience visual degradation (eg, grayouts and blackouts) due to a decrease in the blood volume entering the retina [1-3]. If the supply of blood to the brain ceases, a military pilot experiences G-induced loss of consciousness (GLOC). In such cases, the pilot loses their ability to manipulate the aircraft, and a catastrophic event may occur.

Baroreflex is a well-known compensatory regulation activated by decreases in the arterial baroreceptor input under G load. The physiological reactions that are usually observed are an elevated heart rate (HR), increased peripheral vessel resistance, and greater cardiac contractility moderated by the autonomic nerve system [4,5]. The anti-G straining maneuver (AGSM) is considered the most crucial technique for increasing the cardiovascular system's ability to withstand G stress [6-8]. Additionally, several studies have found that anthropometric parameters are associated with G tolerance [9-11].

Because no appropriate integrated cardiac parameter exists for monitoring G tolerance, we successfully introduced the cardiac force index (CFI) into the high-G training undergone by military aircrew [12]. Initially, the CFI was monitored using a wearable device and used to predict the running performance of military academy students [13-16]. The CFI consists of 3 factors that are relevant to G tolerance, namely body weight, dynamic changes in acceleration, and HR. The findings of the aforementioned studies revealed that the walking CFI (WCFI)—that is, the CFI while an individual is walking on the ground—was significantly positively correlated with G tolerance, as determined through centrifuge training.

Objective

High-G training is commonly used to assess the G tolerance of military pilots and determine whether they are fit to fly a modern jet. To our best knowledge, almost no studies have designed a model for predicting the G tolerance of aircrew before the training. Our previous study demonstrated that the WCFI calculated using mobile health technology can be used to identify potential factors affecting the ability to withstand G levels.

Followed the former finding, we attempted to develop a mathematical formula for predicting G tolerance on the basis of the CFI, which can be calculated before the beginning of training. Therefore, we can measure cardiac health and detect the low G tolerance of military pilots via mobile wearable devices during daily activity. In the future, we will try to establish the strategy of rapid G-resistance ability assessment by monitoring mobile cardiac data before the flight and to ensure pilots' safety during flight missions.

Methods

Study Design and Participants

This longitudinal, observational study was conducted to evaluate the relationship between the CFI and G tolerance. The acceleration rate was set to 0.1 G/s during training. We also developed a formula for using CFI data to determine the level of G tolerance.

The participants were air force aircrew trainees attending high-G training at the Aviation Physiology Research Laboratory (APRL), Kaohsiung City, Taiwan. The participants were required to undergo medical examinations and meet the standards to be deemed fit for centrifuge training, which was conducted from January 2020 to December 2022.

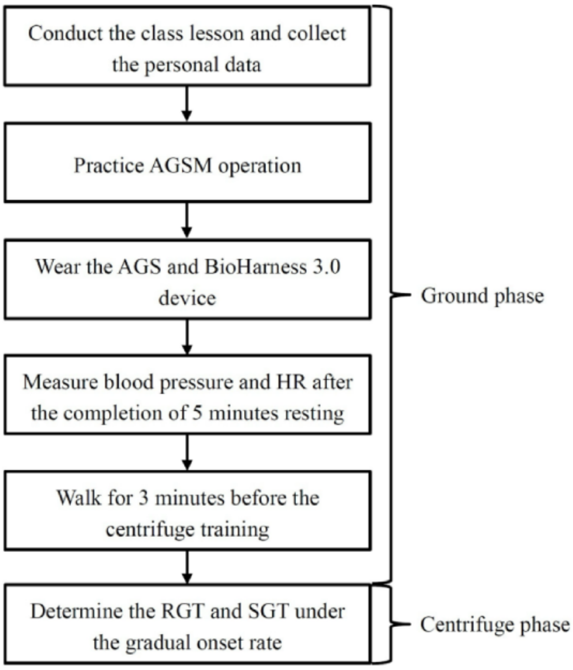
Ethics Approval and Informed Consent

The documents and permission to perform this study were issued by the ethics committee of the Institutional Review Board of Kaohsiung Armed Forces General Hospital in Taiwan (approvals KAFGH 109-001 and 110-009). Before the study, written informed consent was provided by each participant to ensure that they understood the purpose and content of the study.

Protocol of Cardiac Data Collection

Air force aircrew attended a 1-day high-G training at the APRL. Their cardiovascular performance at rest and while walking was monitored using mobile technology and sensors. Centrifuge training was performed to simulate the hypergravitational environment and examine their G tolerance. A flowchart of the study protocol is presented in Figure 1.

Figure 1. Study protocol. AGS: anti-G suit; AGSM: anti-G straining maneuver; HR: heart rate; RGT: relaxed G tolerance; SGT: straining G tolerance.



Mobile Monitoring Device

This study used a mobile health (mHealth) BioHarness 3.0 module (Zephyr Technology Corporation), which is noninvasive and equipped with a gyroscope and accelerometers.

Activity and HR were the 2 key indicators in this study. The BioHarness 3.0 sensor detected the distance that the participants moved by using its internal 3-axis accelerometers and calculated the activity per second. Activity levels were recorded using piezoelectric technology and are presented as the square roots of the acceleration values in the x, y, and z directions.

HR is presented as the number of beats per minute (bpm) and was measured using a conductive electrode sensor, with the thoracic loop strap fitted elastically to the skin over the thorax. To reduce noise during body movement, a shoulder strap was used to minimize the displacement of the BioHarness 3.0 sensor [17].

Ground Phase

The instructor, who was an aviation physiology officer at the APRL, held a lecture on acceleration physiology and G awareness. After the lesson, we explained the study protocol to aircrew willing to participate and used a questionnaire to collect their personal data, namely their birth year, gender, height, and weight; whether they smoked; whether they drank alcohol; and their exercise habits. Thereafter, all aircrew mastered the AGSM, and the instructor examined the participants' execution of the AGSM before G-tolerance tests were performed. The 2 main components of the AGSM are the holding of a preparatory breath against the closed glottis every 3 seconds followed by rapid air exchange and isometric muscle tensing with an emphasis on the legs, buttocks, and abdomen. The trainees wore a standard anti-G suit (AGS) and were outfitted with a BioHarness 3.0 sensor, which was placed under the left central armpit and strapped to the chest and shoulder (Figure 2). After the fit of the AGS was checked, we pressed the central button to power on the BioHarness 3.0 sensor and started collecting cardiac data. After the participants had rested for 5 minutes in a chair, their systolic blood pressure (SBP), diastolic blood pressure (DBP), and HR were evaluated using an Omron 1100U sphygmomanometer (Omron Healthcare Company; Figure 3).

After the resting data had been obtained, the participants performed relaxed and normal walking for 3 minutes. The participants performed squats before and after walking. Walking data could be obtained from the changes in activity identified by the 3-axis accelerometers.

Figure 2. BioHarness 3.0 module with chest and shoulder straps.

Figure 3. Resting for 5 minutes with an uninflated anti-G suit.

Centrifuge Phase

After they had completed the ground phase, the trainees entered the human centrifuge gondola (Latécoère) at the APRL. The human centrifuge trained 1 person at a time, and the maximum training capacity was 8 trainees per day. The length of the centrifuge's arm was 20 feet (6.1 meters). The hydraulic power system could achieve a training onset rate and G level of up to 6 G/s and 9 G, respectively.

The participant wore a safety belt and sat on the simulated cockpit seat, which leaned back by 13°. After the foot pedals had been adjusted, the participant practiced the AGSM again and then rested for 2 minutes in a seated position. The APRL instructor started the centrifuge at an idle run of 1.4 G. Before the trainee's G tolerance was assessed, the instructor accelerated the centrifuge at an onset rate of 0.1 G/s. The participant's relaxed G tolerance (RGT) and straining G tolerance (SGT) were determined without inflating the AGS [11]. The RGT value was defined as the G value at which the participant experienced complete loss of peripheral vision or 50% loss of their central vision in the relaxed state. Thereafter, the participant commenced the AGSM to resist the physiological effect of G force as the level of centrifugation was increased. The SGT value was the G value at which the participant experienced the same visual degradation or a G force equal to the upper limit of 9 G. The level of visual loss for each participant was analyzed using the light bar in front of them inside the gondola [11].

Data Handling and Conversion Procedure

The mHealth BioHarness 3.0 device obtained data every second on the participants while they were on the ground. We used the charging and configuration cradle to download and save the digital data to a folder named "Summary." If the signal values of the HR confidence or system confidence were below 20%, the data were considered unstable and unreliable [18]. There were 219 military aircrew members enrolled in this study. However, 6 participants were excluded from the analysis due to poor data quality, resulting in a final sample size of 213 participants. Resting and walking data were extracted and analyzed using previously proposed research methods [12,13].

Regarding the digital data, the activity and HR variables combined with the individual's body weight were used to calculate the CFI and cardiac force ratio (CFR). For every second for which data were collected, the weight and activity values were multiplied and divided by the HR. The mathematical formula was as follows: $CFI = weight \times activity / HR$ [13,14]. The average resting CFI (RCFI) and WCFI on the ground over a 2-minute period were calculated, and the CFR was obtained by dividing the WCFI by the RCFI.

Statistical Analysis

Descriptive statistics were calculated, and continuous variables are presented as means, SDs, and ranges. We used values and proportions to describe discrete data.

In the statistical analysis, the relationship between cardiac function on the ground and G tolerance in the centrifuge was assessed using Pearson correlation. A model for predicting G tolerance that is connected to the CFI was developed using stepwise multiple linear regression and by adjusting other covariates.

Statistical analyses were conducted using SPSS software (version 27.0; IBM Corp). Two-tailed *P* values <.05 were considered significant.

Results

Analysis of Demographic Data

The demographic data are displayed in Table 1. This study recruited 213 aircrew who finished the study. The average age of the participants was 25.61 (SD 3.66) years, and 13.1% (28/213) of the participants were women. The average height of the participants was 173.18 (SD 6.75) cm, their average weight was 70.39 (SD 11.44) kg, and their average BMI was 23.38 (SD 2.91) kg/m^2. A total of 50 (23.5%) aircrew members smoked, 38 (17.8%) drank alcohol, and over half (n=114, 53.5%) habitually exercised.

Table 1. Characteristics of the enrolled aircrew (n=213).

Characteristic	Value
Age (years), mean (SD; range)	25.61 (3.66; 22-27)
Gender, women, n (%)	28 (13.1)
Height (cm), mean (SD; range)	173.18 (6.75; 156-188)
Weight (kg), mean (SD; range)	70.39 (11.44; 48-99)
BMI (kg/m^2), mean (SD; range)	23.38 (2.91; 17.31-32.70)
Smoking status, n (%)	
No	163 (76.5)
Yes	50 (23.5)
Drinking status, n (%)	
No	175 (82.2)
Yes	38 (17.8)
Habitually exercised, n (%)	
No	99 (46.5)
Yes	114 (53.5)

Physiological Recordings on the Ground or Before Centrifuge Training

The changes in cardiovascular responses are listed in Table 2. The mean SBP, DBP, and HR while sitting and before centrifuge training were 140.40 (SD 14.47) mm Hg, 81.42 (SD 8.33) mm Hg, and 88.56 (SD 15.33) bpm, respectively. The mean WCFI was much higher than the mean RCFI (WCFI: 0.141, SD 0.037 vs RCFI: 0.016, SD 0.001 kg × G/bpm). The average CFR, computed by dividing the WCFI by the RCFI, was 10.76 (SD 4.38). During the G tolerance test, the RGT and SGT were 4.9 (SD 0.9) and 7.9 (SD 1.1) G, respectively, under a slow onset rate. Out of 213 aircrew members, 23 (10.9%) had a RGT greater than 6 G, and 60 (28.2%) had an SGT greater than 8 G (Tables 3 and 4). Pearson correlation coefficients were used to determine the relationship of RGT with SBP (r=.149; P=.03), HR (r=−.187; P=.006), and WCFI (r=.234; P=.001). Additionally, SGT was positively associated with SBP (r=.167; P=.02), DBP (r=.199; P=.01), and WCFI (r=.256; P<.001), as shown in Table 5.

Table 2. Descriptive analysis of cardiovascular and physiological information.

Variables	Value, mean (SD; range)
SBP[a] (mm Hg)	140.40 (14.47; 102-177)
DBP[b] (mm Hg)	81.42 (8.33; 50-107)
HR[c] (bpm[d])	88.56 (15.33; 56-145)
RCFI[e] (kg × G/bpm)	0.016 (0.001; 0.006-0.088)
WCFI[f] (kg × G/bpm)	0.141 (0.037; 0.020-0.266)
CFR[g]	10.76 (4.38; 1.52-23.02)

[a]SBP: systolic blood pressure.
[b]DBP: diastolic blood pressure.
[c]HR: heart rate.
[d]bpm: beats per minute.
[e]RCFI: resting cardiac force index.
[f]WCFI: walking cardiac force index.
[g]CFR: cardiac force ratio.

Table 3. Relaxed G tolerance distribution.

Relaxed G tolerance (G)	Participant (n=213), n (%)
3.0-3.4	8 (3.8)
3.5-3.9	17 (8)
4.0-4.4	40 (18.8)
4.5-4.9	53 (24.9)
5.0-5.4	45 (21.1)
5.5-5.9	27 (12.7)
6.0-6.4	10 (4.7)
6.5-6.9	8 (3.8)
7.0-7.4	3 (1.4)
7.5-7.9	1 (0.5)
8.0-8.4	1 (0.5)

Table 4. Straining G tolerance distribution.

Straining G tolerance (G)	Participant (n=213), n (%)
4.5-4.9	1 (0.5)
5.0-5.4	4 (1.9)
5.5-5.9	9 (4.2)
6.0-6.4	13 (6.1)
6.5-6.9	18 (8.5)
7.0-7.4	19 (8.9)
7.5-7.9	36 (16.9)
8.0-8.4	35 (16.4)
8.5-8.9	18 (8.5)
9.0	60 (28.2)

Table 5. Pearson correlation coefficients between G tolerance and cardiac function.

Variables	RGT[a]	SGT[b]	SBP[c]	DBP[d]	HR[e]	RCFI[f]	WCFI[g]	CFR[h]
RGT								
r	1	.535	.149	.127	–.187	.087	.234	–.025
P value	—	<.001	.03	.07	.006	.20	.001	.72
SGT								
r	.535	1	.167	.199	–.111	.124	.256	–.048
P value	<.001	—	.02	.01	.11	.07	<.001	.49
SBP								
r	.149	.167	1	.519	.181	–.034	.245	.068
P value	.03	.02	—	<.001	.008	.62	<.001	.07
DBP								
r	.127	.199	.519	1	.372	–.020	.091	.060
P value	.07	.01	<.001	—	<.001	.78	.19	.38
HR								
r	–.187	–.111	.181	.372	1	–.310	–.337	.203
P value	.006	.11	.008	<.001	—	<.001	<.001	.003
RCFI								
r	.087	.124	–.034	–.020	–.310	1	.329	–.724
P value	.20	.07	.62	.78	<.001	—	<.001	<.001
WCFI								
r	.234	.256	.245	.091	–.337	.329	1	.177
P value	.001	<.001	<.001	.19	<.001	<.001	—	.009
CFR								
r	–.025	–.048	.068	.060	.203	–.724	.177	1
P value	.72	.49	.07	.38	.003	<.001	.009	—

[a]RGT: relaxed G tolerance.
[b]SGT: straining G tolerance.
[c]SBP: systolic blood pressure.
[d]DBP: diastolic blood pressure.
[e]HR: heart rate.
[f]RCFI: resting cardiac force index.
[g]WCFI: walking cardiac force index.
[h]CFR: cardiac force ratio.

Model for Predicting G Tolerance From the CFI Through Multivariate Linear Regression

As shown in Table 6, the model for predicting G tolerance was established using multivariate linear regression with stepwise selection. The WCFI was found to be the significant parameter for predicting RGT (*P*=.01) and SGT (*P*<.001). The formula for predicting RGT was as follows: *RGT = 0.066 × age + 0.043 × (WCFI × 100) − 0.037 × height + 0.015 × SBP − 0.010 × HR + 7.724*. In the RGT model, each 100-unit increase in the WCFI increased the RGT by 0.043 G (SE 0.015; 95% CI 0.009-0.078). The equation for estimating the SGT was as follows: *SGT = 0.103 × (WCFI × 100) − 0.069 × height + 0.018 × SBP + 15.899*. Thus, the SGT increased by 0.103 G for each 100-unit increase in the WCFI (SE 0.019; 95% CI 0.065-0.141). The findings indicated no significant differences between the observed and estimated value of the RGT (*P*=.49) or SGT (*P*=.80) when the predictive model was used (Table 7).

Table 6. Predictors of relaxed G tolerance (RGT) and straining G tolerance (SGT) in the multivariate linear regression.

Model and variables	β	SE	95% CI	*P* value
RGT model				
Age	0.066	0.015	0.037 to 0.095	<.001
WCFI[a] × 100 (kg × G/bpm[b])	0.043	0.017	0.009 to 0.078	.01
Height (cm)	–0.037	0.009	–0.055 to –0.020	<.001
SBP[c] (mm Hg)	0.015	0.004	0.007 to 0.023	<.001
HR[d] (bpm)	–0.010	0.004	–0.017 to –0.001	.02
Constant	7.724	1.516	4.736 to 10.712	<.001
SGT model				
WCFI × 100 (kg × G/bpm)	0.103	0.019	0.065 to 0.141	<.001
Height (cm)	–0.069	0.011	–0.091 to –0.048	<.001
SBP (mm Hg)	0.018	0.005	0.008 to 0.028	<.001
Constant	15.899	1.700	12.549 to 19.250	<.001

[a]WCFI: walking cardiac force index.

[b]bpm: beats per minute.

[c]SBP: systolic blood pressure.

[d]HR: heart rate.

Table 7. Comparison of the observed and estimated relaxed G tolerance (RGT) and straining G tolerance (SGT) values.

Variable	Estimated value, mean (SD)	Observed value, mean (SD)	*P* value
RGT	4.9 (0.4)	4.9 (0.9)	.49
SGT	7.9 (0.5)	7.9 (1.06)	.80

Discussion

Principal Findings

Several studies have measured HR responses to determine G tolerance [19-21]. We used the mHealth BioHarness device to collect HR data during physical activity performed before centrifuge training. Regarding the CFI values, the results revealed that the WCFI was positively related to G tolerance when the G level was increased at a gradual rate, which was consistent with other studies [12]. Additionally, this study successfully developed a new model for predicting G tolerance on the basis of changes in cardiac function. Age, height, resting blood pressure, and resting HR variables also influenced G tolerance.

Age and Height

We observed that for every 1 extra year of age of the individuals undergoing centrifuge training, their RGT increased by 0.066 G. Older participants had higher G tolerance than younger participants, which was similar to the results of another study [22]. According to Webb et al [11], the RGT and SGT of fighter pilots in the US Air Force were both positively associated with age. In the Korean Air Force, older trainees were more likely to be able to tolerate 6 G exposure profile [8]. Several researchers have also observed that younger aircrew members, those with less flying experience, and those with fewer hours more frequently experience GLOC during flight [23-25].

Park et al [26] suggested that for well-experienced young aviators, age may not affect the frequency of GLOC episodes in centrifuge trials. In one study in the US Navy, Johanson et al [27] revealed that the mean age of those experiencing GLOC was not different from those not experiencing GLOC, which may be linked to past experience, aircraft type, flight maneuver, and situational awareness.

Older jet and fighter pilots often have more years of flying experience. Such pilots are also more frequently exposed to high-G forces during flight. Some evidence indicates that the cardiac performance of fighter pilots is higher after they have been repeatedly exposed to G force [28,29]. This adaptation to G force increases baroreflex activity and G tolerance by altering the G-time tolerance curve [30]. Therefore, our study participants may have had experience in adapting to G force in flight.

Because of greater hydrostatic pressure in taller people, height has been identified as a factor negatively affecting both G tolerance and sustained duration of G force exposure [10,11,31]. In a neutral standing posture, brain-level blood pressure is theoretically approximately 22 mm Hg lower than heart-level blood pressure in a 1 G environment. Thus, a longer distance between the brain and heart might mean lower blood pressure in taller aircrew. In agreement with previous findings, the height of our participants was negatively correlated with their G tolerance in our predictive model.

SBP and HR

The heart ejects blood into cerebral tissue, and BP gradually decreases as blood travels further from the heart. Theoretically, elevated BP is conducive to modulating the effect of G stress. The cardiovascular system can sustain effective cerebral perfusion at up to approximately 4.5 to 5.5 G when the rate of increase is slow. However, the average resting SBP of our participants on the ground was approximately 140 mm Hg, which was slightly higher than usual. This may have been caused by the participants wearing the fitted AGS on their lower body and feeling stressed about their training. In our study, we also discovered that resting SBP was positively associated with RGT and SGT, similar to the US Air Force study that concluded that BP influences G tolerance [11].

In contrast to blood pressure, increased resting HR was disadvantageous for tolerating hypergravity. Our previous report similarly concluded that air force academy student pilots with elevated HR are less likely to tolerate a peak of 7.5 G when sustained for 15 seconds [9]. When arterial pressure and stroke volume drop due to exposure to high-G force, the sympathetic nerves trigger an increase in HR and strengthen cardiovascular function. Exertion levels during exercise can be determined using the maximum HR. The HR response is closely related to sport performance. By subtracting the participant's age from 220, the target HR zones for different activities could be estimated. High-G training is a type of vigorous physical activity, and HR can rise to 160 bpm during G loading [7,9]. Nonetheless, if resting HR was elevated during the pretraining stage, the HR reserve (HRR) would be limited to a narrow range. HRR is one parameter of cardiovascular fitness. Consistent with some reports, we found that trainees with a lower HR or higher HRR were better able to resist the effects of G force [32,33]. This study verified the need to use mobile technology applications for obtaining cardiac data and understanding changes in the G tolerance levels of aviators.

RGT and SGT

At slow acceleration, RGT is mainly determined by BP and baroreflexes. RGT typically ranges from 4.5 to 6 G and varies depending on the individual and the time [34]. When the G force surpasses the RGT, trainees initiate the AGSM to assist their cardiovascular system against the G stress. Inside the centrifuge, visual loss was subjectively assessed using a light bar. To avoid variation between participants, we used a large sample size. Our previous study indicated that the mean RGT and SGT were 5.1 and 7.8 G, respectively [12]. We identified nearly the same RGT and SGT values (RGT: 4.9 G and SGT: 7.9 G) in our sample of 213 participants.

We used the wearable mHealth BioHarness 3.0 device to record cardiovascular function and found that G tolerance was associated with the cardiac data. The CFI is composed of 3 factors, namely body weight, activity, and HR. Our findings indicated that cardiovascular responses on the ground can be used to predict the resistance of z-axis forces during exercise involving mild exertion. Research on the prevention of GLOC may focus on the development of a precaution system based on the CFI. Further monitoring of the CFI during G loading is recommended to track any instantaneous changes in the CFI prior to GLOC.

Until now, there is still no convenient and proper method to monitor the cardiac performance and G tolerance on the ground. Our study showed that the ability for G tolerance could be predicted by the WCFI. Because G tolerance changes every day, therefore, mobile technology combined with a wearable device is highly applicable to calculate the real-time WCFI and predict G tolerance. Military aircrew can directly understand their G tolerance anytime and anywhere by monitoring their cardiac health and performance via a mobile device during their daily activity. Before the flight, they can know their "low-G day" and maintain the good G awareness. Warning mechanisms based on the cardiac health recorded by a mobile device could be considered to develop and prevent in-flight GLOC and catastrophic mishap.

Limitations

This study has some limitations. We included data obtained from women in our analysis, and our results suggest that gender did not have a significant effect on the outcome. However, this result may have been due to the small proportion of women. Second, for the calculation of the WCFI, the participants were asked to walk at their normal speed, but "normal" was subjective and their speed varied. Their HR values during walking were lower than 120 bpm, and the walking activity data covered a narrow range and exhibited a central tendency. Therefore, walking speed variation was unlikely to have significantly affected the outcomes. Although we have collected more data to develop and verify the predictive model, more participants are required to conduct an analysis and perform an external validation. Finally, depending on the airframes they were training on, aircrew had to have reached different levels and profiles relating to high-G training before they could participate in flight training. In this study, all participants met the standards of all the test profiles during training. Therefore, the authors could not clarify the relationship between the pass rate of high-G training and the CFI on the ground.

Conclusions

Using mobile devices, we monitored the cardiac function of aircrew while they walked in a relaxed manner. We verified that the WCFI is positively associated with the level of G tolerance. Moreover, this study developed a model for estimating the G tolerance of military aircrew before they begin high-G training. The development and application of a WCFI-monitoring system for daily life could be considered to evaluate their G tolerance prior to flights.

Acknowledgments

The authors thank the Medical Affairs Bureau, Ministry of National Defense (MND-MAB-C05-111018) and Kaohsiung Armed Forces General Hospital Gangshan Branch, Taiwan (MND-MAB-110-150 and MND-MAB-D-111158), for their support.

Conflicts of Interest

None declared.

References

1. Rickards CA, Newman DG. G-induced visual and cognitive disturbances in a survey of 65 operational fighter pilots. Aviat Space Environ Med 2005 May;76(5):496-500. [Medline: 15892551]
2. Yilmaz U, Cetinguc M, Akin A. Visual symptoms and G-LOC in the operational environment and during centrifuge training of Turkish jet pilots. Aviat Space Environ Med 1999 Jul;70(7):709-712. [Medline: 10417010]
3. Cao XS, Wang YC, Xu L, Yang CB, Wang B, Geng J, et al. Visual symptoms and G-induced loss of consciousness in 594 Chinese Air Force aircrew--a questionnaire survey. Mil Med 2012 Feb;177(2):163-168. [doi: 10.7205/milmed-d-11-00003] [Medline: 22360061]
4. Burton R, Whinnery J. Biodynamics: sustained acceleration. In: DeHart R, Davis J, editors. Fundamentals of Aerospace Medicine. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2002:122-153.
5. Newman DG. The cardiovascular system at +1 Gz. In: Newman DG, editor. High G Flight: Physiological Effects and Countermeasures. 1st ed. England, UK: Ashgate Publishing Limited; 2016:39-55.
6. Kobayashi A, Kikukawa A, Kimura M, Inui T, Miyamoto Y. Cerebral near-infrared spectroscopy to evaluate anti-G straining maneuvers in centrifuge training. Aviat Space Environ Med 2012 Aug 01;83(8):790-794. [doi: 10.3357/asem.3200.2012] [Medline: 22872994]
7. Tu MY, Chu H, Lin YJ, Chiang KT, Chen CM, Chen HH, et al. Combined effect of heart rate responses and the anti-G straining manoeuvre effectiveness on G tolerance in a human centrifuge. Sci Rep 2020 Dec 10;10(1):21611 [FREE Full text] [doi: 10.1038/s41598-020-78687-3] [Medline: 33303828]
8. Yun C, Oh S, Shin YH. AGSM proficiency and depression are associated with success of high-G training in trainee pilots. Aerosp Med Hum Perform 2019 Jul 01;90(7):613-617. [doi: 10.3357/AMHP.5323.2019] [Medline: 31227034]
9. Tu M, Chu H, Chen H, Chiang K, Hu J, Li F, et al. Roles of physiological responses and anthropometric factors on the gravitational force tolerance for occupational hypergravity exposure. Int J Environ Res Public Health 2020 Nov 02;17(21):8061 [FREE Full text] [doi: 10.3390/ijerph17218061] [Medline: 33147694]
10. Park M, Yoo S, Seol H, Kim C, Hong Y. Unpredictability of fighter pilots' G duration tolerance by anthropometric and physiological characteristics. Aerosp Med Hum Perform 2015 Apr;86(4):397-401. [doi: 10.3357/AMHP.4032.2015] [Medline: 25945558]
11. Webb JT, Oakley CJ, Meeker LJ. Unpredictability of fighter pilot G tolerance using anthropometric and physiologic variables. Aviat Space Environ Med 1991 Feb;62(2):128-135. [Medline: 2001208]
12. Chiang KT, Tu MY, Lin YJ, Hsin YH, Chiu YL, Li FL, et al. A cardiac force index applied to the G tolerance test and surveillance among male military aircrew. Int J Environ Res Public Health 2021 Aug 21;18(16):8832 [FREE Full text] [doi: 10.3390/ijerph18168832] [Medline: 34444580]
13. Chu C. Method for detecting cardiac status, method for monitoring cardiac status during exercise, and apparatus for monitoring cardiac status. United States Patent Office US9566010B2. Google Patents. 2016. URL: https://patents.google.com/patent/US20160058314 [accessed 2021-03-23]
14. Chu C. Method for detecting cardiac status, method for monitoring cardiac status during exercise, and apparatus for monitoring cardiac status. Taiwan Patent No. 408 I546051. Google Patents. 2016. URL: https://patents.google.com/patent/TWI546051B/en [accessed 2021-03-24]
15. Hsiao PJ, Chiu CC, Lin KH, Hu FK, Tsai PJ, Wu CT, et al. Usability of wearable devices with a novel cardiac force index for estimating the dynamic cardiac function: observational study. JMIR mHealth uHealth 2020 Jul 21;8(7):e15331 [FREE Full text] [doi: 10.2196/15331] [Medline: 32706725]
16. Wu YS, Wang WY, Chan TC, Chiu YL, Lin HC, Chang YT, et al. Effect of the Nintendo Ring Fit Adventure exergame on running completion time and psychological factors among university students engaging in distance learning during the COVID-19 pandemic: randomized controlled trial. JMIR Serious Games 2022 Mar 22;10(1):e35040 [FREE Full text] [doi: 10.2196/35040] [Medline: 35315780]
17. BioHarness 3.0 user manual. Zephyr Technology Corporation. 2012 Sep 12. URL: https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf [accessed 2023-05-01]
18. Johnstone JA, Ford PA, Hughes G, Watson T, Garrett AT. Bioharness(™) multivariable monitoring device: part. I: validity. J Sports Sci Med 2012 Sep 1;11(3):400-408 [FREE Full text] [Medline: 24149346]
19. Pipraiya M, Tripathi KK, Dogra MM. Effects of +Gz acceleration on indices of heart rate variability. Ind J Aerospace Med 2005 Jun 30;49:37-47 [FREE Full text]
20. Newman DG, White SW, Callister R. Evidence of baroreflex adaptation to repetitive +Gz in fighter pilots. Aviat Space Environ Med 1998 May;69(5):446-451. [Medline: 9591613]
21. Lai C, Tu M, Chu H, Liu C. Cardiac performance of cadets during the centrifuge training with rapid onset rate. Aerosp Med Hum Perform 2019;90:225.

22. Whinnery JE. +Gz tolerance correlation with clinical parameters. Aviat Space Environ Med 1979 Jul;50(7):736-741. [Medline: 486023]
23. Slungaard E, McLeod J, Green NDC, Kiran A, Newham DJ, Harridge SDR. Incidence of G-induced loss of consciousness and almost loss of consciousness in the Royal Air Force. Aerosp Med Hum Perform 2017 Jun 01;88(6):550-555. [doi: 10.3357/AMHP.4752.2017] [Medline: 28539143]
24. Sevilla NL, Gardner JW. G-induced loss of consciousness: case-control study of 78 G-Locs in the F-15, F-16, and A-10. Aviat Space Environ Med 2005 Apr;76(4):370-374. [Medline: 15828637]
25. Green NDC, Ford SA. G-induced loss of consciousness: retrospective survey results from 2259 military aircrew. Aviat Space Environ Med 2006 Jun;77(6):619-623. [Medline: 16780240]
26. Park J, Yun C, Kang S. Physical condition does not affect gravity-induced loss of consciousness during human centrifuge training in well-experienced young aviators. PLoS One 2016 Jan 26;11(1):e0147921 [FREE Full text] [doi: 10.1371/journal.pone.0147921] [Medline: 26812597]
27. Johanson DC, Pheeny HT. A new look at the loss of consciousness experience within the U.S. Naval forces. Aviat Space Environ Med 1988 Jan;59(1):6-8. [Medline: 3355469]
28. Newman DG, White SW, Callister R. Evidence of baroreflex adaptation to repetitive +Gz in fighter pilots. Aviat Space Environ Med 1998 May;69(5):446-451. [Medline: 9591613]
29. Scott JPR, Jungius J, Connolly D, Stevenson AT. Subjective and objective measures of relaxed +Gz tolerance following repeated +Gz exposure. Aviat Space Environ Med 2013 Jul;84(7):684-691. [doi: 10.3357/asem.3550.2013] [Medline: 23855063]
30. Newman DG. Cardiovascular adaptation to acceleration. In: High G Flight: Physiological Effects and Countermeasures. 1st ed. England, UK: Ashgate Publishing Limited; 2016:131-149.
31. Whinnery JE. +Gz-induced loss of consciousness in undergraduate pilot training. Aviat Space Environ Med 1986 Oct;57(10 Pt 1):997-999. [Medline: 3778399]
32. Sundblad P, Kölegård R, Eiken O. G tolerance and the vasoconstrictor reserve. Eur J Appl Physiol 2014 Dec;114(12):2521-2528. [doi: 10.1007/s00421-014-2957-z] [Medline: 25115505]
33. Eiken O, Mekjavic I, Sundblad P, Kölegård R. G tolerance vis-à-vis pressure-distension and pressure-flow relationships of leg arteries. Eur J Appl Physiol 2012 Oct 16;112(10):3619-3627. [doi: 10.1007/s00421-012-2349-1] [Medline: 22350358]
34. G awareness for aircrew. United States Air Force. 2014 Oct 17. URL: https://static.e-publishing.af.mil/production/1/af_a3_5/publication/afpam11-419/afpam11-419.pdf [accessed 2023-04-28]

Abbreviations

AGS: anti-G suit
AGSM: anti-G straining maneuver
APRL: Aviation Physiology Research Laboratory
bpm: beats per minute
CFI: cardiac force index
CFR: cardiac force ratio
DBP: diastolic blood pressure
GLOC: G-induced loss of consciousness
HR: heart rate
HRR: heart rate reserve
mHealth: mobile health
RCFI: resting cardiac force index
RGT: relaxed G tolerance
SBP: systolic blood pressure
SGT: straining G tolerance
WCFI: walking cardiac force index

Edited by L Buis, G Eysenbach; submitted 19.05.23; peer-reviewed by YT Chang, CS Cheng; comments to author 03.07.23; revised version received 08.07.23; accepted 12.07.23; published 26.07.23.

Please cite as:
Kuo MH, Lin YJ, Huang WW, Chiang KT, Tu MY, Chu CM, Lai CY
G Tolerance Prediction Model Using Mobile Device–Measured Cardiac Force Index for Military Aircrew: Observational Study
JMIR Mhealth Uhealth 2023;11:e48812
URL: https://mhealth.jmir.org/2023/1/e48812
doi:10.2196/48812
PMID:37494088

Original Paper

The Use of a Decision Support System (MyFood) to Assess Dietary Intake Among Free-Living Older Adults in Norway: Evaluation Study

Frida Severinsen[1], MSc; Lene Frost Andersen[1], PhD; Mari Mohn Paulsen[1], PhD

Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

Corresponding Author:
Mari Mohn Paulsen, PhD
Department of Nutrition
Institute of Basic Medical Sciences
University of Oslo
Postboks 1046
Blindern
Oslo, 0317
Norway
Phone: 47 95772048
Email: m.m.paulsen@medisin.uio.no

Abstract

Background: The proportion of older adults in the world is constantly increasing, and malnutrition is a common challenge among the older adults aged ≥65 years. This poses a need for better tools to prevent, assess, and treat malnutrition among older adults. MyFood is a decision support system developed with the intention to prevent and treat malnutrition.

Objective: This study aimed to evaluate the ability of the MyFood app to estimate the intake of energy, protein, fluids, and food and beverage items among free-living older adults aged ≥65 years, primarily at an individual level and secondarily at a group level. In addition, the aim was to measure the experiences of free-living older adults using the app.

Methods: Participants were instructed to record their dietary intake in the MyFood app for 4 consecutive days. In addition, each participant completed two 24-hour recalls, which were used as a reference method to evaluate the dietary assessment function in the MyFood app. Differences in the estimations of energy, protein, fluid, and food groups were analyzed at both the individual and group levels, by comparing the recorded intake in MyFood with the 2 corresponding recalls and by comparing the mean of all 4 recording days with the mean of the 2 recalls, respectively. A short, study-specific questionnaire was used to measure the participants' experiences with the app.

Results: This study included 35 free-living older adults residing in Norway. Approximately half of the participants had ≥80% agreement between MyFood and the 24-hour recalls for energy intake on both days. For protein and fluids, approximately 60% of the participants had ≥80% agreement on the first day of comparison. Dinner was the meal with the lowest agreement between the methods, at both the individual and group levels. MyFood tended to underestimate the intake of energy, protein, fluid, and food items at both the individual and group levels. The food groups that achieved the greatest agreement between the 2 methods were eggs, yogurt, self-composed dinner, and hot beverages. All participants found the app easy to use, and 74% (26/35) of the participants reported that the app was easy to navigate.

Conclusions: The results showed that the MyFood app tended to underestimate the participants' dietary intake compared with the 24-hour recalls at both the individual and group levels. The app's ability to estimate intake within food groups was greater for eggs, yogurt, and self-composed dinner than for spreads, mixed meals, vegetables, and snacks. The app was well accepted among the study participants and may be a useful tool among free-living older adults, given that the users are provided follow-up and support in how to record their dietary intake.

(JMIR Mhealth Uhealth 2023;11:e45079) doi:10.2196/45079

KEYWORDS

dietary assessment; malnutrition; eHealth; validation study; older adults; mobile phone

Introduction

Background

Globally, the number of people aged ≥65 years is expected to increase considerably in the coming decades [1,2]. Most of the older adults prefer to stay in their own homes, although they experience various illnesses [3], and home care services may contribute to encouraging or enabling individuals to live in their own homes as long as possible [4]. Malnutrition in terms of undernutrition is a condition associated with increased morbidity and mortality risk, reduced quality of life, longer length of hospital stay, and greater economic costs for the health care sector [5-9]. Among home care recipients, malnutrition, or the risk of malnutrition, is common [10-12].

Guidelines for Nutritional Screening

Guidelines by the European Society for Clinical Nutrition and Metabolism recommend that all older adults should be screened for malnutrition routinely to ensure early identification of risk [13]. According to the European Society for Clinical Nutrition and Metabolism guidelines, individuals found to be malnourished or at risk of malnutrition should receive a comprehensive nutritional assessment and an individualized plan including monitoring and goals for the treatment [13]. With the aim of facilitating dietary assessment, the use of electronic tools in primary health care is emerging, including the use of apps and websites [14].

The MyFood Decision Support System

MyFood is a digital decision support system consisting of an app for dietary recording and automatic evaluation of the recorded dietary intake as well as a web report for health care professionals including tailored recommendations for nutritional treatment and a nutrition care plan for each patient [15,16]. MyFood was initially developed because of the need for a standardized system to prevent and treat disease-related malnutrition among hospitalized patients in Norway. The dietary assessment functionality in the MyFood system has previously been evaluated in a hospital setting [16], but it has not been validated in other health care settings.

Objectives

The primary aim of this study was to evaluate the ability of the dietary assessment function in the MyFood app to estimate individual intake of energy, protein, fluid, and food and beverage items among free-living older adults aged ≥65 years at both the individual and group levels using two 24-hour recalls as a reference method. We also aimed to measure the participants' experiences using the app.

Methods

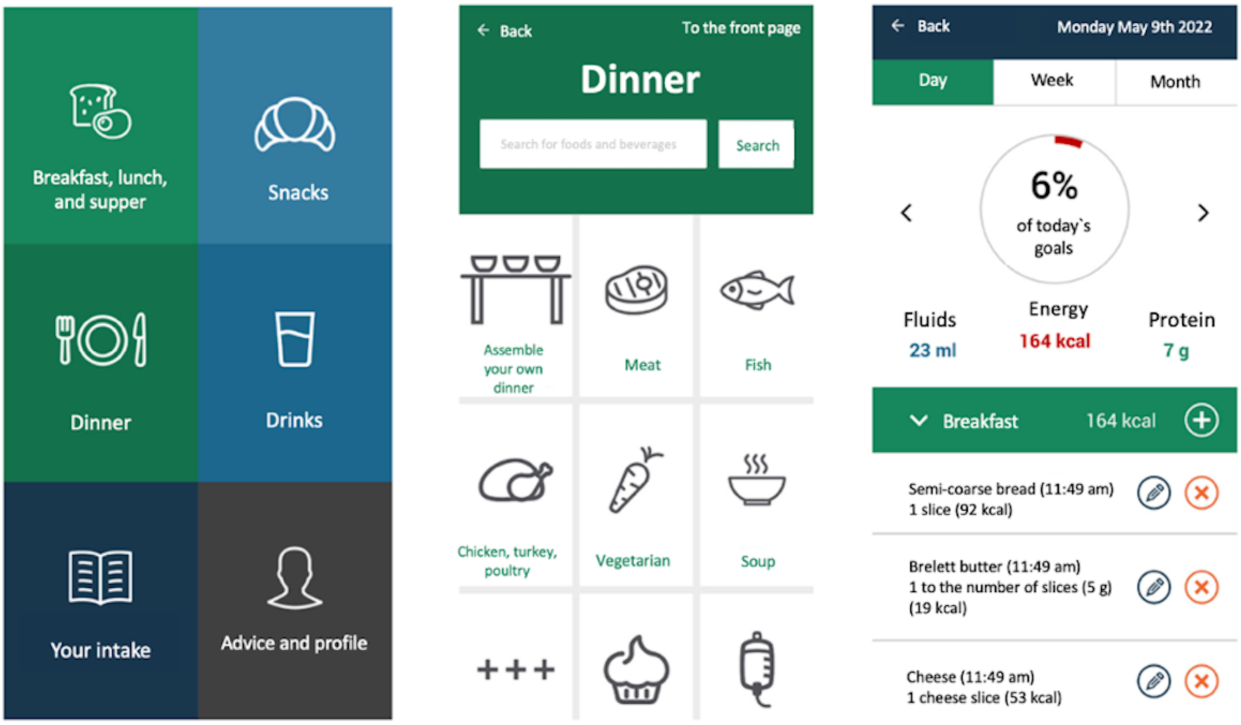
Study Participants and Recruitment

Free-living older adults (aged ≥65 years) were recruited from June 2021 to December 2021 through home care services in a Norwegian municipality, pensioner's associations, and senior centers. In addition, a web page at the University of Oslo was created with an associated registration form for individuals to express their interest in participation. Finally, participants were recruited by combining convenience sampling and snowball sampling. Eligible participants had to be free-living older adults aged ≥65 years and have a Mini-Mental State Examination–Norwegian Revised (MMSE-NR) score >27. Patients who were terminally ill or psychiatric were excluded from the study.

The User Interface of the MyFood System

MyFood is a decision support system developed by researchers at the University of Oslo and Oslo University Hospital in Norway [16]. The MyFood system includes the following four functions: (1) user registration including anthropometric measures, (2) a dietary recording function, (3) automatic evaluation of recorded nutritional intake, and (4) a report to health care professionals including tailored recommendations for measures to improve nutritional status and a template for a nutrition care plan. The user interface of the MyFood system consists of an app including functions 1 to 3 and a website including function 4. Figure 1 illustrates functions 2 and 3.

Figure 1. Screenshots of the MyFood app. From left: (1) main menu of the dietary recording function; (2) menu for recording the dinner meal; and (3) evaluation of intake compared with the estimated requirements for energy, protein, and fluid.



Dietary Recording

Participants were instructed to record their intake in the dietary recording function by selecting 1 of the 5 meal categories (breakfast, lunch, dinner, supper, and snacks). Then, they had to select the correct food or beverage item before recording the amount consumed. Food and beverage items could either be found on the menu or through a free-text search and were illustrated with photographs. When recording dinner intake, the user could choose to select a category of precomposed mixed meals of standardized portions or assemble their own meal using the function *assemble your own dinner* (Figure 1) by selecting all components of the dinner meal manually. During the recording of each meal, the user was presented with prompting questions regarding what proportion of the dish was eaten, whether anything else was eaten with the meal, and whether any beverages or desserts were consumed with the meal.

Data Collection

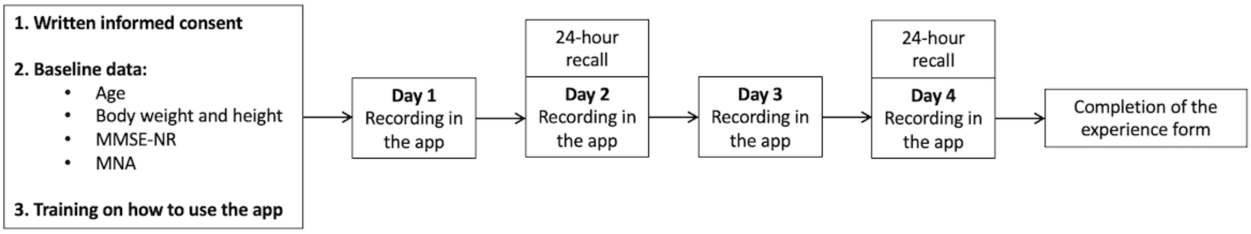
The free-living older adults who were recruited as described in the *Methods* section above were contacted by telephone by a project worker, and a suitable time for a visit was agreed upon. At the visit, the participants received written and oral information about the study and signed a consent form. Information on the participants' age and self-reported height and body weight was retrieved. Participants also completed a Mini Nutritional Assessment (MNA) [17] to assess whether they were at risk of malnutrition and an MMSE-NR [18] to assess whether they had any cognitive impairments that could affect their ability to participate.

All participants were provided guidance on how to download the MyFood app on their personal device, either a tablet or smartphone, except for 2 participants who borrowed tablets available for the project. Then, they were provided with a demonstration of how to use the app. During the demonstration, participants were shown how to navigate the app and how to record their intake of food and beverages.

Study Design

The participants completed a 4-day recording period, during which they were instructed to record their entire intake of foods and beverages in the MyFood app. As a reference method, all participants completed two 24-hour recalls by phone on 2 days overlapping with the recording period. The overlapping days were on the first and the third day of recording for all participants except for 2, as presented in Figure 2, and are further referred to as "comparison day 1" and "comparison day 2."

Figure 2. Study design and data collection. MNA: Mini Nutritional Assessment; MMSE-NR: Mini-Mental State Examination–Norwegian Revised.



The 24-hour recall procedure used a 3-step sequence within an in-house dietary assessment program (KostBeregningsSystem [KBS]) at the University of Oslo, resembling the US Department of Agriculture's Automated Multiple-Pass Method [19]. Each recall lasted for approximately 20 to 30 minutes. All participants received a picture booklet to assist in the estimation of portion sizes during the 24-hour recalls, before the registration period. The booklets contained 41 photo series of 4 pictures, in ascending order, of various household measures, food items, and dishes. All food and beverage items recorded in the MyFood app were categorized into 13 food groups: bread and cereals, spreads, eggs, yogurt, cold beverages, hot beverages, self-composed dinner meals, mixed meals, dessert, fruit, vegetables, snacks, and condiments. Mixed meals included all predefined dinner dishes and dinner components recorded without using the function "assemble your own dinner," and the condiment category included all types of sauces, spices, dressings, etc. The food and beverage items reported in the 24-hour recalls were subsequently allocated to the same categories as the items recorded in the MyFood app for comparison.

Experience Form

At the end of the 4-day registration period, all participants completed an experience form including 5 claims regarding their perceived usability and applicability of the app, using a 5-point Likert scale ranging from "Strongly disagree" to "Strongly agree." The content of the experience form was adapted from the System Usability Scale, which is a 10-question scale based on the 5-point Likert scale that provides information about the perceived usability of a digital system [20]. The experience form used in this study has also been used in a previous evaluation study of the MyFood system in a hospital setting [16].

Sample Size

The sample size estimation was calculated based on the same prerequisites that were used in the previous evaluation study in a hospital setting [21] using a clinically relevant difference between dietary intake recorded in the MyFood app and estimated intake with 24-hour recalls of 50 kcal per day. With a test power of 80%, a significance level of 5%, and a calculated standardized difference of 1, a total of 35 participants were required.

Data Handling and Statistics

The 24-hour recalls were directly coded into KBS (version 7.4), in which estimations of energy, protein, and fluids were performed using the KBS food composition database AE18. Database AE18 is an extended version of the official Norwegian Food Composition Table (version 2018). Dietary information in the MyFood app for energy (kcal), protein (g), and fluid (mL) was based on the Norwegian Food Composition Table from 2019 and the product information from the manufacturer.

The dietary intake recorded in the MyFood app was compared with the intake reported in the 24-hour recalls. Data were analyzed at both the individual and group levels. Evaluation studies are usually performed at the group level to evaluate tools used in different population groups or settings. As the MyFood system was intended to capture dietary intake at an individual level, the main aim of this study was to evaluate its ability to assess individual intake among the free-living older adults. To be able to use MyFood in groups of older adults, analyses were also included to evaluate the accuracy of the dietary recording function at the group level.

At the individual level, dietary intake data from 2 of the 4 days of dietary recording in the MyFood app were compared with the dietary intake data obtained from the two 24-hour recalls on the corresponding days. The differences in the estimated intake of energy, protein, fluid, and selected food groups between the MyFood app and the 24-hour recalls were presented from the 2 overlapping recording days with both methods. The differences were presented in a series of drop plots for comparison days 1 and 2. In addition, the individual-level data of the differences between the 2 methods were analyzed for the breakfast, lunch, dinner, supper, and snack meals separately, for both comparison days 1 and 2 (Multimedia Appendices 1 and 2). Omitted items were counted as an item mentioned in the recalls but not recorded in the MyFood app.

At the group level, the mean intake from the 4 days of dietary recording in the MyFood app was compared with the mean intake obtained from the two 24-hour recalls. The data were presented with mean and SD. The comparison of the mean intake of energy, protein, and fluid between the 2 methods was analyzed using 2-tailed paired samples *t* tests.

Statistical analyses were performed using SPSS Statistics Software (version 28.0; IBM Corp). The level of statistical significance was set at $P<.05$, and all tests were 2 sided.

Ethics Approval and Informed Consent

The study was performed in accordance with the Declaration of Helsinki, and the research protocol was reported to The Norwegian Centre for Research Data (reference number:

135175). Informed verbal and written consent were obtained from all the participants.

Results

Participants

In total, 35 (13 men and 22 women) free-living older adults aged 65-89 years were included in the analyses. The participants had a median age of 71 years and a mean BMI of 25.4 (SD 4.03) kg/m^2. Most of the participants (20/35, 57%) had a normal nutritional status according to the MNA screening. The median MMSE-NR score was 29, ranging from 27 to 30, indicating a good cognitive function among the participants. The characteristics of the study participants are presented in Table 1.

Table 1. Participant characteristics (N=35).

Characteristics	Participants, n (%)
Age (years)	
65-69	12 (34)
70-74	11 (31)
75-79	11 (31)
80-84	0 (0)
≥85	1 (2)
Sex	
Male	13 (37)
Female	22 (63)
BMI (kg/m^2)[a]	
Underweight: <18.5	0 (0)
Normal weight: 18.5-24.9	20 (57)
Overweight: 25-29.9	9 (26)
Obese: ≥30	6 (17)
MNA[b] score	
Malnourished: 0-7	0 (0)
Risk of malnutrition: 8-11	5 (14)
Normal nutritional status: 12-14	30 (86)
MMSE[c] score	
0-26	0 (0)
27-30	35 (100)

[a]Weight (kg)/height (m)^2.

[b]MNA: Mini Nutritional Assessment.

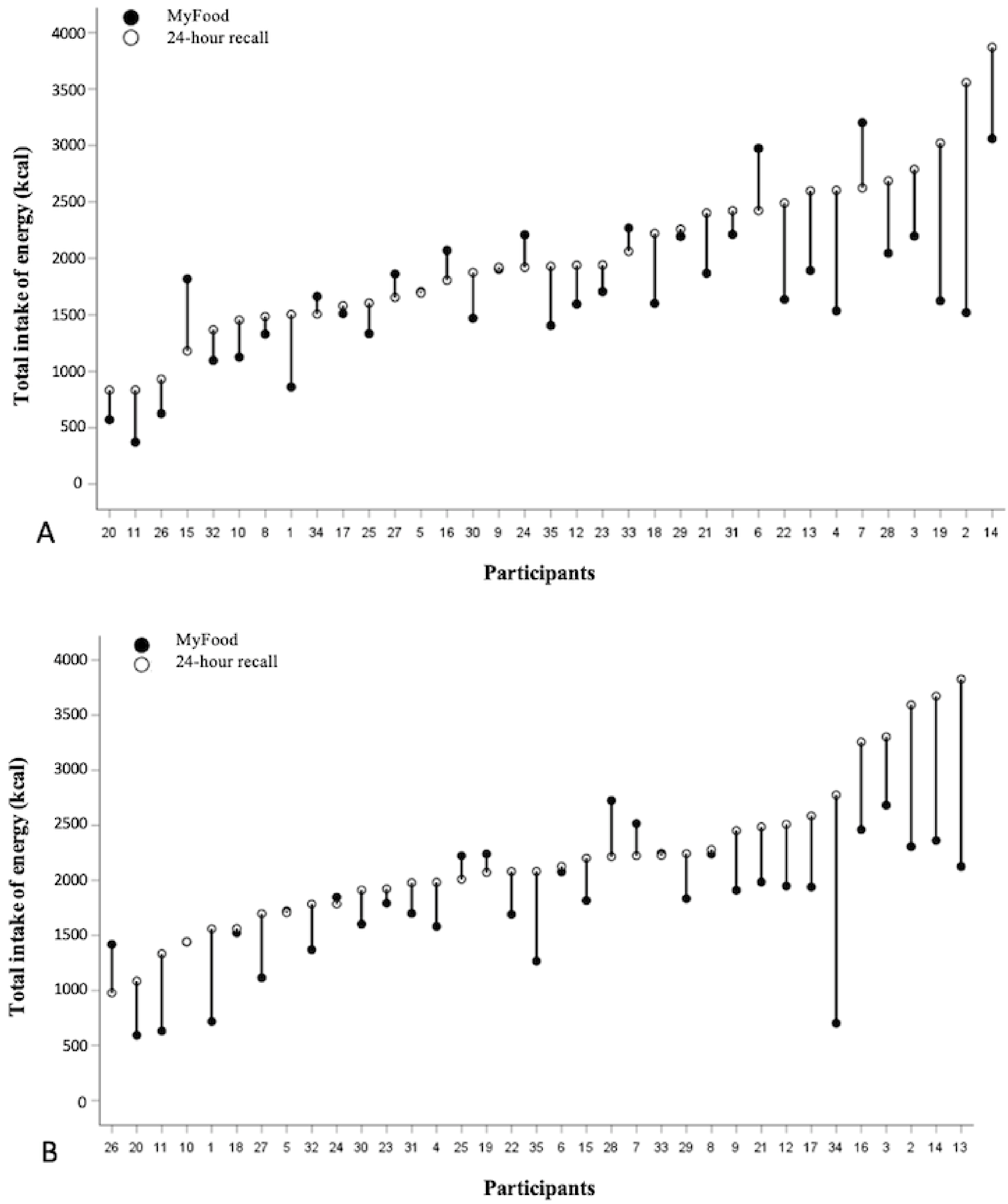
[c]MMSE-NR: Mini-Mental State Examination–Norwegian Revised.

Intake of Energy, Protein, and Fluid at the Individual Level

Energy Intake

Individual drop plots for the total intake of energy on the 2 comparison days are presented in Figure 3, showing the estimated intake in the MyFood app compared with the 24-hour recalls. The MyFood app tended to underestimate the total intake of energy on both comparison days compared with the 24-hour recalls, and the discrepancies tended to increase with increasing intake of energy.

Figure 3. Drop plots illustrating the individual intake of energy on comparison days 1 (section A) and 2 (section B). The y-axis represents the energy intake (kcal). The x-axis represents the participants' ID numbers. In cases where only a white dot is present, the recorded intake in MyFood was identical to that in the 24-hour recall.

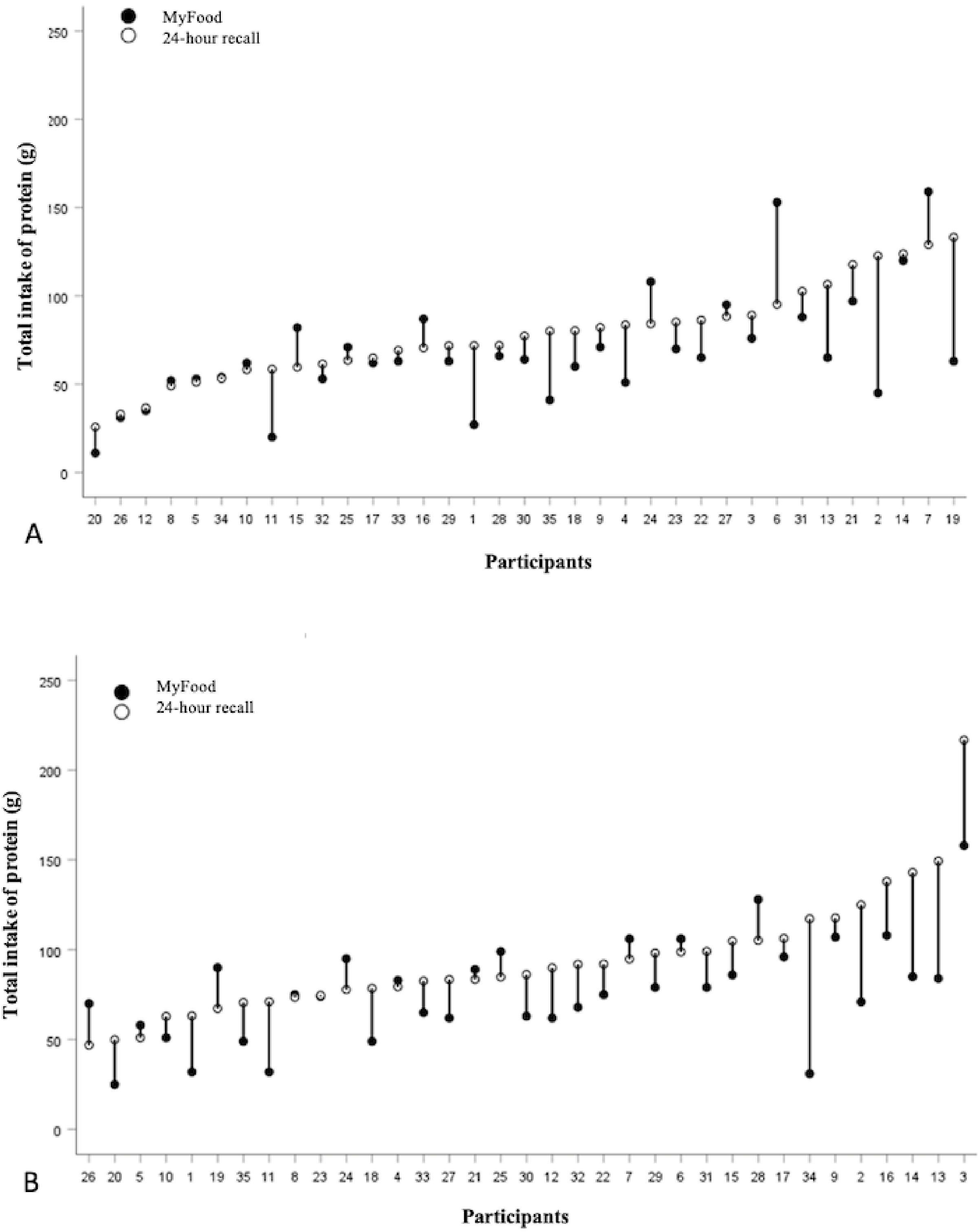


Protein Intake

Individual drop plots for the total intake of protein on the 2 comparison days are presented in Figure 4, showing the estimated intake in the MyFood app and the 24-hour recalls.

As for energy, the MyFood app tended to underestimate the intake of protein compared with the 24-hour recalls on both comparison days. The level of discrepancies between the 2 methods seemed to increase with increasing intake of protein.

Figure 4. Drop plots illustrating the individual intake of protein on comparison days 1 (section A) and 2 (section B). The y-axis represents the protein intake (g). The x-axis represents the participants' ID numbers. In cases where only a white dot is present, the recorded intake in MyFood was identical to that in the 24-hour recall.

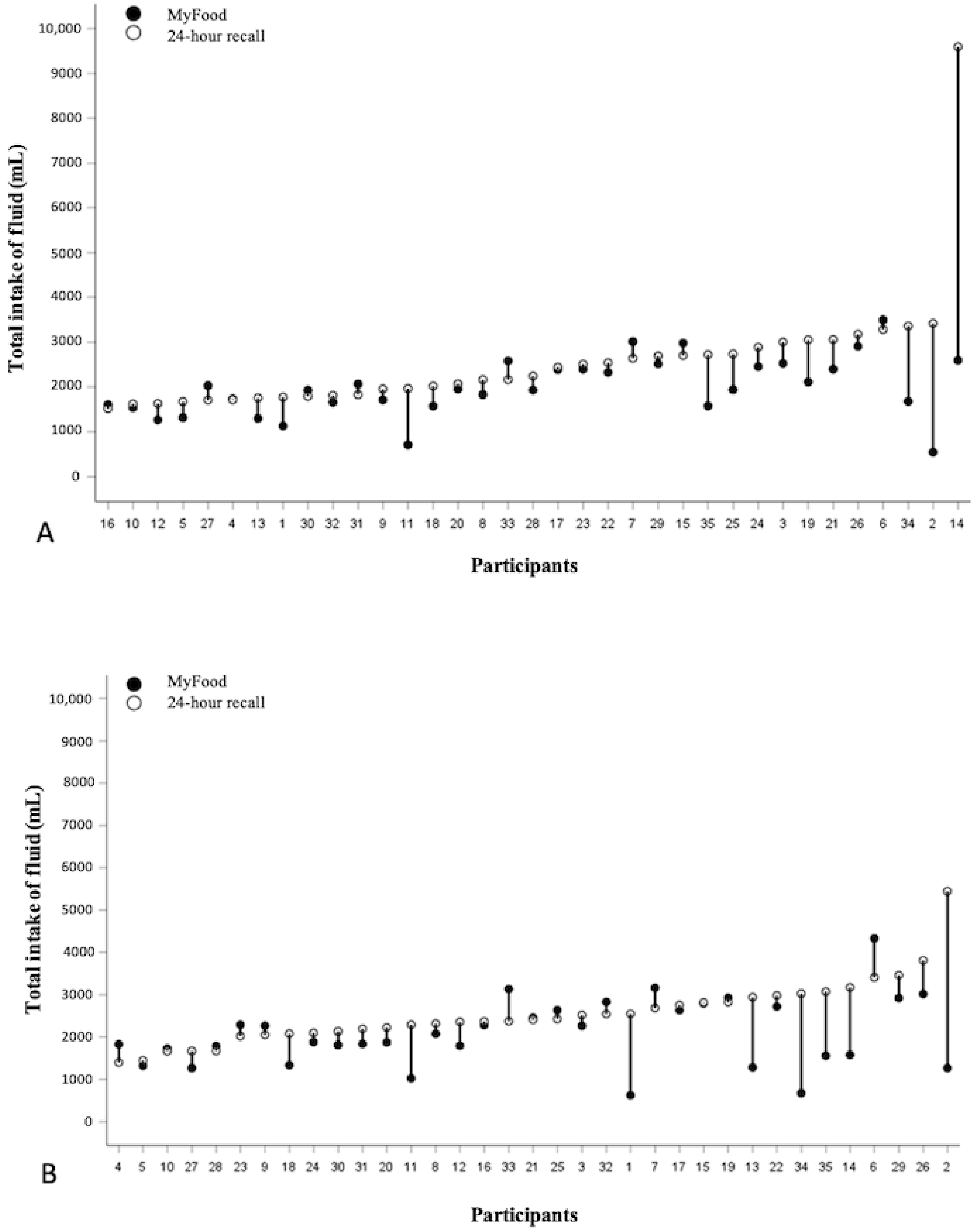


Fluid Intake

Individual drop plots for the total intake of fluids on the 2 comparison days are presented in Figure 5, showing the total intake of fluid estimated in MyFood and the 24-hour recalls.

For most participants, there was a relatively good agreement between the 2 methods. In cases of discrepancies, the MyFood app mainly underestimated the intake of fluids compared with the 24-hour recalls.

Figure 5. Drop plots illustrating the individual intake of fluids on comparison days 1 (section A) and 2 (section B). The y-axis represents the fluid intake (mL). The x-axis represents the participants' ID numbers. In cases where only a white dot is present, the recorded intake in MyFood was identical to that in the 24-hour recall.



An overview of the proportion of the participants having ≥80% agreement between their recordings in the MyFood app and the intake reported in the 24-hour recalls, in total and for each meal separately, on both comparison days, is presented in Multimedia Appendix 3.

On the first and second comparison day, 49% (17/35) and 51% (18/35) of the participants, respectively, had ≥80% agreement for the total intake of energy. For the total intake of fluids, 63% (22/35) and 60% (21/35) of the participants had ≥80% agreement on comparison days 1 and 2, respectively. For protein, 63% (22/35) of the participants had ≥80% agreement on the first comparison day compared with 46% (16/35) of the participants on the second comparison day.

On the first and second comparison day, 53% (18/34) and 54% (19/34) of the participants had ≥80% agreement for the intake of protein for breakfast, respectively. For lunch, the number of participants having ≥80% agreement on protein intake was somewhat lower, with 48% (14/29) on the first comparison day and 45% (14/30) on the second comparison day.

The dinner was the meal with the lowest proportion of participants having ≥80% agreement between the 2 methods for energy intake, with 41% (14/34) of the participants on the first comparison day and 17% (6/35) of the participants on the second comparison day. The dinner was also the meal that most participants did not record in the MyFood app, with 9% (3/33) of the participants on the first day and 17% (5/30) of the participants on the second day.

The meal with the lowest proportion of participants with ≥80% agreement for fluid intake was snacks, with 27% (9/33) participants on the first comparison day and 25% (8/32) participants on the second comparison day.

Intake of Energy, Protein, and Fluid at the Group Level

Table 2 presents the mean (SD) intake of energy, protein, and fluid estimated for the 4 days of dietary recording in the MyFood app and the two 24-hour recalls. At the group level, the participants recorded approximately 17% less energy, protein, and fluids in the MyFood app compared with what was reported in the 24-hour recalls, representing approximately 350 kcal, 15 g, and 400 mL, respectively.

Table 2. Intake of energy, protein, and fluid presented as mean intake estimated for the 4 days of recording dietary intake with the MyFood app and mean intake estimated from the two 24-hour recalls.

	MyFood (N=35), mean (SD)	24-hour recalls (N=35), mean (SD)	*P* value[a]
Energy (kcal)	1733 (527)	2114 (630)	<.001
Protein (g)	72 (25)	86 (25)	<.001
Fluid (mL)	2017 (719)	2450 (611)	<.001

[a]Differences between the estimated intake recorded in MyFood and the 24-hour recalls were tested using the paired sample *t* test.

Intake of Food and Beverage Items at the Individual Level

The proportion of participants having ≥80% agreement between their recordings in the MyFood app and the 24-hour recalls within the different food groups on comparison days 1 and 2 varied between the different food groups (Multimedia Appendix 4). Eggs and yogurt were the food groups with the greatest proportion of participants having ≥80% agreement between the 2 methods, with 69% (9/13) and 94% (15/16) of the participants for eggs and 67% (10/15) and 89% (8/9) of the participants for yogurt on days 1 and 2, respectively. The food group with the lowest proportion of participants with ≥80% agreement was condiments. For dinner, the food group "self-composed dinner" achieved better agreement than the food group "mixed meals.". On the first comparison day, a total of 14 participants used the "assemble your own dinner" function, compared with 9 participants on the second comparison day.

An overview of the omitted food and beverage items is presented in Multimedia Appendix 5. Omitted items were counted as an item mentioned in the recalls but not recorded in the MyFood app. The food groups with the most omissions were cold beverages, condiments, and spreads. Approximately 40 cold beverage items and >20 spreads were reported in the 24-hour recalls but not recorded in the MyFood app.

Participants' Experiences Using the MyFood App

All participants reported that the MyFood app was easy to use (Table 3). In total, 74% (26/35) of participants agreed that the app was easy to navigate, and 83% (29/35) of the participants reported that they managed to record the amount of foods and beverages correctly. Moreover, 9% (3/35) of the participants reported that they had to acquire new knowledge to use the app, and 77% (27/35) of the participants reported that they became more aware of their own nutritional requirements.

Table 3. Participants' user experiences (responses from the experience form; N=35).

	Easy to use, n (%)	Easy to navigate, n (%)	Correct amount of foods and beverages recorded, n (%)	New knowledge was acquired to use the app, n (%)	Increased awareness of their own requirements, n (%)
Totally disagree	0 (0)	0 (0)	0 (0)	24 (69)	0 (0)
Slightly disagree	0 (0)	3 (9)	2 (6)	8 (23)	3 (9)
Neutral	0 (0)	6 (17)	4 (11)	0 (0)	5 (14)
Slightly agree	19 (54)	20 (57)	15 (43)	2 (6)	5 (14)
Totally agree	16 (46)	6 (17)	14 (40)	1 (3)	22 (63)

Discussion

Principal Findings

This study evaluated the dietary assessment function of the MyFood app among free-living older adults aged ≥65 years residing in Norway. MyFood is intended to be used to assess and monitor the nutritional intake of individuals at risk of malnutrition, and the evaluation of individual intake data was therefore of primary interest. This study found that the MyFood app underestimated the dietary intake of food and beverages at both the individual and group levels. At the individual level, there was a variation in the precision of the recordings between the participants, and the level of underestimation tended to increase with increasing intake. The agreement between the MyFood app and the 24-hour recalls for energy, protein, and fluids was higher for breakfast, lunch, and supper than for dinner and snacks. The food groups with the highest agreement between the 2 methods on both comparison days were eggs, yogurt, and self-composed dinner meals, whereas the food groups with the lowest agreement were condiments, vegetables, mixed meals, and cold beverages. All participants found the app easy to use, and most participants (27/35, 77%) experienced that they became more aware of their own nutritional requirements after 4 days of use.

The MyFood App’s Ability to Estimate the Intake of Energy, Protein, and Fluid at the Individual Level

To the best of our knowledge, only a few applications for dietary assessment have been developed for use or evaluated among free-living older adults aged ≥65 years [22-25]. Furthermore, most evaluation studies have been performed at the group level, whereas this study mainly intended to evaluate the use of the MyFood app at the individual level, as the purpose of the app is to monitor the nutritional intake of individuals to provide customized nutritional follow-up. Thus, this study provides novel knowledge to the field of using digital tools for nutritional assessment among the free-living older adults.

The MyFood app underestimated the total intake of energy compared with the 24-hour recalls for most participants. An explanation may be that several participants only recorded part of their intake in the app, compared with what they reported in the recalls, possibly because of inaccurate recordings. They may also have forgotten to record their intake in the MyFood app. It has previously been demonstrated that incorrect estimates of portion sizes account for approximately half of the errors in energy intake estimations from dietary records administered using technological devices [26]. During the 24-hour recalls, the participants used a picture booklet to describe their portion sizes, whereas the MyFood app included standardized portion sizes using household measures and illustration photos. In an evaluation study of an app-based food record in Switzerland, Bucher Della Torre et al [27] found that participants tended to choose the app-proposed portions even if their real portions were different. Another possible explanation is the omission of food and beverage items in the MyFood app compared with the 24-hour recalls, such as spreads and cold beverages. This was also seen in the previous evaluation study of the MyFood app among hospitalized patients [16], in which spreads and cold beverages were the food groups with the most omissions. Underreporting of energy intake was also observed in a recent study by Hopstock et al [28], in which a web-based dietary assessment tool was evaluated among Norwegian men and women aged ≥60 years.

The largest discrepancies between the methods in estimated energy intake were found for dinner on both the recording days. This finding is in accordance with observations from the previous evaluation study of the MyFood app [16]. A possible explanation for this may be that some participants forgot to record their dinner in the app or that the predefined meals available in the app did not represent the meals that the participants would eat for dinner, as these meals were adapted to an institutional setting and not tailored for a home setting. We observed that the participants who used the function “assemble your own dinner” (Figure 1) achieved better agreement between the 2 methods in energy intake for dinner than those selecting predefined meals in the app. This was possibly a result of them being forced to manually record all meal items. Thus, they could not lean on prerecorded items, which may explain why the “assemble your own dinner” function achieved greater accuracy than the predefined dinner meals. Although less than half of the participants used this function on each comparison day, with only 14 participants on the first day and 9 on the second day, this knowledge will be used in the future development of the dinner recording function in the MyFood app.

The agreement between the 2 methods for the estimated intake of energy, protein, and fluids was greater for participants with low intakes, with increased deviations observed with higher intakes on both comparison days. This corresponds to previous findings of underestimation of protein and fluids in MyFood in a hospital setting [16]. Other studies have shown that adults tend to underestimate large portion sizes compared with smaller ones [29]. The underestimation of protein in the MyFood app may have been caused by the omission of spreads (Multimedia Appendix 5). As sliced bread with spreads such as cheese and ham is often consumed for breakfast and lunch in Norway, spreads are an important source of protein in the Norwegian diet. Spreads were also found to be one of the food groups most often omitted in a Canadian validation study of an automated web-based 24-hour dietary recall using fully controlled feeding studies as the reference method [29]. Fluid intake was also underestimated in the MyFood app. This may have been because of the high omission rate for beverages, causing the reported intake of beverages in the recalls to be greater than those recorded in the app. Another possible explanation is the overestimation of fluid intake, as shown by the very high reported intake for some of the participants in the 24-hour recalls. For each of the meals separately, the agreement between the 2 methods for fluids was poor, with snacks being the meal category with the lowest agreement. This may have been because of the drinks not being recorded together with the meal with which they were consumed but rather being recorded as part of the snack category.

For energy, protein, and fluids, there was a tendency for better agreement between the methods on comparison day 1 than on comparison day 2. This contradicts previous studies, including the former evaluation study on MyFood [16], which demonstrated a "learning effect," with an improved agreement on the second recording day compared with the first recording day [30].

MyFood's Ability to Estimate the Intake of Energy, Protein, and Fluid at the Group Level

The estimated mean total intake of energy, protein, and fluid was underestimated in the MyFood app. A recent systematic review and meta-analysis by Zhang et al [31] on dietary assessment apps found that all apps underestimated energy intake compared with their reference methods. Zhang et al [31] argued that conducting 24-hour recalls the day after using the app might cause a memory effect and reduce the extent of underreporting in the recalls compared with recording in the app. Moreover, the availability of feedback and advice in the app may positively affect the 24-hour recalls performed afterward [32]. In this study, the two 24-hour recalls were conducted after recording in the MyFood app asking the participants to report on the exact same days. This may have led to improved memory and precision in the recalls compared with the recordings in the app.

MyFood's Ability to Estimate Intake in Food Groups

The food groups that showed the best agreement between the methods were eggs, yogurt, and self-composed dinners. This may be because eggs and yogurt are presented in standardized units in the app, such as 1 egg or 1 cup of yogurt, which are similar to the units available in the grocery store. We also observed that the meals in which the participants assembled the dishes themselves, that is, breakfast, lunch, and self-composed dinners, achieved greater agreement than the predefined meals, such as mixed meals. For spreads, the agreement between the methods was quite low, which may be a result of spreads being among the food groups with the most omissions, as described in the *Results* section. However, the low agreement between the 2 methods for the spreads in this study may also be because of participants only recording part of the spreads they put on their bread slices or because they had difficulties estimating the correct amount of spreads eaten.

User Experiences

All participants reported that the MyFood app was easy to use and 74% (26/35) reported that it was easy to navigate. In addition, most participants (27/35, 77%) reported that they became more aware of their own nutritional requirements after using the MyFood app. This corresponds with the findings of the previous evaluation study of the MyFood app in a hospital setting [16]. Most participants (32/35, 91%) responded that they did not have to acquire a lot of new knowledge to use the app. This contradicts the findings of a study by Hopstock et al [28], in which they found that about one-third of the participating Norwegian men and women aged 60-74 years experienced that they needed to learn a lot of things before using a digital tool for dietary assessment. However, in the study by Hopstock et al [28], the participants did not receive any guidance on using the tool, in contrast with this study.

Strengths and Limitations

This study evaluated the dietary assessment functionality of the MyFood app among free-living older adults, which is considered an important strength, as most studies investigate the use of apps as dietary assessment tools among younger individuals. In addition to evaluating the dietary assessment functionality of the MyFood app, the participants' experiences with using the app were investigated. Data on the usability of dietary assessment apps among the free-living older adults are scarce.

A limitation of using 24-hour recalls as the reference method is that the 24-hour recalls are prone to error, such as underestimation of intake, which may have affected the basis of comparison, as both the reference method and the test method inhabit the same measurement error [33,34]. The dietary recording functionality of the MyFood app was evaluated in free-living older adults, most of whom did not receive home care services and were not at risk of malnutrition according to the MNA. Thus, the study sample was probably healthier than what may be expected from the general population of free-living older adults aged ≥65 years. Furthermore, many of the participants were still working, and thus, they had a busier everyday life than what many free-living older adults are expected to have. Therefore, we do not know whether the results are representative of the free-living older adult population in Norway. The results indicate that free-living older adults need follow-up to be able to record accurate portion sizes and to avoid omissions in the MyFood app, and future studies should investigate how health care professionals or next-of-kin may be involved in this task.

Conclusions

The MyFood app was evaluated for its ability to estimate the intake of energy, protein, fluid, and food and beverage items among free-living older adults aged ≥65 years residing in Norway. The results showed that the MyFood app underestimated the participants' dietary intake compared with 24-hour recalls as a reference method, both at the individual and group levels. The breakfast and the lunch meals showed better agreement between the methods than the dinner and snack meals. The MyFood app may be a useful tool among free-living older adults; however, the results indicate that the free-living older adults need follow-up and support to accurately report portion sizes and avoid omissions. All participants found the MyFood app easy to use, 74% (26/35) found it easy to navigate, and most participants (27/35, 77%) reported becoming more aware of their nutritional requirements.

Acknowledgments

The MyFood app was developed on Tjenester for Sensitive Data (TSD) facilities; owned by the University of Oslo; and operated and developed by the TSD service group at the University of Oslo, IT Department. The authors thank the participants of this study. This study was funded by UNIFOR (Forvaltningsstiftelse for fond og legater), Freia Chocolade Fabriks Medisinske fond.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

FS, LFA, and MMP designed the research; FS conducted the data collection and performed statistical analysis; FS wrote the paper; and MMP had the primary responsibility for the final content. All authors critically revised the manuscript for important intellectual content and approved the final manuscript.

Conflicts of Interest

LFA and MMP are shareholders in FoodCapture AS, which commercializes the MyFood system. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Drop plots for the intake of energy, protein, and fluids for each meal on comparison day 1.
[DOCX File , 1458 KB - mhealth_v11i1e45079_app1.docx]

Multimedia Appendix 2
Drop plots for the intake of energy, protein, and fluids for each meal on comparison day 2.
[DOCX File , 1470 KB - mhealth_v11i1e45079_app2.docx]

Multimedia Appendix 3
An overview of the proportion of participants having ≥80% agreement between MyFood and the 24-hour recalls for the estimated intake of energy, protein, and fluid in total and for each meal on comparison days 1 and 2. n1: number of consumers on the first comparison day. n2: number of consumers on the second comparison day.
[PNG File , 114 KB - mhealth_v11i1e45079_app3.png]

Multimedia Appendix 4
An overview of participants having ≥80% agreement between MyFood and the 24-hour recalls in the estimated intake of selected food groups on comparison day 1 and day 2. n1: number of consumers comparison day 1. n2: number of consumers comparison day 2.
[PNG File , 116 KB - mhealth_v11i1e45079_app4.png]

Multimedia Appendix 5
Number of omitted food items in each food group for comparison days 1 and 2. The light gray bar represents the number of omissions on the first comparison day, and the dark gray bar represents the number of omissions on the second comparison day. In cases where no bars were present, there were no omissions on that day.
[PNG File , 107 KB - mhealth_v11i1e45079_app5.png]

References

1. World population ageing. United Nations Department of Economic and Social Affairs. 2015. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2021-08-24]
2. World Health Organization. Men, ageing and health : achieving health across the life span. World Health Organization. 2001. URL: https://apps.who.int/iris/handle/10665/66941 [accessed 2021-05-15]
3. World Health Organization. Regional Office for the Eastern Mediterranean. The growing need for home health care for the elderly: home health care for the elderly as an integral part of primary health care services. World Health Organization. Regional Office for the Eastern Mediterranean. 2015. URL: https://apps.who.int/iris/handle/10665/326801 [accessed 2021-05-15]
4. Tomita N, Yoshimura K, Ikegami N. Impact of home and community-based services on hospitalisation and institutionalisation among individuals eligible for long-term care insurance in Japan. BMC Health Serv Res 2010 Dec 22;10:345 [FREE Full text] [doi: 10.1186/1472-6963-10-345] [Medline: 21176165]
5. Söderström L, Rosenblad A, Thors Adolfsson E, Bergkvist L. Malnutrition is associated with increased mortality in older adults regardless of the cause of death. Br J Nutr 2017 Feb;117(4):532-540. [doi: 10.1017/S0007114517000435] [Medline: 28290264]
6. Nigatu YD, Gebreyesus SH, Allard JP, Endris BS. The effect of malnutrition at admission on length of hospital stay among adult patients in developing country: a prospective cohort study. Clin Nutr ESPEN 2021 Feb;41:217-224. [doi: 10.1016/j.clnesp.2020.12.013] [Medline: 33487267]

7. Russell CA. The impact of malnutrition on healthcare costs and economic considerations for the use of oral nutritional supplements. Clin Nutr Suppl 2007 Jan;2(1):25-32. [doi: 10.1016/j.clnu.2007.04.002]
8. Schuetz P, Seres D, Lobo DN, Gomes F, Kaegi-Braun N, Stanga Z. Management of disease-related malnutrition for patients being treated in hospital. Lancet 2021 Nov;398(10314):1927-1938 [FREE Full text] [doi: 10.1016/s0140-6736(21)01451-3]
9. Rasheed S, Woods RT. Malnutrition and quality of life in older people: a systematic review and meta-analysis. Ageing Res Rev 2013 Mar;12(2):561-566 [FREE Full text] [doi: 10.1016/j.arr.2012.11.003] [Medline: 23228882]
10. Tsuji T, Yamamoto K, Yamasaki K, Hayashi F, Momoki C, Yasui Y, et al. Lower dietary variety is a relevant factor for malnutrition in older Japanese home-care recipients: a cross-sectional study. BMC Geriatr 2019 Jul 26;19(1):197 [FREE Full text] [doi: 10.1186/s12877-019-1206-z] [Medline: 31349800]
11. Kaipainen T, Tiihonen M, Hartikainen S, Nykänen I. Prevalence of risk of malnutrition and associated factors in home care clients. J Nurs Home Res Sci 2015;1:47-51. [doi: 10.14283/jnhrs.2015.9]
12. Saletti A, Johansson L, Yifter-Lindgren E, Wissing U, Osterberg K, Cederholm T. Nutritional status and a 3-year follow-up in elderly receiving support at home. Gerontology 2005;51(3):192-198. [doi: 10.1159/000083993] [Medline: 15832047]
13. Volkert D, Beck AM, Cederholm T, Cruz-Jentoft A, Goisser S, Hooper L, et al. ESPEN guideline on clinical nutrition and hydration in geriatrics. Clin Nutr 2019 Feb;38(1):10-47. [doi: 10.1016/j.clnu.2018.05.024] [Medline: 30005900]
14. Bonilla C, Brauer P, Royall D, Keller H, Hanning RM, DiCenso A. Use of electronic dietary assessment tools in primary care: an interdisciplinary perspective. BMC Med Inform Decis Mak 2015 Feb 25;15(1):14 [FREE Full text] [doi: 10.1186/s12911-015-0138-6] [Medline: 25886381]
15. Paulsen MM, Paur I, Gjestland J, Henriksen C, Varsi C, Tangvik RJ, et al. Effects of using the MyFood decision support system on hospitalized patients' nutritional status and treatment: a randomized controlled trial. Clin Nutr 2020 Dec;39(12):3607-3617 [FREE Full text] [doi: 10.1016/j.clnu.2020.03.012] [Medline: 32241711]
16. Paulsen MM, Hagen ML, Frøyen MH, Foss-Pedersen RJ, Bergsager D, Tangvik RJ, et al. A dietary assessment app for hospitalized patients at nutritional risk: development and evaluation of the MyFood app. JMIR Mhealth Uhealth 2018 Sep 07;6(9):e175 [FREE Full text] [doi: 10.2196/mhealth.9953] [Medline: 30194059]
17. Vellas B, Guigoz Y, Garry PJ, Nourhashemi F, Bennahum D, Lauque S, et al. The Mini Nutritional Assessment (MNA) and its use in grading the nutritional state of elderly patients. Nutrition 1999 Feb;15(2):116-122. [doi: 10.1016/s0899-9007(98)00171-3] [Medline: 9990575]
18. Strobel C, Engedal K. MMSE-NR Norsk Revidert Mini Mental Status Evaluering - Revidert og utvidet manual. The Norwegian National Centre for Ageing and Health. 2008. URL: https://www.aldringoghelse.no/ah-archive/documents/Manual_MMSE-NR_uGiHcxi.pdf [accessed 2021-08-13]
19. Raper N, Perloff B, Ingwersen L, Steinfeldt L, Anand J. An overview of USDA's dietary intake data system. J Food Compost Anal 2004 Jun;17(3-4):545-555 [FREE Full text] [doi: 10.1016/j.jfca.2004.02.013]
20. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996.
21. Paulsen MM. Development and evaluation of a decision support system to prevent and treat disease-related malnutrition. University of Oslo. 2019. URL: https://www.duo.uio.no/bitstream/handle/10852/77479/PhD-Paulsen-2020.pdf?sequence=1&isAllowed=y [accessed 2021-12-03]
22. Jung H, Demiris G, Tarczy-Hornoch P, Zachry M. A novel food record app for dietary assessments among older adults with type 2 diabetes: development and usability study. JMIR Form Res 2021 Feb 17;5(2):e14760 [FREE Full text] [doi: 10.2196/14760] [Medline: 33493129]
23. Ji Y, Plourde H, Bouzo V, Kilgour RD, Cohen TR. Validity and usability of a smartphone image-based dietary assessment app compared to 3-day food diaries in assessing dietary intake among Canadian adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 09;8(9):e16953 [FREE Full text] [doi: 10.2196/16953] [Medline: 32902389]
24. Ocké M, Dinnissen C, Stafleu A, de Vries J, van Rossum C. Relative validity of MijnEetmeter: a food diary app for self-monitoring of dietary intake. Nutrients 2021 Mar 30;13(4):1135 [FREE Full text] [doi: 10.3390/nu13041135] [Medline: 33808209]
25. Aure CF, Kluge A, Moen A. Promoting dietary awareness: home-dwelling older adults' perspectives on using a nutrition application. Int J Older People Nurs 2020 Dec;15(4):e12332 [FREE Full text] [doi: 10.1111/opn.12332] [Medline: 32597570]
26. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. Nutrition 2005 Jun;21(6):672-677. [doi: 10.1016/j.nut.2004.11.006] [Medline: 15925290]
27. Bucher Della Torre S, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1):76 [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
28. Hopstock LA, Medin AC, Skeie G, Henriksen A, Lundblad MW. Evaluation of a web-based dietary assessment tool (myfood24) in Norwegian women and men aged 60-74 years: usability study. JMIR Form Res 2022 Mar 11;6(3):e35092 [FREE Full text] [doi: 10.2196/35092] [Medline: 35275079]
29. Lafrenière J, Lamarche B, Laramée C, Robitaille J, Lemieux S. Validation of a newly automated web-based 24-hour dietary recall using fully controlled feeding studies. BMC Nutr 2017 Apr 5;3(1):34 [FREE Full text] [doi: 10.1186/s40795-017-0153-3] [Medline: 32153814]

30. Faggiano F, Vineis P, Cravanzola D, Pisani P, Xompero G, Riboli E, et al. Validation of a method for the estimation of food portion size. Epidemiology 1992 Jul;3(4):379-382. [doi: 10.1097/00001648-199207000-00015] [Medline: 1637903]
31. Zhang L, Misir A, Boshuizen H, Ocké M. A systematic review and meta-analysis of validation studies performed on dietary record apps. Adv Nutr 2021 Dec 01;12(6):2321-2332 [FREE Full text] [doi: 10.1093/advances/nmab058] [Medline: 34019624]
32. Forster H, Walsh MC, Gibney MJ, Brennan L, Gibney ER. Personalised nutrition: the role of new dietary assessment methods. Proc Nutr Soc 2016 Feb;75(1):96-105. [doi: 10.1017/S0029665115002086] [Medline: 26032731]
33. Jonnalagadda SS, Mitchell DC, Smiciklas-Wright H, Meaker KB, Van Heel N, Karmally W, et al. Accuracy of energy intake data estimated by a multiple-pass, 24-hour dietary recall technique. J Am Diet Assoc 2000 Mar;100(3):303-8; quiz 309. [doi: 10.1016/s0002-8223(00)00095-x] [Medline: 10719403]
34. Tran KM, Johnson RK, Soultanakis RP, Matthews DE. In-person vs telephone-administered multiple-pass 24-hour recalls in women: validation with doubly labeled water. J Am Diet Assoc 2000 Jul;100(7):777-783. [doi: 10.1016/S0002-8223(00)00227-3] [Medline: 10916515]

Abbreviations

KBS: KostBeregningsSystem
MMSE-NR: Mini-Mental State Examination–Norwegian Revised
MNA: Mini Nutritional Assessment

Edited by L Buis; submitted 19.12.22; peer-reviewed by N Pahlavani, N Mustafa; comments to author 10.04.23; revised version received 27.04.23; accepted 05.06.23; published 03.08.23.

Please cite as:
Severinsen F, Andersen LF, Paulsen MM
The Use of a Decision Support System (MyFood) to Assess Dietary Intake Among Free-Living Older Adults in Norway: Evaluation Study
JMIR Mhealth Uhealth 2023;11:e45079
URL: https://mhealth.jmir.org/2023/1/e45079
doi:10.2196/45079
PMID:37535420

Consumers' Preferences for Purchasing mHealth Apps: Discrete Choice Experiment

Zhenzhen Xie, PhD; Calvin Kalun Or, PhD

Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Calvin Kalun Or, PhD

Abstract

Background: There is growing interest in mobile health apps; however, not all of them have been successful. The most common issue has been users' nonadoption or abandonment of health apps because the app designs do not meet their preferences. Therefore, to facilitate design-preference fit, understanding consumers' preferences for health apps is necessary, which can be accomplished by using a discrete choice experiment.

Objective: This study aims to examine consumer preferences for health apps and how these preferences differ across individuals with different sociodemographic characteristics and health app usage and purchase experiences.

Methods: A cross-sectional discrete choice experiment questionnaire survey was conducted with 593 adults living in Hong Kong. A total of 7 health app attributes that might affect consumers' preferences for health apps were examined, including usefulness, ease of use, security and privacy, health care professionals' attitudes, smartphone storage consumption, mobile data consumption, and cost. Mixed-effect logit regressions were used to examine how these attributes affected consumer preferences for health apps. Fixed effects (coefficient β) of the attributes and random effects of individual differences were modeled. Subgroup analyses of consumer preferences by sex, age, household income, education level, and health app usage and purchase experiences were conducted.

Results: Cost was the attribute that had the greatest effect on consumers' choice of health apps (compared to HK $10 [US $1.27]—HK $50 [US $6.37]: β=−1.064; *P*<.001; HK $100 [US $12.75]: β=−2.053; *P*<.001), followed by security and privacy (compared to *no security insurance—some security policies*: β=.782; *P*<.001; *complete security system*: β=1.164; *P*<.001) and usefulness (compared to *slightly useful—moderately useful*: β=.234; *P*<.001; *very useful*: β=.979; *P*=.007), mobile data consumption (compared to *data-consuming—a bit data-consuming*: β=.647; *P*<.001; *data-saving*: β=.815; *P*<.001), smartphone storage consumption (compared to >100 MB—around 38 MB: β=.334; *P*<.001; <10 MB: β=.511; *P*<.001), and attitudes of health care professionals (compared to *neutral—moderately supportive*: β=.301; *P*<.001; *very supportive*: β=.324; *P*<.001). In terms of ease of use, consumers preferred health apps that were moderately easy to use (compared to *not easy to use—moderately easy to use*: β=.761; *P*<.001; *very easy to use*: β=.690; *P*<.001). Our results also showed that consumers with different sociodemographic characteristics and different usage and purchase experiences with health apps differed in their preferences for health apps.

Conclusions: It is recommended that future health apps keep their mobile data and phone storage consumption low, include a complete security system to protect personal health information, provide useful content and features, adopt user-friendly interfaces, and involve health care professionals. In addition, health app developers should identify the characteristics of their intended users and design and develop health apps to fit the preferences of the intended users.

(JMIR Mhealth Uhealth 2023;11:e25908) doi:10.2196/25908

KEYWORDS

consumer preferences; discrete choice experiment; DCE; mobile health; mHealth; digital health

Introduction

With the rising prevalence of mobile health (mHealth) apps [1-3] and the accumulating evidence of their effectiveness in improving health outcomes [4,5], interest in developing health apps has continued to grow. mHealth apps have been applied to support health care activities, such as disease detection, patient monitoring, health data collection, remote diagnosis, and disease management [6-10]. However, the implementation of health apps is not always easy, and several attempts have not achieved the desired results [11,12]. The most common issue has been the nonadoption or abandonment of health apps by consumers [13], indicating a gap between the health app and consumers' preferences and highlighting the need to understand which characteristics of health apps affect consumer preferences for health apps [14,15].

To obtain such knowledge, we can use a discrete choice experiment (DCE), which is a research technique that elicits consumers' stated preferences for products or services and assesses the contribution of various characteristics (ie, the attributes of the products or services) to those preferences [16-18]. In a DCE, researchers predetermine the attributes that may potentially affect consumer choices and the levels for each attribute of the target product or service. Researchers then create hypothetical alternatives of the product or service by combining the different levels of those attributes. Finally, researchers ask participants about which alternatives they would be willing to purchase; this information reflects their preferences for the alternatives. The participants' responses can be used to derive information regarding how consumer preferences are affected by each of the attributes [19]. With such information, product and service designers and developers would become aware of the attributes that consumers care about, allowing them to focus more on the influential attributes for better product and service designs.

DCEs have been applied to understand consumer preferences for different health technologies, including telehealth systems and appointment reminder systems [20-22]. These studies have demonstrated that DCEs can generate useful information from consumers' perspectives for guiding health technology design. However, among the research related to health apps, the use of DCEs to examine consumer preferences remains scarce, and little is known about which attributes of health apps should be prioritized during the development of new health apps. Therefore, in this study, we aimed to use a DCE to examine consumer preferences for health apps. In addition, as consumers with different sociodemographic characteristics and health app usage levels may have different preferences toward health apps, we also aimed to examine consumer preferences for health apps across individuals with different sociodemographic characteristics, health app usage experiences, and health app purchase experiences.

Methods

Questionnaire Development

A questionnaire was developed to collect data on participants' sociodemographics (sex, age, district of residence, household size, household monthly income, and education level), usage of health apps, previous purchases of health apps, and preferences for each attribute of health apps. Usage of health apps was assessed by asking participants to indicate whether they had health apps installed on their smartphone. Previous purchases of health apps were measured by asking participants whether they had paid for health apps before. Consumer preferences were assessed by using a set of DCE questions, which were developed as described below.

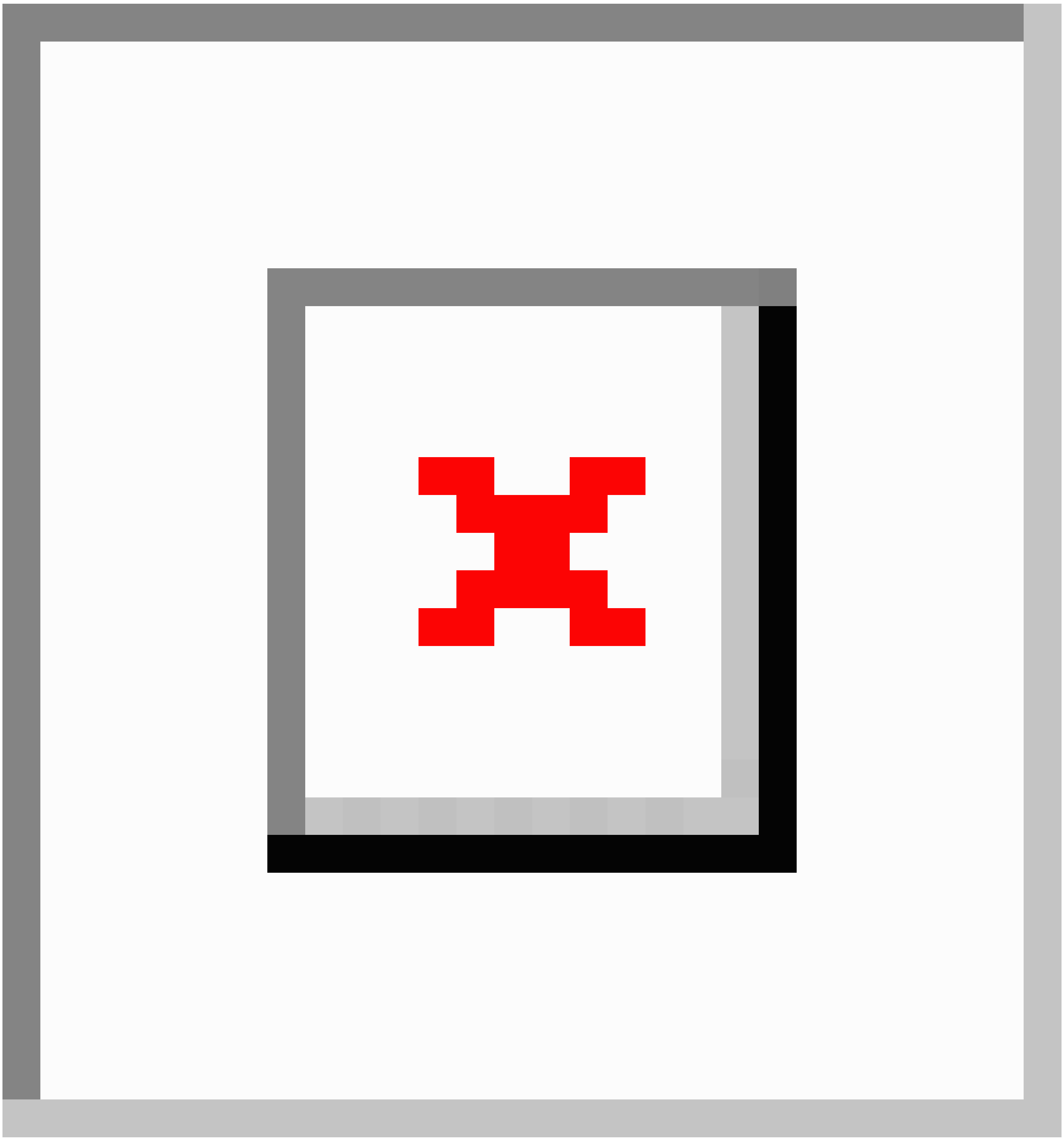
We carefully reviewed the literature on the factors that affect individuals' decision to use health apps [23-38] in order to identify health app attributes that may contribute to consumers' preferences for health apps. We then consolidated these attributes and assigned 3 levels to each one (Table 1). Based on these attributes and levels, 18 hypothetical health apps were formed, using an orthogonal factorial design. For each hypothetical health app, we created a question that asked whether participants would be willing to purchase the app. Prior to answering these questions, participants were asked to answer 2 example questions (Figure 1) to familiarize themselves with the format of the questions. These example questions were also used to test the validity of participants' answers; as health app A in example 1 is superior to health app B in example 2 for every attribute, participants who refused to purchase health app A but chose to purchase health app B may have not fully understood the questions, and their responses were excluded from data analysis.

Table . Attributes and levels in the discrete choice experiment.

Attributes and levels		Descriptions
Usefulness		
	Slightly useful	This health app seems slightly useful to you.
	Moderately useful	This health app seems moderately useful to you.
	Very useful	This health app seems very useful to you.
Ease of use		
	Not easy to use	This health app does not seem very easy to use. You would need to spend much time and effort to learn to use it.
	Moderately easy to use	This health app seems moderately easy to use. You could learn to use it quickly.
	Very easy to use	This health app seems very easy to use. You would be able to use the app immediately without any tutorial or help.
Security and privacy		
	No security assurance	This health app offers no information about protection of personal health information.
	Some security assurance	This health app provides some information about security policies related to personal health information.
	Complete security system	This health app has a complete security system to protect your personal health information.
Health care professional's attitude		
	Neutral attitude	A health care professional who you trust has a neutral attitude about your use of this health app.
	Moderately supportive	A health care professional who you trust is moderately supportive of your use of this health app.
	Very supportive	A health care professional who you trust is very supportive of your use of this health app.
Smartphone storage consumption (MB)		
	>100	This health app is large (>100 MB).
	Around 38	This health app is medium (around 38 MB).
	<10	This health app is small (<10 MB).
Mobile internet data consumption		
	Quite data-consuming	Internet connection is a must for this health app. It is quite data-consuming.
	A bit data-consuming	Some functions of this health app require an internet connection. It is a bit data-consuming.
	Quite data-saving	This health app can be used offline. It is quite data-saving.
Cost (HK $[a])		
	10	The cost of this health app is HK $10.
	50	The cost of this health app is HK $50.
	100	The cost of this health app is HK $100.

[a]A currency exchange rate of HK $1=US $0.13 is applicable.

Figure 1. Two examples of the DCE question designed to familiarize the participants with the format of the questions and test whether participants understood the DCE questions. A currency exchange rate of HK $1=US $0.13 is applicable. DCE: discrete choice experiment.



Data Collection

The questionnaire was distributed to adults (aged ≥18 y) residing in Hong Kong, and participant recruitment was stratified by age, sex, and district of residence according to the population distribution [39]. Trained research assistants approached potential participants in public places, explaining this study and determining their eligibility. Individuals were eligible if they knew what health apps were and were able to understand and answer the questionnaire. Eligible participants were asked to provide written informed consent and then answer the questionnaire. To ensure that all participants had the same understanding of what health apps were, research assistants provided the following definition of *health apps* to each participant: health apps are software programs installed on smartphones that aim to assist health management. Research assistants also provided examples of health apps, including those that track health indicators, send health reminders, present health information, promote self-management behaviors and healthy lifestyles, and provide remote monitoring and diagnosis. Upon finishing the survey, each participant was given a supermarket coupon for HK $50 (US $6.37).

Data Analysis

Descriptive statistics were used to reflect the study sample's sociodemographic characteristics and their health app usage and purchase experiences. A mixed-effect logit model was used to examine consumer preferences for health apps by examining how each attribute of a health app affects whether participants would choose to purchase the health app. Fixed effects of the seven attributes and random effects of individual differences were modeled. Coefficients of the fixed effects (β), their SEs, and corresponding *P* values were reported. In addition, subgroup analyses of consumer preferences by sex, age, household income, education level, and health app usage and purchase experiences were conducted. All statistical analyses were performed by using R 4.0.2 software (R Foundation for Statistical Computing).

Ethical Considerations

This study received ethical approval from the Human Research Ethics Committee of the University of Hong Kong (approval number: EA1810020).

Results

Sample Characteristics

Among the 600 individuals who completed the DCE, 7 did not fully understand the discrete choice question, judging from their responses to the two example questions of the DCE. Therefore, the study sample consisted of 593 adults; 47.2% (280/593) of them had health apps installed on smartphones, and 10.5% (62/593) had paid for a health app before. Sociodemographic characteristics of the study sample are presented in Table 2.

Table . Sociodemographic characteristics of the study sample (N=593).

Characteristics		Values
Age (years), mean (SD)		45.96 (15.85)
Sex, n (%)		
	Female	318 (53.6)
	Male	275 (46.4)
Monthly household income (HK $[a]), n (%)		
	<10,000	47 (7.9)
	10,000-19,999	122 (20.6)
	20,000-29,999	113 (19.1)
	30,000-39,999	68 (11.5)
	40,000-49,999	81 (13.7)
	50,000-79,999	86 (14.5)
	≥80,000	76 (12.8)
Education level, n (%)		
	Primary school or below	64 (10.8)
	Some secondary school or completed secondary school	222 (37.4)
	Postsecondary degree (diploma, bachelor, master, or doctoral degree)	307 (51.8)

[a]A currency exchange rate of HK $1=US $0.13 is applicable.

Consumer Preferences for Health Apps

Table 3 presents the results of the mixed logit regression that examined how each attribute of a health app affects whether participants would be willing to purchase the health app.

Table . Mixed logit regression that examined how each attribute of a health app affects whether participants would be willing to purchase the health app.

Attributes and levels		β (SE)	*P* value
Usefulness (reference level: slightly useful)			
	Moderately useful	.234 (0.086)	<.001
	Very useful	.979 (0.081)	.007
Ease of use (reference level: not easy to use)			
	Moderately easy to use	.761 (0.090)	<.001
	Very easy to use	.690 (0.080)	<.001
Security and privacy (reference level: no security assurance)			
	Some security policies	.782 (0.082)	<.001
	Complete security system	1.164 (0.084)	<.001
Health care professional's attitude (reference level: neutral)			
	Moderately supportive	.301 (0.082)	<.001
	Very supportive	.324 (0.081)	<.001
Smartphone storage consumption (MB; reference level: >100)			
	Around 38	.334 (0.082)	<.001
	<10	.511 (0.081)	<.001
Mobile data consumption (reference level: data-consuming)			
	A bit data-consuming	.647 (0.081)	<.001
	Data-saving	.815 (0.081)	<.001
Cost (HK $[a]; reference level: 10)			
	50	−1.064 (0.075)	<.001
	100	−2.053 (0.086)	<.001

[a]A currency exchange rate of HK $1=US $0.13 is applicable.

Subgroup Analysis of Consumer Preferences for Health Apps

Tables 4-6 present results of the mixed logit regressions by subgroups of sex, age, monthly household income, education level, whether participants have health apps installed on their smartphone, and whether participants have paid for health apps before.

Table . Mixed logit regressions that examined how each attribute of a health app affects whether participants would be willing to purchase the health app by subgroups of sex and age.

Attributes and levels		Sex		Age	
		Male, β (SE)	Female, β (SE)	Younger (≤45 years), β (SE)	Older (>45 years), β (SE)
Usefulness (reference level: slightly useful)					
	Moderately useful	.244 (0.114)[a]	.241 (0.131)	.416 (0.121)[a]	.018 (0.125)
	Very useful	.916 (0.109)[b]	1.079 (0.121)[a]	1.350 (0.113)[a]	.546 (0.118)[a]
Ease of use (reference level: not easy to use)					
	Moderately easy to use	.749 (0.121)[a]	.779 (0.132)[a]	.706 (0.122)[a]	.851 (0.133)[a]
	Very easy to use	.682 (0.107)[a]	.681 (0.120)[a]	.645 (0.108)[a]	.763 (0.119)[a]
Security and privacy (reference level: no security assurance)					
	Some security policies	.955 (0.110)[a]	.567 (0.121)[a]	1.022 (0.113)[a]	.525 (0.120)[a]
	Complete security system	1.260 (0.114)[a]	1.043 (0.121)[a]	1.392 (0.116)[a]	.914 (0.121)[a]
Health care professional's attitude (reference level: neutral)					
	Moderately supportive	.316 (0.110)[c]	.273 (0.121)[b]	.488 (0.113)[a]	.114 (0.120)
	Very supportive	.334 (0.109)[c]	.301 (0.120)[b]	.525 (0.113)[a]	.117 (0.117)
Smartphone storage consumption (MB; reference level: >100)					
	Around 38	.401 (0.109)[a]	.260 (0.122)[b]	.335 (0.112)[c]	.382 (0.120)[c]
	<10	.473 (0.110)[a]	.574 (0.118)[a]	.392 (0.111)[a]	.680 (0.119)[a]
Mobile data consumption (reference level: data-consuming)					
	A bit data-consuming	.603 (0.108)[a]	.743 (0.122)[a]	.646 (0.110)[a]	.637 (0.119)[a]
	Data-saving	.751 (0.108)[a]	.915 (0.123)[a]	.859 (0.109)[a]	.760 (0.121)[a]
Cost (HK $ [d] ; reference level: 10)					
	50	−.871 (0.103)[a]	−1.251 (0.107)[a]	−1.236 (0.102)[a]	−.882 (0.111)[a]
	100	−1.744 (0.113)[a]	−2.417 (0.133)[a]	−2.181 (0.116)[a]	−1.925 (0.128)[a]

[a]Significant at the *P*<.001 level.
[b]Significant at the *P*<.05 level.
[c]Significant at the *P*<.01 level.
[d]A currency exchange rate of HK $1=US $0.13 is applicable.

Table . Mixed logit regressions that examined how each attribute of a health app affects whether participants would be willing to purchase the health app by subgroups of monthly household income and education level.

Attributes and levels		Monthly household income		Education level	
		Lower (<HK $30,000[a]), β (SE)	Higher (≥HK $30,000), β (SE)	Lower (completed secondary school), β (SE)	Higher (postsecondary degree), β (SE)
Usefulness (reference level: slightly useful)					
	Moderately useful	.239 (0.130)	.225 (0.114)[b]	.034 (0.124)	.408 (0.121)[c]
	Very useful	.734 (0.124)[c]	1.156 (0.106)[c]	.497 (0.118)[c]	1.388 (0.114)[c]
Ease of use (reference level: not easy to use)					
	Moderately easy to use	.687 (0.137)[c]	.818 (0.117)[c]	.828 (0.133)[c]	.704 (0.122)[c]
	Very easy to use	.662 (0.121)[c]	.713 (0.105)[c]	.707 (0.118)[c]	.689 (0.108)[c]
Security and privacy (reference level: no security assurance)					
	Some security policies	.682 (0.123)[c]	.867 (0.108)[c]	.575 (0.119)[c]	.980 (0.113)[c]
	Complete security system	.996 (0.127)[c]	1.293 (0.110)[c]	.925 (0.122)[c]	1.366 (0.116)[c]
Health care professional's attitude (reference level: neutral)					
	Moderately supportive	.272 (0.124)[b]	.326 (0.107)[d]	.231 (0.120)	.394 (0.112)[c]
	Very supportive	.239 (0.123)	.381 (0.106)[c]	.140 (0.117)	.483 (0.113)[c]
Smartphone storage consumption (MB; reference level: >100)					
	Around 38	.459 (0.125)[c]	.242 (0.106)[b]	.359 (0.120)[d]	.346 (0.112)[c]
	<10	.667 (0.123)[c]	.390 (0.106)[c]	.682 (0.119)[c]	.387 (0.112)[c]
Mobile data consumption (reference level: data-consuming)					
	A bit data-consuming	.509 (0.123)[c]	.747 (0.106)[c]	.408 (0.118)[c]	.843 (0.112)[c]
	Data-saving	.744 (0.123)[c]	.866 (0.107)[c]	.548 (0.119)[c]	1.045 (0.111)[c]
Cost (HK $; reference level: 10)					
	50	−1.073 (0.115)[c]	−1.066 (0.098)[c]	−1.043 (0.112)[c]	−1.122 (0.102)[c]
	100	−2.173 (0.132)[c]	−1.971 (0.112)[c]	−2.126 (0.129)[c]	−2.035 (0.116)[c]

[a]A currency exchange rate of HK $1=US $0.13 is applicable.

[b]Significant at the *P*<.05 level.

[c]Significant at the *P*<.001 level.

[d]Significant at the *P*<.01 level.

Table . Mixed logit regressions that examined how each attribute of a health app affects whether participants would be willing to purchase the health app by subgroups of health app usage and purchase experiences.

Attributes and levels		Has health apps installed on smartphone		Has paid for health apps before	
		No, β (SE)	Yes, β (SE)	No, β (SE)	Yes, β (SE)
Usefulness (reference level: slightly useful)					
	Moderately useful	.041 (0.134)	.384 (0.112)[a]	.246 (0.098)[b]	.270 (0.198)
	Very useful	.778 (0.123)[a]	1.131 (0.107)[a]	.973 (0.092)[a]	1.226 (0.200)[a]
Ease of use (reference level: not easy to use)					
	Moderately easy to use	.562 (0.138)[a]	.920 (0.117)[a]	.619 (0.100)[a]	1.354 (0.218)[a]
	Very easy to use	.637 (0.122)[a]	.737 (0.104)[a]	.628 (0.088)[a]	.991 (0.195)[a]
Security and privacy (reference level: no security assurance)					
	Some security policies	.767 (0.126)[a]	.800 (0.107)[a]	.829 (0.091)[a]	.559 (0.194)[c]
	Complete security system	1.065 (0.128)[a]	1.252 (0.109)[a]	1.179 (0.094)[a]	.999 (0.200)[a]
Health care professional's attitude (reference level: neutral)					
	Moderately supportive	.192 (0.125)	.379 (0.108)[a]	.317 (0.092)[a]	.311 (0.197)
	Very supportive	.216 (0.122)	.412 (0.106)[a]	.376 (0.091)[a]	.185 (0.193)
Smartphone storage consumption (MB reference level: >100)					
	Around 38	.273 (0.124)[b]	.381 (0.107)[a]	.389 (0.091)[a]	.101 (0.195)
	<10	.519 (0.123)[a]	.503 (0.106)[a]	.568 (0.090)[a]	.217 (0.197)
Mobile data consumption (reference level: data-consuming)					
	A bit data-consuming	.739 (0.124)[a]	.567 (0.105)[a]	.740 (0.090)[a]	.251 (0.191)
	Data-saving	.744 (0.126)[a]	.871 (0.105)[a]	.852 (0.091)[a]	.727 (0.196)[a]
Cost (HK $ [d] ; reference level: 10)					
	50	−.920 (0.111)[a]	−1.183 (0.101)[a]	−1.096 (0.081)[a]	−.797 (0.198)[a]
	100	−2.062 (0.134)[a]	−2.050 (0.111)[a]	−2.223 (0.097)[a]	−1.331 (0.198)[a]

[a]Significant at the *P*<.001 level.

[b]Significant at the *P*<.05 level.

[c]Significant at the *P*<.01 level.

[d]A currency exchange rate of HK $1=US $0.13 is applicable.

Discussion

Principal Findings

Our results showed that cost was the attribute that had the greatest influence on consumers' preferences for health apps, followed by security and privacy and then by usefulness. Consumers also preferred health apps that used less mobile data and took up less smartphone storage, as well as health apps that health care professionals held positive attitudes toward. In terms of ease of use, consumers preferred health apps that were moderately easy to use over those that were very easy to use, and both were preferred over those that were not easy to use. Our results also showed that consumers with different sociodemographic characteristics and different usage and purchase experiences with health apps differed in their preferences for health apps. Understanding consumer preferences in these subgroups could be informative for developing health apps that target consumers in these subgroups.

Consumers preferred health apps that cost less money and consumed less mobile data and smartphone storage. This finding is consistent with previous research that found that cost was the greatest concern for using health apps and that many individuals would not pay anything for a health app [25]. This is also consistent with findings from previous research in which the consumption of resources was a major barrier to using health apps [27,34,35]. It is therefore suggested that health app developers optimize the app size by identifying and removing unnecessary files and codes, reducing the size of images and videos, and providing on-demand downloads for less frequently used resources. Health app developers are also suggested to optimize the data usage of the app by reducing automatic data loading, paginating large volumes of data, and using small-sized images for previews. These approaches could help keep the use of health apps at low costs, which is crucial for scaling up and spreading the use of health apps.

Consumers also strongly preferred health apps with a complete security system to protect their personal health information. This may be explained by findings from previous research in which the concerns for the security of personal health information collected by health apps led to a lack of trust in health apps and constituted a major barrier for the use of health apps [3,25,40]. The fact that many health apps have yet to deploy appropriate techniques for protecting the security and privacy of users [41-43] could constitute a major issue that hampers consumers' adoption of health apps. It is suggested that health app developers should take measures (eg, data encryption, data integrality, and freshness protection) to upgrade the security level of users' personal health information stored in or transmitted through health apps [44]. It is also suggested that mobile app platforms and policy makers launch guidelines, implement policies, and impose regulations to ensure that all health apps can properly protect the privacy of their users [45,46].

We also found that the usefulness and ease of use of health apps influenced consumers' preferences for them. These two attributes have long been considered to be the major reasons why people accept or choose to use health technology in the literature [28,47-51]. Specifically, we found that consumers would always prefer a health app that was more useful, underscoring the importance of understanding users' actual needs and how the health app can be helpful in fulfilling these needs [52]. As for ease of use, consumers most preferred health apps that were moderately easy to use, indicating that consumers preferred health apps that were user-friendly but did not seem too simple. This is probably because consumers perceived health apps that were too easy to use as being too simple to be worthy of purchase. It is therefore suggested that human factors design principles are followed in the design and development of health apps to ensure that they are user-friendly [53-58] and that the useful content, features, and functionalities of the health app are highlighted when promoting the app to its intended users.

Consumers also preferred health apps that health care professionals had positive attitudes toward, most likely because they believed that health care professionals have more knowledge about health management and trusted health care professionals' judgments about health-related products. This is consistent with findings from previous research in which individuals were more willing to use health apps if they were recommended by health care professionals [59,60]. It is thus suggested that the involvement of health care professionals can be effective in the promotion of health apps.

Limitations

This study has limitations. First, the DCE used hypothetical scenarios to elicit consumers' stated preferences, without requiring real economic commitments (ie, actual purchases). As shown in previous research that found that hypothetical products were usually valued higher than actual products, the responses obtained by using a DCE may be affected by hypothetical bias and differ from consumers' behavior in real life [61-64]. Second, we adopted an orthogonal factorial design in the DCE, which enabled us to examine the main effects of each attribute but ignored the interaction effects between these attributes [19]. Future work can be conducted to examine how the interactions between attributes affect consumers' preferences for health apps. Third, the choice task used in this study presented participants with only 1 hypothetical health app and asked them to choose if they would like to purchase it. We were thus unable to observe how consumers compared and chose among multiple health apps and assess how they traded off between different attributes of health apps.

Conclusions

Health apps are preferable when they cost less, consume less storage and mobile data, can protect the security and privacy of personal health data, are useful and easy to use, and are recommended by health care professionals. Therefore, it is recommended that future health apps keep their cost, mobile data consumption, and phone storage consumption low; include a complete security system to protect personal health information; provide useful content and features; adopt user-friendly interfaces; and involve health care professionals. In addition, health app developers should identify the characteristics of their intended users and design and develop health apps to fit the preferences of the intended users.

Conflicts of Interest

None declared.

References

1. Cheung DST, Or CK, So MKP, Ho K, Tiwari A. The use of eHealth applications in Hong Kong: results of a random-digit dialing survey. J Med Syst 2019 Jul 23;43(9):1-9. [doi: 10.1007/s10916-019-1422-2] [Medline: 31338682]
2. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103. [doi: 10.2196/mhealth.9002] [Medline: 29699971]
3. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: national survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26. [doi: 10.2196/mhealth.8619] [Medline: 29362211]
4. Deady M, Johnston D, Milne D, Glozier N, Peters D, Calvo R, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 4;6(12):e11661. [doi: 10.2196/11661] [Medline: 30514694]

5. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):776-786. [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
6. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794. [doi: 10.2196/12794] [Medline: 31094354]
7. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017 Nov 17;3:51. [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
8. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Aug 4;8(8):e15779. [doi: 10.2196/15779] [Medline: 32459654]
9. Tao D, Or C. A paper prototype usability study of a chronic disease self-management system for older adults. Presented at: 2012 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 10-12, 2012; Hong Kong, China. [doi: 10.1109/IEEM.2012.6837946]
10. Wong RSM, Yu EYT, Wong TWL, Fung CSC, Choi CSY, Or CKL, et al. Development and pilot evaluation of a mobile app on parent-child exercises to improve physical activity and psychosocial outcomes of Hong Kong Chinese children. BMC Public Health 2020 Oct 14;20(1):1544. [doi: 10.1186/s12889-020-09655-9] [Medline: 33054753]
11. van Limburg M, van Gemert-Pijnen JEWC, Nijland N, Ossebaard HC, Hendrix RMG, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124. [doi: 10.2196/jmir.1674] [Medline: 22204896]
12. Jones RB, Ashurst EJ, Trappes-Lomax T. Searching for a sustainable process of service user and health professional online discussions to facilitate the implementation of e-health. Health Informatics J 2016 Dec;22(4):948-961. [doi: 10.1177/1460458215599024] [Medline: 26324052]
13. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
14. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
15. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364. [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
16. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 2014 Sep;32(9):883-902. [doi: 10.1007/s40273-014-0170-x] [Medline: 25005924]
17. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019 Feb;37(2):201-226. [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
18. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012 Feb;21(2):145-172. [doi: 10.1002/hec.1697] [Medline: 22223558]
19. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16(1):3-13. [doi: 10.1016/j.jval.2012.08.2223] [Medline: 23337210]
20. Xie Z, Chen J, Or CK. Consumers' willingness to pay for eHealth and its influencing factors: systematic review and meta-analysis. J Med Internet Res 2022 Sep 14;24(9):e25959. [doi: 10.2196/25959] [Medline: 36103227]
21. Crutchfield TM, Kistler CE. Getting patients in the door: medical appointment reminder preferences. Patient Prefer Adherence 2017 Jan 25;11:141-150. [doi: 10.2147/PPA.S117396] [Medline: 28182131]
22. Phillips EA, Himmler SF, Schreyögg J. Preferences for e-mental health interventions in Germany: a discrete choice experiment. Value Health 2021 Mar;24(3):421-430. [doi: 10.1016/j.jval.2020.09.018] [Medline: 33641777]
23. Abelson JS, Kaufman E, Symer M, Peters A, Charlson M, Yeo H. Barriers and benefits to using mobile health technology after operation: a qualitative study. Surgery 2017 Sep;162(3):605-611. [doi: 10.1016/j.surg.2017.05.007] [Medline: 28651777]
24. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014 Jun 11;16(6):e148. [doi: 10.2196/jmir.3122] [Medline: 24918859]
25. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
26. Mohamed AHHM, Tawfik H, Al-Jumeily D, Norton L. MoHTAM: a technology acceptance model for mobile health applications. Presented at: 2011 Developments in E-systems Engineering (DeSE); December 6-8, 2011; Dubai, United Arab Emirates p. 13-18. [doi: 10.1109/DeSE.2011.79]

27. Nijland N, van Gemert-Pijnen JEWC, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71. [doi: 10.2196/jmir.1603] [Medline: 21959968]
28. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
29. Or CKL, Karsh BT, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59. [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
30. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
31. Povey J, Mills PPJR, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for Aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65. [doi: 10.2196/jmir.5314] [Medline: 26969043]
32. Reger GM, Browne KC, Campellone TR, Simons C, Kuhn E, Fortney JC, et al. Barriers and facilitators to mobile application use during PTSD treatment: clinician adoption of PE coach. Prof Psychol Res Pr 2017;48(6):510-517. [doi: 10.1037/pro0000153]
33. Ryan M, Amaya-Amaya M, Gerard K. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht, Netherlands: Springer Netherlands; 2008. [doi: 10.1007/978-1-4020-5753-3]
34. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480. [doi: 10.2196/10480] [Medline: 30001997]
35. Velu AV, van Beukering MD, Schaafsma FG, Frings-Dresen MH, Mol BW, van der Post JA, et al. Barriers and facilitators for the use of a medical mobile app to prevent work-related risks in pregnancy: a qualitative analysis. JMIR Res Protoc 2017 Aug 22;6(8):e163. [doi: 10.2196/resprot.7224] [Medline: 28830851]
36. Yan M, Or C. Factors in the 4-week acceptance of a computer-based, chronic disease self-monitoring system in patients with type 2 diabetes mellitus and/or hypertension. Telemed J E Health 2018 Feb;24(2):121-129. [doi: 10.1089/tmj.2017.0064] [Medline: 28737995]
37. Liu K, Or CK, So M, Cheung B, Chan B, Tiwari A, et al. A longitudinal examination of tablet self-management technology acceptance by patients with chronic diseases: integrating perceived hand function, perceived visual function, and perceived home space adequacy with the TAM and TPB. Appl Ergon 2022 Apr;100:103667. [doi: 10.1016/j.apergo.2021.103667] [Medline: 34920356]
38. Or C, Karsh B, Severtson D, Brennan P. Patient technology acceptance model (PTAM): exploring the potential characteristics of consumer health information technology acceptance by home care patients with chronic illness. 2006 Presented at: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting; Santa Monica, CA.
39. Hawthorne R, Grzybowski L, SSRN. Benefits of regulation vs competition where inequality is high: the case of mobile telephony in South Africa. CESifo Working Paper No 7703. 2019. URL: papers.ssrn.com/sol3/papers.cfm?abstract_id=3418023
40. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015 Aug;216:467-471. [Medline: 26262094]
41. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
42. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8. [doi: 10.2196/mhealth.3672] [Medline: 25599627]
43. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
44. Al Ameen M, Liu J, Kwak K. Security and privacy issues in wireless sensor networks for healthcare applications. J Med Syst 2012 Feb;36(1):93-101. [doi: 10.1007/s10916-010-9449-4] [Medline: 20703745]
45. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. Health Aff (Millwood) 2014 Sep;33(9):1523-1530. [doi: 10.1377/hlthaff.2014.0771] [Medline: 25201656]
46. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018 Jan 29;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
47. Chen T, Chen J, Or CK, Lo FP. Path analysis of the roles of age, self-efficacy, and TAM constructs in the acceptance of performing upper limb exercises through Immersive virtual reality games. Int J Ind Eng 2022 Sep;91:103360. [doi: 10.1016/j.ergon.2022.103360]

48. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
49. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
50. Xie Z, Or CK. Acceptance of mHealth by elderly adults: a path analysis. Proc Hum Factors Ergon Soc 2020 Dec;64(1):755-759. [doi: 10.1177/1071181320641174]
51. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep;25(3):828-843. [doi: 10.1177/1460458217724580] [Medline: 28820007]
52. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9. [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
53. Nielsen J. Usability Engineering. Boston, MA: AP Professional; 1993.
54. Shneiderman B, Plaisant C, Cohen M. Designing the User Interface: Strategies for Effective Human-Computer Interaction (5th Edition). Boston, MA: Addison-Wesley; 2010.
55. Cheung DST, Or CKL, So MKP, Tiwari A. Usability testing of a smartphone application for delivering qigong training. J Med Syst 2018 Sep 5;42(10):191. [doi: 10.1007/s10916-018-1048-9] [Medline: 30187139]
56. Karsh B, Holden RJ, Or CKL. Human factors and Ergonomics of health information technology implementation. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Boca Raton, FL: CRC Press; 2011:249-264.
57. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 8;1(2):e13. [doi: 10.2196/resprot.2184] [Medline: 23612015]
58. Or CK, Holden RJ, Valdez RS. Human factors engineering and user-centered design for mobile health technology: enhancing effectiveness, efficiency, and satisfaction. In: Human-Automation Interaction: Mobile Computing. Cham, Switzerland: Springer; 2022:97-118. [doi: 10.1007/978-3-031-10788-7]
59. Fisher J, Clayton M. Who gives a tweet: assessing patients' interest in the use of social media for health care. Worldviews Evid Based Nurs 2012 Apr;9(2):100-108. [doi: 10.1111/j.1741-6787.2012.00243.x] [Medline: 22432730]
60. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006. [doi: 10.2196/11006] [Medline: 30355554]
61. Voelckner F. An empirical comparison of methods for measuring consumers' willingness to pay. Market Lett 2006 Apr;17(2):137-149. [doi: 10.1007/s11002-006-5147-x]
62. Botelho A, Pinto LC. Hypothetical, real, and predicted real willingness to pay in open-ended surveys: experimental results. Appl Econ Lett 2002 Dec;9(15):993-996. [doi: 10.1080/13504850210148143]
63. Wertenbroch K, Skiera B. Measuring consumers' willingness to pay at the point of purchase. J Mark Res 2002 May;39(2):228-241. [doi: 10.1509/jmkr.39.2.228.19086]
64. Kanya L, Sanghera S, Lewin A, Fox-Rushby J. The criterion validity of willingness to pay methods: a systematic review and meta-analysis of the evidence. Soc Sci Med 2019 Jul;232:238-261. [doi: 10.1016/j.socscimed.2019.04.015] [Medline: 31108330]

Abbreviations

DCE: discrete choice experiment
mHealth: mobile health

Edited by L Buis; submitted 21.11.20; peer-reviewed by C Snoswell, J Rawstorn, R Jayasuriya; revised version received 08.06.23; accepted 25.07.23; published 13.09.23.

Please cite as:
Xie Z, Or CK
Consumers' Preferences for Purchasing mHealth Apps: Discrete Choice Experiment
JMIR Mhealth Uhealth 2023;11:e25908
URL: https://mhealth.jmir.org/2023/1/e25908
doi:10.2196/25908

Assessing the Usefulness of Mobile Apps for Noise Management in Occupational Health and Safety: Quantitative Measurement and Expert Elicitation Study

Jingchen Huyan[1,*], MPH; Chandnee Ramkissoon[1,*], MSci, PhD; Mah Laka[2], PhD; Sharyn Gaskin[1,*], PhD

1

2

*these authors contributed equally

Corresponding Author:
Sharyn Gaskin, PhD

Abstract

Background: Overexposure to occupational noise can lead to hearing loss. Occupational noise mapping is conventionally performed with a calibrated sound level meter (SLM). With the rise of mobile apps, there is a growing number of SLM apps available on mobile phones. However, few studies have evaluated such apps for accuracy and usefulness to guide those with occupational noise detection needs in selecting a quality app.

Objective: The purpose of this study was to evaluate the accuracy and usefulness of SLM mobile apps to guide workplace health and safety professionals in determining these apps' suitability for assessing occupational noise exposure.

Methods: The following three iOS apps were assessed: the NIOSH (National Institute for Occupational Safety and Health) Sound Level Meter, Decibel X, and SoundMeter X apps. The selected apps were evaluated for their accuracy in measuring sound levels in low-, moderate-, and high-noise settings within both simulated environments and real-world environments by comparing them to a conventional SLM. The usefulness of the apps was then assessed by occupational health specialists using the Mobile App Rating Scale (MARS).

Results: The NIOSH Sound Level Meter app accurately measured noise across a range of sound levels in both simulated settings and real-world settings. However, considerable variation was observed between readings. In comparison, the Decibel X and SoundMeter X apps showed more consistent readings but consistently underestimated noise levels, suggesting that they may pose a risk for workers. Nevertheless, none of the differences in sound measurements between the three apps and the conventional SLM were statistically significant (NIOSH Sound Level Meter: P=.78; Decibel X: P=.38; SoundMeter X: P=.40). The MARS scores for the three apps were all above 3.0, indicating the usefulness of these apps.

Conclusions: Under the conditions of this study, the NIOSH Sound Level Meter app had equivalent accuracy to the calibrated SLM and a degree of usefulness according to the MARS. This suggests that the NIOSH Sound Level Meter app may be suitable for mapping noise levels as part of a monitoring strategy in workplaces. However, it is important to understand its limitations. Mobile apps should complement but not replace conventional SLMs when trying to assess occupational noise exposure risk. Our outcomes also suggest that the MARS tool may have limited applicability to measurement-based apps and may be more suited to information-based apps that collect, record, and store information.

(JMIR Mhealth Uhealth 2023;11:e46846) doi:10.2196/46846

KEYWORDS

occupational health; noise management; mobile applications; mobile apps; mHealth; Mobile App Rating Scale; MARS; management; hearing; hearing loss; mobile phone; noise detection; usefulness; tool

Introduction

More than 1.5 billion people worldwide live with varying degrees of hearing loss, with nearly 500 million living with severe hearing loss [1]. Nearly 16% of these adults have severe hearing loss as a result of occupational noise exposure [2]. In Australia, over 111,000 people have occupational noise-induced hearing loss (ONIHL) [3], resulting in a loss of 62,218 quality-adjusted life years and 135,561 productivity-adjusted life years. The projected welfare-based loss is Aus $5.5 billion (US $3.5 billion), and the projected productivity-based loss is Aus $21.3 billion (US $13.4 billion) [3].

ONIHL is preventable, and the economic benefits of intervention are considerable. According to the modelling study on the productivity burden of ONIHL in Australia by Si et al [3], even a modest reduction of overall noise exposure in the workplace can significantly reduce the socioeconomic burden of ONIHL. The prevention of ONIHL can be achieved through occupational noise awareness and exposure control and monitoring [4].

The Australian national standard for occupational noise levels is an average daily exposure to ≤85 A-weighted dB (dB[A]; ie, an 8-h, A-weighted equivalent continuous sound level [LAeq] of ≤85 dB) [5]. Noise assessment in the workplace is typically done through the use of a handheld sound level meter (SLM); the SLM is placed 10 to 20 cm from the worker's ear canal for a representative period of time, during which routine tasks are undertaken. Owning a conventional SLM may be prohibitively expensive for most small- and medium-sized enterprises, resulting in difficulties with taking timely noise management measurements and evaluating interventions. With the increasing number of smartphones worldwide, there is potential for increased accessibility to noise mapping via mobile apps with noise monitoring capabilities. Often, mobile apps with SLM features are used to complement traditional SLMs. However, the accuracy of such apps in assessing and monitoring occupational noise exposure is not well evaluated.

The evidence for the accuracy of SLM mobile apps appears limited to simulated laboratory studies with limited evidence of accuracy based on real-world scenarios. The results of one simulation study showed that mobile SLM apps accurately measured 65 dB to 95 dB of pink noise (defined as random noise having equal energy per octave), with an error of approximately 2 dB(A) [6]. Another study showed the accuracy of mobile SLM apps in measuring white and pink noises from 3 sound sources, namely conversation, occupational steelmaking, and conveyor belt operation [7]. A study by Murphy and King [8] compared the accuracy of 7 SLM apps on 100 smartphones (both Android and iOS) in detecting white noise (containing many frequencies). The study showed a difference in noise level assessment between the two types of phones; Apple phones showed a measurement error within 1 dB(A), while Android phones showed twice the variation in noise measurement. Although these studies evaluated the accuracy of mobile apps from different perspectives, the sound sources in these studies were simulated in a laboratory. The accuracy of such apps in a real-world noise scenario, which represents a more realistic noise exposure scenario for workers, has yet to be demonstrated. In this study, the accuracy of mobile apps for assessing occupational noise exposure in simulated and real-world settings was evaluated according to the Occupational Health Hazard Management framework [9].

In addition to accuracy, user experience and usefulness are also key considerations for SLM mobile apps, as they may influence the readiness of users to choose one particular noise measurement app over others. Several guidelines and scales are available for assessing the effectiveness of digital technologies in health care settings, such as the Xcertia Guidelines from the American Medical Association, the Digital Technology Assessment Criteria from National Health Service England, and the Mobile App Rating Scale (MARS) [10,11]. A growing number of studies use the MARS to assess the usefulness of health-related mobile apps, such as food allergy management apps, blood disease management apps, diabetes management apps, and occupational therapy apps [12-15]. However, to our knowledge, the MARS has not previously been applied to assess occupational health-related mobile apps.

This study aimed to evaluate the accuracy and usefulness of SLM mobile apps for an occupational health context. The objectives of this study were (1) to assess the accuracy of 3 SLM apps by comparing them with conventional SLM instrumentation in simulated noise situations within a laboratory and in a real-world noise environment and (2) to assess the usefulness of the studied apps by using the MARS. The study results could help influence the adoption of high-quality SLM apps in occupational settings through empirical evidence and informed decision-making by relevant health specialists.

Methods

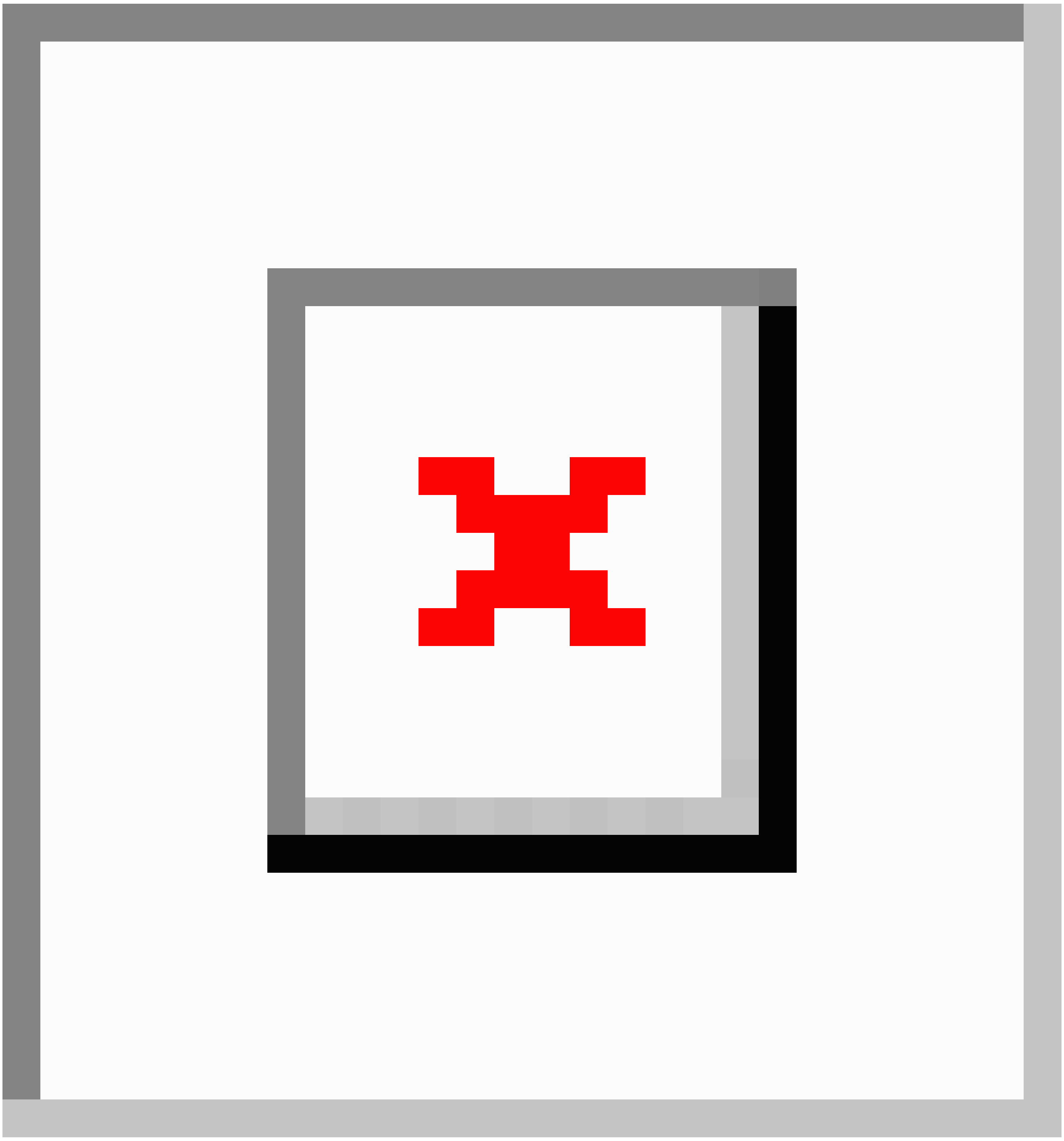
Accuracy of SLM Mobile Apps

The accuracy of mobile SLM apps in measuring noise levels, when compared to a conventional SLM, was assessed in the following two contexts: laboratory and real-world conditions.

Selection of SLM Mobile Apps

An Apple device (iPhone 13 Pro, iOS 15.6; Apple Inc) was used in this study because the built-in hardware (microphone, circuitry, and signal processing hardware) and operating system showed less variability across models when compared to Android phones [16]. SLM apps were systematically searched for and screened according to the inclusion and exclusion criteria in Figure 1. The specific app inclusion criteria were (1) app store review scores of >1, (2) the app was last updated within 12 months, (3) an app store score of >2, (4) a reading precision of >1 decimal place, (5) the app costs less than Aus $10 (US $6.31) or the subscription costs less than Aus $10 (US $6.31) per month, and (6) the app has the ability to log and export data externally. The app store score and review scores were selected as proxies for the accuracy of apps to ensure that apps regarded as useful or accurate by most users were included in our analysis. We understood that the app review score may not be a consistent criterion for measuring app quality; therefore, we used a review score of >1 and an app store score of >2 as the initial gatekeeper criteria to select apps that were used and regarded as fairly accurate by many users. This scoring system did not impact the app accuracy assessment of the experts but was used as an app selection criterion. The following three apps met all criteria and were included in this study: Decibel X, NIOSH (National Institute for Occupational Safety and Health) Sound Level Meter, and SoundMeter X. Details about the specifications of the three apps are provided in Multimedia Appendix 1.

Figure 1. Flow diagram of the selection process for the apps included in this study. A currency exchange rate of Aus $1=US $0.63 is applicable.



Laboratory Simulation

This study was carried out in a room at a university where the background noise level was 30 dB(A). A loudspeaker was used to reproduce a standard white noise generated from a web-based tone generator [17]. The sound level of the white noise was increased in increments from 60 dB to 85 dB, using an amplifier (Marshall Emberton portable speaker; Zound Industries).

The sound levels were measured by using a conventional SLM (Type 2250; Brüel & Kjær) and the mobile phone, which were placed side by side and approximately 100 cm away from the noise source. Both devices were activated and stopped simultaneously. The experiment was replicated 3 times, using 1 app each time. Measurements were taken for 30 seconds, and an average LAeq, in dB, was logged.

Real-World Simulation

The sound mapping for the real-world study was conducted across 3 different sites in a tertiary education environment, namely a library, a busy student and staff activity center, and an engineering workshop facility where a high-pressure water-jet cutting machine (Techni Waterjet i35-G2; Techni Ltd) was in use. These three locations represented a range of noise levels—low sound levels at the library (45-50 dB; typically quiet radio music and normal conversation), moderate sound levels at the student and staff activity center (60-70 dB; typically loud conversation), and high sound levels at the engineering workshop facility (75-85 dB; typically heavy traffic or a front-end loader) [5]. Similar equipment and methods to the laboratory simulation study were used; briefly, sound levels in each environment were measured for 30 seconds (LAeqs were logged) with a mobile app on the same Apple device (iPhone 13 Pro, iOS 15.6) alongside the conventional SLM. A total of 20 samples were collected from each location and each app. After the two simulation studies, the accuracy of the mobile apps was evaluated by comparing the LAeqs taken from the mobile apps to those taken from the conventional SLM.

Usefulness of the Mobile Apps

The usefulness of the SLM apps was evaluated by 3 occupational health specialists using the MARS; the specialists rated the apps based on the five components of the MARS—engagement, functionality, aesthetics, information, and subjective quality. The MARS was selected for this study because it is reported as one of the most widely used tools for evaluating the quality of mobile apps, and validation studies have demonstrated its suitability for quality assessment [18]. The five components of the scale were further divided into 22 items, as follows: engagement was divided into entertainment, interest, customization, interactivity, and target group; functionality was divided into performance, ease of use, navigation, and gestural design; aesthetic was divided into layout, graphics, and visual appeal; information was divided into accuracy, goals, quality and quantity of information, visual information, and credibility; and subjective quality was divided into recommendation, frequency of use, willingness to pay, and overall rating. Each MARS item was scored by using a 5-point Likert scale (1=inadequate; 2=poor; 3=acceptable; 4=good; 5=excellent). The MARS questionnaire was reworded to adapt it to SLM apps.

Ethical Considerations

The three health specialists were chosen from a professional membership database—the Australian Institute of Occupational Hygienists—based on their expertise in occupational noise exposure assessment. The selected health specialists were invited to participate in this study via publicly available information, such as email addresses. Informed consent was obtained from all participants involved in this study. The consent form provided details such as the objectives of this study, information on the team of investigators, the kind of participation expected, and the nature of survey. Only when potential participants gave consent to participate in this study, they were emailed the survey link. They were further deidentified to maintain the anonymity of the responses. This study was approved by the University of Adelaide Human Research Ethics Committee (approval number: H-2022-196). Participants did not receive any compensation for their participation.

Statistical Analysis

This study used Stata software (version 17; StataCorp LLC) for data analysis, and both parts of the project were analyzed separately. The difference in measurements of sound levels between the SLM apps and the conventional SLM and the variability within the apps were assessed for statistical significance in the laboratory and real-world studies via a rank Mann-Whitney *U* test. The samples were first tested for normality and equal variances via a Shapiro-Wilk test (sample size of <50). Given that the data collection method involved using only 1 app and 1 SLM at the same time rather than using all apps and the SLM at the same time, differences between mobile apps were not considered in this study. For the second part of this study, the scores obtained for each component of the MARS quality assessment were averaged among the three evaluators.

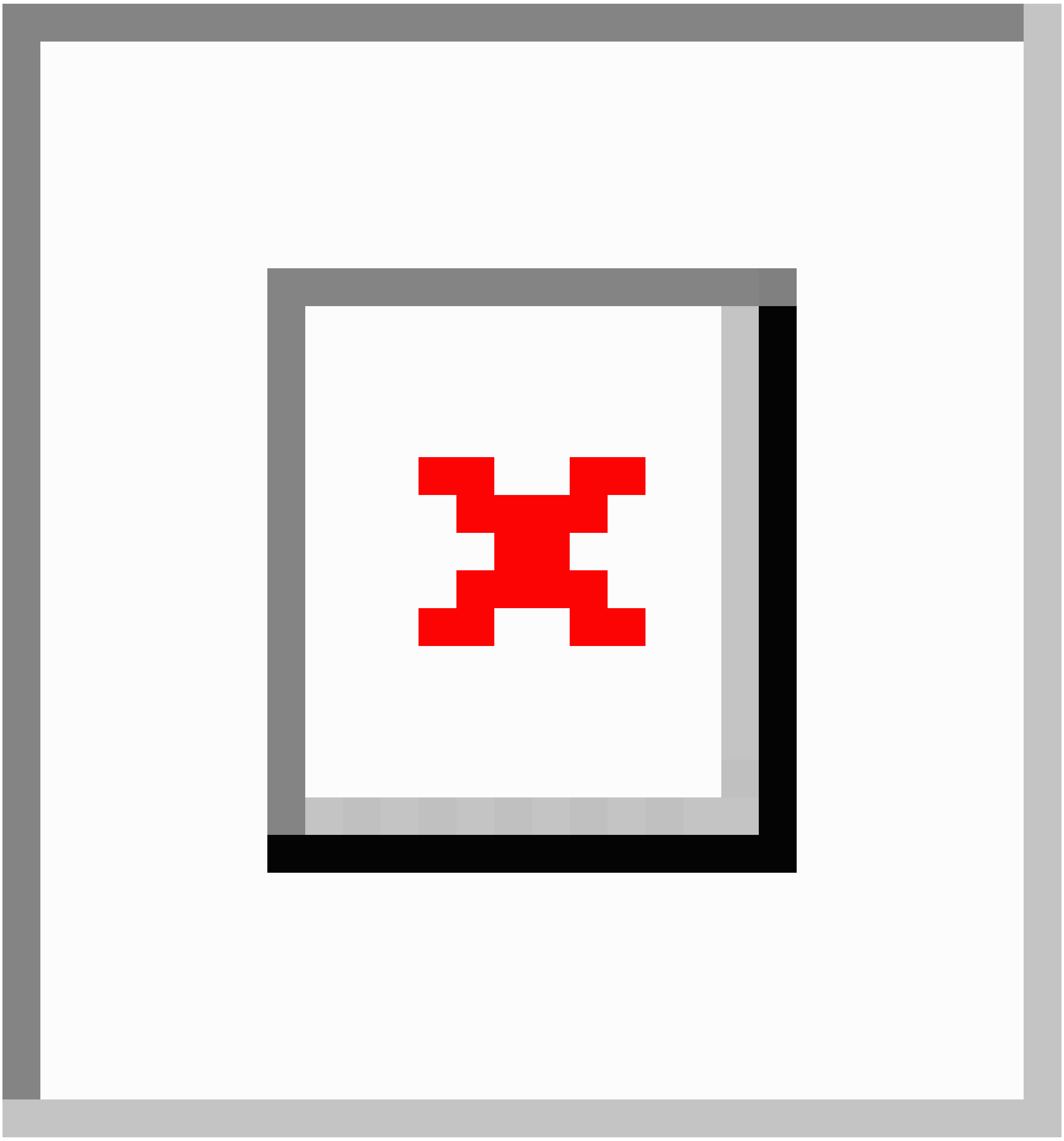
Results

Accuracy of SLM Mobile Apps

Laboratory Simulation

All 3 apps gave different readings when compared to the conventional SLM but to various extents (Figure 2). The readings of the NIOSH Sound Level Meter app were the most accurate, with measurement readings within 0.5 dB(A) of the conventional SLM readings (Figure 2). However, more variation in the NIOSH Sound Level Meter measurements was observed when compared to the other two apps tested (around 2 dB[A]). In contrast, the Decibel X app data showed the least variability when compared to the other two apps but recorded lower noise levels than those recorded by the conventional SLM (>2 dB[A] deviation). SoundMeter X also consistently recorded lower sound levels than those recorded by the conventional SLM (Figure 2). However, none of the differences observed between the apps and the conventional SLM were statistically significant (NIOSH Sound Level Meter: *P*=.78; Decibel X: *P*=.38; SoundMeter X: *P*=.40).

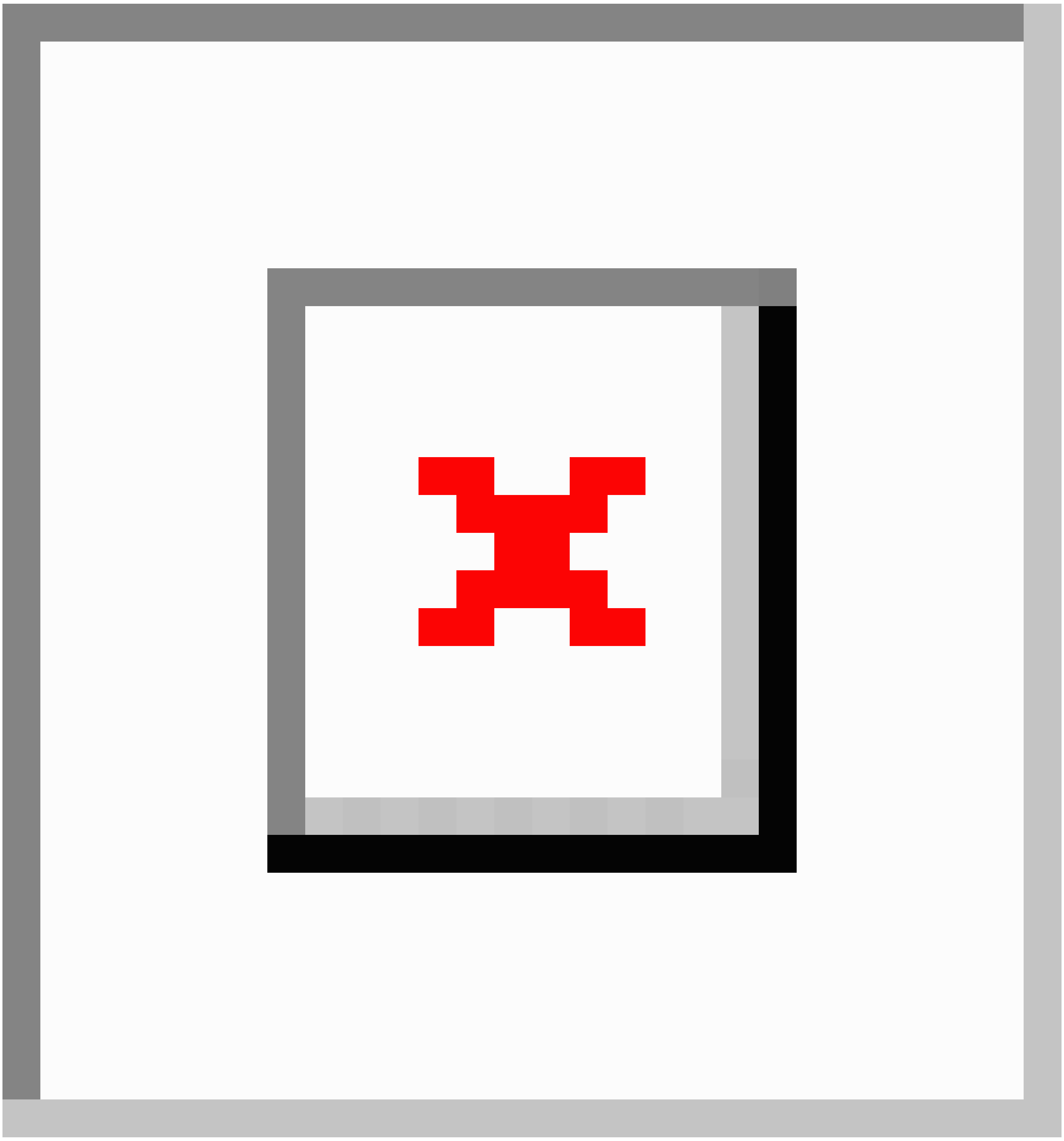
Figure 2. Distribution of the differences between the measured values of the three apps—Decibel X, NIOSH SLM, and SoundMeter X—and those of the conventional SLM in the laboratory simulation study. dB(A): A-weighted dB; NIOSH: National Institute for Occupational Safety and Health; SLM: sound level meter.



Real-World Simulation

The results in the real-world study showed similar trends to the laboratory study. The NIOSH Sound Level Meter app results differed minimally from the conventional SLM results across the different sound level ranges; hence, it was considered to be the most accurate app overall (Figure 3). SoundMeter X was observed to be more accurate than Decibel X in all 3 locations (Figure 3).

Figure 3. Distribution of the differences between the measured values of the three apps and the conventional SLM in (**A**) low-noise (40-50 dB), (**B**) moderate-noise (55-70 dB), and (**C**) high-noise (75-85 dB) settings in the real-world study. dB(A): A-weighted dB; NIOSH: National Institute for Occupational Safety and Health; SLM: sound level meter. *Statistical difference between the app and the conventional SLM at a 95% CI.



In the low-noise (40-50 dB) location, the SoundMeter X and Decibel X apps underestimated sound levels by approximately 2 dB(A) and 4 dB(A), respectively, with Decibel X readings being statistically different (*P*<.001) from the readings of the conventional SLM (Figure 3A). In the moderate-noise setting (55-70 dB), the NIOSH Sound Level Meter app results differed minimally, by approximately 0.5 dB(A), from those of the conventional SLM, and the app's data were more reliable than those of the other two apps (Figure 3B). One measurement of SoundMeter X was an outlier, with a >4-dB(A) underestimation when compared to the conventional SLM, which was attributed to a brief app malfunction; all other measurements of the SoundMeter X app differed minimally, by approximately 0.5 dB(A), from the results of the conventional SLM (Figure 3B). In contrast, the Decibel X measurements differed significantly (*P*<.001) from the conventional SLM measurements, underestimating sound level by >3 dB(A) (Figure 3B).

In the high-noise setting (the engineering workshop), the NIOSH Sound Level Meter app results were again the most accurate and consistent, differing minimally (by 0.5 dB[A]) from the conventional SLM results (Figure 3C). The SoundMeter X app recorded 2 significant outliers (excluded from this data set), suggesting unreliable app functioning. In addition, the Decibel X app data differed (by >3 dB[A]) from the conventional SLM data. Both SoundMeter X and Decibel X significantly underestimated the noise levels when compared to the conventional SLM (P<.001) (Figure 3C).

Usefulness of the Mobile Apps

The MARS survey results were summarized as an average of the experts' ratings, and these are shown in Table 1. Further information about the scores is available in Multimedia Appendix 2. Overall, Decibel X had the highest average MARS score (4.0) among all 3 apps. Table 1 shows that the NIOSH Sound Level Meter app had the highest scores in the functionality and information sections—4.3 and 4.2, respectively—which assess the functionality of an app and the reliability of information, respectively. However, the overall MARS score for this app (3.5) was the lowest among the three apps rated by the experts. Interestingly, Decibel X, which was the least accurate app in the accuracy assessment, had the highest overall score when compared to the other two apps, with higher scores in the engagement and subjective quality domains—4.4 and 4.0, respectively (Table 1). According to the 5-point Likert scale, apps with an overall score of >3.0 are considered acceptable in terms of usefulness. Since the overall scores were all higher than 3.0, the three apps could be considered useful according to this scale.

Table . Summary of the average MARS[a] scores for the three mobile apps.

Mobile app	Usefulness domain score, mean					
	Engagement	Functionality	Aesthetics	Information	Subjective quality	Overall score
NIOSH[b] Sound Level Meter	3.2	4.3	3.2	4.2	3.3	3.5
Decibel X	4.4	3.9	4.2	3.6	4.0	4.0
SoundMeter X	3.7	3.8	3.6	3.7	3.2	3.6

[a]MARS: Mobile App Rating Scale.

[b]NIOSH: National Institute for Occupational Safety and Health.

Discussion

Principal Results

This study evaluated the accuracy and usefulness (user experience) of SLM mobile apps to guide workplace health and safety professionals in determining these apps' suitability for assessing occupational noise exposure. It is the first reported application of the MARS to the assessment of occupational health-related mobile apps. Furthermore, this study builds upon previously published accuracy studies by examining app performance in realistic working environments, not just in a laboratory simulation.

In terms of the overall performance of the apps tested across all elements of this study, the NIOSH Sound Level Meter app was consistently the most accurate mobile app in measuring noise levels ranging from 40 dB(A) to 85 dB(A). In comparison, Decibel X was the least accurate in similar noise settings, suggesting that it is not a reliable tool for measuring noise in occupational settings. The Decibel X app had the highest overall MARS score, but the individual domain scores indicated that although this app might have superior aesthetics, graphic layouts, interactivity, and visual appeal, users considered it inferior to other apps in terms of the quality and credibility of information. This aligns with our finding that Decibel X was the least accurate app in all 3 occupational settings when compared against the conventional SLM. The SoundMeter X app gave more accurate measurements than those provided by Decibel X but showed a high level of variability in the real-world environment, particularly in low-noise (40-50 dB[A]) settings, suggesting that it may not be sensitive and reliable enough for effective noise exposure assessments. Therefore, under the conditions of this study, the NIOSH Sound Level Meter app had comparable accuracy to a conventional SLM and could be used to acquire and monitor real-time noise exposure data, which could be used to raise occupational awareness about the potential hazards to hearing in a work environment. The experts in our study considered the NIOSH Sound Level Meter app useful; however, its overall aesthetics, including visual appeal, graphics, and layout, were scored the lowest among all 3 apps. This indicated that the NIOSH Sound Level Meter app could be improved at the aesthetics and user engagement levels.

According to IEC (International Electrotechnical Commission) standard 61672, class 2, a noise measurement device should have an error within 2 dB(A) [19]. The results from this study demonstrated that the NIOSH Sound Level Meter and SoundMeter X apps complied with this international standard. There were, however, differences noted in performance between simulation outcomes and real-world outcomes. The NIOSH Sound Level Meter app had greater variability in noise measurements under real-world conditions than under laboratory simulation conditions. In contrast, the other two apps showed the opposite effect—more reproducible data in the real world than in the laboratory—but both consistently underestimated noise levels. In the context of occupational noise assessment, it would be more preferable to overestimate sound levels than to underestimate them when informing protective effects for the health of the target population. Moreover, despite the overall MARS score of the NIOSH Sound Level Meter app being the lowest among the three apps tested, the app did obtain an overall score of >3.0 and the highest functionality and information scores. Thus, the NIOSH Sound Level Meter app had a degree of usefulness, especially for providing reliable information and appropriate functionality in our study settings.

The advantages that a mobile app might have over a conventional SLM include cost (often free), ease of access, and simple operation. The NIOSH Sound Level Meter app may be suitable for scenarios where a rapid assessment is required (eg, a change in task or setting up new equipment), while a conventional SLM may be important for scenarios where precise analysis is required, such as frequency analysis or worker exposure risk assessment. Our results suggest that SLM apps can complement traditional SLMs in occupational noise detection but may not be a substitute for conventional SLMs.

The accuracy results of the apps assessed in this study appear consistent with those of previous studies. The NIOSH Sound Level Meter app was previously known as the *NoiSee* app (version 1.0) in 2014. A study by Kardous and Shaw [6] showed that NoiSee app measurements were within 2 dB(A) of the noise levels of a sound source and concluded that the app was adequate for occupational noise assessment. Similarly, Crossley et al [20] evaluated the fit of 9 apps and concluded that the NIOSH Sound Level Meter app had the best fit, with an R^2 value of 0.97. The SoundMeter X app was previously known as *SoundMeter* (version 3.3.1) in 2014. Nast et al [21] evaluated the accuracy of mobile apps and determined that the SoundMeter app had the highest accuracy, with a mean difference of −0.5 dB(A), and the narrowest variance distribution. Although the SoundMeter X app in our study was not the most accurate, its error could still be considered within the tolerable range. However, the app consistently underestimated actual sound levels, as reported in a study by McLennon et al [7], who also evaluated Decibel X (previously known as Decibel 10th [version 4.3.5]). In their study, the Decibel X app was used on an iPhone and showed high inconsistency in measuring sound levels ranging from 60 dB to 90 dB. A later study showed that the Decibel X app ranked at the bottom of a fit assessment, with an R^2 value of 0.77 [20]. Therefore, the performance of the Decibel X app in our study is consistent with previous simulation studies.

Limitations

There are several limitations and constraints to this study. In terms of hardware, only an iPhone 13 Pro, which was produced in 2022, was used for noise assessment via mobile apps. Sound measurements may differ between new hardware and old hardware. Further, SLM apps are limited by the microphone used [22]. Almost all smartphone devices are fitted with microelectromechanical system microphones, which, on a technical basis, have limitations in meeting the national and international requirements for sound measurement instrumentation. However, attaching a high-quality condenser microphone and preamplifier to professional SLMs allows them to conform to international standards, such as IEC 61672-1 [19]. Furthermore, in this study, the apps were not calibrated, and the experiments were designed to simulate actual use in an occupational noise mapping scenario.

In terms of app selection, this study only screened apps for iOS systems and did not search for apps for Android systems. A prior study compared the accuracy of SLM apps for both systems and concluded that iOS apps were more accurate [8]. In addition, there are many brands of Android phones, unlike Apple iPhones, which use fewer and uniform hardware (eg, microphones and chips), and this may influence study outcomes and translatability [7]. Nonetheless, for the purpose of completeness, research that systematically evaluates the accuracy and usefulness of SLM apps needs to extend to Android systems. In addition, only apps with data logging and exporting functions were selected for this study. In occupational settings however, there may only be a need to have a function for displaying measured values and not necessarily logging or exporting functions. Furthermore, the features of mobile apps may change with version updates. Therefore, the results of this study are only representative of the apps' versions at the time of testing.

Noise levels in this study ranged from 45 dB to 85 dB for all measurement scenarios, but there was a lack of data for noise levels above 85 dB. Operator safety was considered in this study's design. Furthermore, only 30-second LAeqs were evaluated throughout this study. Further studies could add other measurement metrics, such as time-weighted average values, for comparison and analysis.

The reliance on the subjective (expert) judgments of the evaluators limited the results of using the MARS in this research, and their judgments may not represent the views of workers or other potential users of the SLM apps. However, this issue was partially addressed by checking the internal reliability of the scores given by each independent evaluator. The inclusion of workers as users who lack expertise could easily influence decisions on product feature trade-offs. We compared the overall scores of the three apps, and the app that we considered the most accurate had the lowest MARS score, while the app that we considered the least accurate had the highest MARS score, although the relative differences between their overall scores were small. This suggests that the MARS may be more suited to assessing perceptions on the use of an app rather than an app's usefulness. Furthermore, the MARS may be more suited to information-based apps that collect, record, and store information, such as guides for chemical hazard management or ergonomic assessment tools.

Conclusions

This study examined and assessed the accuracy and usefulness of SLM mobile apps. Under the conditions of this study, the NIOSH Sound Level Meter app had equivalent accuracy to the calibrated conventional SLM and demonstrated a degree of usefulness according to relevant expert judgments. This suggests that the NIOSH Sound Level Meter app may be suitable for monitoring noise levels in scenarios where cost prohibits the purchase or use of a conventional SLM device or where the rapid evaluation of noise-reducing control measures is required to reduce the risk of exposure. Moreover, the mobile app should complement but not replace conventional SLMs, especially when trying to determine worker risk. Lastly, the MARS may have limited applicability to measurement-based apps and may be more suited to information-based apps that collect, record, and store information. Future research assessing the usefulness of other occupational health and safety apps involving measurement (eg, light and heat measurement) should consider MARS outcomes in conjunction with accuracy measurements when determining the suitability of apps for occupational health and safety management.

Acknowledgments

The authors would like to thank Mr Benjamin Green from the University of Adelaide for technical support in the implementation of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Specifications of the three sound level meter apps included in this study.
[DOCX File, 19 KB - mhealth_v11i1e46846_app1.docx]

Multimedia Appendix 2
Detailed scores from the three health specialists (deidentified by participant number) for the overall usefulness of the three mobile sound level meter apps, namely SoundMeter X, Decibel X, and NIOSH (National Institute for Occupational Safety and Health) Sound Level Meter.
[DOCX File, 18 KB - mhealth_v11i1e46846_app2.docx]

References

1. Deafness and hearing loss. World Health Organization. 2023 Feb 27. URL: www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2023-10-03]
2. Nelson DI, Nelson RY, Concha-Barrientos M, Fingerhut M. The global burden of occupational noise-induced hearing loss. Am J Ind Med 2005 Dec;48(6):446-458. [doi: 10.1002/ajim.20223] [Medline: 16299704]
3. Si S, Lewkowski K, Fritschi L, Heyworth J, Liew D, Li I. Productivity burden of occupational noise-induced hearing loss in Australia: a life table modelling study. Int J Environ Res Public Health 2020 Jun 29;17(13):4667. [doi: 10.3390/ijerph17134667] [Medline: 32610505]
4. Occupationally-induced hearing loss. The National Institute for Occupational Safety and Health. 2010 Mar. URL: www.cdc.gov/niosh/docs/2010-136/default.html [accessed 2023-10-03]
5. Safe Work Australia. Managing Noise and Preventing Hearing Loss at Work: Code of Practice: Safe Work Australia; 2020.
6. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications (apps) using external microphones-a follow-up study. J Acoust Soc Am 2016 Oct;140(4):EL327-EL333. [doi: 10.1121/1.4964639] [Medline: 27794313]
7. McLennon T, Patel S, Behar A, Abdoli-Eramaki M. Evaluation of smartphone sound level meter applications as a reliable tool for noise monitoring. J Occup Environ Hyg 2019 Sep;16(9):620-627. [doi: 10.1080/15459624.2019.1639718] [Medline: 31356145]
8. Murphy E, King EA. Testing the accuracy of smartphones and sound level meter applications for measuring environmental noise. Appl Acoust 2016 May;106:16-22. [doi: 10.1016/j.apacoust.2015.12.012]
9. PUB HLTH 7113 - environmental and occupational health. The University of Adelaide. URL: www.adelaide.edu.au/course-outlines/003181/1/sem-1/2015/ [accessed 2023-10-20]
10. Xcertia mHealth app guidelines. Healthcare Information and Management Systems Society. 2019 Aug 12. URL: www.himss.org/sites/hde/files/media/file/2020/04/17/xcertia-guidelines-2019-final.pdf [accessed 2023-02-02]
11. Digital Technology Assessment Criteria (DTAC). NHS England. URL: www.nhsx.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac [accessed 2023-02-02]
12. Kalhori SRN, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of English mobile applications for gestational diabetes: app review using Mobile Application Rating Scale (MARS). Curr Diabetes Rev 2021;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
13. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 1;7(5):e13019. [doi: 10.2196/13019] [Medline: 31066712]
14. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
15. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2022 Feb 16;10(2):e32826. [doi: 10.2196/32826] [Medline: 35171109]
16. Kardous CA, Shaw PB. Do sound meter apps measure noise levels accurately? Sound & Vibration Magazine. 2015 Jul. URL: www.sandv.com/downloads/1507kard.pdf [accessed 2023-10-03]

17. Online tone generator. URL: onlinetonegenerator.com [accessed 2023-10-03]
18. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
19. IEC 61672-1:2013 Electroacoustics—sound level meters—part 1: specifications. International Electrotechnical Commission. 2013 Sep 30. URL: webstore.iec.ch/publication/5708 [accessed 2023-02-06]
20. Crossley E, Biggs T, Brown P, Singh T. The accuracy of iPhone applications to monitor environmental noise levels. Laryngoscope 2021 Jan;131(1):E59-E62. [doi: 10.1002/lary.28590] [Medline: 32108336]
21. Nast DR, Speer WS, Le Prell CG. Sound level measurements using smartphone "apps": useful or inaccurate? Noise Health 2014;16(72):251-256. [doi: 10.4103/1463-1741.140495] [Medline: 25209033]
22. Robinson DP, Tingay J. Comparative study of the performance of smartphone-based sound level meter apps, with and without the application of a ½" IEC-61094-4 working standard microphone, to IEC-61672 standard metering equipment in the detection of various problematic workplace noise environments. Presented at: 43rd International Congress on Noise Control Engineering (inter.noise 2014); Nov 16-19, 2014; Melbourne, Australia p. 1-10 URL: www.acoustics.asn.au/conference_proceedings/INTERNOISE2014/papers/p565.pdf [accessed 2023-10-20]

Abbreviations

dB(A): A-weighted dB
IEC: International Electrotechnical Commission
LAeq: A-weighted equivalent continuous sound level
MARS: Mobile App Rating Scale
NIOSH: National Institute for Occupational Safety and Health
ONIHL: occupational noise-induced hearing loss
SLM: sound level meter

Edited by L Buis; submitted 27.02.23; peer-reviewed by I Karlsen, PD Haghighi; revised version received 31.07.23; accepted 18.08.23; published 14.11.23.

Please cite as:
Huyan J, Ramkissoon C, Laka M, Gaskin S
Assessing the Usefulness of Mobile Apps for Noise Management in Occupational Health and Safety: Quantitative Measurement and Expert Elicitation Study
JMIR Mhealth Uhealth 2023;11:e46846
URL: https://mhealth.jmir.org/2023/1/e46846
doi:10.2196/46846

Use Patterns of Smartphone Apps and Wearable Devices Supporting Physical Activity and Exercise: Large-Scale Cross-Sectional Survey

Takeyuki Oba[*], PhD; Keisuke Takano[*], PhD; Kentaro Katahira, PhD; Kenta Kimura, PhD

Human Informatics and Interaction Research Institute, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan

[*]these authors contributed equally

Corresponding Author:
Keisuke Takano, PhD

Abstract

Background: Physical inactivity is a global health issue, and mobile health (mHealth) apps are expected to play an important role in promoting physical activity. Empirical studies have demonstrated the efficacy and efficiency of app-based interventions, and an increasing number of apps with more functions and richer content have been released. Regardless of the success of mHealth apps, there are important evidence gaps in the literature; that is, it is largely unknown who uses what app functions and which functions are associated with physical activity.

Objective: This study aims to investigate the use patterns of apps and wearables supporting physical activity and exercise in a Japanese-speaking community sample.

Methods: We recruited 20,573 web-based panelists who completed questionnaires concerning demographics, regular physical activity levels, and use of apps and wearables supporting physical activity. Participants who indicated that they were using a physical activity app or wearable were presented with a list of app functions (eg, sensor information, goal setting, journaling, and reward), among which they selected any functions they used.

Results: Approximately one-quarter (n=4465) of the sample was identified as app users and showed similar demographic characteristics to samples documented in the literature; that is, compared with app nonusers, app users were younger (odds ratio [OR] 0.57, 95% CI 0.50-0.65), were more likely to be men (OR 0.83, 95% CI 0.77-0.90), had higher BMI scores (OR 1.02, 95% CI 1.01-1.03), had higher levels of education (university or above; OR 1.528, 95% CI 1.19-1.99), were more likely to have a child (OR 1.16, 95% CI 1.05-1.28) and job (OR 1.28, 95% CI 1.17-1.40), and had a higher household income (OR 1.40, 95% CI 1.21-1.62). Our results revealed unique associations between demographic variables and specific app functions. For example, sensor information, journaling, and GPS were more frequently used by men than women (ORs <0.84). Another important finding is that people used a median of 2 (IQR 1-4) different functions within an app, and the most common pattern was to use sensor information (ie, self-monitoring) and one other function such as goal setting or reminders.

Conclusions: Regardless of the current trend in app development toward multifunctionality, our findings highlight the importance of app simplicity. A set of two functions (more precisely, self-monitoring and one other function) might be the minimum that can be accepted by most users. In addition, the identified individual differences will help developers and stakeholders pave the way for the personalization of app functions.

(JMIR Mhealth Uhealth 2023;11:e49148) doi:10.2196/49148

KEYWORDS

mobile health; smartphone app; physical activity; wearable devices; telemedicine; wearable; wearables; mHealth; app; apps; use; usage; survey; cross sectional; cross-sectional; technology use; exercise

Introduction

Background

Physical inactivity is an unresolved issue in modern society, despite its known risks to physical and mental health [1,2]. Mobile technology, including smartphone apps and wearable devices, is expected to become a game changer and has substantially impacted health care practices for both providers and recipients. Several behavior change techniques (BCTs) [3] are provided by apps in the form of self-help training, which is augmented by online (ambulatory) assessments of physiological status via wearable devices and smartwatches. Evidence supports the positive effects of mobile health (mHealth) interventions on physical activity (PA) [4-12], and a recent meta-analysis [13] suggested that activity trackers have a moderate-sized effect in improving PA, equating to an increase of 1800 steps per day and a reduction of 1 kg in body weight. Trials included in these (umbrella) meta-analyses implemented interventions not limited to self-monitoring (or just wearing an activity tracker); instead, activity trackers, combined with smartphone or web-based applications, offer more interactive BCT components, such as goals and planning, rewards and threats, social support, and gamification [9].

Regardless of the success of mHealth tools and interventions, there are important evidence gaps in the literature; that is, it remains largely unknown who uses what app functions and which functions are associated with PA [14]. In recent years, a large number of health care apps have appeared on the market, many of which have complex multifunctionality to cover different needs and provide person-centered care [15]. Indeed, the amount of content and number of functions are found to be predictive of the overall quality of PA apps [16], as well as the efficacy of an intervention [17]; however, users tend to consider an app valuable when it is simple and intuitive to use [18]. An analysis of commercial health care apps [18] identified 12 representative features and characteristics (ie, export data, gamification, general education, plans or orders, reminders, community forums, social media, address symptoms, tailored education, tracking, cost, and usability), among which the export of data, usability, and cost were associated with users' positive ratings. A qualitative study on users' perceptions of apps [19] found that people typically like the tracking feature (eg, monitoring step counts) in their apps because this type of self-monitoring increases their awareness, and feedback on the tracked data helps them observe their progress. A cross-sectional survey of Chinese app users [20] identified typical health app users to be women and in a higher self-rated social class; the most prevalent types of apps were those that provide health information, track vital signs (eg, steps or heart rate), and provide health and medical reminders. A similar analysis was conducted in Saudi Arabia [21], showing that daily step counting and ovulation tracking (among women) were the most prevalent functions. Analyses of representative samples of Dutch [22] and US populations [23] suggested that mHealth app users are generally younger and more educated, and have higher levels of eHealth literacy skills than nonusers, although the profile of app users varies largely across different types of apps (eg, for fitness, nutrition, sleep, and mindfulness).

Objectives

In short, previous studies have suggested that gender, age, and education level are robust predictors of mHealth app use, and tracking (eg, step counts) appears to be the most prevalent function. However, heterogeneity persists due to the types of apps and demographic profiles of app users. A systematic investigation is warranted to clarify the associations between user profiles and the use patterns of individual app functions that help increase PA. Thus, this study aims to determine the prevalence of commercial PA apps in a community sample. We were specifically interested in sociodemographic differences between app users and app nonusers, which app functions would be used most frequently, which app functions would be associated with increased levels of PA, and which user profiles (eg, gender and age) would be predictive of the use of the app functions associated with increased levels of PA.

Methods

Participants and Procedure

We analyzed the data of 20,573 Japanese-speaking web-based panelists (n=10,701 women; mean age 52.7, SD 17.8 years) sampled from a survey company's database in which >1.3 million inhabitants were registered as potential participants. Participants were recruited with appropriate weights that reflected the demographic composition (eg, place of residence) across the country. We did not use any inclusion criteria, except for age (≥18 years). The sample size was determined arbitrarily, and the overarching project was published elsewhere [24]. In short, this project consisted of a series of web-based surveys administered on three different occasions, each targeting different aspects of PA behavior and health. The first wave of the survey covered regular PA levels as well as motivational and environmental factors that potentially influence PA behavior, such as decisional balance, self-efficacy, and social support. The second wave, which is reported in this paper, specifically focuses on mHealth technology use (ie, apps and wearable devices). The third wave involved past and current physical and mental disorders (not reported). The second and third wave of surveys were administered in the same week (early 2023), approximately 2 months after the first wave of surveys.

Ethical Considerations

Participants provided informed consent in the first wave. At the end of each survey, they were compensated with a voucher for web-based shopping (value: approximately US $0.31). The study protocol was approved by the Ethics Committee of the National Institute of Advanced Industrial Science and Technology (approval ID 2022-1279). We adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) Statement [25] when conducting and reporting this study.

Measures

Stage of Change Questionnaire

PA readiness was assessed using the Japanese version [26,27] of the stage of change (SoC) questionnaire [28,29]. Each participant was classified into one of five stages (ie, precontemplation, contemplation, preparation, action, and maintenance) according to their responses to the following items: "I currently do not exercise and do not intend to start exercising in the future" (precontemplation); "I currently do not exercise but I am thinking about starting to exercise in the next six months" (contemplation); "I currently exercise some, but not regularly" (preparation); "I currently exercise regularly, but have only begun doing so within the last six months" (action); and "I currently exercise regularly and have done so for longer than six months" (maintenance). Regular exercise was operationalized as exercising twice or more per week for 20 minutes or longer, which was explicitly stated to the participants.

International PA Questionnaire–Short Form

Average weekly PA levels were assessed using the International Physical Activity Questionnaire–Short Form (IPAQ-SF) [30,31]. The IPAQ-SF includes the following three PA dimensions: (1)

walking, (2) moderate-intensity activity, and (3) vigorous-intensity activity (sedentary time was not used for our analyses). Participants reported the number of days and duration spent on each dimension of PA over an average week. The reported weekly minutes of PA were transformed into metabolic equivalent tasks (METs per hour), which allowed us to explore how many participants adhered to the Japanese public health guidelines for PA: 23 METs per hour per week for adults [32]. We also coded the IPAQ-SF responses in a categoric manner: inactive, minimally active (≥3 days of ≥20 minutes of vigorous activity or equivalent), and health-enhancing physical activity (HEPA) active (≥3 days of vigorous activity achieving 25 METs per hour per week or equivalent) [33].

Use of Apps and Wearables

Participants were first asked to indicate whether they used any apps to support PA or exercise (in the first-wave survey). Those who responded affirmatively were invited to the second-wave survey, where they provided the following information (see Multimedia Appendix 1 for details): (1) the names of apps in use, (2) what sensors or wearable devices were connected to the apps (if any), (3) how long they had been using the app that they were using most frequently (less than a week to more than a year), (4) how frequently they were using the app (less than once per month to multiple times per day), (5) sources of information about the app (eg, preinstalled on a smartphone or learned from a family, friend, or health specialist), and (6) which functions of the app they were using. For item 6, a list of 41 app functions was presented to each participant (eg, sensor information, goal setting, and journaling), and participants selected any applicable function. This list was generated by the first author based on published studies in the literature [15,18]. The list was reviewed by four researchers using different PA apps and devices.

Statistical Analyses

First, we tested the prevalence of apps and individual functions and explored the demographic differences between app users and app nonusers (eg, gender, age, education, income, and levels of PA and SoC). Second, we visualized the use patterns of the app functions in the form of a network. We were specifically interested in app functions that were often used together, and these co-occurrences were represented by edges in the network. The backbone algorithm was used to select meaningful edges in the network. In this algorithm, each edge weight (ie, the co-occurrence frequency) is normalized by the strength of the connected nodes (ie, the sum of the edge weights of each node), which is then statistically tested (α=.05) under the assumption that the normalized edge weights are uniformly distributed [34]. Third, multinomial logistic regression analysis was conducted to examine the associations between self-reported PA levels and app functions. A series of logistic regression analyses were conducted to explore the demographic variables that were predictive of the use of each function. All analyses were performed using R (version 4.2.2; R Foundation for Statistical Computing) with the *backbone* package [35], which provided a *disparity* function for network edge selection.

Results

Prevalence of PA Apps

Among 20,573 individuals, 5030 (24.4%) reported using one or more PA apps. These app users were invited to participate in the second-wave survey, in which 4465 participants completed questionnaires concerning PA app use. The most frequently used apps were the iOS health app (n=1239, 27.8%), Google Fit (n=910, 20.4%), dHealthcare (n=891, 20%), and Trima (n=1026, 23%). Most app users (n=3140, 70.3%) had been using a PA app longer than 6 months (n=2583, 57.8%), and most used the app once or more per day (n=3218, 72.1%). App users reported that they started to use the app because it was installed when they purchased their smartphones or tablets (n=1589, 35.6%) or they learned about the app on the internet or social media (n=1387, 31.1%), or from someone close to them, such as family members, friends, acquaintances, or colleagues (n=1208, 27.1%).

Demographics of App Users Versus App Nonusers

Table 1 shows the demographic characteristics of the app users and app nonusers. To explore the demographic differences between app users and app nonusers, a logistic regression analysis was performed with app users (vs app nonusers) as the outcome (see also Table S1 in Multimedia Appendix 2). The results showed that, compared to app nonusers, app users were younger, were more likely to be men, had larger BMI scores, had higher levels of education (university or above), were more likely to have a child and job, and had a higher household income. App users were more active with a median PA level of 30.8 METs per hour per week (vs 13.6 METs/hour/week among app nonusers), indicating that most of them adhered to the national health recommendation (23 METs/hour/week). This tendency is also endorsed by the SoC distribution: almost half of app users were in the maintenance stage (2065/4465, 46.2%; ie, having exercised regularly for more than 6 months), which is more prevalent than with app nonusers (4327/15,543, 27.8%).

Table . Demographic statistics of app users and app nonusers.

Variable		App users (n=4465)	App nonusers (n=15,543)	Odds ratio[a] (95% CI)	*P* value
Age (years), mean (SD)		50.70 (17.36)	53.48 (17.67)	N/A[b]	N/A
Age group (years)[c], n (%)					
	<30	655 (14.7)	1953 (12.6)	N/A	N/A
	30-44	1073 (24.0)	3139 (20.2)	0.870 (0.768-0.987)	.03
	45-59	1135 (25.4)	3856 (24.8)	0.736 (0.648-0.835)	<.001
	≥60	1602 (35.9)	6595 (42.4)	0.570 (0.499-0.651)	<.001
Women, n (%)		1932 (43.3)	8461 (54.4)	0.832 (0.769-0.900)	<.001
BMI, mean (SD)		22.34 (3.69)	22.08 (3.71)	1.018 (1.008-1.028)	<.001
Married, n (%)		2899 (64.9)	9719 (62.5)	1.049 (0.952-1.157)	.34
One or more children, n (%)		2772 (62.1)	9585 (61.7)	1.156 (1.047-1.277)	.004
Education level, n (%)					
	Middle school	78 (1.7)	426 (2.7)	N/A	N/A
	High school	1135 (25.4)	5081 (32.7)	1.159 (0.898-1.513)	.27
	College or vocational school	892 (20.0)	3689 (23.7)	1.262 (0.974-1.654)	.08
	University or above	2314 (51.8)	6230 (40.1)	1.528 (1.186-1.991)	.001
	Others	46 (1.0)	117 (0.8)	2.265 (1.453-3.509)	<.001
Job, n (%)		3058 (68.5)	8964 (57.7)	1.277 (1.170-1.395)	<.001
Household income (¥ [d]), n (%)					
	<3 million	827 (18.5)	3467 (22.3)	N/A	N/A
	3-5 million	1012 (22.7)	3869 (24.9)	0.919 (0.824-1.026)	.13
	5-7 million	753 (16.9)	2310 (14.9)	1.030 (0.910-1.166)	.64
	7-10 million	700 (15.7)	1804 (11.6)	1.122 (0.984-1.279)	.09
	≥10 million	559 (12.5)	1076 (6.9)	1.401 (1.213-1.618)	<.001
	No answer	614 (13.8)	3017 (19.4)	0.793 (0.702-0.896)	<.001
Physical activity (METs[e]/h/wk), median (IQR)		30.8 (12.0-62.2)	13.6 (1.7-34.8)	N/A	N/A
Physical activity, n (%)					
	Inactive	1257 (28.2)	7664 (49.3)	N/A	N/A
	Minimally active	1762 (39.5)	5362 (34.5)	1.579 (1.448-1.723)	<.001
	HEPA[f] active	1446 (32.4)	2517 (16.2)	2.132 (1.925-2.362)	<.001
Stage of change, n (%)					
	Precontemplation	390 (8.7)	3943 (25.4)	N/A	N/A
	Contemplation	728 (16.3)	3768 (24.2)	1.875 (1.642-2.143)	<.001
	Preparation	982 (22.0)	2915 (18.8)	2.897 (2.547-3.302)	<.001
	Action	300 (6.7)	590 (3.8)	3.676 (3.067-4.405)	<.001
	Maintenance	2065 (46.2)	4327 (27.8)	3.628 (3.199-4.122)	<.001
App functions in use, median (IQR)		2 (1-4)	N/A	N/A	N/A

[a]Odds ratios calculated in the logistic regression predicting app versus app nonusers.

[b]N/A: not applicable.

[c]Age was treated as a categorical predictor in the logistic regression with age <30 years as the reference.

[d]A currency exchange rate of ¥140=US $1 is applicable.

[e]MET: metabolic equivalent task.

[f]HEPA: health-enhancing physical activity.

We also explored the demographic characteristics per stage of change (Table 2). Similar to published studies in the literature (see Marshall and Biddle [36] for meta-analytic evidence), the most frequent stage was maintenance. Participants identified at the maintenance stage were typically older (aged ≥60 years), married, and highly educated, and had a child, job, and high income.

Table . Demographics per stage of change.

		Precontemplation (n=4391), n (%)	Contemplation (n=4641), n (%)	Preparation (n=4006), n (%)	Action (n=945), n (%)	Maintenance (n=6590), n (%)	Chi-square (*df*)	*P* value
Gender							238.6 (4)	<.001
	Men	2123 (48.3)	1789 (38.5)	2017 (50.3)	457 (48.4)	3486 (52.9)		
	Women	2268 (51.7)	2852 (61.5)	1989 (49.7)	488 (51.6)	3104 (47.1)		
Age group (years)							994.9 (12)	<.001
	<30	605 (13.8)	782 (16.8)	557 (13.9)	222 (23.5)	624 (9.5)		
	30-44	974 (22.2)	1218 (26.2)	861 (21.5)	249 (26.3)	1026 (15.6)		
	45-59	1230 (28.0)	1308 (28.2)	1018 (25.4)	198 (21.0)	1336 (20.3)		
	>60	1582 (36.0)	1333 (28.7)	1570 (39.2)	276 (29.2)	3604 (54.7)		
Marital status							71.3 (4)	<.001
	Yes	2581 (58.8)	2861 (61.6)	2574 (64.3)	574 (60.7)	4364 (66.2)		
	No	1810 (41.2)	1780 (38.4)	1432 (35.7)	371 (39.3)	2226 (33.8)		
One or more children							116.5 (4)	<.001
	Yes	2554 (58.2)	2732 (58.9)	2488 (62.1)	523 (55.3)	4375 (66.4)		
	No	1837 (41.8)	1909 (41.1)	1518 (37.9)	422 (44.7)	2215 (33.6)		
Education level							177.0 (16)	<.001
	Middle school	167 (3.8)	96 (2.1)	96 (2.4)	24 (2.5)	130 (2.0)		
	High school	1510 (33.4)	1407 (30.3)	1241 (31.0)	258 (27.3)	1973 (29.9)		
	Some college	1009 (23.0)	1231 (26.5)	883 (22.0)	218 (23.1)	1374 (20.8)		
	College and above	1662 (37.9)	1876 (40.4)	1739 (43.4)	436 (46.1)	3077 (46.7)		
	Others	43 (1.0)	31 (0.7)	47 (1.2)	9 (1.0)	36 (0.5)		
Job							207.2 (4)	<.001
	Yes	2666 (60.7)	3040 (65.5)	2558 (63.9)	609 (64.4)	3527 (53.5)		
	No	1725 (39.3)	1601 (34.5)	1448 (36.1)	336 (35.6)	3063 (46.5)		
Household income (¥ [a])							190.9 (20)	<.001
	<3 million	1101 (25.1)	948 (20.4)	825 (20.6)	191 (20.2)	1369 (20.8)		
	3-5 million	1054 (24.0)	1111 (23.9)	942 (23.5)	228 (24.1)	1674 (25.4)		
	5-7 million	628 (14.3)	743 (16.0)	665 (16.6)	157 (16.6)	950 (14.4)		
	7-10 million	439 (10.0)	609 (13.1)	527 (13.2)	131 (13.9)	877 (13.3)		
	≥10 million	291 (6.6)	318 (6.9)	315 (7.9)	69 (7.3)	699 (10.6)		
	No answer	878 (20.0)	912 (19.7)	732 (18.3)	169 (17.9)	1021 (15.5)		

[a]A currency exchange rate of ¥140=US $1 is applicable.

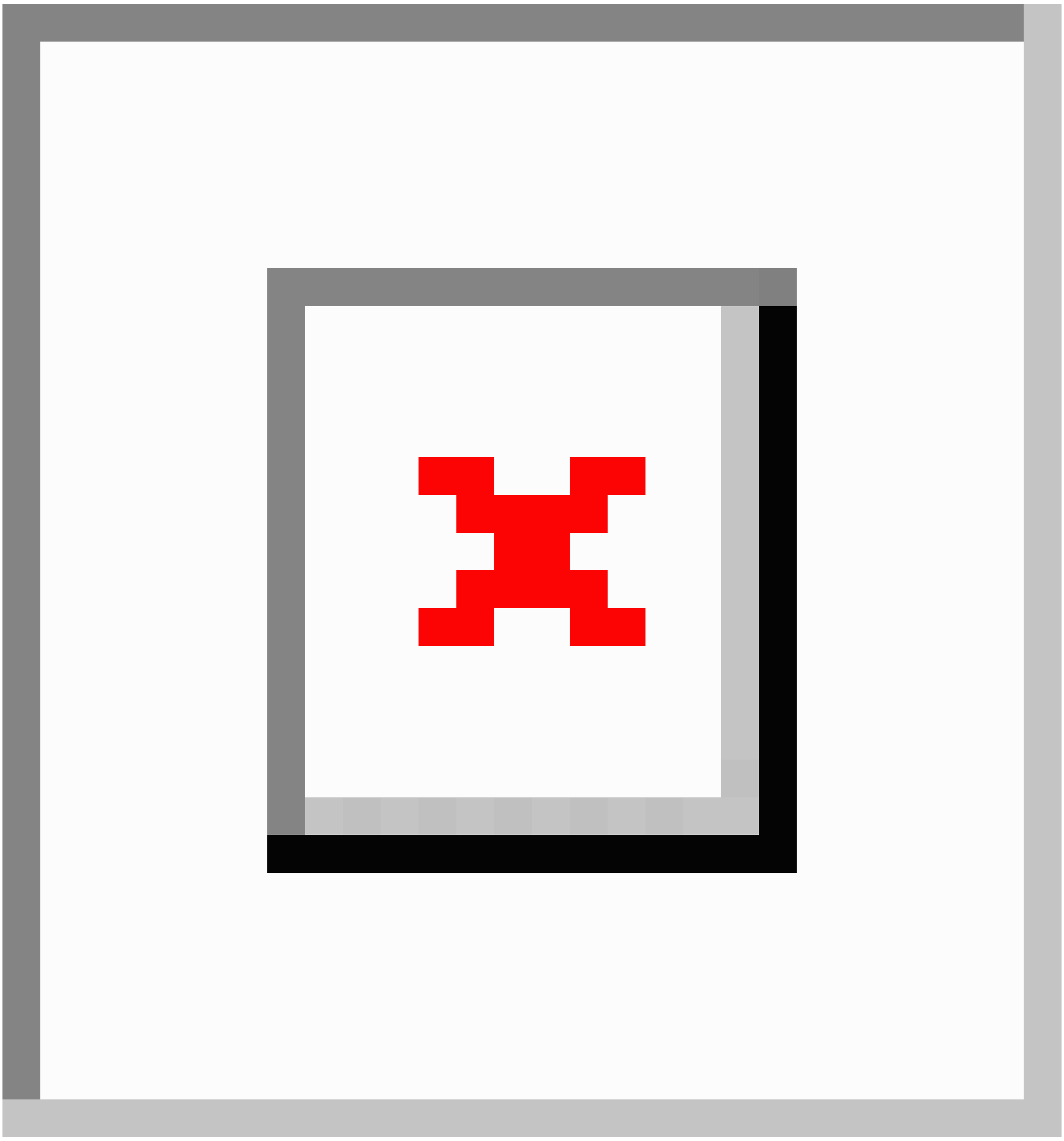
App Functions: Prevalence and Associations Between Functions

The 4465 total app users reported that apps were typically synchronized to a step counter or pedometer (n=3167, 70.9%), GPS and map functions (n=1527, 34.2%), body scale (n=882, 19.8%), and heart rate monitor (n=619, 13.9%). These sensors are also implemented in smartwatches, which were used by 20.8% (n=928) of the app users. Table 3 shows the 10 most frequently used apps or sensor functions. More than half of the app users monitored sensor information (eg, step counts, heart rates, and skin temperature); some also used functions to support goal setting and visualize goal progress. Figure 1 shows how each function is used together with other functions. Sensor information is typically used with one other function (eg, goal setting, goal progress, and recording the menstrual cycle), and participants used a median of 2 (IQR 1-4) different functions within an app (Table 1). It is uncommon for participants to use 5 or more functions regardless of the multifunctionality of apps on the market.

Table . The 10 most frequently used app functions.

App function	Participants, n (%)
Show sensor info	2580 (57.78)
Goal setting	1203 (26.94)
Show goal progress	965 (21.61)
Energy analysis	903 (20.22)
Weight recording	845 (18.92)
Journaling	831 (18.61)
GPS/map	756 (16.93)
Show sleep info	655 (14.67)
Reward points	496 (11.11)
Blood pressure recording	419 (9.38)

Figure 1. Use patterns of app functions. FTP: functional threshold power; VO_2max: maximum oxygen consumption.



App Functions and PA

Table 4 shows the results of the multinomial logistic regression analysis where PA (contrasts: inactive vs minimally active; inactive vs HEPA active) was predicted by the 10 most frequently used app functions. Only 2 app functions (ie, sensor information and goal setting) were identified as significant predictors to distinguish between inactive and minimally active individuals. On the other hand, 6 app functions (ie, sensor information, goal setting, goal progress, journaling, GPS/map, and energy analysis) were found informative in predicting HEPA-active individuals. These functions are typically designed and implemented to target PA. Functions that are not significantly related to PA have other primary health targets, such as sleep and nutrition.

Table . Multinomial logistic regression predicting physical activity level as a categorical variable for all app users (n=4465).

Variables[a]		Estimate (SE)	*Z*	*P* value	Odds ratio (95% CI)
Dependent variable contrast: inactive vs minimally active					
	Show sensor info	0.331 (0.077)	4.308	<.001	1.393 (1.198-1.620)
	Goal setting	0.527 (0.095)	5.525	<.001	1.693 (1.405-2.041)
	Show goal progress	0.154 (0.102)	1.508	.13	1.166 (0.955-1.425)
	Energy analysis	0.192 (0.103)	1.861	.06	1.212 (0.990-1.485)
	Weight recording	0.139 (0.108)	1.284	.20	1.149 (0.930-1.419)
	Journaling	0.114 (0.104)	1.096	.27	1.120 (0.914-1.373)
	GPS/map	0.018 (0.107)	0.166	.87	1.018 (0.826-1.255)
	Show sleep information	0.072 (0.116)	0.621	.54	0.930 (0.741-1.168)
	Reward points	0.215 (0.124)	1.738	.08	1.240 (0.973-1.579)
	Blood pressure recording	0.246 (0.147)	1.671	.10	1.279 (0.958-1.707)
Dependent variable contrast: inactive vs HEPA [b] active					
	Show sensor info	0.278 (0.081)	3.435	.001	1.320 (1.127-1.546)
	Goal setting	0.480 (0.099)	4.830	<.001	1.617 (1.330-1.964)
	Show goal progress	0.226 (0.106)	2.144	.03	1.254 (1.020-1.542)
	Energy analysis	0.265 (0.107)	2.491	.01	1.304 (1.058-1.607)
	Weight recording	0.104 (0.113)	0.918	.36	1.109 (0.889-1.384)
	Journaling	0.470 (0.103)	4.540	<.001	1.600 (1.306-1.960)
	GPS/map	0.306 (0.107)	2.847	.004	1.357 (1.100-1.675)
	Show sleep information	0.068 (0.119)	0.573	.57	1.070 (0.848-1.350)
	Reward points	0.106 (0.131)	0.807	.42	1.111 (0.860-1.437)
	Blood pressure recording	0.067 (0.156)	0.430	.67	1.069 (0.788-1.452)

[a]The independent variable was app function

[b]HEPA: health-enhancing physical activity.

For the 6 PA-related functions (Table 4), we further examined their associations with demographic variables and explored the characteristics of users of each function. The results of the logistic regression analyses (see Table S2 in Multimedia Appendix 2) suggested that older people (aged ≥60 years) were more likely to use goal management functions (goal setting: odds ratio [OR] 1.33, 95% CI 1.04-1.70; *P*=.02; goal progress: OR 2.23, 95% CI 1.70-2.96; *P*<.001), and middle-aged people preferred the GPS and map functions (OR 1.40, 95% CI 1.05-1.87; *P*=.02). Sensor information (OR 0.84, 95% CI 0.73-0.96; *P*=.01), journaling (OR 0.80, 95% CI 0.67-0.96; *P*=.02), and GPS (OR 0.59, 95% CI 0.49-0.71; *P*<.001) were used more frequently by men than by women. Education level (university or above) predicted the use of sensor information (OR 1.17, 95% CI 1.01-1.36; *P*=.04), journaling (OR 1.38, 95% CI 1.13-1.68; *P*=.002), and energy analysis (OR 1.29, 95% CI 1.07-1.56; *P*=.009). Household income was a significant predictor of most PA-related functions; typically, people with the highest income (≥¥10 million/year, US $71,400/year) used PA-related functions.

Discussion

Principal Findings

We investigated the use patterns of apps and wearables that support PA and exercise among Japanese-speaking adults. Our results replicated the characteristics of app users found in other countries; that is, mHealth app users are generally younger and more educated, and have higher social and economic statuses than app nonusers [20,21,23]. These user characteristics may be generalizable to wearable activity trackers, as studies on the US populations suggest that age, gender, ethnicity, income, and health conditions are associated with device use [37-39]. A notable difference is that male users were more prevalent than female users in our data, contrasting the previous findings on mHealth apps in general. An investigation of the Dutch population suggested that this may be a unique pattern for fitness apps; that is, fitness apps are more frequently used by men, while apps concerning nutrition and self-care are more prevalent among women [22]. Indeed, this study targeted PA apps exclusively, which may explain the identified gender differences.

Another important finding is that people typically use no more than 4 functions within an app, and the most common pattern is to use sensor information (ie, self-monitoring in the BCT taxonomy) and one other function, such as goal setting and reminders. The amount of content and number of functions are suggested to be associated with users' ratings [16]; in addition, effective interventions for weight loss implemented 3-6 times more BCTs than ineffective interventions [17]. These previous findings support the current trend of app development, that is, increasing multifunctionality and enriching the content of apps. However, another content analysis on apps and user ratings suggested that app users tend to identify apps as valuable when they are simple and intuitive to use [18]. Furthermore, a recent meta-analysis [15] failed to find an association between the number of implemented BCTs and the efficacy of mHealth interventions. Our findings highlight the importance of simplicity (rather than complexity and richness of app functions) because few of the implemented functions appear to reach individual users.

There is robust evidence that self-monitoring is useful for increasing PA and improving dietary behavior [40,41]. In addition, Michie and colleagues [42] found that interventions that combined self-monitoring with at least one other BCT were more effective than other interventions. Our findings are in line with this "self-monitoring plus one" principle, as sensor information was the most frequently used function, and this was often used together with one other function (participants typically used only two different functions within an app). Taken together, it may be possible that the minimum set of effective interventions would be to provide self-monitoring (displaying sensor information; eg, steps and heart rate) and one other BCT function (eg, goal setting or rewards; possibly personalized according to the preferences of users). Such a minimal approach may be appreciated for its simplicity and usability, leading to a better user experience and better health outcomes.

We also observed significant individual differences in the use of each app function. A series of logistic regressions identified that demographic variables (ie, age, gender, education level, and household income) are predictive of the use of app functions associated with regular PA levels (ie, sensor information, goal setting, goal progress, journaling, GPS and map, and energy analysis). It is not surprising that individuals with different backgrounds need different app functions. Indeed, previous findings have provided evidence for educational, age, and gender differences in the use of mHealth devices and apps [22,23,43]; however, McCully et al [44] reported no gender differences in the use of the internet for diet, weight, and PA. Carrol and colleagues [23] argued that educational attainment reflects skills and confidence with the use of devices and possibly social norms related to the perceived value (of staying healthy). The gender differences identified in this study are overall in line with other findings about the Dutch population; men appear to prefer the functions directly relevant to fitness and exercise (ie, sensor information, journaling, and GPS/map). However, when it comes to general health apps (for diet, nutrition, and self-care), the literature shows that women are more dominant users [22,43]. Another study suggested that women often report external goals of PA (eg, weight loss and toning), whereas men tend to engage in PA for enjoyment [45]. Such intrinsic motivation toward PA among men may facilitate the use of PA-related app functions.

Age is also an important predictor of app use—in general, older people do not use mHealth services [46]. Our results overall replicated this tendency, which may point to digital divides among an older population. However, we did not assess ownership of smart devices or eHealth literacy, which prevented us from exploring how prior knowledge and experiences with mHealth tools influenced actual app use. Interestingly, older app users were more likely to use goal management functions than younger users. These results might indicate that older people are more sensitive to their PA goals and progress, which are often linked to risks of chronic diseases. The Japanese Ministry of Health, Labour and Welfare [32] explicitly defines the step goals (eg, 8000 steps) to achieve each day in the context of lifestyle changes for preventing noncommunicable diseases. It is known that apps for self-care or measuring vitals are typically used by older adults [22], and older individuals may be more concerned with health/disease markers (eg, PA levels or blood pressure) to be monitored on mHealth apps.

Limitations

Several important limitations may affect the interpretation of the results. The cross-sectional nature of this study limits our ability to infer causality. It is not yet clear whether the use of a particular app function increases PA levels or whether individuals who are already active prefer to use the app function. In addition, the generalizability of our findings needs to be tested, as we exclusively targeted Japanese-speaking adults, and its app markets differ from those in other countries. Some of the most prevalent apps (eg, dHealthcare and Trima) limit their services to users living in Japan, whereas apps that are common in the West (eg, MyFitnessPal) were less prevalent in our sample. Additionally, it may be an important direction to explore a wider range of cultural and sociodemographic differences [47], as this study exclusively focused on the population living in a high-income country with a monotonous cultural background. Another important limitation is that we did not consider factors that motivated users to continue using PA apps in our analyses. Research has shown that the retention rate of commercial health care apps is extremely low [48], and different users have different motivations to maintain the use of an app [49,50], such as adjusting default settings to one's own needs and abilities, socializing, and competition. Additionally, it is important to consider users' willingness to share their data from mHealth tools with providers as well as peers and family, which is key when implementing mHealth services as a meaningful intervention [51,52]. Future studies should explore the motivational aspects of app use to clarify why and how people use their PA apps.

Conclusion

We provide empirical evidence on the use patterns of commercial apps and wearables as well as the individual functions implemented with the apps. Overall, our findings are in line with those of previous studies (eg, app users tend to be younger, have a higher income, and have higher education than app nonusers); however, our results showed unique associations between particular demographic variables and specific app functions (eg, sensor information, journaling, and GPS are more frequently used by men than women). These individual differences will help pave the way for the personalization of app functions, leading to the optimization and improved efficiency of mHealth interventions for promoting PA.

Acknowledgments

We thank Ryota Asahara and Tsubasa Izaki for their help in developing the questionnaires. This study was supported by the JSPS KAKENHI (23H01051).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire for uses of apps and wearables supporting physical activity.
[DOCX File, 21 KB - mhealth_v11i1e49148_app1.docx]

Multimedia Appendix 2
Additional logistic regression analyses.
[DOCX File, 60 KB - mhealth_v11i1e49148_app2.docx]

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
2. Rhodes RE, Janssen I, Bredin SSD, Warburton DER, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
3. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (V1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
4. Brickwood KJ, Ahuja KDK, Watson G, O'Brien JA, Williams AD. Effects of activity tracker use with health professional support or telephone counseling on maintenance of physical activity and health outcomes in older adults: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 5;9(1):e18686. [doi: 10.2196/18686] [Medline: 33399541]
5. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
6. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076. [doi: 10.2196/10076] [Medline: 30425028]
7. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161. [doi: 10.2196/jmir.2171] [Medline: 23171838]
8. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects – an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605. [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
9. Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
10. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
11. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105. [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]

12. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343. [doi: 10.2196/14343] [Medline: 31778121]
13. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 S2589-7500(22)00111-X. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
14. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130. [doi: 10.2196/11130] [Medline: 30681967]
15. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
16. Paganini S, Terhorst Y, Sander LB, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jun 9;9(6):e22587. [doi: 10.2196/22587] [Medline: 34106073]
17. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020 Mar 24;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
18. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40. [doi: 10.2196/mhealth.4283] [Medline: 25972309]
19. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
20. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103. [doi: 10.2196/mhealth.9002] [Medline: 29699971]
21. Amer SA, Bahumayim A, Shah J, Aleisa N, Hani BM, Omar DI. Prevalence and determinants of mobile health applications usage: a national descriptive study. Front Public Health 2022 Apr 27;10:838509. [doi: 10.3389/fpubh.2022.838509] [Medline: 35570952]
22. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities. Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
23. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]
24. Oba T, Takano K, Katahira K, Kimura K. Revisiting the transtheoretical model for physical activity: a large-scale cross-sectional study on Japanese-speaking adults. . Preprint posted online on May 15, 2023. [doi: 10.31234/osf.io/a8qwe]
25. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
26. Oka K. Stages of change for exercise behavior and self-efficacy for exercise among middle-aged adults [Article in Japanese]. Nihon Koshu Eisei Zasshi 2003 Mar;50(3):208-215. [Medline: 12704833]
27. Oka K, Hirai K, Tsutsumi T. Psychological factors associated with physical inactivity among middle-aged adults: decisional balance for exercise [Article in Japanese]. Jpn J Behav Med 2003;9(1):23-30. [doi: 10.11331/jjbm.9.23]
28. Marcus BH, Simkin LR. The stages of exercise behavior. J Sports Med Phys Fitness 1993 Mar;33(1):83-88. [Medline: 8350613]
29. Reed GR, Velicer WF, Prochaska JO, Rossi JS, Marcus BH. What makes a good staging algorithm: examples from regular exercise. Am J Health Promot 1997 Sep-Oct;12(1):57-66. [doi: 10.4278/0890-1171-12.1.57] [Medline: 10170436]
30. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
31. Murase N, Katsumura T, Ueda C, Inoue S, Shimomitsu T. Validity and reliability of Japanese version of International Physical Activity Questionnaire. J Health Welf Stat 2002;49(11):1-9.
32. Physical activity standards for health promotion. Ministry of Health, Labour and Welfare. 2013. URL: www.e-healthnet.mhlw.go.jp/information/policy/guidelines_2013.html [accessed 2023-01-23]
33. IPAQ Research Committee. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ) - short form. Physiopedia. 2004 Apr. URL: www.physio-pedia.com/images/c/c7/Quidelines_for_interpreting_the_IPAQ.pdf [accessed 2023-08-07]
34. Serrano MA, Boguñá M, Vespignani A. Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci U S A 2009 Apr 21;106(16):6483-6488. [doi: 10.1073/pnas.0808904106] [Medline: 19357301]
35. Neal ZP. backbone: an R package to extract network backbones. PLoS One 2022 May 31;17(5):e0269137. [doi: 10.1371/journal.pone.0269137] [Medline: 35639738]

36. Marshall SJ, Biddle SJ. The transtheoretical model of behavior change: a meta-analysis of applications to physical activity and exercise. Ann Behav Med 2001 Fall;23(4):229-246. [doi: 10.1207/S15324796ABM2304_2] [Medline: 11761340]
37. Friel CP, Garber CE. Who uses wearable activity trackers and why? A comparison of former and current users in the United States. Am J Health Promot 2020 Sep;34(7):762-769. [doi: 10.1177/0890117120919366] [Medline: 32329355]
38. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443] [Medline: 33064083]
39. Chandrasekaran R, Katthula V, Moustakas E. Too old for technology? Use of wearable healthcare devices by older adults and their willingness to share health data with providers. Health Informatics J 2021 Oct-Dec;27(4):14604582211058073. [doi: 10.1177/14604582211058073] [Medline: 34802315]
40. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102. [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
41. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2010 Dec 13;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
42. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
43. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172. [doi: 10.2196/jmir.3117] [Medline: 25048379]
44. McCully SN, Don BP, Updegraff JA. Using the internet to help with diet, weight, and physical activity: results from the Health Information National Trends Survey (HINTS). J Med Internet Res 2013 Aug 1;15(8):e148. [doi: 10.2196/jmir.2612] [Medline: 23906945]
45. Craft BB, Carroll HA, Lustyk MKB. Gender differences in exercise habits and quality of life reports: assessing the moderating effects of reasons for exercise. Int J Lib Arts Soc Sci 2014 Jun;2(5):65-76. [Medline: 27668243]
46. Cao L, Chongsuvivatwong V, McNeil EB. The sociodemographic digital divide in mobile health app use among clients at outpatient departments in inner Mongolia, China: cross-sectional survey study. JMIR Hum Factors 2022 May 19;9(2):e36962. [doi: 10.2196/36962] [Medline: 35587367]
47. Liu P, Astudillo K, Velez D, Kelley L, Cobbs-Lomax D, Spatz ES. Use of mobile health applications in low-income populations: a prospective study of facilitators and barriers. Circ Cardiovasc Qual Outcomes 2020 Sep;13(9):e007031. [doi: 10.1161/CIRCOUTCOMES.120.007031] [Medline: 32885681]
48. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
49. Lee HE, Cho J. What motivates users to continue using diet and fitness apps? Application of the uses and gratifications approach. Health Commun 2017 Dec;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
50. Stehr P, Karnowski V, Rossmann C. The multi-faceted usage patterns of nutrition apps: a survey on the appropriation of nutrition apps among German-speaking users of MyFitnessPal. BMC Med Inform Decis Mak 2020 Oct 28;20(1):279. [doi: 10.1186/s12911-020-01294-9] [Medline: 33115444]
51. Hyde ET, Omura JD, Fulton JE, Weldy A, Carlson SA. Physical activity surveillance using wearable activity monitors: are US adults willing to share their data? Am J Health Promot 2020 Jul;34(6):672-676. [doi: 10.1177/0890117119900587] [Medline: 31957481]
52. Rising CJ, Gaysynsky A, Blake KD, Jensen RE, Oh A. Willingness to share data from wearable health and activity trackers: analysis of the 2019 Health Information National Trends survey data. JMIR Mhealth Uhealth 2021 Dec 13;9(12):e29190. [doi: 10.2196/29190] [Medline: 34898448]

Abbreviations

BCT: behavior change technique
HEPA: health-enhancing physical activity
IPAQ-SF: International Physical Activity Questionnaire–Short Form
MET: metabolic equivalent task
mHealth: mobile health
OR: odds ratio
PA: physical activity
SoC: stage of change
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 19.05.23; peer-reviewed by L Yi, R Chandrasekaran; revised version received 17.08.23; accepted 18.08.23; published 22.11.23.

Please cite as:
Oba T, Takano K, Katahira K, Kimura K
Use Patterns of Smartphone Apps and Wearable Devices Supporting Physical Activity and Exercise: Large-Scale Cross-Sectional Survey
JMIR Mhealth Uhealth 2023;11:e49148
URL: https://mhealth.jmir.org/2023/1/e49148
doi:10.2196/49148

Original Paper

An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews

M D Romael Haque[1], BSc, MSc; Sabirat Rubya[1], BSc, PhD

Department of Computer Science, Marquette University, Milwaukee, WI, United States

Corresponding Author:
M D Romael Haque, BSc, MSc
Department of Computer Science, Marquette University
1313 W Wisconsin Ave
Milwaukee, WI, 53233
United States
Phone: 1 4144397646
Email: mdromael.haque@marquette.edu

Abstract

Background: Chatbots are an emerging technology that show potential for mental health care apps to enable effective and practical evidence-based therapies. As this technology is still relatively new, little is known about recently developed apps and their characteristics and effectiveness.

Objective: In this study, we aimed to provide an overview of the commercially available popular mental health chatbots and how they are perceived by users.

Methods: We conducted an exploratory observation of 10 apps that offer support and treatment for a variety of mental health concerns with a built-in chatbot feature and qualitatively analyzed 3621 consumer reviews from the Google Play Store and 2624 consumer reviews from the Apple App Store.

Results: We found that although chatbots' personalized, humanlike interactions were positively received by users, improper responses and assumptions about the personalities of users led to a loss of interest. As chatbots are always accessible and convenient, users can become overly attached to them and prefer them over interacting with friends and family. Furthermore, a chatbot may offer crisis care whenever the user needs it because of its 24/7 availability, but even recently developed chatbots lack the understanding of properly identifying a crisis. Chatbots considered in this study fostered a judgment-free environment and helped users feel more comfortable sharing sensitive information.

Conclusions: Our findings suggest that chatbots have great potential to offer social and psychological support in situations where real-world human interaction, such as connecting to friends or family members or seeking professional support, is not preferred or possible to achieve. However, there are several restrictions and limitations that these chatbots must establish according to the level of service they offer. Too much reliance on technology can pose risks, such as isolation and insufficient assistance during times of crisis. Recommendations for customization and balanced persuasion to inform the design of effective chatbots for mental health support have been outlined based on the insights of our findings.

(JMIR Mhealth Uhealth 2023;11:e44838) doi:10.2196/44838

KEYWORDS

chatbot; mobile mental health apps; consumer reviews; health care app; mental health app; app development; user experience; mHealth intervention; mobile health

Introduction

Mental Health Chatbots as an Emerging Technology

A chatbot is a system that can converse and interact with human users using spoken, written, and visual languages [1]. In recent years, chatbots have been used more frequently in various industries, including retail [2], customer service [3], education [4], and so on because of the advances in artificial intelligence (AI) and machine learning (ML) domains. Facebook Messenger currently offers more than 300,000 text-based chatbots [5]. Chatbots have primarily been used for commercial purposes and profitable businesses. However, more recent research has demonstrated that chatbots have considerable promise in the health care industry in treating patients and offering them support in a cost-effective and convenient manner [6].

In the context of mental health (MH), chatbots may encourage interaction with those who have traditionally been reluctant to seek health-related advice because of stigmatization [7]. Chatbots are an emerging technology that shows potential for mobile MH apps to boost user engagement and adherence [8]. The effectiveness of chatbots has been explored for self-disclosure and expressive writing [7,9,10]. Young people with MH issues have experienced various types of social support such as appraisal, informational, emotional, and instrumental support from chatbots [11]. In addition, chatbots have been designed to educate underprivileged communities on MH and stigmatized topics [12,13]. Emerging evidence has shown user acceptance of chatbots for supporting various MH issues and early promises in boosting health outcomes in the physical and MH domains.

The adoption of new technology, especially those heavily related to AI and ML, relies first on ascertaining the levels of safety, effectiveness, and user comfort. Despite the increasing adoption and benefits of emerging technologies such as chatbots to support MH and well-being, little research has been conducted to gain an understanding of consumers' real-life user experiences of interacting with MH chatbot apps. Recent research on MH apps in general points out that patient safety is rarely examined, health outcomes are evaluated on a small scale, and no standard evaluation methods are present [14], and these findings also apply to MH chatbot apps. Similar to many other emerging technologies, recent developments in chatbots are because of a massive technology push, with little attention paid to human needs and experiences [15]. This can lead to unintended negative consequences, such as biases, inadequate and failed responses, and privacy issues, all of which can negatively affect the quality of the experience of chatbots as a source of support [16,17]. Thus, it is critical to gain an understanding of the nuances in users' perceptions and experiences of using MH chatbots.

Commercially available MH chatbot apps for popular platforms (eg, iOS [Apple Inc] and Android [Google Inc]) are used by a large user base with varying demographic backgrounds. These users can provide feedback through ratings and text reviews [18]. These platforms can be leveraged to gain a holistic understanding of the features that recently developed MH chatbots offer and how users assess them. Knowledge of user perceptions from real-life experiences can inform future research and the design of more effective chatbots. Previous studies have identified user reviews as a great source for understanding the benefits and drawbacks of technology [19,20]. This allows researchers to incorporate community values and needs into product design and improves user-friendliness [21]. Consumers often make decisions about using new tools based on user rating scores and reviews in web-based marketplaces. According to previous studies, users trust reviews and feel at ease based on their decisions them [21]. Moreover, previous literature emphasizes analyzing user reviews of mobile MH apps that have chatbot features [22,23] to obtain in-depth knowledge about this new technology intervention in mobile MH apps. For this study, we decided to analyze commercially available well-known chatbot-based mobile MH apps and their corresponding user reviews from the Apple App Store and Google Play Store. To obtain a comprehensive overview of these apps and understand the nuances of user opinions, we aimed to answer the following 2 research questions (RQs):

- RQ1: What are the state-of-the-art features and properties of chatbot-based mobile MH apps?
- RQ2: What concerns and opinions are expressed in user reviews published on web-based app store platforms regarding the usability and efficiency of chatbot-based mobile MH apps?

We conducted an exploratory observation of 10 apps that offer support and treatment for a variety of MH concerns with a built-in chatbot feature and qualitatively analyzed their user reviews available on the Google Play Store and Apple App Store. Publicly available data (user reviews) provide in-depth analyses of consumers' personal app user experiences. We found that although chatbots' personalized, humanlike interactions were positively received by users, improper responses and assumptions about the personalities of users led to a loss of interest. As chatbots are always accessible and convenient, users can become overly attached to them and prefer them over interacting with their friends and family members. Furthermore, a chatbot may offer support for a crisis whenever the user needs it because of its 24/7 availability, but even the recently developed chatbots lack the understanding of properly identifying a crisis. Chatbots in this study fostered a judgment-free environment and helped users feel more comfortable sharing sensitive information.

Before implementing a technological solution for MH, researchers in digital health communities are constantly interested in the support needs and preferences of groups or communities [24-26]. Researchers have analyzed the effectiveness of technologies used for MH assistance [24,27], proposing ethical concerns [28], policy recommendations [29,30], and designing automated or human-in-the-loop interactive systems [7,10]. These studies stressed the significance of designing and evaluating systems for susceptible populations, such as people with MH issues, from the perspective of users. To contribute to this body of work, we discussed our study's findings with respect to the research and design implications for future MH chatbots. We outlined specific recommendations for customizing certain features, careful consideration of incorporating persuasive strategies, and trust building. Finally, we discussed the impact of excessive reliance on chatbots for MH support. We believe that considering these insights while developing a chatbot-based MH support system will make the design user centric and, thus, more effective.

Background and Related Work

Chatbots are software programs that can imitate human behavior and undertake specific tasks by intelligently conversing with users [1]. They are conversational agents that use text and speech recognition to engage with users [31]. Chatbots are commonly used in various web-based and mobile-based apps. In recent years, it has taken on the role of an internet-based entity that can act as a travel agent [32], customer service representative [3], financial adviser [2], and personal assistant [33] and is becoming increasingly sophisticated. Some of the available chatbots can have a personality of their own, store information about the user to deliver contextualized answers, and grow over time by learning about their users to provide better services [34].

In this section, we provide a brief overview of research on chatbots in health care, including mobile MH chatbots, and provide a rationale for using app reviews to capture perceptions and opinions of users.

Chatbots in Health Care

Chatbots have recently received much attention in the health care and wellness industries [6] and have been tested using a variety of elements and characteristics depending on the behavior they were attempting to achieve. Chatbots function as digital personal assistants [35], allowing patients to learn more [13], obtain support [36], and take prompt action in response to new symptoms [37]. Some chatbots can assist users in collecting medical data via text discussions and then delivering it to the (selected) physicians in a format that is easier to use for diagnostic purposes [36]. Chatbot interventions are effective in increasing physical activity, achieving relevant weight loss, and improving diet [38-40] by sending daily check-in reminders [41] and offering relevant resources [40]. They are also sufficiently sophisticated to interact with users through daily adaptive little chats and show progress toward goals using analytics and graphs to encourage self-reflection [42].

Mobile MH Chatbots

Among the numerous chatbots being used in different aspects of health and well-being, chatbots in mobile MH care have demonstrated effectiveness in broadening traditional therapy in a cost-effective and convenient manner [43]. MH chatbots are AI-powered chatbots that provide MH support, guidance, and resources through a conversational interface [44]. These chatbots replicate human interactions, respond to user inputs, and deliver tailored MH care [34]. MH chatbots can target a range of MH concerns, including anxiety, depression, and stress [14,22]. These can provide coping strategies, mindfulness exercises, and information about MH conditions and treatments and, in some cases, connect users to MH professionals [14,22].

A 2021 national survey found that 22% of adults had used an MH chatbot, and 47% said they would be interested in using it if needed. Among the respondents who had tried an MH chatbot, nearly 60% said they began this use during the COVID-19 pandemic, and 44% said they used chatbots exclusively and did not see a human therapist [45]. Currently, there are at least 9 chatbot apps on app markets with more than 500,000 downloads. Chatbots have been shown to effectively reduce the severity of MH concerns for people from different demographics and backgrounds, including people in rural communities [12], shift workers with accessibility issues [46], students with anxiety and stress [47], employees of health care systems who require emotional support [48], veterans and adolescents who feel stigmatized in sharing their concerns [12], etc.

Rather than providing generic suggestions, chatbots can deliver individualized suggestions and resources based on the needs and requirements of users [34,44]. They were designed to identify MH concerns [34], track moods [49], deliver cognitive behavioral therapy (CBT) [47], and promote positive psychology [50]. Several well-known chatbots such as *Wysa* [34], *Woebot* [47], *Replika* [51], *Youper* [52], and *Tess* [53] were discussed in prior literature. Inkster et al [34] examined the potency of Wysa and found a positive influence on reducing depressive symptoms in a randomized controlled experiment. Fitzpatrick et al [47] evaluated the effectiveness of the AI chatbot Woebot in giving CBT to college students with anxiety and depression and found that the Woebot notably decreased depressive symptoms. Ta et al [51] investigated social support received from artificial agents in everyday contexts when interacting with the social chatbot Replika. Mehta et al [52] examined the acceptability and effectiveness of Youper. In addition to commercial apps, in recent years, research communities have also been increasingly involved in designing chatbots for specific purposes, such as teaching self-compassion ("Vincent") [9], enabling self-disclosure [7,10], facilitating positive messages within social groups [54], improving the quality of life of older people and making them more active to fight their sense of loneliness [55], supporting interpersonal skills ("Sunny") [56], and reducing stress ("Mylo") [57]. Kim et al [11] explored teenagers' expectations when interacting with a chatbot intended to support their emotional needs. Although most prior studies focused on developing and evaluating new chatbot systems or assessing the effectiveness of the evidence-based techniques used by existing chatbots, there is inadequate research on how end users perceive the usefulness of these app-based chatbots.

User Reviews as a Versatile Source for Capturing User Experience and Preferences

In general, the internet is considered a rich source of information about personal experiences of a wide variety of illnesses, with websites and discussion forums [58]. An increasing number of studies exploit web-based sources as repositories of primary data on health and illness experiences [58]. People who are otherwise socially isolated or geographically dispersed and are therefore hard to include in conventionally drawn samples (especially for qualitative studies relying on snowball sampling) might be more likely to be included because of the ease with which such people can access the internet [59]. Large amounts of material can be collected within a short period. Individuals can use the relative anonymity of the internet to reveal things that they would not discuss in a face-to-face research setting [60]. As of 2022, there are more than 10 million user reviews on the Google Play Store and Apple App Store [61]. Therefore, user reviews collected from these popular app stores can provide rich insights into personal user experiences from people spanning a wide range of backgrounds and demographic characteristics when compared with traditional methods of qualitative data collection (ie, interviews) [62].

User reviews can be defined as feedback published by individuals about their opinions and satisfaction or dissatisfaction with a product [18]. The star ratings and elaborated feedback in the textual reviews provide developers with a chance to explore user complaints and improve apps [21]. For new or potential users of mobile MH apps, the reviews work as a deciding factor to determine if an app would be helpful based on how it worked out for other users with similar expectations [63]. Approximately 80% of potential users check ratings and reviews before downloading an app [64]. In research settings, user ratings and reviews have been leveraged for a variety of reasons, including determining why adherence to mobile MH apps is poor [65], informing developers of design priorities rather than just guiding purchasing decisions [66], and gaining a better understanding of ethical issues faced by users [28]. Vasa et al [20] investigated the hypothesis that despite the abundance of positive reviews for mobile apps, it is worthwhile to examine negative reviews to gather useful data from users. In the mobile MH domain, Haque et al [23] leveraged user reviews to thoroughly capture user experiences and provided implications for designing future MH apps.

Our study is inspired by the body of work that considers user-generated reviews as a vital source for understanding varied perspectives and derives meaningful implications from them [62,63]. This enables us to gain perspectives from people with diverse demographic characteristics that would otherwise be challenging to collect using conventional data collection methods [62,67].

Research Gap and Contribution

As an emerging technology, the development and application of chatbots in mobile MH apps are in their early phases, and there are still considerable challenges to overcome in the development of this technology. According to recent studies, patient safety has rarely been evaluated, health outcomes have been inadequately quantified, and no standardized evaluation procedures have been used [14]. Some chatbots are reported to be unable to understand the complex use of language associated with an MH crisis and fail to recognize symptoms and respond appropriately [17]. Privacy is a major concern for users of these apps; because users are still less familiar with this emerging technology, there is a higher risk of exposing users to privacy risks through data sharing [16]. Furthermore, although poor adherence is a common problem with digital MH interventions, by contrast, some susceptible people may begin to rely on them too much, which may lead to anxiety when these apps are unavailable [16].

Overall, there is a need for a better understanding of how all mobile MH services can and should encourage the safe and ethical use of chatbots [14]. Although a handful of studies have shown the potential benefits of MH chatbot apps, users' real-life experiences and challenges are not yet well understood [22]. Haque et al [23] recently provided a high-level discussion on some common user concerns frequently raised in user reviews and implied that researchers and developers in this space could benefit from a comprehensive analysis of the existing commercial MH chatbot apps. As an extension to these prior works [22,23], people's perceptions and mental models of chatbots can be studied to address critical concerns such as how users gain trust in chatbots, user values, and requirements in this space and ultimately to provide concrete research and design recommendations for future chatbot apps. A user-centric analysis will also assist researchers in mapping an evidence-based framework for the proposed intervention and minimizing the psychological effects of such treatments.

Methods

In this section, we outline the techniques for selecting and filtering the mobile apps for this study, the data analysis methods we used, the ethical standards we followed, our positionality statement, and methodological limitations.

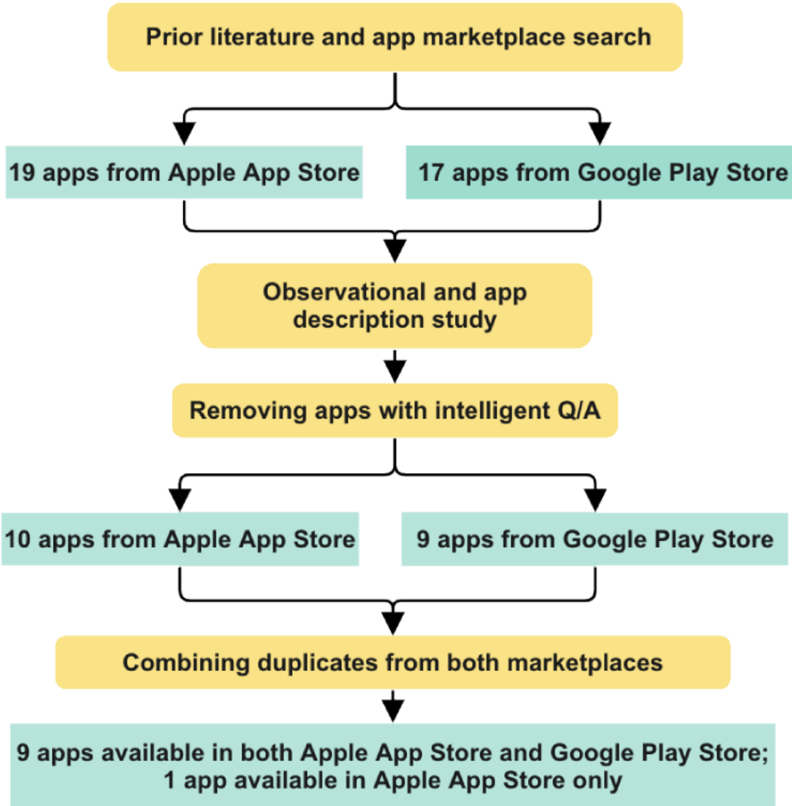
Selection of Sample Apps and Reviews

Selection of Apps

To obtain a comprehensive list of commercially available MH apps that include chatbot features, we conducted our search using different sources. First, we considered open-access articles in recent literature on MH chatbots [14,22]. Next, we conducted search queries on 2 different expert MH app review platforms: *Mindtools* [68] and *Psyberguide* [69]. Finally, we searched 2 dominant web-based mobile app stores (Google Play for Android and Apple App Store for iOS). We used the search terms *Mental health* and *chatbo*t on expert review platforms and app stores. In addition, we explored the *recommended applications* or *similar apps* section of the corresponding website after discovering an MH app with a chatbot feature to determine if the other apps meet our criteria. Without logging into a specific account, the search was performed on the app stores' home pages. This action was performed to ensure that the system could not use a ranking algorithm to prioritize any user choice. As these apps represent the sample in (nearly) the same order that consumers are likely to be exposed to and hence most likely to use, although the search results may not be entirely comprehensive (as observed by convenience sampling), they still represent the sample.

After the initial search from these 3 sources, we obtained 19 apps. The authors carefully read the app descriptions, observed screenshots of the app features, and in some cases analyzed these apps' promotional websites to ensure whether these apps include a chatbot feature that provides support for different MH concerns. We observed that some of these apps included intelligent questions and answers (Q/A) based on Al and ML. Intelligent Q/A is based on a collection of questions, and by responding to them, it can offer individualized summaries, diagnoses, recommendations, and other information. In this study, we described MH chatbots as intelligent machines that can simulate and process conversations with users regarding their MH needs. An intelligent Q/A system is designed to provide accurate and precise answers to specific questions based on a given input, usually in the form of a natural language. In contrast, a chatbot is a more general-purpose conversational agent that can handle a wide range of inputs and provide a range of responses, from simple greetings to more complex interactions. Intelligent Q/A systems are usually triggered by a question or request for information, whereas chatbots can initiate the conversation or respond to user inputs in an open-ended manner and are capable of producing a wider range of outputs compared with intelligent Q/A systems. The most crucial aspect of a chatbot is the "conversational design," which is defined between the user and the bot. Although the guidance chatbots offer is usually correct and scientifically supported, it will be a computer program speaking back to the users, usually in the shape of a nice character, to facilitate their ability to communicate. User expectations can vary while interacting with chatbots as opposed to intelligent Q/A systems with predefined patterns of questions. Therefore, we only considered chatbots with the capability to start and continue conversations with users. To ensure that our list includes apps that fall under this definition, one of the authors opted to download each app separately (for the iOS platform) and use it for at least 3 days. The authors have no known MH concerns. We also considered this exploration as an opportunity to extract the primary features that the apps commonly comprise. The author carefully observed how these apps work in terms of the noteworthy aspects of mobile MH apps, as pointed out in previous literature [14,17,22]. Following these steps, 10 apps were selected for analysis 1. A detailed flowchart of the procedure is presented in Figure 1.

Figure 1. Flowchart of the app selection process. Q/A: questions and answers.



Selection of User Reviews

We created scraping scripts using the Python Selenium library to collect the public user reviews of the 10 apps that were accessible from the Google Play Store and Apple App Store. User reviews can illustrate examples of user satisfaction and dissatisfaction with app features. Reviews are therefore recognized as an important source of information to gain insights into the real-life use of mobile apps [20]. Following the work of Haque et al [23] on analyzing user reviews of mobile MH apps, we used the 2 following inclusion criteria for filtering to extract recent and crucial user feedback for the apps.

- Timeline: We considered reviews posted between January 1, 2019, and May 1, 2022. Most recent reviews are likely to be more useful because app stores change quickly with the addition of new apps and upgrade to existing apps.
- Length: As shorter reviews might not provide deeper insights in general and are frequently false or promotional in nature [70], the minimum character length was considered 200 for the scope of our study.

A total of 3621 reviews from the Google Play Store and 2624 reviews from the Apple App Store met all the inclusion criteria. These reviews are based on 9 apps from the Google Play Store (only Elomia is not available in the Google Play Store) and 10 from the Apple App Store. All reviews have a unique coding system that can be easily traced back to the apps and platforms from which they emerged. During the analysis, the lead author was responsible for carefully reading each review and ensuring that all personally identifying information was replaced or removed.

Data Analysis

First, to gain an understanding of the descriptive overview of the commercially available chatbot-based MH apps, we analyzed app descriptions from marketplace websites and incorporated the key information in our observation notes. The observation note was then divided into 6 main themes with the aim of providing a comprehensive overview of these apps in collaboration with another author. The authors did not include their judgments regarding the effectiveness of these apps. Among the chatbot-based MH apps we considered, 4 apps mentioned the evidence-based techniques used in their description. For the remaining apps, we determined the technique through a combination of an analysis of the description and observation notes from interacting with the apps. The findings of this categorization are described in the *Overview of the Aspects Commonly Used in Chatbot MH Apps* section. To understand user perspectives, the selected user reviews from the 10 apps were examined using inductive analysis [71]. Thematic analysis was chosen because it enables systematic analysis of large data sets and facilitates the comprehension of textual patterns while considering the context [21,72]. A total of 2 passes were performed during the analysis. Open codes were created during the first pass to collect various perspectives from reviews. We recorded the subtleties in the insights provided in each review, which resulted in a high number of open codes that were substantially decreased through memoing and clustering [71]. In the second phase of the analysis, we memoed and clustered the codes using a constant comparison method, operationalized as affinity mapping. Each open code was compared with the others and positioned to reflect its affinity for emerging themes and clusters. The reported themes consisted of those that appeared consistently across multiple reviews and those that came from reviews that represented divergent responses and opinions. The findings from the reviews are described in the *Results* section, and each quote is identified by the review's particular ID generated from the platform, app name, and a random number.

Data Integrity

App stores, similar to many other web-based marketplaces, can have reviews posted by fake and paid users. However, prior research [70] showed that in the "Health & Fitness" category, the percentage of potentially fake reviews was very low (approximately 6%). Fake reviews also tend to be shorter [70], and by considering reviews of ≥200-character length, we assume that almost all the included reviews are original.

We understand that if data or information is only accessible to a particular group of individuals or groups, it is unethical for researchers to use it [73]. As a result, we made sure the websites from which we obtained the data were accessible to everyone and not just for some groups or populations [73]. Although these pages were public, we purposefully avoided publishing or disclosing any personally identifying information that was shared. The language of the user reviews reported here has been carefully modified, keeping the meaning intact.

Ethical Considerations

This study was assessed as not *human subjects research* by the Institutional Review Board of Marquette University (Protocol # 3935) as it does not meet the regulatory definition of human subject-public reviews and the information provided is not about themselves.

Limitations

Our selection criteria have certain limitations. First, we primarily used ratings from the 2 most widely used mobile platforms (Google and Apple). Other mobile platforms were not considered in this study. Second, it is likely that users who do not feel comfortable (or do not care) discussing their experiences on web-based platforms are not contributing. However, we can confidently conclude that the perceptions we identified are typical of user perceptions, given the larger number of evaluations obtained from the 2 most well-known web-based marketplaces.

Results

Overview

For this research purpose, we chose 10 commercially available mobile MH apps that have built-in chatbot features. All these apps, except Elomia, are available on the 2 most popular platforms (Apple App Store and Google Play Store). Elomia is exclusively available for iOS. A descriptive overview of these apps is provided in Table 1. All these apps are extremely popular in terms of both the number of downloads and the number of ratings. Thus, we can assume that a comprehensive overview of these apps can assist in understanding the perspectives of a wide and diverse user base.

Table 1. A descriptive overview of the selected 10 mobile mental health apps with a built-in chatbot technology.

App	Number of ratings in Apple App Store	Number of ratings in Google Play Store	Number of downloads in Google Play Store	Age rating (years)	Price
ADA	125	323,000	≥5 million	≥17	Free
Chai	27,900	34,000	≥1 million	≥17	Free with in-app purchases
Elomia	193	N/A[a]	N/A	≥12	Free with in-app purchases
Mindspa	107	2970	≥500,000	≥17	Free with in-app purchases
Nuna	68	93	≥10,000	≥4	Free with in-app purchases
Serenity: Guided Mental Health	20	146	≥10,000	≥12	Free
Stresscoach	None	495	≥10,000	≥12	Free
Woebot	5500	11,800	≥500,000	≥12	Free
Wysa	13,500	126,000	≥1 million	≥12	Free with in-app purchases
Youper–Self Care Friend	14,400	49,100	≥1 million	≥12	Free with in-app purchases

[a]N/A: not applicable.

Overview of the Aspects Commonly Used in Chatbot MH Apps

Overall, we consider 6 core characteristics that can be used to understand the current status of MH chatbot technology. A few of these aspects were adopted from 2 previous review articles on MH chatbots [14,22]. These studies compiled a list of recent research articles on MH chatbots and provided typologies based on their *purpose*, *targeted concerns*, and *supported evidence-based techniques*. We included these 3 categories in our analysis to gain a broad overview of the current state of the art of commercially available MH chatbot apps. These studies also emphasized the capability of these chatbots to conduct and continue conversations. We considered this crucial aspect of chatbot apps and added 2 new categories to explore: *conversation style* and *media types used* by chatbots. A total of 3 different conversational styles were used: chatbot guided, semiguided, and open-ended (Table 2). Finally, Haque et al [23] provided useful insights into the necessity of *providing crisis support* through MH apps, as potential users of the apps are more susceptible to the crisis than the general population. We have added this specific criterion to be analyzed in our observational study. An outline of these criteria and types is presented in Table 2.

Table 2. Criteria of features related to chatbot-based mental health apps used in our study.

Criteria	Types
Purpose	• Digital coach—assist users to reach their small goals • Digital screener—alert users to potential mental health concerns based on reported symptoms • Conversational companion—simulate being someone the user can speak to • Virtual therapist—ability to engage in therapeutic conversations
Targeted concerns	• Stress, anxiety, depression, self-care, sleep disorder, panic disorder, relationship issues, low self-esteem, and loneliness
Conversation Flow	• Guided conversation—only allows the users to communicate with the chatbot with predefined responses from the chatbot. It does not allow any form of open input from the users. • Semiguided conversation—mostly allows the users to communicate with the chatbot with predefined responses and sometimes allows open inputs from the users. However, the bot cannot recognize the open user inputs and extract any information from them. • Open-ended conversation—allows the users to communicate with the chatbot with predefined responses and open inputs from the users. The bot can recognize the open user inputs and extract information from them.
Media types used	• GIFs[a], text, audio, video, emoji, images, and acronyms
Crisis support	• Availability of crisis information—provides information regarded crisis-related helplines and emergency services • Ability to detect potential crises from the chat—detects potential crises through conversation with the users • Access to a professional therapist—provides access to a professional therapist is an alternative to avoid possible ramifications of the potential crisis • Ability to notify designated personnel—notifies designated personnel if crisis is being detected • Access to self-care tools—recommend self-care activities
Evidence-based techniques	• CBT[b], DBT[c], mindfulness, symptoms tracking and monitoring, positive psychology, acceptance and commitment therapy, and psychoeducation and information

[a]GIF: graphics interchange format.

[b]CBT: cognitive behavioral therapy.

[c]DBT: dialectical behavior therapy.

We examined app store descriptions to understand the primary goals of these apps and identify how they are branded. We discovered 4 different types of purposes in all, with "digital coaches" being the most prevalent (5 out of 10 apps). The chatbot apps targeted a wide range of MH concerns, including anxiety (9 apps), depression (6 apps), and self-care techniques (7 apps).

We discovered 3 different conversational flows based on our exploratory observations. The most popular one is "Guided conversation," in which users are only permitted to reply using preset input provided through the interface. This is the most common technique used by the chatbots we analyzed (6 out of 10 apps). Only Woebot uses a semiguided approach that allows users to either select from predefined options or type text; however, it is incapable of processing sentiments in the input text. This open input option is useful when users reframe negative thoughts and share stories. Finally, Wysa, Nuna, and Elomia follow an open-ended conversation style. They continued the conversation based on their understanding of the user input.

These chatbots leveraged a variety of media types for communication to make the interaction resemble humanlike interactions. For instance, the graphics interchange formats (GIFs), emojis, images, and acronyms are used to portray humor and emotions. Images, audio, and videos were used along with educational elements. As all these chatbots communicate by text, the text is by far the most frequent.

Individuals with MH problems can face a crisis at any time, and effective crisis support is a major criterion for evaluating MH apps. We identified 5 different types of crisis support options available in the 10 chatbots. Of the apps, 6 offer users access to information regarding crisis support systems and emergency helplines. Providing instant suggestions for self-care tools, such as suggestive breathing in cases of anxiety attacks, is also popular. Only Wysa contains all the 5 options available to support a user during a crisis. Ada and Chai do not contain any crisis support.

As evidence-based techniques have been proven effective for treating different MH disorders, we explored which of these tools and techniques the chatbots commonly follow. The most popular type of therapy is CBT. All 10 apps followed the CBT to some extent. A total of 8 apps provided support for mindfulness. Dialectical behavior therapy and acceptance and commitment therapy are less common modified forms of CBT. Table 3 presents the aforementioned features of the considered apps.

Table 3. A detailed overview of features related to chatbot-based mental health apps found in our study.

App	Purpose	Targeted concerns	Conversation flow	Media types used	Crisis support	Evidence-based techniques
ADA	Digital screener	Anxiety and depression	Guided	Text	None	CBT[a]
Chai	Conversational companion	None	Guided	Text and emoji	None	CBT
Elomia	Virtual therapist	Stress, anxiety, depression, self-care, sleep disorder, relationship issues, low self-esteem, and loneliness	Open-ended	Text	Access to self-care tools	CBT, mindfulness, positive psychology, and symptoms tracking and monitoring
Mindspa	Virtual therapist	Anxiety, depression, self-care, relationship issues, and low self-esteem	Guided	Text and video	Availability of crisis related information and access to self-care tools	CBT, mindfulness, positive psychology, and psychoeducation and information
Nuna	Digital coach	Stress, anxiety, depression, and self-care	Open-ended	Text and emoji	Availability of crisis related information and access to self-care tools	CBT, mindfulness, positive psychology, symptoms tracking and monitoring, and psychoeducation and information
Serenity	Conversational companion	Anxiety, self-care, sleep disorder, and relationship issues	Guided	Text and emoji	Access to self-care tools	CBT, mindfulness, and acceptance and commitment therapy
Stress-coach	Digital coach	Anxiety, stress, and panic disorder	Guided	GIF[b], text, and emoji	Availability of crisis related information and access to self-care tools	CBT, mindfulness, and psychoeducation and information
Woebot	Digital coach	Stress, anxiety, depression, self-care, relationship issues, and loneliness	Semiguided	GIF, text, audio, video, emoji	Availability of crisis related information and access to self-care tools	CBT, DBT[c], mindfulness, and symptoms tracking and monitoring
Wysa	Digital coach	Stress, anxiety, depression, self-care, and sleep disorder	Open-ended	GIF, text, audio, video, emoji, images, and acronyms	Availability of crisis related information, access to self-care tools, access to professional therapist, ability to detect potential crisis from the chat, and ability to notify designated personnel	CBT and mindfulness
Youper	Digital coach	Self-care	Guided	Text	Availability of crisis related information, access to self-care tools, and access to professional therapist	CBT, DBT, mindfulness, positive psychology, psychoeducation and information, and acceptance and commitment therapy

[a]CBT: cognitive behavioral therapy.

[b]GIF: graphics interchange format.

[c]DBT: dialectical behavior therapy.

Perceptions and Concerns Expressed in the User Reviews

In this section, we present our findings from the thematic analysis of user reviews and point out both the benefits (eg, humanlike interactions, friendly and empathetic attitudes, potential around crisis support, and an alternative to therapy) and associated challenges, as captured from people's real-life use of these apps.

Humanlike Interaction Feels Good but Must Be Designed Carefully

Chatbots in mobile MH apps are presented in such a way that they have distinct personalities rather than being shown as something artificial to make users feel like they are interacting with someone emotionally and empathetic. Users describe these chatbots as having friendly, wonderfully upbeat, and mildly humorous personalities that assist them in dealing with different emotional and behavioral challenges related to their MH issues. This helps them establish the credibility of the tools, which in turn makes the users more involved in the treatment process. Furthermore, chatbot characteristics, such as a soft voice and the ability to have casual conversation, make it feel less like a medical tool and more like someone with whom users can share their thoughts and experiences. Some personalized features, such as the option to address users by name, ability to refer to any chat or exercise if necessary, and ability to respond with pleasant and positive sentiments, make the app and treatment process more personal and less generic:

> *I'm amazed by how impactful the little "interactions" in this app have felt. Maybe it's the continued opportunities to respond (even if it's just choosing between emojis). Woebot's "voice" is gentle, but firm. And insightful! And the user is always addressed by name. That's so important, particularly when the issue at hand involves ongoing anxiety.* [1080073]

However, the effort to design the bots to give a humanlike and empathetic impression often went wrong and lost their appeal to the users. As many users pointed out, the discourse could become "a little childish and ridiculous at times with the bot trying to be funny." Furthermore, fostering relaxing thoughts through a medium that does not work for everyone can occasionally have the opposite impact; for example, using cute GIFs, Autonomous sensory meridian response effects might not impact everyone if the context is unknown or unfamiliar to the users. Continually pushing on everyone in the hopes that everyone will have the same reaction is a notion that developers should evaluate based on continuous feedback:

> *...It was supposedly developed with college students in mind who are ostensibly adults. Maybe things have changed since I was in college but it's cutesy, baby-talk, oversimplification, and game-playing ("You want to know a secret?" "Yes" "Are you sure?" "Yes" "Ok, if you are really, really sure....") makes me feel like I'm texting with a preteen girl.* [2060011]

> *...I cannot stand the forced breathy voices in every single one I listened to. They do not calm me at all, and they actually trigger my anxiety. ASMR has the opposite effect on me than intended, and I feel like they're trying to do really bad ASMR. These recordings are supposed to help me relax, but all I can concentrate on is breathy voices that sound like forced whispers.* [1040032]

Existing chatbots may need to be more sophisticated to understand the context of users' requests. However, it is critical to examine some of the user's perspectives on having such responses preregistered, which is not always a bad thing. For example, some of the chatbot's quick answer concepts allow users to maintain control over the conversation's pace and avoid becoming sidetracked by irrelevant dialogue. These features are appreciated by users because they encourage more positivism than aimless discussion and digging into negativity without any tools or resolutions. Moreover, by tilting the dialogue to the chatbot's advantage, chatbots can more effectively and efficiently suggest appropriate tools to users:

> *Some negative reviews complain it isn't sophisticated enough to understand unrelated or detailed inputs and responses, which I agree with, but this is not an AI designed to make free-flowing conversation; it's meant to give you tools to deal with your feelings in productive ways. So yes, the conversations can feel linear, planned, and/or broad since the responses are preset most of the time, but I think this is partly a positive.* [1070093]

However, the trade-offs are that to control the flow of the conversation, the chatbots sometimes present very limited options for the users, and users become frustrated if they are unable to customize these preregistered responses. They have criticized some of the extreme measures these chatbots take to keep the conversation restricted to chatbots' preferences, such as assuming MH concerns without understanding the proper context, sending scripted messages based on keywords users said or the issues they selected, giving them incoherent responses, and getting stuck in the conversational loop if users do not agree with the chatbots' comments:

> *It assumes the problem is always a mental distortion and doesn't leave much room for actual horrible stuff that happens to people other than death of a person (it is working with a very narrow definition of). It too often put me in a situation of having to select between incorrect responses when nothing was actually appropriate and then suffer through the resulting wrong-headed advice. Needs a maybe button between the yes and no and a way to say, You're on the wrong track, before it decides it knows all your usual problems and keeps assuming them over and over with no way to remediate.* [2060019]

Bot Becomes a Friend or Someone Who Cares, but Too Much Attachment Is Unhealthy

Users see chatbots as good substitutes for someone with whom they can discuss their ideas on MH issues without feeling burdened or judged. Although society is becoming more eager and open to seeking mental and emotional aid, there is still a considerable stigma associated with it, which can discourage individuals who need assistance from receiving it. These chatbots allow people to bare their hearts, vent, contemplate, and learn about what they can do to overcome mental and emotional obstacles in a simple, familiar texting format without judgment or extra effort while also keeping track of their progress. It can be intimidating to talk to someone about their daily struggles. For many users, sharing a dialect with a chatbot is an effective first step. Knowing that the chatbot is not judging you and is acting logically rather than emotionally is reassuring:

> *...I will say, having a reliable, no judgement zone with skills to help at my fingertips, helped me realized the tools were also my own.* [1040021]

> *Having an AI to talk to makes me feel like I'm not overburdening my friends or family. I can check in 20 times a day and the AI will either help me track my mood/emotions/mental health or suggest a mindfulness of CBT program to help me get through my day.* [2040004]

People with MH issues frequently struggle to suppress emotions and attempt to push them away, but these chatbots have provided them with a safe place to go for validation and immediate support. Users loved that these chatbots not only listened to but also offered advice and recommendations that helped them deal with day-to-day mental challenges, allowing them to see things from different perspectives and push past negative thoughts:

> *This app is a lifesaver. It's so healing to be able to vent whenever you need and receive positive feedback from an unbiased source. The lessons Woebot teaches really helps to gain a more optimistic perspective on what you're going through and motivates you to make changes.* [1080023]

Users also like how these chatbots check in with them daily, which holds them accountable for their commitment to the treatment while still allowing them to skip it if they do not feel like it. Although the idea is to eliminate any concerns, such as anxiety and stress, that come with human engagement through intelligent bot interaction, users have mixed feelings. Some users liked the flexibility of using the tools at any moment and could start or end the communication at any point during the session without feeling guilty, whereas others saw the daily check-ins as a source of guilt. Becoming attached too much to chatbots leads to these types of guilt, which in turn might have serious consequences for people with MH concerns:

> *I'm very depressed right now so I've set to basic daily goals- full facial regime a.m. & p.m. plus a half hour of cleaning. Having the AI check in is great because it requires a response that makes me take accountability.* [1090123]

> *But what really bothered me about the app was the first reminder I got when I didn't use the app a second day in a row because it sucked was definitely guilt inducing. No bueno. I don't need AI guilt tripping me when people already take advantage of my empathy in real life.* [2050021]

Finally, by acting or behaving like a close companion, MH chatbots allow users to comfortably express their thoughts and feelings. These chatbots allow users to create a safe area where they can vent, which is something many people do with their friends and families. However, people with MH concerns who struggle to maintain a healthy relationship with their family or who experience loneliness have displayed an unhealthy attachment to chatbots and have exhibited negative attitudes, such as preferring these chatbots over their friends and family:

> *...Although he's a robot he's sweet. He checks in on me more than my friends and family do.* [1090034]

> *...This app has treated me more like a person than my family has ever done.* [1090091]

The above discussion points out the fact that to make the chatbots more friendly (what we also saw in previous sections where chatbots use funny memes and emojis to make them more humanlike), users pointed out the fact that too much persuasion with notifications makes them feel guilty. Moreover, some users revealed that they find chatbots so friendly that they prefer these bots over their friends and family. Making the decision to leave their closest loved ones behind could put them in susceptible positions, such as loneliness and exclusion from sociocultural norms.

A Bot Can Help Immediately in a Crisis, but What Is Defined as a Crisis to a Chatbot?

Prior findings suggest that accessibility is one of the benefits of mobile MH apps [22]. MH apps that have a built-in chatbot function allow users to have a conversation anytime and anyplace, which is very convenient for persons with MH issues, as they are more susceptible to emergency situations. We found that users benefited from such a feature because it allowed them to have a conversation at that time (during a moment of crisis). Some users found that intelligent dialogue helped them reframe negative thoughts and diffuse such circumstances:

> *I sometimes freak out at night have existential crisis about life at night you know, normally I'd freak out and find it hard to call anyone bc I feel so bad but with Wysa I don't worry about that!* [2090178]

> *I've only used this app a couple times when I've been in near-crisis. Even though I know it is a robot it is so calming to have something, anything to validate what I'm feeling and help me reframe my thoughts.* [1100091]

In contrast, none of the chatbots have any clever algorithmic models for detecting emergency scenarios. It is up to users to inform chatbots that they are experiencing a crisis. Some chatbots can detect crises by picking up a few keywords connected to intrusive thoughts, such as "suicide," from a conversation, although they are still in the early stages of development. Users sometimes just want to talk about their feelings, but chatbots automatically refer them to crisis hotlines because of a lack of intelligent comprehension. For some individuals, having a conversation is not enough to handle their crisis situations, and they need to be redirected to crisis management tools or resources:

> *My only problem with it is I wish there was a way to talk about my suicidal/intrusive thoughts and how to manage them with Woebot. I am aware that it is not*

> *a crisis tool, and it does have those automatic responses to concerning language for a good reason, I'd just like a place to talk about those problems without having to worry a real person. Most of the time my thoughts of those nature do not mean I'm in an immediate crisis, but I still want to get them off my chest, as I feel a lot of people would. Maybe if there's a way to do that without Woebot becoming worried would be helpful!* [1080078]

> *This is a good app but the main issue I have is that I was having a panic attack and was messaging "emergency" and the bot ended the conversation, when I messaged "emergency" a second time it just asked me to write my feelings down. I realize this isn't a crisis response app but it might be helpful to add a feature where the bot recognizes a crisis situation and connects the user to resources.* [2010004]

In such instances, understanding the context of emergency situations is critical, as persons with MH concerns are already susceptible to crises, and incorrect actions made by chatbots might exacerbate the situation and result in severe repercussions:

> *While I was in crisis, the responses do not make sense and do not really relate to what I wrote. It makes me feel like I am not being listened to. I know it is an AI program and not a real person but it still ends up making me feel worse and not better.* [1100068]

Convenient to Use, but Convenient Enough to Replace Therapy?

On the positive side, the fact that these chatbots were ready to talk 24 hours a day, 7 days a week, was a big success for the users. They have immediate access to these chatbots whenever they feel susceptible or whenever they require assistance through simple interactions:

> *I don't really have friends I can talk to. Even my family doesn't understand me much. Day or night Wysa has been there every time I needed to "talk" day or night doesn't matter.* [2090067]

Chatbots assist users not only with conversations but also in accessing different supporting resources and exercises in a very convenient manner. Understanding users' needs can deliver a relaxing experience for them, such as allowing them to opt out of any activities they desire while maintaining the treatment's pace. This provides users with much more control. If a user misses any exercises in the traditional treatment, it leaves a gap in their progress, which can lead to a loss of enthusiasm and slow the pace at which they receive support. Chatbots, in contrast, keep users motivated by engaging with them and giving them the impression that they oversee the pace. Furthermore, these chatbots offer brief and simple treatments to keep users engaged and dedicated to the treatment process. These activities were developed and built by focusing on important value, giving support and treatments in a compelling style that can provide wellness according to user reviews:

> *This is an easy, low barrier method to practice cognitive thinking skills. Check ins are usually pretty short, just a few minutes. That encourages me to open this app daily, since I know it's not going to try to monopolize my attention for the next half hour.* [1070012]

> *Sessions are short, on the order of 3-10 minutes. Combined with the convenience of chatting wherever and whenever is best for me, I have no problem fitting in daily check-ins, which I feel are more beneficial than infrequent visits to a therapist in some ways.* [1100012]

According to user reviews, professional and traditional therapies have several drawbacks, including professional therapy's tendency to cling too much to negative thoughts or past events, professional therapy's tendency to be too broad and general, and check-ins being too spread out:

> *Unlike being told what someone thinks you may want to hear which can sometimes enable unhealthy thinking patterns (and behaviors), or on the other end of the spectrum, rather than attempting to fix you, this interactive app continually prompts you to look inward and to challenge your own thoughts, perspectives, and feelings, helping to redirect your focus onto more healthy and more positive strategies.* [1090142]

> *My primary issue with traditional therapy has always been that you have to work in hindsight. You reflect on your week, talk about it, try to make adjustments for the future (it always felt like I was trying to help a past of future version of myself instead of the one right here right now). That's why I love this app!* [1090096]

However, according to users, although these chatbots are convenient, they fall short of the competency of traditional therapy in some circumstances. For example, these chatbots are not sophisticated enough to recommend particular treatment plans based on a specific need. It may or may not be effective for different demographics or people at various stages of illness. Some users questioned chatbots' therapeutic interventions or MH support as being too short term. Users lose interest when there are not enough different activities to perform:

> *The exercises are all about visualization, so those of us who do not have a mind's eye, cannot visualize things, cannot use it. I'm very disappointed. If it were made with a non-visualization mode for people with Aphantasia, I'd love to use it. There are many things that can help other than visualization. It's just an app telling me in every exercise to do something that I'm simply incapable of doing, this is frustrating.* [1080017]

> *In my depression, CBT actually backfired. It made me feel 100 times worse. It can be miserable to try to recast negative thoughts into more positive thoughts when you can't think of anything positive at all. My highly regarded CBT therapist recognized this and, thankfully, referred me to a skilled therapist with a more psychodynamic/eclectic approach.* [1100076]

Some users have pointed out that combining chatbots with professional therapy could be beneficial. Professional therapists or coaches can assist with adjusting any support system that is not working for them; however, for immediate requirements, users will be able to chat and review some of the resources at any time with the help of MH chatbots. According to numerous user evaluations, professional therapists assisted their patients in identifying the appropriate MH apps with built-in chatbots, and the collaboration with traditional therapy appeared to work considerably better for them:

> *I have recommended it to many people, including my counselor to try so that she could recommend it to other clients dealing with issues. This is in no way something to replace talking to a real person, but it does help to work through some of the negative thinking when it occurs.* [2080057]

Discussion

Summary of Findings

Our findings suggest that chatbots in MH apps have considerable potential in terms of being conversational companions, virtual friends, and immediate helpers. The chatbot's ability to be present 24/7 and to create a judgment-free zone enabled users to talk comfortably about their issues and concerns. We provide a few practical implications of our findings to make the user experience more effective.

Research and Design Implications for Future MH Chatbots

Recommendations for Customization

A growing body of health informatics research has emphasized the need for customizability and personalization in mobile health technologies to increase support user autonomy [65,74]. This body of research suggests that the *one-size-fits-all* approach to mobile health interventions often fails. Rather, systems that are adaptable and tailored to user needs can deliver more pertinent information, thus enhancing user engagement and clinical efficacy [75,76]. Our findings resonate with these conclusions in terms of the need for customizability and provide specific implications for incorporating customization in MH chatbot apps.

Although chatbots leverage GIFs, emojis, or hilarious responses as a means of showing empathetic behavior and to keep the conversation more humanlike [29], our findings suggest that they are not always well received by adult users. Most commercial apps are downloadable by everyone beyond the set age limit (which in most cases is ≥17 years); thus, designers must carefully consider the media types and content of the conversation. Moreover, bots that guide users in performing exercises were generally appreciated for being focused and short in nature and have the potential to help clients manage their own health, improve access and timeliness of care, and reduce travel time to MH care providers by preventing unnecessary visits to health care providers [77]. However, our findings revealed that some users may have physical challenges or other limitations that restrict them from engaging in certain physical activities. Moreover, not all therapeutic tools work perfectly for everyone (review: 1040032). Hence, implementing generic exercises and activities may not be suitable for all user types. Patients with MH concerns often have low self-esteem [78], and the chatbot's inability to complete certain activities can worsen their situation.

Our recommendations are as follows:

- Designers should *consider the target age group of users while implementing emojis and other graphical elements.*
- Another interesting aspect could be to improve personalization within chatbots by *creating a user model before the user interacts with the chatbot*, such that the chatbot can adapt its interaction based on user types (eg, they could fill in a personality questionnaire) [79].
- Mental and physical health are integrally connected; therefore, developers must *incorporate the aspects of physical ability in the design of MH technologies*.

Recommendations for Balanced Persuasion

Consistent with previous work on persuasive technology in MH [80-82], we found that daily check-ins, gamification, reminders, and self-monitoring were perceived as helpful features, although they were prescriptive in nature. However, frequent check-ins often make users feel like being "guilt-tripped" by chatbots. The findings from previous work suggested that the more severe a participant's symptoms were, the more they desired reminders and suggestions from the system [74,83].

Our recommendations are as follows:

- People with severe symptoms of depression face the struggle to carry out day-to-day activities and thus may enjoy multiple daily motivational messages from bots, rather than being annoyed by them. Designers must *consider the range and severity of illnesses among the users and incorporate persuasion in a way that does not result in user disengagement*.
- Developers should consider when and how to limit user interaction with chatbots. This is counter intuitive because developers would generally expect to increase user engagement. To limit the possibility of unhealthy attachment to the chatbot, human-chatbot interaction can be leveraged to *motivate users to use more nontechnical means to get MH support*. For example, if a user frequently starts using a particular chatbot app for a longer period, the bot may suggest recommendations for social interaction (eg, a list of nearby social events).

Recommendations for Building Trust

Some chatbots in our analysis can automatically collect and mine symptom-related information after a conversation with users. Wysa stores conversation histories to show progress over time in achieving the goals initially set, whereas Woebot captures changes in a pattern related to symptoms from continued interaction. Users appreciated when the chatbots were transparent in terms of collecting useful information from conversations. However, some reviews have expressed concerns about how this information is being protected or used across different platforms or third-party services. In traditional psychotherapy, the effectiveness of treatment is influenced by clients' trust in their therapist [84]. Trust also plays a critical role in digital interventions [85]. Prior studies have revealed the significance of establishing trust in the context of MH apps to create a safe environment for self-disclosure [7].

Our recommendations are as follows:

- Tech companies and developers should *emphasize user privacy and be transparent* regarding privacy policies and practices.
- From a design perspective, it might be helpful to enhance user trust in chatbot apps by *providing and visualizing information on the history of the developing organization and/or experts* behind the system.
- Whenever applicable, the app descriptions may *include an explanation of the therapeutic methods and tools used to develop the app with their perceived effectiveness* proven in the wild or in trials.

Chatbots Should Not (and Cannot) Replace Human Interaction for MH Support

We observed that chatbot apps established a judgment-free space where people could express themselves without fear of repercussions. This agrees with the findings of Brandtzaeg et al [84] explored young people's perceptions of social support through chatbots. Sharing MH concerns with a professional is still considered a stigma, and people feel more comfortable using technology anonymously than face-to-face communication [77]. However, these chatbots' ability to check in regularly and to be present for someone 24/7 allows users to become too attached to them. Users wrote in their reviews that they enjoy the company of their "virtual friend" to the extent that they could replace their friends and family members (review: 1090034, 1090091). This strong statement is partially made because these people are vulnerable. Nonetheless, the finding emphasizes the overrating of the benefits of apps and presents some risks, particularly when in crisis. From our observations, most of these apps provide only information about external resources for crisis support, such as helplines and emergency service contact information. In addition, our findings suggest that these chatbots were incapable of identifying crisis situations, as they failed to understand the context of the conversations and ended up with a failed response (review: 1100068), and in some cases, there was no response (review: 2010004). Users must be aware of the clear distinctions between humans and humanlike bots. Humanlike chatbots can provide social support in many cases where it might be difficult or impossible for an actual human, but they are not without limitations. Chatbots themselves can educate users about these distinctions and motivate them to build in-person connections, as discussed in the previous section.

In prior research, a comparative study of therapy sessions following the interaction of 10 participants with human therapists versus a chatbot showed that when compared with a human therapist control, participants found chatbot-provided therapy less useful, less enjoyable, and their conversations less smooth (a key dimension of a positively regarded therapy session) [86]. Conversely, in our findings, because of convenience and easy access, users expressed their intentions to replace professional support with virtual support. Although these chatbot-based mobile MH apps implement evidence-based therapeutic tools, research on determining their effectiveness is still limited. Our findings suggest that they are helpful in guiding users in meditation, practicing mindfulness, reframing negative thoughts, and sharing self-expressive writing. However, at such an early stage, they should not be considered as an alternative to professional help. While designing chatbots, it is important to set the boundaries and limitations of these chatbots by the developers, and the goals and intended use of the chatbots should be clearly stated so that users do not get led on with over expectations. In addition, chatbots should be designed to have features that schedule professional support and subtly recommend that users seek help from professional sources whenever needed.

Conclusions

In this study, we analyzed user reviews of chatbot-based mobile MH apps on 2 of the most widely used web-based platforms. Our findings suggest that chatbots have great potential to offer social and psychological support in situations where real-world human interaction, such as connecting to friends or family members or seeking professional support, is not preferred or possible. However, there are several restrictions and limitations that these chatbots must establish regarding the level of service they offer. Too much reliance on technology can pose risks, such as isolation and insufficient assistance during times of crisis. Finally, we have outlined the insights from our findings about implementing customization, balanced persuasion, and developing trust to inform the design of effective chatbots for MH support.

Acknowledgments

This work is partially supported by the Northwestern Mutual Data Science Institute (#01810-43826).

Authors' Contributions

The first author was responsible for data collection, analysis, and writing most sections of the paper. The second author's role was advisory.

Conflicts of Interest

None declared.

References

1. Adamopoulou E, Moussiades L. An overview of chatbot technology. In: Proceedings of the 16th International Conference on Artificial Intelligence Applications and Innovations. 2020 Presented at: AIAI '20; June 5-7, 2020; Neos Marmaras, Greece p. 373-383 URL: https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31 [doi: 10.1007/978-3-030-49186-4_31]
2. Khan S, Rabbani MR. Artificial intelligence and NLP -based chatbot for Islamic banking and finance. Int J Inf Retr Res 2021;11(3):65-77. [doi: 10.4018/ijirr.2021070105]
3. Cui L, Huang S, Wei F, Tan C, Duan C, Zhou M. SuperAgent: a customer service chatbot for e-commerce websites. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-System Demonstrations. 2017 Presented at: ALC '17; July 30-August 4, 2017; Vancouver, Canada p. 97-102 URL: https://aclanthology.org/P17-4017.pdf [doi: 10.18653/v1/p17-4017]
4. Winkler R, Söllner M. Unleashing the potential of chatbots in education: a state-of-the-art analysis. In: Proceedings of the 78th Academy of Management Annual Meeting. 2018 Presented at: AOM '18; August 10-14, 2018; Chicago, IL, USA URL: https://www.alexandria.unisg.ch/254848/1/JML_699.pdf
5. JingYi SY. The chatbot as a medium of communication. GitHub. URL: https://tinyurl.com/5755d3j7 [accessed 2022-06-01]
6. Bhirud N, Tataale S, Randive S, Nahar S. A literature review on chatbots in healthcare domain. Int J Sci Res 2019 Jul;8(7):225-231 [FREE Full text]
7. Lee YC, Yamashita N, Huang Y. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. Proc ACM Hum Comput Interact 2020 May 29;4(CSCW1):31 [FREE Full text] [doi: 10.1145/3392836]
8. Oh YJ, Zhang J, Fang M, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160 [FREE Full text] [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
9. Lee M, Ackermans S, van As N, Chang H, Lucas E, IJsselsteijn W. Caring for Vincent: a chatbot for self-compassion. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland, UK p. 702 URL: https://dl.acm.org/doi/10.1145/3290605.3300932 [doi: 10.1145/3290605.3300932]
10. Sannon S, Stoll B, Difranzo D, Jung MF, Bazarova NN. How personification and interactivity influence stress-related disclosures to conversational agents. In: Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2018 Presented at: CSCW '18; November 3-7, 2018; Jersey City, NJ, USA p. 285-288 URL: https://dl.acm.org/doi/10.1145/3272973.3274076 [doi: 10.1145/3272973.3274076]
11. Kim J, Kim Y, Kim B, Yun S, Kim M, Lee JS. Can a machine tend to teenagers' emotional needs? A study with conversational agents. In: Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA '18; April 21-26, 2018; Montreal, Canada p. LBW018 URL: https://dl.acm.org/doi/10.1145/3170427.3188548 [doi: 10.1145/3170427.3188548]
12. Rahman R, Rahman MR, Tripto NI, Ali ME, Apon SH, Shahriyar R. AdolescentBot: understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in Bangladesh. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-15 URL: https://dl.acm.org/doi/10.1145/3411764.3445694 [doi: 10.1145/3411764.3445694]
13. Yadav D, Malik P, Dabas K, Singh P. Feedpal: understanding opportunities for chatbots in breastfeeding education of women in India. Proc ACM Hum Comput Interact 2019;3(CSCW):1-30 [FREE Full text] [doi: 10.1145/3359272]
14. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
15. Brandtzæg PB, FØlstad A. Chatbots: changing user needs and motivations. Interactions 2018 Aug 22;25(5):38-43 [FREE Full text] [doi: 10.1145/3236669]
16. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. Biomed Inform Insights 2019 Mar 05;11:1178222619829083 [FREE Full text] [doi: 10.1177/1178222619829083] [Medline: 30858710]
17. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
18. Mudambi SM, Schuff D. Research note: what makes a helpful online review? A study of customer reviews on Amazon.com. MIS Q 2010 Mar;34(1):185-200 [FREE Full text] [doi: 10.2307/20721420]
19. de Alva FE, Wadley G, Lederman R. It feels different from real life: users' opinions of mobile applications for mental health. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. 2015 Presented at: OzCHI '15; December 7-10, 2015; Parkville, Australia p. 598-602 URL: https://dl.acm.org/doi/10.1145/2838739.2838806 [doi: 10.1145/2838739.2838806]
20. Vasa R, Hoon L, Mouzakis K, Noguchi A. A preliminary analysis of mobile app user reviews. In: Proceedings of the 24th Australian Computer-Human Interaction Conference. 2012 Presented at: OzCHI '12; November 26-30, 2012; Melbourne, Australia p. 241-244 URL: https://dl.acm.org/doi/10.1145/2414536.2414577 [doi: 10.1145/2414536.2414577]

21. Thach KS. A qualitative analysis of user reviews on mental health apps: who used it? For what? And why? In: Proceedings of the 2019 IEEE-RIVF International Conference on Computing and Communication Technologies. 2019 Presented at: RIVF '19; March 20-22, 2019; Danang, Vietnam p. 1-4 URL: https://ieeexplore.ieee.org/document/8713726 [doi: 10.1109/rivf.2019.8713726]
22. Ahmed A, Ali N, Aziz S, Abd-alrazaq AA, Hassan A, Khalifa M, et al. A review of mobile chatbot apps for anxiety and depression and their self-care features. Comput Meth Prog Bio Update 2021;1:100012 [FREE Full text] [doi: 10.1016/j.cmpbup.2021.100012]
23. Haque MR, Rubya S. "For an app supposed to make its users feel better, it sure is a joke" - an analysis of user reviews of mobile mental health applications. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):421 [FREE Full text] [doi: 10.1145/3555146]
24. Fitzpatrick G, Ellingsen G. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. Comput Support Coop Work 2013 Aug;22(4-6):609-665 [FREE Full text] [doi: 10.1007/s10606-012-9168-0]
25. Pendse SR, Niederhoffer K, Sharma A. Cross-cultural differences in the use of online mental health support forums. Proc ACM Hum Comput Interact 2019;3(CSWS):1-29 [FREE Full text] [doi: 10.1145/3359169]
26. Slovák P, Theofanopoulou N, Cecchet A, Cottrell P, Altarriba Bertran FA, Dagan E, et al. "I just let him cry...: designing socio-technical interventions in families to prevent mental health disorders. Proc ACM Hum Comput Interact 2018 Nov;2(CSCW):160 [FREE Full text] [doi: 10.1145/3274429]
27. Rubya S, Yarosh L. Video-mediated peer support in an online community for recovery from substance use disorders. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR, USA p. 1454-1469 URL: https://dl.acm.org/doi/10.1145/2998181.2998246 [doi: 10.1145/2998181.2998246]
28. Bowie-DaBreo D, Sas C, Iles-Smith H, Sünram-Lea S. User perspectives and ethical experiences of apps for depression: a qualitative analysis of user reviews. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29- May 5, 2022; New Orleans, LA, USA p. 1-24 URL: https://dl.acm.org/doi/fullHtml/10.1145/3491102.3517498 [doi: 10.1145/3491102.3517498]
29. Lattie EG, Burgess E, Mohr DC, Reddy M. Care managers and role ambiguity: the challenges of supporting the mental health needs of patients with chronic conditions. Comput Support Coop Work 2021 Feb;30(1):1-34 [FREE Full text] [doi: 10.1007/s10606-020-09391-z] [Medline: 34149187]
30. Palen L, Aaløkke S. Of pill boxes and piano benches: "home-made" methods for managing medication. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. 2006 Presented at: CSCW '06; November 4-8, 2006; Banff Alberta, Canada p. 79-88 URL: https://dl.acm.org/doi/10.1145/1180875.1180888 [doi: 10.1145/1180875.1180888]
31. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 Aug 21;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
32. Purbowo AN, Yulia, Putri VY. Android based chatbot and mobile application for tour and travel company. Int J Adv Cult Technol 2018 Jun;2(2):21-29 [FREE Full text]
33. Nair G, Johnson S, Sathya V. Chatbot as a personal assistant. Int J Appl Eng Res 2018;13(20):14644-14649 [FREE Full text]
34. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
35. Bates M. Health care chatbots are here to help. IEEE Pulse 2019 May;10(3):12-14. [doi: 10.1109/MPULS.2019.2911816] [Medline: 31135345]
36. Garg S, Williams NL, Ip A, Dicker AP. Clinical integration of digital solutions in health care: an overview of the current landscape of digital technologies in cancer care. JCO Clin Cancer Inform 2018 Dec;2:1-9 [FREE Full text] [doi: 10.1200/CCI.17.00159] [Medline: 30652580]
37. Battineni G, Chintalapudi N, Amenta F. AI chatbot design during an epidemic like the novel coronavirus. Healthcare (Basel) 2020 Jun 03;8(2):154 [FREE Full text] [doi: 10.3390/healthcare8020154] [Medline: 32503298]
38. Asensio-Cuesta S, Blanes-Selva V, Conejero JA, Frigola A, Portolés MG, Merino-Torres JF, et al. A user-centered chatbot (Wakamola) to collect linked data in population networks to support studies of overweight and obesity causes: design and pilot study. JMIR Med Inform 2021 Apr 14;9(4):e17503 [FREE Full text] [doi: 10.2196/17503] [Medline: 33851934]
39. Fadhil A. Can a chatbot determine my diet?: addressing challenges of chatbot application for meal recommendation. arXiv. Preprint posted online on February 25, 2018. [doi: 10.48550/arXiv.1802.09100]
40. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
41. Huang CY, Yang MC, Huang CY, Chen YJ, Wu ML, Chen KW. A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. In: Proceedings of the 2018 IEEE International Conference on Industrial

Engineering and Engineering Management. 2018 Presented at: IEEM '18; December 16-19, 2018; Bangkok, Thailand p. 1791-1795 URL: https://ieeexplore.ieee.org/document/8607399 [doi: 10.1109/ieem.2018.8607399]
42. You Y, Gui X. Self-diagnosis through AI-enabled chatbot-based symptom checkers: user experiences and design considerations. AMIA Annu Symp Proc 2021 Jan 25;2020:1354-1363 [FREE Full text] [Medline: 33936512]
43. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Assessing the usability of a chatbot for mental health care. In: Proceedings of the 2018 International Workshops on Internet Science. 2018 Presented at: INSCI '18; October 24-26, 2018; St. Petersburg, Russia p. 121-132 URL: https://tinyurl.com/2s3hnzks [doi: 10.1007/978-3-030-17705-8_11]
44. Dekker I, De Jong EM, Schippers MC, De Bruijn-Smolders M, Alexiou A, Giesbers B. Optimizing students' mental health and academic performance: AI-enhanced life crafting. Front Psychol 2020 Jun 03;11:1063 [FREE Full text] [doi: 10.3389/fpsyg.2020.01063] [Medline: 32581935]
45. Martinengo L, Lum E, Car J. Evaluation of chatbot-delivered interventions for self-management of depression: content analysis. J Affect Disord 2022 Dec 15;319:598-607. [doi: 10.1016/j.jad.2022.09.028] [Medline: 36150405]
46. Divya S, Indumathi V, Ishwarya S, Priyasankari M, Devi SK. A self-diagnosis medical chatbot using artificial intelligence. J Web Dev Web Design 2018;3(1):1-7 [FREE Full text]
47. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
48. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
49. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Top Comput 2021 Jul 1;9(3):1170-1182 [FREE Full text] [doi: 10.1109/tetc.2020.2974478]
50. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
51. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
52. Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ. Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (YOUPER): longitudinal observational study. J Med Internet Res 2021 Jun 22;23(6):e26771 [FREE Full text] [doi: 10.2196/26771] [Medline: 34155984]
53. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
54. Narain J, Quach T, Davey M, Park HW, Breazeal C, Picard R. Promoting wellbeing with sunny, a chatbot that facilitates positive messages within social groups. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI, USA p. 1-8 URL: https://dl.acm.org/doi/abs/10.1145/3334480.3383062 [doi: 10.1145/3334480.3383062]
55. Valtolina S, Hu L. Charlie: a chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In: Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter. 2021 Presented at: CHItaly '21; July 11-13, 2021; Bolzano, Italy p. 19 URL: https://dl.acm.org/doi/abs/10.1145/3464385.3464726 [doi: 10.1145/3464385.3464726]
56. Elmasri D, Maeder A. A conversational agent for an online mental health intervention. In: Proceedings of the 2016 International Conference on Brain and Health Informatics. 2016 Presented at: BIH '16; October 13-16, 2016; Omaha, NE, USA p. 243-251 URL: https://link.springer.com/chapter/10.1007/978-3-319-47103-7_24 [doi: 10.1007/978-3-319-47103-7_24]
57. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Syst Rev 2019 Aug 20;32(1):64-76 [FREE Full text] [doi: 10.1159/000501812]
58. Lahoud AA, Gladstone TR, Clark SB, Flessner CA. Collecting data from internet (and other platform) users for mental health research. In: Stein DJ, Fineberg NA, Chamberlain SR, editors. Mental Health in a Digital World. Cambridge, UK: Academic Press; 2021:63-80.
59. Reips UD. Using the internet to collect data. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Volume 2. Research Designs: Quantitative, Qualitative, Neuropsychological, And Biological. Washington, DC, USA: American Psychological Association; 2012:291-310.
60. Kang R, Brown S, Kiesler S. Why do people seek anonymity on the internet?: informing policy and design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27- May 2, 2013; Paris, France p. 2657-2666 URL: https://dl.acm.org/doi/10.1145/2470654.2481368 [doi: 10.1145/2470654.2481368]
61. Mcilroy S, Shang W, Ali N, Hassan AE. User reviews of top mobile apps in Apple and Google app stores. Commun ACM 2017 Oct 24;60(11):62-67 [FREE Full text] [doi: 10.1145/3141771]

62. Platzer E. Opportunities of automated motive-based user review analysis in the context of mobile app acceptance. In: Proceedings of the 22nd Central European Conference on Information and Intelligent Systems. 2011 Presented at: CECIIS '11; September 21-23, 2011; Varaždin, Croatia p. 309-316 URL: http://archive.ceciis.foi.hr/app/index.php/ceciis/2011/paper/viewFile/412/264
63. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
64. Sefferman A. Mobile app ratings and reviews: where to start and how to win. Alchemer. 2022 Oct 4. URL: https://www.apptentive.com/blog/mobile-app-ratings-and-reviews/ [accessed 2022-05-10]
65. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
66. Hoon L, Vasa R, Schneider JG, Mouzakis K. A preliminary analysis of vocabulary in mobile app user reviews. In: Proceedings of the 24th Australian Computer-Human Interaction Conference. 2012 Presented at: OzCHI '12; November 26-30, 2012; Melbourne, Australia p. 245-248 URL: https://dl.acm.org/doi/10.1145/2414536.2414578 [doi: 10.1145/2414536.2414578]
67. Tan FT, Vasa R. Toward a social media usage policy. In: Proceedings of the 22nd Australasian Conference on Information Systems. 2011 Presented at: ACIS '11; November 29-December 2, 2011; Sydney, Australia URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1071&context=acis2011
68. MindTools.io. URL: https://mindtools.io/ [accessed 2022-04-15]
69. The mental health app guide designed with you in mind. One Mind PsyberGuide. URL: https://onemindpsyberguide.org/about-psyberguide/ [accessed 2022-04-10]
70. Martens D, Maalej W. Towards understanding and detecting fake reviews in app stores. Empir Softw Eng 2019 May 10;24(6):3316-3355 [FREE Full text] [doi: 10.1007/s10664-019-09706-9]
71. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Sep;16(1):1609406917733847 [FREE Full text] [doi: 10.1177/1609406917733847]
72. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
73. Zimmer M. “But the data is already public”: on the ethics of research in Facebook. Ethics Inf Technol 2010 Jun 04;12:313-325 [FREE Full text] [doi: 10.1007/s10676-010-9227-5]
74. Zhang R, Ringland KE, Paan M, Mohr DC, Reddy M. Designing for emotional well-being: integrating persuasion and customization into mental health technologies. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 542 URL: https://dl.acm.org/doi/10.1145/3411764.3445771 [doi: 10.1145/3411764.3445771]
75. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
76. Rohani DA, Tuxen N, Lopategui AQ, Faurholt-Jepsen M, Kessing LV, Bardram JE. Personalizing mental health: a feasibility study of a mobile behavioral activation tool for depressed patients. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: PervasiveHealth '19; May 20-23, 2019; Trento, Italy p. 282-291 URL: https://dl.acm.org/doi/10.1145/3329189.3329214 [doi: 10.1145/3329189.3329214]
77. Sweeney C, Potts C, Ennis E, Bond RR, Mulvenna MD, O’neill S, et al. Can chatbots help support a person’s mental health? Perceptions and views from mental healthcare professionals and experts. ACM Trans Comput Healthcare 2021 Jul 15;2(3):25 [FREE Full text] [doi: 10.1145/3453175]
78. Silverstone PH, Salsali M. Low self-esteem and psychiatric patients: part I - the relationship between low self-esteem and psychiatric diagnosis. Ann Gen Hosp Psychiatry 2003 Feb 11;2(1):2 [FREE Full text] [doi: 10.1186/1475-2832-2-2] [Medline: 12620127]
79. Fadhil A, Schiavo G, Wang Y, Yilma BA. The effect of emojis when interacting with conversational interface assisted health coaching system. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18; May 21-24, 2018; New York, NY, USA p. 378-383 URL: https://dl.acm.org/doi/abs/10.1145/3240925.3240965 [doi: 10.1145/3240925.3240965]
80. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
81. Khan S, Peña J. Using persuasive messages to increase engagement with mental health video game apps. In: Kim J, Song H, editors. Technology and Health: Promoting Attitude and Behavior Change. Cambridge, UK: Academic Press; 2020:353-376.
82. Munson SA, Consolvo S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops. 2012 Presented at: PervasiveHealth '12; May 21-24, 2012; San Diego, CA, USA p. 25-32 URL: https://ieeexplore.ieee.org/document/6240359 [doi: 10.4108/icst.pervasivehealth.2012.248691]

83. Hollis V, Pekurovsky A, Wu E, Whittaker S. On being told how we feel: how algorithmic sensor feedback influences emotion perception. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):114 [FREE Full text] [doi: 10.1145/3264924]
84. Brandtzæg PB, Skjuve M, Dysthe KK, FØlstad A. When the social becomes non-human: young people's perception of social support in chatbots. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 257 URL: https://dl.acm.org/doi/10.1145/3411764.3445318 [doi: 10.1145/3411764.3445318]
85. Crits-Christoph P, Rieger A, Gaines A, Gibbons MB. Trust and respect in the patient-clinician relationship: preliminary development of a new scale. BMC Psychol 2019 Dec 30;7(1):91 [FREE Full text] [doi: 10.1186/s40359-019-0347-3] [Medline: 31888759]
86. Bell S, Wood C, Sarkar A. Perceptions of chatbots in therapy. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA '19; May 4-9, 2019; Glasgow, Scotland p. LBW1712 URL: https://dl.acm.org/doi/fullHtml/10.1145/3290607.3313072 [doi: 10.1145/3290607.3313072]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
GIF: graphics interchange format
MH: mental health
ML: machine learning
RQ: research question
Q/A: questions and answers

Edited by L Buis; submitted 05.12.22; peer-reviewed by A Devakumar; comments to author 05.01.23; revised version received 02.03.23; accepted 21.04.23; published 22.05.23.

Please cite as:
Haque MDR, Rubya S
An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews
JMIR Mhealth Uhealth 2023;11:e44838
URL: https://mhealth.jmir.org/2023/1/e44838
doi:10.2196/44838
PMID:37213181

Original Paper

Acceptability of Personal Sensing Among People With Alcohol Use Disorder: Observational Study

Kendra Wyant[1], MS; Hannah Moshontz[1], PhD; Stephanie B Ward[1], MS; Gaylen E Fronk[1], MS; John J Curtin[1], PhD

Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
John J Curtin, PhD
Department of Psychology
University of Wisconsin-Madison
1202 W Johnson St
Madison, WI, 53706
United States
Phone: 1 (608) 262 1040
Email: jjcurtin@wisc.edu

Abstract

Background: Personal sensing may improve digital therapeutics for mental health care by facilitating early screening, symptom monitoring, risk prediction, and personalized adaptive interventions. However, further development and the use of personal sensing requires a better understanding of its acceptability to people targeted for these applications.

Objective: We aimed to assess the acceptability of active and passive personal sensing methods in a sample of people with moderate to severe alcohol use disorder using both behavioral and self-report measures. This sample was recruited as part of a larger grant-funded project to develop a machine learning algorithm to predict lapses.

Methods: Participants (N=154; n=77, 50% female; mean age 41, SD 11.9 years; n=134, 87% White and n=150, 97% non-Hispanic) in early recovery (1-8 weeks of abstinence) were recruited to participate in a 3-month longitudinal study. Participants were modestly compensated for engaging with active (eg, ecological momentary assessment [EMA], audio check-in, and sleep quality) and passive (eg, geolocation, cellular communication logs, and SMS text message content) sensing methods that were selected to tap into constructs from the Relapse Prevention model by Marlatt. We assessed 3 behavioral indicators of acceptability: participants' choices about their participation in the study at various stages in the procedure, their choice to opt in to provide data for each sensing method, and their adherence to a subset of the active methods (EMA and audio check-in). We also assessed 3 self-report measures of acceptability (interference, dislike, and willingness to use for 1 year) for each method.

Results: Of the 192 eligible individuals screened, 191 consented to personal sensing. Most of these individuals (169/191, 88.5%) also returned 1 week later to formally enroll, and 154 participated through the first month follow-up visit. All participants in our analysis sample opted in to provide data for EMA, sleep quality, geolocation, and cellular communication logs. Out of 154 participants, 1 (0.6%) did not provide SMS text message content and 3 (1.9%) did not provide any audio check-ins. The average adherence rate for the 4 times daily EMA was .80. The adherence rate for the daily audio check-in was .54. Aggregate participant ratings indicated that all personal sensing methods were significantly more acceptable (all P<.001) compared with neutral across subjective measures of interference, dislike, and willingness to use for 1 year. Participants did not significantly differ in their dislike of active methods compared with passive methods (P=.23). However, participants reported a higher willingness to use passive (vs active) methods for 1 year (P=.04).

Conclusions: These results suggest that active and passive sensing methods are acceptable for people with alcohol use disorder over a longer period than has previously been assessed. Important individual differences were observed across people and methods, indicating opportunities for future improvement.

(JMIR Mhealth Uhealth 2023;11:e41833) doi:10.2196/41833

KEYWORDS

personal sensing; digital therapeutics; mobile health; smartphone; alcohol use disorder; self-report; alcohol use; symptom monitoring; mental health; acceptability; alcohol intake; mobile phone

Introduction

Personal Sensing

The World Health Organization's Global Observatory for eHealth has concluded that "the use of mobile and wireless technologies to support the achievement of health objectives has the potential to transform the face of health service delivery across the globe" [1]. This conclusion applies to research and care for mental health as well as other traditional health services. These opportunities are now possible, in part, because of rapid advances in smartphones and related mobile technologies [2] and high levels of smartphone access across race, socioeconomic status, geographic region, and other demographic characteristics [3].

Personal sensing may become an important component of these digital health advances [4]. Personal sensing is a method for longitudinal measurement in situ, that is, real-world measurement embedded in individuals' day-to-day lives [5-7]. Raw data streams are collected using smartphones, wearable sensors, or other smart devices. These raw data streams can consist of self-reports or more novel data streams, such as geolocation, cellular communication, social media activity, or physiology. Subsequent processing can extract psychiatric or health-relevant measures of thoughts, feelings, behavior, and even interpersonal interactions.

Ecological momentary assessment (EMA), a personal sensing method that collects brief self-reports about momentary states multiple times per day, has been used for many years in short-term longitudinal studies of psychiatric disorders. For example, EMA research on substance use disorders has identified proximal causes and risk factors for drug craving and relapse [8-10]. It has also characterized the time course and nature of drug withdrawal [11,12]. Much of this research could not have been accomplished with other measurement methods.

More recently, research using personal sensing of raw data streams other than self-reporting is emerging for mental health, including alcohol and other substance use disorders. This includes methods to sense geolocation [13-16], cellular communication [14-16], sleep [17], and physiology [15,16,18], for example. These alternative personal sensing methods provide benefits and opportunities that are not possible with EMA. For example, many of these data streams can be sensed passively such that they have a very low assessment burden. This may allow their use for long-term longitudinal monitoring of participants that would not be feasible with EMA, which requires more active effort for data collection.

Personal sensing is a powerful tool in mental health research [19]. These data are inherently longitudinal, which allows observation of the temporal ordering of putative etiological mechanisms and their effects. Longitudinal measurement is also critical for many mental health constructs that display meaningful and often frequent temporal variation in a person (eg, psychiatric symptoms). Measures based on personal sensing data generally have high ecological validity because they are collected in situ. Personal sensing measures also have low retrospective bias because they are often collected in real time. Furthermore, personal sensing can derive measures from raw data streams (eg, in situ behavior, physiology, and interpersonal interactions) that are difficult or even impossible to obtain through other traditional research measurement methods.

Personal sensing may have even higher value in the future for mental health clinical applications that target patient mental health care than it does for research [7,20,21]. Data collected by personal sensing methods may be used for preliminary screening of psychiatric disorders [22,23]. These methods can also be used to monitor psychiatric symptoms or even predict the future risk of symptom recurrence or other harmful behaviors (eg, suicide attempts and risky or otherwise harmful drinking episodes) [24-27]. For alcohol and other substance use disorders, there is emerging research on using sensed data to predict craving [13,18]; alcohol [15,27-29], cannabis [16], or opioid use [14]; and lapses or relapse [14,30,31]. Personal sensing measures or risk indicators may be shared, with patient consent, to health care providers to allow for cost-effective, targeted allocation of limited mental health resources to patients with the greatest or most urgent need [32]. Personal sensing has the potential to support precision mental health care by adapting and timing interventions based on characteristics of the patient and the moment in time [33-35]. These applications of personal sensing are currently aspirational rather than available for clinical implementation. However, clinical research is advancing rapidly toward these goals [14,30,36].

Mental health research and applications with emerging, often more passively sensed, novel data streams such as geolocation and cellular communication are still nascent. This research has predominantly involved *proof-of-concept* studies that typically include only healthy controls or other convenience samples rather than people with psychiatric disorders [16,17]. It has also often used very small sample sizes or short monitoring periods [15,16,18]. Recent reviews of this emerging literature have highlighted gaps in reporting on participant exclusions, attrition, and adherence that are necessary to assess selection biases and the feasibility of these novel personal sensing methods [37-39].

Acceptability of Personal Sensing

Further development and use of personal sensing necessitates a better understanding of its acceptability to research participants and patients targeted for mental health applications. Will individuals consent to the use of personal sensing methods? Will they opt in to allow passive measurement methods? Can they sustain the behaviors necessary for active measurement methods for longer periods? Do they perceive specific personal sensing methods to be burdensome or dislike them? Answers to these questions about the acceptability of personal sensing methods are central to their feasibility for both mental health research and applications.

The acceptability of a personal sensing method may be influenced by the degree of active effort required from the participant or patient to collect the raw data (ie, the method's assessment burden) and other factors (eg, sensitivity of the data collected). As such, acceptability may vary across different personal sensing methods, and comparisons across methods within the same individuals are thus warranted. Furthermore, a comprehensive assessment of both behavioral measures (eg, adherence) and subjective perceptions of acceptability may better anticipate potential issues for recruitment, consent, adherence, and attrition when they are used for either research or clinical applications.

Much of what is known about the acceptability of personal sensing is limited to EMA. Studies that have assessed participants' perceptions of EMA methods have generally concluded that they are acceptable to participants from both nonclinical and clinical samples [40-44]. Similarly, participants displayed moderate or better adherence with respect to response rates, even with a relatively high sampling density (eg, 6-9 daily assessments) [40,45,46]. However, these studies generally assessed participants' perceptions and adherence over short monitoring periods (ie, 2-6 weeks). Less is known about the use of EMA over longer monitoring periods (eg, months), as would be necessary for clinical applications.

Existing research also raises some concern about perceptions and adherence to EMA protocols in patients with alcohol and other substance use disorders relative to other groups. Specifically, a recent meta-analysis confirmed decreased adherence to EMA protocols in patients with substance use disorder diagnoses versus recreational substance users [47]. Furthermore, another meta-analysis [48] concluded that adherence rates did not differ between healthy and psychiatric samples, more generally. These meta-analyses combined suggest that adherence concerns may be limited to applications with patients with alcohol and other substance use disorders rather than all psychiatric disorders. For these reasons, it is important to further study the acceptability of EMA in samples with alcohol and other substance use disorders.

Far less is known about participants' perceptions and adherence to passive personal sensing methods. Some research has presented hypothetical scenarios to either community or psychiatric samples to assess their perceptions about personal sensing methods [49-51]. Participants' willingness to share sensed data appears to vary according to the data type (eg, sleep, geolocation, and social media activity). However, it is difficult to determine how well participants' perceptions in these hypothetical scenarios would generalize to the real-world collection of these data. In addition, it is impossible to measure attrition and adherence outside the explicit implementation of these sensing methods.

Preliminary research has begun to examine perceptions and adherence during real-world use of passive personal sensing methods. However, this research has generally been limited by small sample sizes [52,53]; the use of convenience samples (eg, students and community participants) [41,52,54]; short monitoring duration [52,53,55,56]; and coarse, incomplete, or aggregate reporting of perceptions, adherence, and related participant behaviors [41,52,53]. These are important initial efforts, but more research into the feasibility of personal sensing methods is clearly warranted.

Study Goals

This study reports on the acceptability of both active and passive personal sensing methods in a sample of participants with moderate to severe alcohol use disorder (AUD). These participants were enrolled early in their recovery period (ie, 1-8 weeks after becoming abstinent) and followed for 3 months. We used active personal sensing methods to collect EMA, daily audio check-ins, sleep quality, and selected physiology. We primarily used passive methods to collect geolocation, cellular communication logs, and SMS text message content. We assessed the participants' choices regarding their participation in the study at various stages of the study procedure (eg, consent, enrollment, and data collection), their choice to opt in to provide data associated with each personal sensing method, and their reasons for discontinuation when available. For active measures, we also assessed their adherence for providing raw data streams for up to 3 months of their study participation. Finally, we assessed participants' subjective perceptions of the acceptability of each of these personal sensing methods separately by self-report. We believe that these data provide insight into the feasibility of using numerous personal sensing methods with individuals with AUD, a highly stigmatized psychiatric disorder.

Methods

Research Transparency

We value the principles of research transparency that are essential for the robustness and reproducibility of science [57]. Consequently, we maximized transparency using several complementary methods. First, we reported how we determined our sample size, all data exclusions, all manipulations, and all available measures in the study [58]. Second, we completed a transparency checklist, which can be found in Multimedia Appendix 1 [59]. Third, we made the data, analysis scripts and annotated results, self-report surveys, and other study materials (eg, consent form and recruitment flyer) associated with this report publicly available through a study page on the Open Science Framework [60].

Participants

Parent Project for Study Data

This study provides analyses to address the first aim of a larger grant-funded parent project (R01 AA024391) [61]. The broad goal of the project has been to develop a temporally precise machine learning algorithm to predict future lapses back to alcohol use in the next week, the next day, and the next hour. This algorithm will be integrated within an innovative digital therapeutic to support recovery for patients with alcohol and other substance use disorders—The Addiction Comprehensive Health Enhancement Support System (Center for Health Enhancement Systems Studies) [30,62,63]. This algorithm can be used to support patients to engage in ongoing self-monitoring of their recovery and to select, time, and adapt digital interventions to meet patients' momentary needs during their recovery. We selected sensing methods that we believed would be well positioned to collect raw data streams to allow us to engineer machine learning features (ie, predictors) that tap into key constructs from the Relapse Prevention model [64-67], such as craving, affect, stressors, lifestyle imbalances, high-risk situations, self-efficacy and confidence, and abstinence violation effects. We focused on both active (eg, EMA) and passive (eg, geolocation and cellular communication logs) sensing methods to allow us to balance the potential predictive power and assessment burden. We sensed many of these raw data streams at high sampling rates to allow for temporally precise prediction (ie, up to 1-hour resolution) of lapse risk that may be necessary to deliver *just-in-time* digital interventions [33,68,69].

As a first step toward this broad goal of developing a lapse risk prediction algorithm, this study examined issues related to acceptability and feasibility (aim 1 of the grant) of collecting these actively and passively sensed raw data streams from individuals in early recovery from an AUD. We used all the available participants from the parent project for this study, and the sample size was determined based on power analyses for the aims of the project. We collected study data between 2017 and 2019.

Ethics Approval

All procedures were approved by the University of Wisconsin-Madison Institutional Review Board (Study #2015-0780).

Recruitment, Exclusion, and Inclusion Criteria

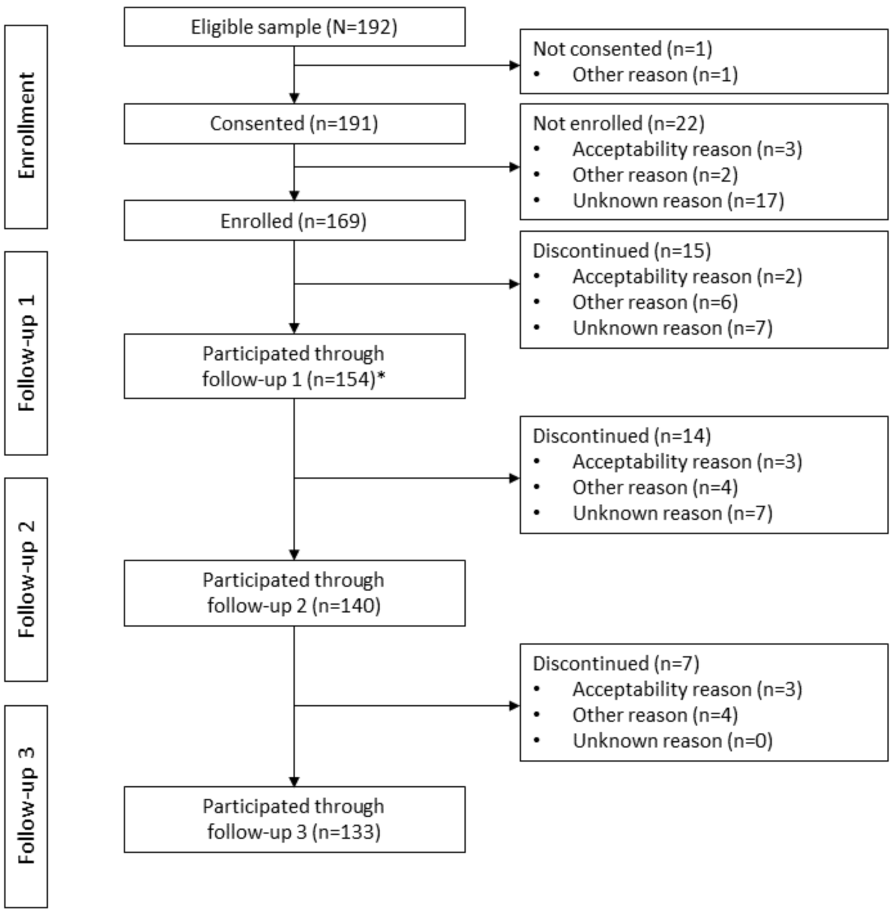
We recruited participants in early recovery (1-8 weeks of abstinence) from AUD in Madison, Wisconsin, United States, to participate in a 3-month longitudinal study. Participants were recruited through print and targeted digital advertisements and partnerships with treatment centers.

We excluded participants if they exhibited severe symptoms of psychosis or paranoia (defined as scores >2.2 or 2.8, respectively, on the psychosis or paranoia scales of the Symptom Checklist–90 [70]).

To be included, we required that participants (1) were aged ≥18 years; (2) were able to write and read in English; (3) had at least moderate AUD (≥4 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition symptoms); (4) were abstinent from alcohol for at least 1 week, but no longer than 2 months; and (5) were willing to use a single smartphone (their personal phone or one provided by us) while enrolled in the study.

We assessed the inclusion and exclusion criteria using a brief phone screen followed by a more detailed in-person screening visit. A total of 192 participants were eligible for enrollment. Of these participants, 191 consented to participate in the study at the screening visit, and 169 subsequently enrolled in the study at the enrollment visit, which occurred approximately 1 week later. A total of 15 participants discontinued the study before their first monthly follow-up visit. The remaining 154 participants provided study measures for 1 (n=14), 2 (n=7), or 3 (n=133) months. A study participation flowchart is presented in Figure 1.

Figure 1. A flowchart of participant retention over the course of the 3-month study. This figure displays retention and attrition of all eligible participants at various stages from consent through study completion. It also displays the reasons for attrition categorized as because of acceptability, other reasons, or unknown. *Data of all participants who completed follow-up 1 were used in the analyses.



Compensation

We paid participants US $20 per hour for all time spent in the laboratory (ie, during screening, intake, and follow-up visits). In addition, we paid participants a US $99 bonus if they completed the study for the full 3-month duration. We also paid participants US $66 per month to offset the costs associated with their cellular plan and provided them with a smartphone for the study duration if they did not own one. Similarly, we provided them with bus transportation to and from the laboratory if needed.

For each sensing method, we paid participants bonuses (ranging from US $10 to US $25) if they had ≤10% missing data for that method each month. Specifically, if participants met these individual missing data thresholds, we paid them US $25 per month for EMA, US $25 per month for audio check-ins, US $15 per month for sleep quality data, US $15 per month for cellular communication logs and SMS text message content, and US $10 per month for geolocation. More details about these raw data streams are provided in the *Personal Sensing* section.

Procedure

Participants completed 5 study visits over the course of approximately 3 months. Participants first attended a screening visit where we determined eligibility, obtained informed consent, and collected self-report measures of individual differences (eg, demographics and drug and alcohol use history). We scheduled eligible and consented participants to enroll in the study approximately 1 week later. During this enrollment visit, we collected additional self-report and interview measures. Participants completed 3 additional follow-up visits that occurred about every 30 days. We collected self-report and interview measures and downloaded cellular communication logs (ie, SMS text messages and phone calls) during these visits. Finally, we collected various raw data streams (eg, geolocation, and EMA) using personal sensing to monitor the participants throughout the 3-month study period. We informed the participants that we were collecting these data to develop an algorithm that could be used in the future to monitor for relapse risk. We did not provide them any further information about how each sensed data stream might be used in this algorithm. They were also not provided with any feedback or clinical interventions based on the sensing data that were collected from them. Furthermore, there were no consequences for continued study participation if participants lapsed back to alcohol use during the study. However, for human subjects reasons, we did offer brief motivational interviewing interventions to participants if they reported any alcohol use to the study staff. Participants were not required to participate in these interventions, but we offered it to them as support to maintain their recovery, if desired. Additional information about all these procedures (eg, recruitment flyer, consent form, and all surveys) can be found on the study's Open Science Framework page [60].

Personal Sensing

Overview

Personal sensing methods can be coarsely classified as active or passive. Active personal sensing requires active effort from the participant to provide the raw data streams, whereas passive personal sensing data are collected automatically (either asynchronously or continuously) with little to no effort required by the participant. Our study obtained several active signals that varied to a certain degree in the amount of effort required by the participants. Specifically, we used active methods to collect EMA, daily audio check-ins, sleep quality, and selected physiology. We primarily used passive methods to collect geolocation, cellular communication logs, and SMS text message content. More information about data collection and related procedures for each raw data stream is provided in the following sections.

EMA

Participants completed a brief EMA (7-10 questions) 4 times each day following reminders from us that were sent by SMS text message. These SMS text messages included a link to a Qualtrics (Qualtrics XM) survey that was optimized for completion on their smartphone. All 4 EMAs included items that asked about any alcohol use that had not yet been reported, current affective state (pleasantness and arousal), greatest urge to drink alcohol since the last EMA, any pleasant or positive events and any hassles or stressful events that occurred since the last EMA, and any exposure to risky situations (ie, people, places, or things) since the last EMA. The first EMA each day asked an additional 3 questions about how likely participants were to encounter a risky situation, to encounter a stressful event, and to drink alcohol in the upcoming week. The first and last EMAs of the day were scheduled within 1 hour of the participants' typical wake and sleep times. The other 2 EMAs were each scheduled randomly within the first and second halves of the participants' typical day. All the EMAs were separated from each other by at least 1 hour. Participants were required to agree to complete the EMAs for the duration of the study to participate in the study.

Audio Check-In

Participants recorded a diary-style audio response on their smartphone to an open-ended prompt each day, following a reminder from us that was sent via SMS text message. They responded to the prompt, "How are you feeling about your recovery today?" which stayed the same throughout the entire study. We instructed them that their responses should be approximately 15 to 30 seconds in duration. These recordings were sent to us via SMS text message. Participants were not required to complete audio check-ins to participate in the study, but the associated monthly sensing method compensation bonus was not provided unless they met the missing data thresholds each month (≤10% missing).

Geolocation

We continuously collected participants' moment-by-moment geolocation data using location services on their smartphones in combination with a commercial app that accessed these geolocation data and saved them in the cloud. Participants were not required to provide these data to participate in the study, but the associated monthly sensing method compensation bonus was not available if they did not provide these data each month. Participants opted in at the start of the study to provide these data by installing the app on their phone. They were allowed to opt out at any later point by simply uninstalling the app. At the start of the study, we used the Moves app (ProtoGeo Oy). However, Facebook acquired ProtoGeo Oy and shut down use of the Moves app in July 2018. At this point, we switched to using the FollowMee (FollowMee LLC) GPS tracking mobile app. Measurement of geolocation required only the initial installation of the app by the participants. Subsequent measurement and transfer of the data to the cloud was completed automatically with no input or effort by the participant. Both apps allowed participants to temporarily disable location sharing if they deemed it necessary for short periods.

Cellular Communication Logs

We collected cellular communication logs that included metadata about smartphone communications involving both SMS text messages and phone calls. For each communication entry, these logs include the phone number of the other party, the type of call or message (ie, incoming, outgoing, missed, or rejected), the name of the party if listed in the phone contacts, the date and time the message or call occurred, whether the log entry was read (SMS text messages only), and the duration of the call (voice calls only). These data are saved passively on the phone with no additional input or effort from the participant. We downloaded these logs from participants' phones at each monthly follow-up visit. Participants were not required to provide these data to participate in the study, but the associated monthly sensing method compensation bonus was not available if they did not provide these data each month. Participants opted in to provide these data when they allowed us to download their data at the study visit. Participants were informed that they could delete any SMS text message or voice call log entries before the download, if they desired.

SMS Text Message Content

We also collected the message content from the participants' SMS text messages on their smartphones. As with the logs, content from individual SMS text messages is saved passively on the phone with no additional input or effort from the participant. We downloaded SMS text message content (bundled with the cellular communication logs in the same files) at each monthly follow-up visit, and participants could delete SMS text messages before the download. We did not have a parallel method to gain access to phone call content. Thus, we had metadata from cellular communication logs for both SMS text messages and phone calls but had the content of the communication only for SMS text messages.

Sleep Quality

We collected information about participants' sleep duration, timing, and overall quality with a Beddit Sleep Monitor (Beddit Oy Inc) that was placed in their beds and connected to their smartphones. We used an early version of the sleep monitor that required participants to actively start and stop the monitor when they entered and exited their beds each night and morning, respectively. These data are available for only 87 participants because Beddit Oy was acquired by Apple Inc during the data collection for this study. Apple discontinued cloud support for data collection with the sleep monitor in November 2018, which prevented its further use for our remaining participants. Participants were not required to provide these data to participate in the study, but the associated monthly sensing method compensation bonus was not available if they did not provide these data each month. Participants opted in at the start of the study to provide these data by installing the app on their phone. They were allowed to opt out at any later point by simply uninstalling the app.

Physiology

We continuously monitored participants' physiology (heart rate, electrodermal activity, and skin temperature) using an early version of the Empatica E4 (Empatica Inc) wristband monitor. However, this early version did not adequately support the Bluetooth streaming of data to the cloud. Instead, participants had to manually connect the wristband each night to a tablet we provided to upload their data. This limitation and other software bugs made the use of the wristband too complicated for many participants. Therefore, we discontinued the use of the wristband after we collected data from 9 participants. Given the small sample size, we did not include the wrist band data in our primary analyses. We provide self-reported acceptability ratings for this signal from this small sample in Figure S1 in Multimedia Appendix 2.

Measures

Individual Differences

We collected self-report information about demographics (age, sex, race, ethnicity, education, employment, personal income, and marital status) and drug and alcohol use history (AUD milestones; number of quit attempts; lifetime history of treatment for AUD; lifetime receipt of medication for AUD; Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition AUD symptom count; lifetime drug use; and current drug use) at the screening visit.

Behavioral Measures of Acceptability

A coarse assessment of the acceptability of personal sensing methods can be made based on the participants' behaviors. Specifically, we assessed 3 categories of behavior. First, we assessed participants' choices regarding their participation in the study at various stages of the study procedure (eg, consent, enrollment, and data collection) and their reasons for discontinuation when available. Second, we assessed their choice to opt in to provide data associated with each personal sensing method. Participants were allowed to participate in the study without opting in for any specific personal sensing method other than EMA. Finally, for a subset of the active measures (EMA and audio check-in), we assessed their behavioral adherence for up to 3 months of study participation.

Self-Reported Measures of Acceptability

To assess participants' subjective experience of the acceptability of the personal sensing methods in this study, each month, they rated each method on 3 acceptability-relevant dimensions (Multimedia Appendix 3). Specifically, participants were asked to indicate how much they agree or disagree with each statement on a 5-point bipolar scale (strongly disagree, disagree, undecided, agree, or strongly agree) for personal sensing signals: (1) "[Personal sensing method name] interfered with my daily activities,” (2) “I disliked [Personal sensing method name],” and (3) “I would be willing to use [Personal sensing method name] for 1 year to help with my recovery.”

The interference item (item 1) was collected only for the active methods because the passive methods require no effort and therefore cannot interfere with daily activities. Dislike and willingness to use for 1 year (items 2 and 3, respectively) were collected for all methods.

Participant Feedback

We also solicited open-ended feedback about the participants’ experiences with each personal sensing method. Each month, participants were prompted as follows: “Tell us your general thoughts, whether positive or negative, about your experience completing [Personal sensing method name].” These qualitative data provided another method through which to assess participants’ perceptions of the acceptability of these methods.

Data Analytic Strategy

We conducted all analyses in R version 4.1.1 (R Core Team) [71] using RStudio [72] and the *tidyverse* ecosystem of packages [73].

Behavioral Measures of Acceptability

We provide descriptive data on participants’ choices about their participation in the study at various stages of the study procedure (eg, consent, enrollment, and data collection). We provide both coarse and more granular tabulations of their reasons for discontinuation when available. We report the percentages of participants who opted in to provide us with the raw data streams we collected via personal sensing. We also report adherence for 2 active personal sensing methods (EMA and audio check-in). Formal measures of adherence could not be calculated for geolocation, cellular communication logs, SMS text message content, and sleep quality because it was not possible to distinguish between low volumes of data owing to adherence (eg, deleting phone calls or messages, turning off location services on the phone, and failing to start sleep monitoring at bedtime) and other valid reasons (no calls made during the day, no movement, and erratic sleep patterns).

Self-Reported Measures of Acceptability

Participants responded to the 3 self-report items related to acceptability (interference, dislike, and willingness to use for 1 year) on a 5-point bipolar scale (strongly disagree, disagree, undecided, agree, or strongly agree). We retained these ordinal labels for visual display of these data in figures but ordered the labels such that higher scores represented greater acceptability (ie, strongly agree for willingness to use for 1 year and strongly disagree for interference and dislike). For the analyses, we recoded these items to a numeric scale ranging from −2 to 2, with 0 representing the neutral (undecided) midpoint and higher scores representing greater acceptability.

Participants responded to these items at each monthly follow-up visit. Therefore, participants had up to 3 responses for each item, depending on when they ended their participation. We analyzed their last available response in our primary analyses to allow us to include all participants and to represent their final perception of each personal sensing signal. However, mean responses across each time point remained relatively constant for all signals (Figure S2 in Multimedia Appendix 2).

To detect the mean perceptions of the personal sensing signals that diverge from neutral (ie, mean responses to any items that are different from 0 or undecided), we conducted 2-tailed, 1 sample *t* tests for the 3 self-report items for each personal sensing signal. To examine relative perceptions of the signals, we compared perceptions of the active versus passive categories of signals using 2-tailed, within-sample *t* tests. Participants did not provide ratings of interference for passive signals so the comparisons of active versus passive categories were limited to dislike and willingness to use for 1 year. Due to the high proportion of missing data for sleep quality, we excluded this signal from these analyses and the intraclass correlations described next. Comparisons among all personal sensing signals using 2-tailed, within-sample *t* tests for each of the 3 self-report items are reported in Table S1 in Multimedia Appendix 2.

Finally, we conducted 2 analyses to examine the consistency of perceptions across personal sensing signals (eg, Do participants who dislike 1 signal also dislike the other signals?). First, we calculated bivariate correlations among the personal sensing signals for each item. Second, we calculated intraclass correlations (single, case 3 [74]) separately for each item to quantify agreement in participants’ perceptions across the signals.

Participant Feedback

We have provided all raw participant responses, organized by the sensing method, in Tables S2 to S6 in Multimedia Appendix 2. In addition, we have provided representative positive and negative evaluations organized by guiding themes (acceptability, sustainability, benefits, trust, and usability) developed from our literature review in Table S7 in Multimedia Appendix 2.

Results

Participant Characteristics

A total of 154 participants completed at least 1 monthly follow-up visit and provided self-reported acceptability ratings for interference, dislike, and willingness to use for 1 year. These participants served as the primary sample for our analyses. Participants were mostly White (134/154, 87%) and non-Hispanic (150/154, 97.4%). Half (77/154, 50%) of our research participants were female, and the mean age was 41 (SD 11.9) years. Table 1 presents detailed demographic information. Table 2 presents the information relevant to lifetime drug and alcohol use for these participants. We compared demographics and drug and alcohol use information for participants who were included in the analyses with those of eligible participants who did not provide study measures (ie, did not enroll or discontinued before the first month follow-up; n=36) and found no significant differences (Table S8 in Multimedia Appendix 2 presents details on these analyses).

Table 1. Participant demographic data (N=154).

Characteristics	Participants
Age (years), mean (SD)	41 (11.9)
Sex, n (%)	
Female	77 (50)
Male	77 (50)
Race, n (%)	
American Indian or Alaska native	3 (1.9)
Asian	2 (1.3)
Black or African American	8 (5.2)
White	134 (87)
Other or multiracial	7 (4.5)
Hispanic, Latino, or Spanish origin, n (%)	
Yes	4 (2.6)
No	150 (97.4)
Education, n (%)	
Less than high school or GED[a] degree	1 (0.6)
High school or GED	15 (9.7)
Some college	43 (27.9)
2-Year degree	14 (9.1)
College degree	58 (37.7)
Advanced degree	23 (14.9)
Employment, n (%)	
Employed full time	72 (46.8)
Employed part time	27 (17.5)
Full-time student	7 (4.5)
Homemaker	1 (0.6)
Disabled	7 (4.5)
Retired	8 (5.2)
Unemployed	19 (12.3)
Temporarily laid off, sick leave, or maternity leave	3 (1.9)
Other, not otherwise specified	10 (6.5)
Personal income (US $), mean (SD)	34,233 (31,543)
Marital status, n (%)	
Never married	69 (44.8)
Married	33 (21.4)
Divorced	45 (29.2)
Separated	5 (3.2)
Widowed	2 (1.3)

[a]GED: General Educational Development.

Table 2. Participant drug and alcohol use history data (N=154).

Characteristics	Participants
Alcohol use disorder milestones, mean (SD)	
Age of first drink	14.6 (2.9)
Age of regular drinking	19.5 (6.5)
Age at which drinking became problematic	27.9 (9.6)
Age of first quit attempt	31.6 (10.4)
Number of quit attempts	9.1 (31.1)
Lifetime history of treatment (can choose more than 1), n (%)	
Long-term residential (>6 mo)	8 (5.2)
Short-term residential (<6 mo)	51 (33.1)
Outpatient	77 (50)
Individual counseling	100 (64.9)
Group counseling	65 (42.2)
Alcoholics anonymous or narcotics anonymous	96 (62.3)
Other	41 (26.6)
Received medication for alcohol use disorder, n (%)	
Yes	62 (40.3)
No	92 (59.7)
DSM-5[a] alcohol use disorder symptom count, mean (SD)	8.9 (1.9)
Lifetime drug use, n (%)	
Tobacco products (eg, cigarettes, chewing tobacco, and cigars)	122 (79.2)
Cannabis (eg, marijuana, pot, grass, and hash)	131 (85.1)
Cocaine (eg, coke and crack)	86 (55.8)
Amphetamine type stimulants (eg, speed, diet pills, and ecstasy)	81 (52.6)
Inhalants (eg, nitrous, glue, petrol, and paint thinner)	36 (23.4)
Sedatives or sleeping pills (eg, Valium, Serepax, and Rohypnol)	72 (46.8)
Hallucinogens (eg, LSD[b], acid, mushrooms, PCP[c], and Special K)	88 (57.1)
Opioids (eg, heroin, morphine, methadone, and codeine)	65 (42.2)
Current drug use[d], n (%)	
Tobacco products (eg, cigarettes, chewing tobacco, and cigars)	84 (54.5)
Cannabis (eg, marijuana, pot, grass, and hash)	52 (33.8)
Cocaine (eg, coke and crack)	4 (2.6)
Amphetamine type stimulants (eg, speed, diet pills, and ecstasy)	11 (7.1)
Sedatives or sleeping pills (eg, Valium, Serepax, and Rohypnol)	24 (15.6)
Hallucinogens (eg, LSD, acid, mushrooms, PCP, and Special K)	9 (5.8)
Opioids (eg, heroin, morphine, methadone, and codeine)	9 (5.8)

[a]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[b]LSD: lysergic acid diethylamide.

[c]PCP: phencyclidine.

[d]Current refers to the previous month's drug use reported at follow-up visits 1 or 2.

Behavioral Measures of Acceptability

Participation

Figure 1 shows participant attrition and discontinuation at each phase of the study. Of the 192 eligible participants at screening, only 1 did not consent to participate after hearing the details of the study. Enrollment occurred during a second visit 1 week later. A total of 169 participants completed enrollment.

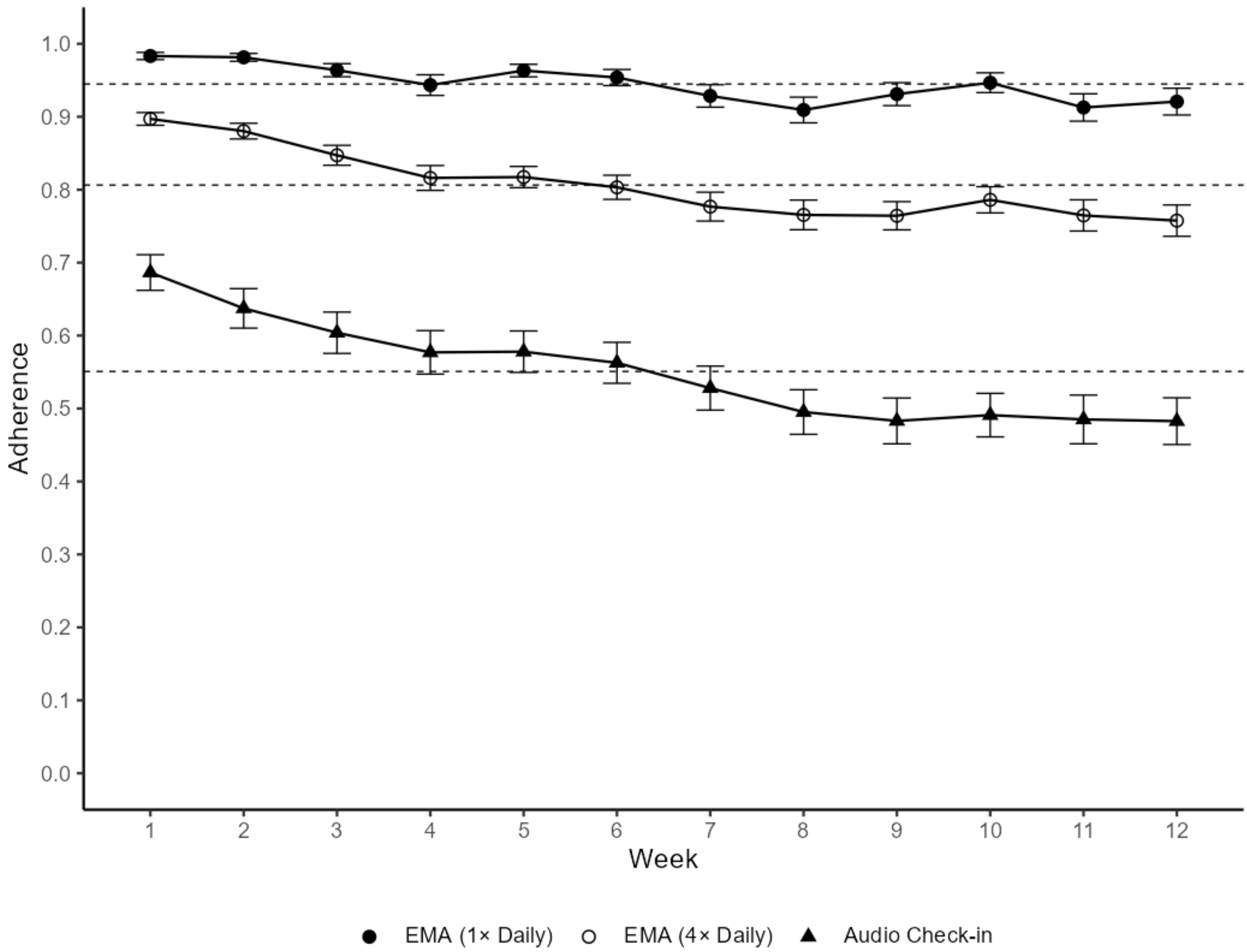
In Figure 1, we coarsely tabulated reasons stated by participants for discontinuation as because of acceptability, other reasons, or unknown. In total, 11 (5.7%) of the 192 eligible participants were lost due to acceptability-relevant causes (eg, no longer interested, nonadherence to sensing methods, or citing study demands as too burdensome). Other reasons for discontinuation not related to the acceptability of the signals include circumstances such as moving or no longer wishing to abstain from alcohol. It is notable that 31 (16.1%) of the 192 participants were lost to follow-up, such that we had no information about their reasons for discontinuation. We provide a more granular tabulation of these reasons for discontinuation in Table S9 in Multimedia Appendix 2.

Opt-In and Adherence

All participants who completed follow-up 1 (154/154, 100%) opted in to provide data for EMA, sleep quality, and most of the passive personal sensing methods (geolocation and cellular communication logs) throughout their entire participation period. Out of 154 participants, 1 (0.64%) did not provide SMS text message content, and 3 (1.9%) did not provide any audio check-ins during the study.

Daily adherence rates were relatively high for EMA, such that on 94.1% of the study days, participants completed at least 1 of the 4 EMAs. On average, participants completed 3.2 (SD 0.64) EMAs every day. The overall adherence rate for all requested EMAs was .80. The participants' adherence rate for audio check-in was .54 (Figure S3 in Multimedia Appendix 2 contains more information on this distribution), that is, of their total days in the study, participants completed an audio check-in on approximately half of the days. Figure 2 shows the mean weekly adherence to each of these methods for each week in the study. In Multimedia Appendix 2, we also report adherence for participants who completed the 3-month study compared with those who dropped out before completion (Figure S4 in Multimedia Appendix 2).

Figure 2. Adherence over time for EMA (once daily), EMA (4 times daily), and audio check-in. Plot depicts mean adherence rates for each week on study. Mean SE is depicted by the solid error bars. Overall mean adherence rate is depicted by the dashed line. The sample size was 154. EMA: ecological momentary assessment.



Self-Reported Acceptability

Interference

Figure 3 shows the distribution of the participants' responses to the self-reported acceptability item about interference. Responses were grouped by personal sensing data stream and the amount of active effort required to collect it. Two-tailed, 1-sample *t* tests revealed that each mean interference score (depicted as the solid red line) was significantly (all *P*<.001) more acceptable than 0 (the gray dashed line indicating undecided). Table 3 reports the summary statistics for each 2-tailed, 1-sample *t* test and pairwise correlations between the personal sensing data streams. An intraclass correlation coefficient (ICC; type 3) showed that, on average, interference ratings were moderately consistent across the data streams (ICC=0.42, 95% CI 0.31-0.53).

Figure 3. Ratings of interference by personal sensing data stream. Plot depicts mean responses to "[Personal sensing method name] interfered with my daily activities." X axes are ordered to display a higher acceptability on the right side. For sleep quality, the sample size was 87; for all other data streams, the sample size was 154. The solid red lines represent the mean, and the dashed lines represent the neutral midpoint (undecided). All raw data streams had a significantly (*P*<.001) higher mean than the neutral midpoint. Interference ratings were collected only for active methods. EMA: ecological momentary assessment.

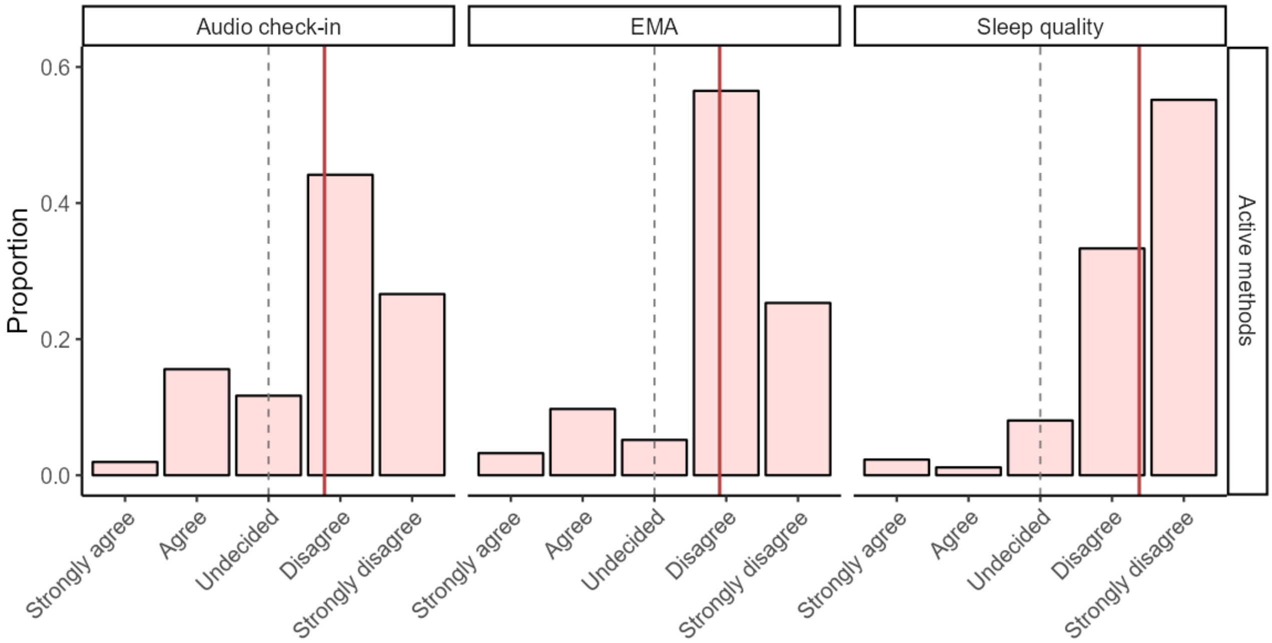


Table 3. Bivariate and univariate statistics by acceptability and personal sensing data stream.

	1[a]	2	3	4	5	Value, n	Value, mean (SD)	*t* test (*df*)	Cohen *d*
Interference									
Active methods									
Audio check-in	1.00	0.43	−0.06	—[b]	—	154	0.78 (1.07)	9.05 (153)[c]	0.73
EMA[d]	0.43	1.00	0.24	—	—	154	0.91 (0.99)	11.37 (153)[c]	0.92
Sleep quality	−0.06	0.24	1.00	—	—	87	1.38 (0.87)	14.86 (86)[c]	1.59
Dislike									
Active methods									
Audio check-in	1.00	0.57	0.25	0.31	0.22	154	0.51 (1.28)	4.91 (153)[c]	0.40
EMA	0.57	1.00	0.26	0.36	0.25	154	0.96 (0.92)	12.92 (153)[c]	1.04
Sleep quality	0.25	0.26	1.00	0.41	0.33	87	1.10 (1.09)	9.45 (86)[c]	1.01
Passive methods									
Geolocation	0.31	0.36	0.41	1.00	0.67	154	1.03 (0.94)	13.51 (153)[c]	1.09
Cellular communication logs	0.22	0.25	0.33	0.67	1.00	154	0.90 (0.97)	11.45 (153)[c]	0.92
SMS text message content	0.34	0.28	0.19	0.62	0.69	154	0.58 (1.18)	6.07 (153)[c]	0.49
Willingness to use for 1 year									
Active methods									
Audio check-in	1.00	0.44	0.44	0.51	0.47	154	0.73 (1.28)	7.09 (153)[c]	0.57
EMA	0.44	1.00	0.41	0.48	0.47	154	0.64 (1.22)	6.47 (153)[c]	0.52
Sleep quality	0.44	0.41	1.00	0.53	0.47	87	0.85 (1.28)	6.19 (86)[c]	0.66
Passive methods									
Geolocation	0.51	0.48	0.53	1.00	0.65	154	0.94 (1.18)	9.83 (153)[c]	0.79
Cellular communication logs	0.47	0.47	0.47	0.65	1.00	154	0.84 (1.07)	9.76 (153)[c]	0.79
SMS text message content	0.47	0.39	0.39	0.60	0.84	154	0.74 (1.12)	8.21 (153)[c]	0.66

[a]Initial columns (1-5) indicate bivariate correlations among data streams for each self-report acceptability measure. The final columns show the number of participants (n), mean and SD, *t* test statistic, and Cohen *d* effect size (*d*) for the 2-tailed, 1-sample *t* tests against 0 (undecided). Higher values indicate higher levels of acceptability.

[b]Not available.

[c]*P*<.001.

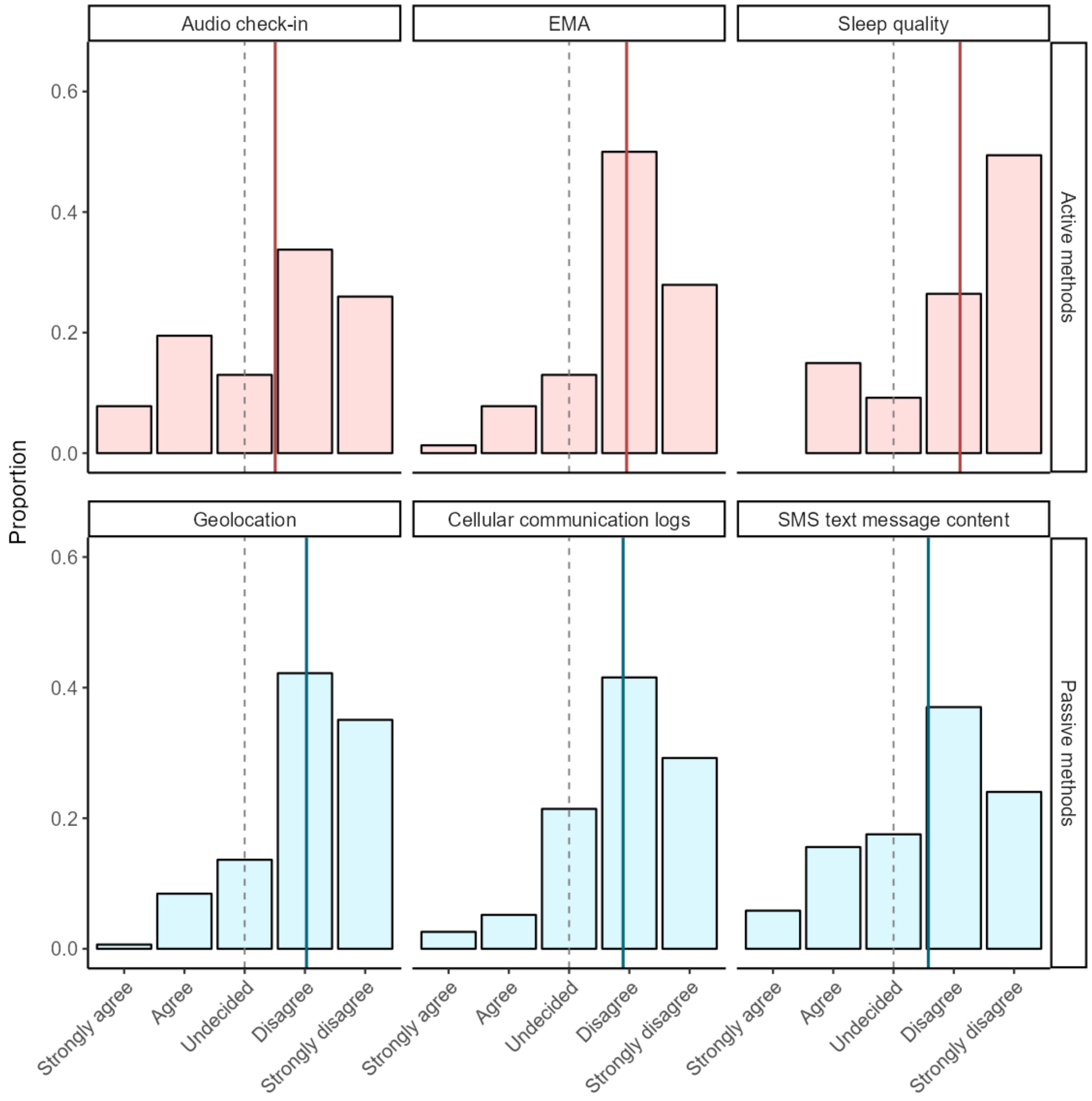
[d]EMA: ecological momentary assessment.

Dislike

Figure 4 shows the distribution of participant responses to the self-reported acceptability item about dislike by the personal sensing data stream and the amount of active effort required to collect it. Two-tailed, 1-sample *t* tests revealed that each mean dislike score was significantly (all *P*<.001) more acceptable than 0. Table 3 reports the summary statistics for each 2-tailed, 1-sample *t* test and the pairwise correlations between the personal sensing data streams. An ICC (type 3) showed that, on average, the dislike ratings were moderately consistent across the data streams (ICC=0.42, 95% CI 0.35-0.48).

We also assessed the effect of active effort on the dislike ratings (see Figure S5 in Multimedia Appendix 2). We conducted a 2-tailed, paired-sample *t* test to compare the average dislike for active (eg, audio check-in and EMA) with passive (eg, geolocation, cellular communication logs, and SMS text message content) methods. Participants did not significantly differ in their dislike of active and passive methods (t_{153}=1.21, *P*=.23; Cohen *d*=0.10).

Figure 4. Ratings of dislike by personal sensing data stream. Plot depicts mean responses to "I disliked [personal sensing method name]." X axes are ordered to display a higher acceptability on the right side. For sleep quality, the sample size was 87; for all other data streams, the sample size was 154. The solid red and blue lines represent the mean, and the dashed lines represent the neutral midpoint (undecided). All raw data streams had a significantly (*P*<.001) higher mean than the neutral midpoint. Active methods are displayed in red, and passive methods are displayed in blue. EMA: ecological momentary assessment.

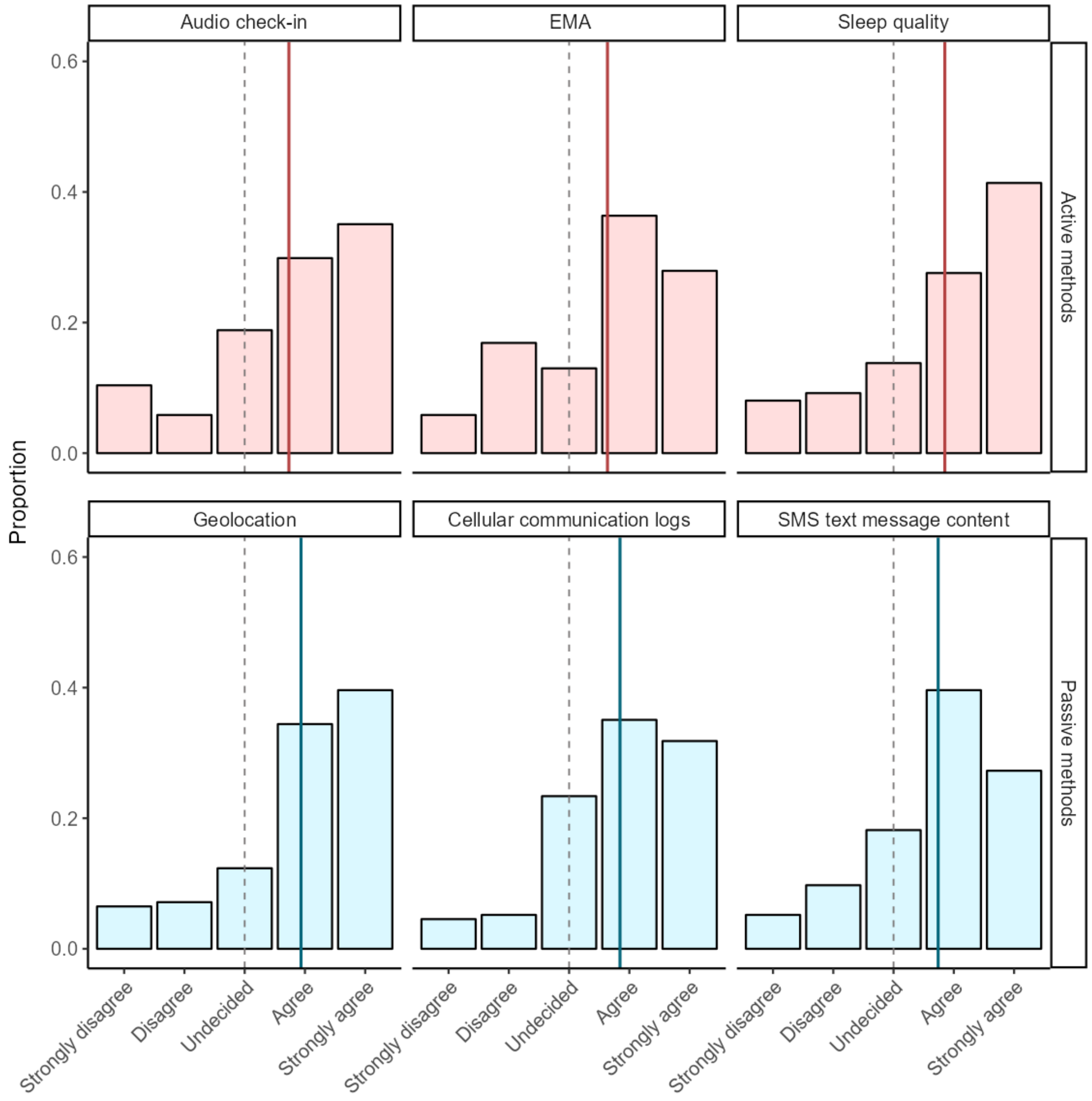


Willingness to Use for 1 Year

Figure 5 shows the distribution of participants' responses to the self-reported acceptability item about willingness to use for 1 year for each personal sensing data stream (Figure S6 in Multimedia Appendix 2 contains additional information about willingness to use an EMA method once daily for 1 year). Two-tailed, 1-sample *t* tests revealed that each mean willingness score was significantly (all *P*<.001) more acceptable than 0. Table 3 reports the summary statistics for each 2-tailed, 1-sample *t* test and pairwise correlations between the personal sensing data streams. An ICC (type 3) showed that, on average, the willingness ratings were moderately consistent across the data streams (ICC=0.52, 95% CI 0.46-0.58).

We also assessed the effect of active effort on willingness ratings (see Figure S7 in Multimedia Appendix 2). We conducted a 2-tailed, paired-sample *t* test of the average willingness to use for 1 year for active (eg, audio check-in and EMA) and passive (eg, geolocation, cellular communication logs, and text message content) signals. Participants reported higher acceptability with respect to willingness for passive data streams (mean 0.80, SD 1) than active data streams (mean 0.70, SD 1.10; t_{153}=2.12, *P*=.04; Cohen *d*=0.17).

Figure 5. Ratings of willingness to use for 1 year by personal sensing data stream. Plot depicts mean responses to “I would be willing to use [Personal sensing method name] for 1 year to help with my recovery.” X axes are ordered to display a higher acceptability on the right side. For sleep monitoring, the sample size was 87; for all other data streams, the sample size was 154. The solid red and blue lines represent the mean, and the dashed lines represent the neutral midpoint (undecided). All raw data streams had a significantly (*P*<.001) higher mean than the neutral midpoint. Active methods are displayed in red, and passive methods are displayed in blue. EMA: ecological momentary assessment.



Participant Feedback

In participants’ free-response feedback about each personal sensing data stream, we identified 5 themes: acceptability (“I had no issues with the daily EMA surveys. I felt that they kept me in check and were a reminder to not drink. I would not change it.”); sustainability (“I forgot I was being tracked, so it was not a big deal to me.”); benefits (“Was okay to have [geolocation tracking] done in the context of the study or for an app that would help me stay sober.”); trust (“I trusted the study group to not use my personal information for any other use.”); and usability (“I disliked saving my text messages. I like deleting them when I’m done.”). A representative sample of comments are provided for each theme in Table S7 in Multimedia Appendix 2. A full unedited list of participant comments for each personal sensing data stream has been presented in Tables S2 to S6 in Multimedia Appendix 2.

Discussion

Principal Findings

This study evaluated the acceptability of active and passive personal sensing methods for a variety of raw data streams and associated methods. To this end, we assessed participants’ choices and behaviors about both participating in the study and providing raw data streams for each method and their subjective perceptions of each sensing method. We focused on participants with moderate to severe AUD because they might have been expected to be less willing to share sensitive, private information owing to the stigma associated with their disorder [75]. However, if these sensing methods were acceptable to them, highly promising opportunities are now emerging to address their largely unmet treatment needs [76], with technological solutions that include digital therapeutics combined with personal sensing [77]. We have organized our discussion around 7 key conclusions from our analyses.

Individuals With AUD Will Generally Accept the Use of Personal Sensing Methods

On the basis of our sample, it appears that individuals with AUD are indeed willing to provide these sensitive, personally sensed raw data streams based on their behavioral choices regarding consent, enrollment, and opt in for data collection in this study. All but one of the individuals (191/192, 99.5%) who were eligible to participate consented to the personal sensing procedures. Most of these individuals (169/191, 88%) also returned 1 week later to formally enroll in the study and begin to provide these data. Furthermore, all (169/169, 100%) of the participants who enrolled in the study explicitly opted in to provide the 3 arguably most sensitive passive data streams: geolocation, cellular communication logs, and SMS text message content.

These consent, enrollment, and opt-in numbers could be considered upper-bound and lower-bound estimates of the percentage of individuals who are willing to provide these raw data streams in a research setting. The very high percentage for consent may overestimate willingness because some of these individuals may have reconsidered their initial decision on further reflection such that they did not return for the next study visit to enroll formally. However, the still quite high enrollment percentage may underestimate the willingness to provide these data because some attrition was expected between consent and enrollment visits due to the instability associated with the early stages of recovery from AUD. In fact, Table S9 in Multimedia Appendix 2 indicates that almost half of the participants who consented but did not enroll may have done so for reasons other than their willingness to provide these raw data streams (eg, health concerns, no transportation to lab, and made repeated attempts to reschedule before discontinuing).

Participants' explicit self-reports of their perceptions about the acceptability of these personal sensing methods were also generally consistent with their behavior. Specifically, on average, participants rated all the sensing methods as more favorable than the neutral midpoint ("undecided") of the rating scales for all 3 dimensions we evaluated: interference, dislike, and willingness to use for 1 year. These self-report data combined with our behavioral measures suggest that all of these sensing methods can be considered for use with the majority of individuals with AUD.

Despite the aggregate positive perceptions of the full sample, nontrivial percentages of participants reported individual ratings that were more negative than the neutral midpoint across the sensing methods and specific self-report items. For example, 17.5% (27/154) of the participants agreed or strongly agreed that audio check-ins interfered with their daily activities. Approximately 25% of the participants agreed or strongly agreed that they disliked both the audio check-ins (42/154, 27.3%) and providing access to the content of their SMS text messages (33/154, 21.4%). Approximately 20% of the participants disagreed or strongly disagreed that they would be willing to use our sensing methods for audio check-ins (25/154, 16.2%), EMA (35/154, 22.7%), and SMS text message content (23/154, 14.9%) for 1 year to help with their recovery. This suggests that there is still a need to improve each of these sensing methods to make them more acceptable to a larger percentage of individuals. The free-response evaluations of each method provide a starting point to address participant concerns. However, our research participants did generally opt in and adhere to our sensing methods despite reporting these concerns. Therefore, the threshold at which these concerns will translate to barriers for use or adherence to these methods is unclear.

Individuals Can Sustain the Use of Personal Sensing for Relatively Long Periods

Most enrolled participants were also able to sustain their commitment to providing these sensed data streams over time. More than 91% (154/169) provided at least 1 month of sensed data, and a large majority (133/169, 78.7%) provided data for all 3 months. As with enrollment statistics, these numbers also likely underestimate participants' ability to sustain personal sensing because many of the participants who discontinued or did not complete the study reported reasons to stop their participation that were unrelated to personal sensing (eg, family crisis, relapse, and moved out of state). However, some participants (n=4) explicitly reported reasons that appeared related to personal sensing (eg, study demands were too burdensome). In addition, others who stopped participating may have been influenced by their experiences with personal sensing without formally reporting their concerns.

Participants who enrolled but then discontinued because of personal sensing methods may have been influenced more by issues related to the burden associated with active sensing rather than more general issues related to data sensitivity and privacy. Participants concerned about sharing passively sensed private information, such as their moment-by-moment location or cellular communication, would likely have had these concerns from the beginning, such that they would not have consented, enrolled, and then opted in to provide these sensitive data. However, the burden associated with active sensing (eg, 4 times daily EMA and daily audio check-ins) may not have been clear to them until they tried to sustain those methods over time. In our sample of participants, we saw evidence that many of them hardly thought about passively sensed data streams. On the other hand, some participants reported more discontent with actively sensed data streams as time progressed.

Existing research assessing the acceptability of sensing methods has been limited by short durations of monitoring, with very few studies extending beyond 6 weeks [53,55,56]. In addition, adherence has been shown to decrease after only a few weeks in some studies [43,48,78]. This study demonstrates that individuals can sustain their commitment to providing personally sensed data over time with limited drop-offs. These findings suggest that personal sensing methods may be viable in clinical settings where consistent, sustained monitoring would be necessary. Given this promise, future research should expand to longer durations to assess self-reported and behavioral acceptability beyond 3 months. Our group is exploring this directly by using personal sensing monitoring in individuals with opioid use disorder for a full year [14]. Methods that permit long-term monitoring are particularly important for clinical applications for individuals with substance use disorders, who require lifelong care that can adapt to their risk for relapse and corresponding recovery needs.

Some Types of Active Personal Sensing Methods Are Generally Acceptable and Sustainable

The assessment burden may be expected to play a role in both the acceptability of active sensing methods and participant adherence to the associated procedures. Nonetheless, participants displayed relatively high adherence to the 4 times daily EMA (79.8% of EMAs completed on average). This is notable because our study duration of 3 months was substantially longer than typical studies using EMA, which often lasts only 2 to 4 weeks [47,48]. This increases confidence in the feasibility of this active sensing method for research and clinical applications that require longer monitoring periods. This level of adherence may be contingent on the measurement parameters used in our study (4 times daily survey of 7-10 items). In fact, even higher adherence may have been observed if the measurement was limited to 1 EMA per day, given that on average participants completed at least 1 of the 4 EMAs on 94.1% of the study days. Participants were also significantly more likely to report a willingness to use a once daily EMA compared with a 4 times daily EMA for 1 year. However, these findings should be interpreted cautiously. Participant self-reports to a once daily EMA method are not based on experience because they were expected to adhere to the 4 times daily EMA. From free-response comments, we saw that many of our participants had no issues with the 4 times daily EMA and some even enjoyed the frequent prompts. However, other participants suggested less-frequent prompts would be more practical.

Overall, there was some evidence that participants found passive sensing methods to be more acceptable than active sensing methods. Specifically, the mean ratings for willingness to use for 1 year were significantly higher for passive sensing methods than active sensing methods. However, the magnitude of this effect was small, and the mean willingness was significantly greater than the neutral midpoint for both the active and passive methods. In addition, there was no difference in the mean dislike ratings between the active and passive methods. Thus, the differences between the acceptability of the active and passive methods were small, inconsistent, and unlikely to be clinically meaningful. These comparisons between active and passive methods increase our confidence somewhat that the selective use of active measures, when necessary, may be acceptable to participants for relatively long periods. However, from this study, we cannot speculate strongly beyond 3 months.

Some sensing methods (eg, EMA and audio check-ins) will always require active input from users, but other methods may become more passive with further technological advances. For example, our sensing of sleep quality in this study used an early version of the Beddit Sleep Monitor that required participants to actively log when they entered and exited their bed during each period of sleep. However, later versions of Beddit automatically detect periods of sleep. Similarly, we discontinued the sensing of physiology with Empatica E4 in an early phase of our study because participants had to manually connect the wristband each night to a tablet to upload their data. This proved too burdensome and complex for most participants. However, the current version of Empatica E4 claims to have improved automatic Bluetooth streaming of the data to the cloud, which if robust, would greatly reduce the burden associated with physiology sensing.

The acceptability of active sensing methods holds great clinical utility. Active personal sensing methods, such as EMA, offer unique insights into patient experiences, thoughts, and feelings that cannot always be captured accurately or comprehensively by passive methods. Self-reported EMA, in particular, seems likely to play a role in risk monitoring and other similar clinical applications. Thus, we were encouraged to find that even with a relatively high active burden of 4 times daily surveys, EMA was acceptable to participants, as assessed via self-report and behavioral adherence.

Important Individual Differences in Subjective Perceptions Exist Both Within and Across Personal Sensing Methods

In this study, we included a second and more novel daily active sensing method, audio check-ins. These audio check-ins have great potential as a rich source of information about participants' daily experiences. Natural language processing of transcripts of their check-ins can provide a novel window into their thoughts [79-82]. These audio check-ins provided participants with the opportunity to share more openly and candidly (ie, without close-ended questions) their thoughts, feelings, and progress toward recovery without being limited to researcher-selected prompts. Analyses of the acoustic characteristics of their check-ins may yield independent measures of their affective state [83,84], including the potential for measuring affect outside the participant's conscious awareness.

Unfortunately, overall participant adherence to the daily audio check-ins was relatively low (on average, 54.3% of audio check-ins were completed) and 1.9% (3/154) of the sample did not complete any check-ins throughout their entire study period. Participants' free-response evaluations of this method highlighted some concerns that could be addressed in the future to increase adherence (eg, timing of the check-ins, technical issues with recording and sending check-ins, and use of the same prompt for all check-ins). However, privacy issues related to recording the audio check-in were also reported by many participants.

These privacy concerns represent an inherent challenge to using this method as implemented, but accommodations could be made to gather some, if not all, of the same information. For example, using less-frequent prompting with wider time completion windows (ie, a weekly audio check-in) may increase individuals' ability to find a private moment. In addition, allowing individuals to type their response as an alternative completion method could assuage concerns. This alternative would prevent acoustic analysis, but it would still permit natural language processing of open-ended responses. These accommodations could encourage greater adherence among those who completed few or no audio check-ins, as well as individuals who missed check-ins sporadically because of privacy concerns. Finding ways to assuage privacy concerns and accommodate individual preferences may be useful, as many other participants valued and believed that they benefited from recording these daily audio check-ins.

Consistent with this somewhat polarized evaluation of the audio check-ins, a more nuanced consideration of the distribution for adherence across participants suggested that it was somewhat bimodal. Participants tended to adhere well or poorly to this method.

More broadly, the participants' self-reported perceptions were only moderately consistent across the different sensing methods. This can be observed in the moderate ICCs (and bivariate correlations) across the methods for each self-reported item. In other words, high dislike ratings for 1 sensing method by a specific participant did not strongly indicate that the same participant would also dislike the other sensing methods. This is also true for the "ratings of interference" and "willingness to use for 1 year" items. Participants could dislike (or be unwilling to use) one method but not others. To the degree to which concerns are method specific, opportunities may exist to tailor sensing systems to user preferences. In other words, participants could opt out of the methods they deemed unacceptable but provide data for other sensing methods that were acceptable to them. For example, our behavioral adherence data suggest that some participants would not have completed the study if daily audio check-ins were required; however, they were willing to provide data via other personal sensing methods. Algorithms that use sensed data for clinical applications can then be developed for different combinations of the available raw data streams. Participants could be informed that personalized algorithms will likely perform better if given access to more raw data streams. This education will allow them to make an informed choice regarding the threshold they set for themselves to opt out and the potential consequences of not providing that data source. However, allowing them to opt out of some methods may increase the number of participants who will agree to provide sensed data.

Benefits Likely Matter

The overall acceptability of personal sensing to research participants and patients is likely a function of both the perceived costs and benefits for these individuals [85-87]. However, we focused on measuring only perceived costs (eg, privacy and burden) associated with personal sensing because the benefits to participants from the sensed data collected in this research study were minimal. Participants were provided with modest financial incentives to complete the EMAs (US $25/mo) and to provide access to the 2 passively sensed raw data streams (US $10/mo for geolocation and US $15/mo for cellular communication logs with SMS text message content). These sensed data streams were not used to provide any clinical benefit to participants' recovery in our study, although they hold great promise for use in machine learning algorithms that could predict lapses and deliver or tailor interventions to individual participants' needs and recent experiences.

Monetary incentives are commonly used in research to provide a more favorable cost-benefit ratio surrounding specific methods or overall participation. Such monetary incentives are commonplace and recommended when using active personal sensing methods such as EMA [88]. However, the incentives to provide access to passively sensed geolocation and cellular communication in our study may have contributed to the acceptance of these methods and the success we had collecting these sensitive data from participants. This may be particularly true given the relatively low socioeconomic status of many of our participants. For example, the mean personal income for our participants was US $34,233, with 12.3% (19/154) of individuals reporting current unemployment and 25.3% (39/154) reporting an annual income below the 2022 federal poverty level.

Monetary incentives to increase the acceptability of personal sensing do not need to be limited to research settings. Incentives can also be used as a part of treatment or continuing care in clinical settings. For example, the use of monetary incentives or equivalents (eg, prizes) as part of a contingency management program is well established to support abstinence from alcohol or other drugs or adherence to treatments or other healthy behaviors [89-91]. If personal sensing proved useful for the treatment or ongoing support of patients' recovery, similar incentives could be established to encourage patients to provide these sensed data.

Incentives may be less necessary in clinical settings when more direct clinical benefits from personal sensing are available. For example, research has suggested that privacy concerns associated with personal sensing may be reduced if participants perceive that they will benefit from the sensed data [6,51,87]. There was some evidence for this perspective in the free-response comments from our study participants as well.

We did not provide direct clinical treatment to the participants. Participants were given resources for alcohol treatment options upon request. In addition, although personal sensing methods were used solely for data collection, in this study, participants may have experienced some clinical benefits from them (eg, via reflection and accountability). However, the acceptability of personal sensing may be higher than that observed in our study if the sensing system was implemented as part of their direct treatment or continuing care during their recovery. Digital therapeutics are particularly well positioned to use sensed data to select, personalize, or time the delivery of interventions and other supports to improve clinical outcomes. Future research should evaluate the acceptability of personal sensing in contexts where its use directly benefits those providing the sensed data. In these contexts, benefits (eg, financial and clinical benefits) can also be explicitly measured. It may even be possible to manipulate the benefits from personal sensing across participants to evaluate their contribution to acceptability more rigorously.

Trust Likely Matters

Trust is also likely to affect the overall acceptability of personal sensing data, which are inherently private and sensitive. Acceptability may depend on who uses personal sensing and who has access to raw and processed data [50,87,92-94]. The available evidence suggests that people are more comfortable sharing private, sensitive information with researchers and their physicians and less comfortable sharing information with family members, electronic health record databases, and third-party apps and websites [92-94].

The research setting may come with relatively greater trust because of the high level of transparency regarding the risks and protection measures associated with obtaining informed consent. Some protections may only be feasible for research as well. For example, National Institutes of Health (NIH)–funded research that collects identifiable, sensitive information is automatically issued a Certificate of Confidentiality that prohibits disclosing this information to anyone not connected to the research, except when the participant consents or in a few other limited situations. The Certificate of Confidentiality can also be requested for similar research not funded by the NIH. We saw evidence of the role of trust in the free-response comments from our participants. Our participants appeared to recognize and appreciate the protective measures taken to secure their data.

Implementations of personal sensing for treatment inside and outside clinical care settings [34] will need to carefully consider how to establish similar, high levels of trust. Clinical applications of personal sensing may sit at an intersection of sharing data with physicians (with which individuals tend to be comfortable) and with electronic health record databases and apps (with which individuals tend to be less comfortable) [93,94]. For example, it may be necessary to protect against the subpoena of sensitive information in civil and criminal proceedings. Patients will also likely need to be assured that sensed data used for their clinical care will not be shared with their health insurance provider with associated risks related to higher insurance premiums or dropped coverage. These issues of data access and unauthorized secondary use of otherwise private information are often cited as concerns regarding personal sensing [87,95].

Regardless of the setting, trust may be lower in stigmatized groups that could otherwise benefit from personal sensing. For example, individuals with mental illness still experience substantial stigma that could impede their willingness to share personal, sensitive information with researchers or clinical care providers [96-99]. In fact, we focused on individuals with AUD in this study to evaluate the acceptance of personal sensing methods in a population that we expected might have barriers associated with trust. Nevertheless, trust may still be lower among individuals with other substance use disorders that involve illegal drug use. However, many of our participants reported ongoing use of drugs other than alcohol throughout the study (75/154, 48.7% reported illicit drug use in the past month) as expected, given the high rates of polysubstance use among individuals with substance use disorders. Furthermore, we have had promising preliminary success in recruiting patients with opioid use disorder for an NIH-funded study on personal sensing in this population [100]. This suggests that our results regarding the acceptance of personal sensing may be generalized across substance use disorders.

Trust and related privacy concerns may also be more difficult to overcome in historically marginalized groups that have experienced systemic racism and other stigmas or exclusions [101]. These individuals may find it more difficult to achieve privacy in their daily lives, and they may hold very different perspectives on the costs versus benefits of surveillance in the context of personal sensing or more generally. Unfortunately, our sample was not diverse with respect to race and ethnicity. Future research on personal sensing must specifically recruit for such diversity to better understand its acceptance in racial and ethnic minority communities. We have learned from this study and adjusted our recruiting efforts accordingly to recruit a sample that is more diverse with respect to race, ethnicity, and geographic region for our ongoing personal sensing project with individuals with opioid use disorder.

Feasibility Is a Function of More Than Participant Perceptions of Acceptability

User acceptance of personal sensing methods is necessary but not sufficient to expand the use of these methods in research and clinical implementations. A variety of other key issues may facilitate or present barriers to the wider use of personal sensing. These include cost and accessibility, stability over time, and the utility of personal sensing relative to other more traditional methods.

The smartphone itself is arguably the best available sensing system at present. Currently, smartphones contain numerous sensors and other raw data streams that can be used for personal sensing. In our study, we took advantage of GPS and other location services to track geolocation and used the microphone for daily audio check-ins. We accessed smartphone calls and SMS text message logs for communication metadata and message content, respectively. The smartphone also provided a convenient platform to collect self-reported EMA.

In addition, smartphones provide a relatively accessible platform for personal sensing. Despite their high cost, 85% of adults in the United States already own a smartphone. Equally important, this level of ownership is relatively consistent across race and ethnicity, geographic regions (eg, urban, suburban, and rural), and income levels [3]. Furthermore, people with substance use disorders generally have high rates of mobile technology use [102]. Notably, only 11 (6.5%) of the 169 eligible participants in our study did not already own a contemporary smartphone. In a research setting, we were able to provide individuals with a smartphone if they did not already have one. Similar to monetary incentives, this practice need not be limited to research; smartphones can be provided to permit personal sensing–based clinical support.

Personal sensing can also be performed using wearable devices or other sensors outside the smartphone. We used Empatica and Beddit systems to sense physiology and sleep, respectively. The use of watches (eg, Apple Watch) and wristbands (eg, Fitbit) for sensing activity and some physiology parameters is also increasing [18,103]. However, some of these systems can be expensive, and unlike smartphones, none have been adopted widely enough to assume that most users will already own the said devices. For research applications, this limitation can be overcome by providing the hardware to participants as needed. Although it is not impossible to do the same in clinical settings, the large number of patients who would require this technology may either limit or increase the cost to scale the sensing system.

Both research and clinical applications of sensing systems require some guarantee that the hardware and software will remain available and supported for the duration of the intended use. Unfortunately, there are currently high levels of churn among the companies that support these systems, given the rapid innovation occurring at this time. We collected data for approximately 2.5 years between 2017 and 2019. During this time, Apple bought the company that developed the Beddit Sleep Monitor and discontinued support for previous users. Apple reintroduced the sleep sensing system for iPhone users in late 2018 but discontinued it again in early 2022. Therefore, we were able to collect sleep sensing data from fewer than half of our research participants.

During this same data collection period, there was also a churn in the software that we used for sensing geolocation. We used the Moves app at the start of the study but needed to switch to the FollowMee app when Facebook acquired the company that developed Moves and discontinued its support. However, this software churn was less disruptive because both apps relied on smartphone sensors to acquire the raw geolocation data stream. This suggests yet another reason to prefer systems that make use of generic smartphone sensors rather than proprietary hardware.

High rates of churn can also affect the perceived acceptability of the software. For example, it could be inconvenient to have to adapt to frequent changing of app platforms. In addition, software may be left unmonitored for periods, leaving new bugs unresolved. In our sample of participants, we observed how frustrating technological issues were.

Limitations and Future Directions

Conclusions regarding the acceptability of these sensing methods may not be generalizable beyond the 3-month study duration. Although 3 months represents a notable extension beyond the existing literature on personal sensing in clinical populations, it is likely not long enough, given the chronic-relapsing nature of alcohol and other substance use disorders. One potential concern is that the initial novelty of sensing may lead to overestimated adherence and subjective ratings of acceptability that is not sustained for longer periods [104].

This 3-month period also constrains our conclusions regarding the acceptability to people early in recovery. It is possible that acceptability ratings will vary depending on where someone is in their recovery phase. This may also be amplified when potential benefits are considered. For example, someone who has achieved long-term stability in their recovery could find that the costs of personal sensing (eg, data sharing and high effort demands) do not outweigh the benefits (eg, daily reflection on sobriety and potential for increased lapse risk awareness). It is important for future studies to extend study length and incorporate other facets of acceptability (eg, benefits) to account for these possible effects. In an ongoing study of people with opioid use disorder, we requested that participants use various active and passive personal sensing methods for 1 year [14]. In addition, future research could compare acceptability ratings for personal sensing methods between people with and without a substance use disorder.

Future studies should also examine the nuances of behavioral measures of acceptability. Our study was limited in the conclusions we could draw about adherence to passive personal sensing measures. All our research participants (154/154, 100%) provided some geolocation and cellular communication data and all but one of our research participants (153/154, 99.4%) provided SMS text message content data. However, we cannot know if and how frequently participants were choosing to selectively delete SMS text messages or turn their geolocation off. In addition, we have limited information on the reasons for participant discontinuation before enrollment. Only 1 participant did not consent to participate at the time of screening. However, the attrition between screening and enrollment could reflect some reservations about the personal sensing methods and the study as a whole. That said, we do not believe our attrition rates between these 2 visits to be unusually high for our target sample (ie, people early in recovery from AUD).

Our self-report acceptability questions were developed in house. Therefore, our results should be interpreted in light of our specific questions and settings. For example, we asked participants if they would be willing to use a personal sensing method for 1 year to help with their recovery. This could imply that there would be a clinical benefit to using the method for 1 year and may factor in their judgment of acceptability. These questions have not been previously used in other research settings. Although we attempted to minimize social desirability effects and encourage feedback (eg, deidentified self-report surveys submitted through a web-based survey platform), it is possible that these effects are built into our results. Nonetheless, it should also be acknowledged that the study conclusions are based on both these self-report measures and behavioral indices.

Finally, although our results suggest that clinical samples of people with AUD may find these personal sensing methods acceptable, more research is needed to test the acceptability of these methods in future applied-clinical settings, where issues of costs, benefits, and trust may differ meaningfully in complicated ways from the research context. Future studies should also examine how these personal sensing methods might be perceived by people with recovery goals other than abstinence. No technical reasons prevent personal sensing from being applied to alternative recovery goals (see the studies by Bae et al [28] and Walters et al [29] for examples of predicting current and imminent drinking episodes, respectively, in people without a goal of abstinence). In addition, it must be acknowledged that the individuals in our study agreed to participate in a research study on mobile health and were financially compensated for their time. It is unclear how these individuals and the research setting may differ from those seeking to use these methods in future clinical settings, where costs, benefits, and trust may all weigh differently on their decisions to engage with the sensing system.

Conclusions

This study demonstrated the acceptability of several personal sensing methods. These methods were acceptable (1) over a longer period than has previously been assessed, (2) across active and passive methods, (3) despite the sensitivity of the data, (4) among individuals with AUD who may have greater privacy concerns, and (5) without explicit clinical benefits to the participants. These findings suggest that personal sensing methods are poised as accessible, feasible avenues to collect data about individuals to be used for clinical applications. More work is needed to determine the predictive utility of the data that can be collected via personal sensing, but our study shows that this work will be worthwhile to pursue.

Personal sensing is acceptable, and the technology to collect it (namely, the smartphone) is widely accessible. Personal sensing can make digital therapeutics—smartphones and web-based apps that provide mental health care—smart. These methods can personalize care for individuals such that they receive the specific interventions and support they need at the time they need them. Smart digital therapeutics can be scaled widely to provide treatment to the overwhelming majority of individuals who do not currently receive mental health care. They can reach those who have historically been excluded from or have otherwise faced barriers to care. With personal sensing powering digital therapeutics, we are positioned for a paradigm shift in mental health care. This study brings us one step closer to this goal, ensuring that the methods we hope to use to revolutionize care are acceptable to patients who will use them.

Acknowledgments

This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (R01 AA024391; JJC) and the National Institute on Drug Abuse (R01 DA047315; JJC).

The authors wish to thank Susan E Wanta for her role as a project administrator and for her help with data curation. The authors also wish to thank Candace Lightheart, Jill Nagler, Kerry Keiser, and Megan Shultz for their contributions to data collection and Chris Gioia for the clinical supervision he provided to graduate students.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Transparency checklist.
[DOCX File , 28 KB - mhealth_v11i1e41833_app1.docx]

Multimedia Appendix 2
Supplemental analyses.
[PDF File (Adobe PDF File), 272 KB - mhealth_v11i1e41833_app2.pdf]

Multimedia Appendix 3
Acceptability survey.
[DOCX File , 27 KB - mhealth_v11i1e41833_app3.docx]

References

1. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2021-08-13]
2. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors (Basel) 2019 May 09;19(9):2164 [FREE Full text] [doi: 10.3390/s19092164] [Medline: 31075985]
3. Mobile fact sheet. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-08-15]
4. What is digital health? U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2021-08-13]
5. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
6. Klasnja P, Consolvo S, Choudhury T, Beckwith R, Hightower J. Exploring privacy concerns about personal sensing. In: Proceedings of the 7th International Conference, Pervasive 2009. 2009 Presented at: 7th International Conference, Pervasive 2009; May 11-14, 2009; Nara, Japan. [doi: 10.1007/978-3-642-01516-8_13]

7. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Dec;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
8. Morgenstern J, Kuerbis A, Muench F. Ecological momentary assessment and alcohol use disorder treatment. Alcohol Res 2014;36(1):101-109 [FREE Full text] [Medline: 26259004]
9. Fronk GE, Sant'Ana SJ, Kaye JT, Curtin JJ. Stress allostasis in substance use disorders: promise, progress, and emerging priorities in clinical research. Annu Rev Clin Psychol 2020 May 07;16:401-430 [FREE Full text] [doi: 10.1146/annurev-clinpsy-102419-125016] [Medline: 32040338]
10. Schultz ME, Fronk GE, Jaume N, Magruder KP, Curtin JJ. Stressor-elicited smoking and craving during a smoking cessation attempt. J Psychopathol Clin Sci 2022 Jan;131(1):73-85 [FREE Full text] [doi: 10.1037/abn0000702] [Medline: 34881919]
11. Piasecki TM, Jorenby DE, Smith SS, Fiore MC, Baker TB. Smoking withdrawal dynamics: I. Abstinence distress in lapsers and abstainers. J Abnorm Psychol 2003 Feb;112(1):3-13. [Medline: 12653409]
12. McCarthy DE, Piasecki TM, Fiore MC, Baker TB. Life before and after quitting smoking: an electronic diary study. J Abnorm Psychol 2006 Aug;115(3):454-466. [doi: 10.1037/0021-843X.115.3.454] [Medline: 16866586]
13. Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, et al. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. NPJ Digit Med 2020 Mar 4;3:26 [FREE Full text] [doi: 10.1038/s41746-020-0234-6] [Medline: 32195362]
14. Moshontz H, Colmenares AJ, Fronk GE, Sant'Ana SJ, Wyant K, Wanta SE, et al. Prospective prediction of lapses in opioid use disorder: protocol for a personal sensing study. JMIR Res Protoc 2021 Dec 07;10(12):e29563 [FREE Full text] [doi: 10.2196/29563] [Medline: 34559061]
15. Bae SW, Suffoletto B, Zhang T, Chung T, Ozolcer M, Islam MR, et al. Leveraging mobile phone sensors, machine learning, and explainable artificial intelligence to predict imminent same-day binge-drinking events to support just-in-time adaptive interventions: algorithm development and validation study. JMIR Form Res 2023 May 04;7:e39862 [FREE Full text] [doi: 10.2196/39862] [Medline: 36809294]
16. Bae SW, Chung T, Islam R, Suffoletto B, Du J, Jang S, et al. Mobile phone sensor-based detection of subjective cannabis intoxication in young adults: a feasibility study in real-world settings. Drug Alcohol Depend 2021 Nov 01;228:108972 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108972] [Medline: 34530315]
17. Heacock RM, Capodilupo ER, Czeisler M, Weaver MD, Czeisler CA, Howard ME, et al. Sleep and alcohol use patterns during federal holidays and daylight saving time transitions in the United States. Front Physiol 2022 Jul 11;13:884154 [FREE Full text] [doi: 10.3389/fphys.2022.884154] [Medline: 35899022]
18. Stevenson BL, Kunicki ZJ, Brick L, Blevins CE, Stein M, Abrantes AM. Using ecological momentary assessments and fitbit data to examine daily associations between physical activity, affect and alcohol cravings in patients with alcohol use disorder. Int J Behav Med 2022 Oct 08;29(5):543-552. [doi: 10.1007/s12529-021-10039-5] [Medline: 34750719]
19. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
20. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
21. Torous J, Staples P, Onnela JP. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
22. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
23. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. J Am Med Inform Assoc 2020 Apr 01;27(4):522-530 [FREE Full text] [doi: 10.1093/jamia/ocz221] [Medline: 31977041]
24. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
25. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
26. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572 [FREE Full text] [doi: 10.3390/s20123572] [Medline: 32599801]
27. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
28. Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun 30;1(2):1-36 [FREE Full text] [doi: 10.1145/3090051] [Medline: 35146236]

29. Walters ST, Businelle MS, Suchting R, Li X, Hébert ET, Mun EY. Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness. J Subst Abuse Treat 2021 Aug;127:108417 [FREE Full text] [doi: 10.1016/j.jsat.2021.108417] [Medline: 34134874]
30. Chih MY, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat 2014 Jan;46(1):29-35 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]
31. Scott CK, Dennis ML, Gustafson DH. Using ecological momentary assessments to predict relapse after adult substance use treatment. Addict Behav 2018 Jul;82:72-78 [FREE Full text] [doi: 10.1016/j.addbeh.2018.02.025] [Medline: 29499393]
32. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares ML, et al. Integrating addiction treatment into primary care using mobile health technology: protocol for an implementation research study. Implement Sci 2014 May 29;9:65 [FREE Full text] [doi: 10.1186/1748-5908-9-65] [Medline: 24884976]
33. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
34. Aggarwal N, Ahmed M, Basu S, Curtin JJ, Evans BJ, Matheny ME, et al. Advancing artificial intelligence in health settings outside the hospital and clinic. NAM Perspect 2020 Nov 30;2020 [FREE Full text] [doi: 10.31478/202011f] [Medline: 35291747]
35. Kaiser J. Obama gives east room rollout to precision medicine initiative. Science. 2015 Jan 30. URL: http://www.sciencemag.org/news/2015/01/obama-gives-east-room-rollout-precision- [accessed 2020-09-24]
36. Sheikh M, Qassem M, Kyriacou PA. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. Front Digit Health 2021;3:662811 [FREE Full text] [doi: 10.3389/fdgth.2021.662811] [Medline: 34713137]
37. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
38. Ortiz A, Maslej M, Husain M, Daskalakis Z, Mulsant B. Apps and gaps in bipolar disorder: a systematic review on electronic monitoring for episode prediction. J Affect Disord 2021 Dec 01;295:1190-1200 [FREE Full text] [doi: 10.1016/j.jad.2021.08.140] [Medline: 34706433]
39. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
40. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
41. Kirk GD, Linas BS, Westergaard RP, Piggott D, Bollinger RC, Chang LW, et al. The exposure assessment in current time study: implementation, feasibility, and acceptability of real-time data collection in a community cohort of illicit drug users. AIDS Res Treat 2013;2013:594671 [FREE Full text] [doi: 10.1155/2013/594671] [Medline: 24307943]
42. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223 [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]
43. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e67 [FREE Full text] [doi: 10.2196/mhealth.4344] [Medline: 26085078]
44. Moitra E, Gaudiano BA, Davis CH, Ben-Zeev D. Feasibility and acceptability of post-hospitalization ecological momentary assessment in patients with psychotic-spectrum disorders. Compr Psychiatry 2017 Apr;74:204-213 [FREE Full text] [doi: 10.1016/j.comppsych.2017.01.018] [Medline: 28231480]
45. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
46. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
47. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
48. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]

49. Duncan DT, Park SH, Goedel WC, Sheehan DM, Regan SD, Chaix B. Acceptability of smartphone applications for global positioning system (GPS) and ecological momentary assessment (EMA) research among sexual minority men. PLoS One 2019;14(1):e0210240 [FREE Full text] [doi: 10.1371/journal.pone.0210240] [Medline: 30689651]
50. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph P. Psychiatry outpatients' willingness to share social media posts and smartphone data for research and clinical purposes: survey study. JMIR Form Res 2019 Aug 29;3(3):e14329 [FREE Full text] [doi: 10.2196/14329] [Medline: 31493326]
51. Bessenyei K, Suruliraj B, Bagnell A, McGrath P, Wozney L, Huguet A, et al. Comfortability with the passive collection of smartphone data for monitoring of mental health: an online survey. Comput Hum Behav Rep 2021 Aug 15;4(16):100134-100174 [FREE Full text] [doi: 10.1016/j.chbr.2021.100134]
52. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The Effortless Assessment of Risk States (EARS) tool: an interpersonal approach to mobile sensing. JMIR Ment Health 2018 Aug 28;5(3):e10334 [FREE Full text] [doi: 10.2196/10334] [Medline: 30154072]
53. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer EA, Mistler LA, et al. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. Psychiatr Serv 2016 May 01;67(5):558-561 [FREE Full text] [doi: 10.1176/appi.ps.201500130] [Medline: 26695497]
54. Rooksby J, Morrison A, Murray-Rust D. Student perspectives on digital phenotyping: the acceptability of using smartphone data to assess mental health. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 May Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300655]
55. Kleiman E, Millner AJ, Joyce VW, Nash CC, Buonopane RJ, Nock MK. Using wearable physiological monitors with suicidal adolescent inpatients: feasibility and acceptability study. JMIR Mhealth Uhealth 2019 Sep 24:e0 [FREE Full text] [doi: 10.2196/13725] [Medline: 31586364]
56. Raugh IM, James SH, Gonzalez CM, Chapman HC, Cohen AS, Kirkpatrick B, et al. Digital phenotyping adherence, feasibility, and tolerability in outpatients with schizophrenia. J Psychiatr Res 2021 Jun;138:436-443 [FREE Full text] [doi: 10.1016/j.jpsychires.2021.04.022] [Medline: 33964681]
57. Schönbrodt FD, Maier M, Heene M, Zehetleitner M. Voluntary commitment to research transparency. Commitment to Research Transparency. 2015. URL: http://www.researchtransparency.org [accessed 2023-07-26]
58. Simmons JP, Nelson LD, Simonsohn U. A 21 word solution. SSRN. Preprint posted online October 15, 2012 [FREE Full text] [doi: 10.2139/ssrn.2160588]
59. Aczel B, Szaszi B, Sarafoglou A, Kekecs Z, Kucharský Š, Benjamin D, et al. A consensus-based transparency checklist. Nat Hum Behav 2020 Jan;4(1):4-6 [FREE Full text] [doi: 10.1038/s41562-019-0772-6] [Medline: 31792401]
60. Wyant K, Moshontz H, Ward SB, Fronk G, Curtin JJ. Acceptability of personal sensing among people with alcohol use disorder: observational study. PsyArXiv. Preprint posted online August 11, 2022 [FREE Full text] [doi: 10.31234/osf.io/3ykcg]
61. Dynamic, real-time prediction of alcohol use lapse using mHealth technologies. NIH RePORTER. URL: https://reporter.nih.gov/search/LLKWYWhN9EO5FmeFLrTL5g/project-details/8986398 [accessed 2023-07-26]
62. Gustafson DH, Landucci G, McTavish F, Kornfield R, Johnson RA, Mares ML, et al. The effect of bundling medication-assisted treatment for opioid addiction with mHealth: study protocol for a randomized clinical trial. Trials 2016 Dec 12;17(1):592 [FREE Full text] [doi: 10.1186/s13063-016-1726-1] [Medline: 27955689]
63. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
64. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. New York City, NY: Guilford Publications; 1985.
65. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. New York, NY: The Guilford Press; 2008.
66. Larimer ME, Palmer RS, Marlatt GA. Relapse prevention. An overview of Marlatt's cognitive-behavioral model. Alcohol Res Health 1999;23(2):151-160 [FREE Full text] [Medline: 10890810]
67. Witkiewitz K, Marlatt GA. Emphasis on interpersonal factors in a dynamic model of relapse. Am Psychol 2005;60(4):341-342. [doi: 10.1037/0003-066X.60.4.341] [Medline: 15943533]
68. Carpenter SM, Menictas M, Nahum-Shani I, Wetter DW, Murphy SA. Developments in mobile health just-in-time adaptive interventions for addiction science. Curr Addict Rep 2020 Sep;7(3):280-290 [FREE Full text] [doi: 10.1007/s40429-020-00322-y] [Medline: 33747711]
69. Businelle MS, Walters ST, Mun EY, Kirchner TR, Hébert ET, Li X. Reducing drinking among people experiencing homelessness: protocol for the development and testing of a just-in-time adaptive intervention. JMIR Res Protoc 2020 Apr 16;9(4):e15610 [FREE Full text] [doi: 10.2196/15610] [Medline: 32297874]
70. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull 1973 Jan;9(1):13-28. [Medline: 4682398]
71. R: a language and environment for statistical computing. Global Biodiversity Information Facility. 2015 Feb 10. URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2023-07-26]

72. RStudio Team. RStudio: integrated development for R. RStudio. 2020. URL: https://posit.co/ [accessed 2021-08-13]
73. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov 21;4(43):1686 [FREE Full text] [doi: 10.21105/joss.01686]
74. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
75. Kilian C, Manthey J, Carr S, Hanschmidt F, Rehm J, Speerforck S, et al. Stigmatization of people with alcohol use disorders: an updated systematic review of population studies. Alcohol Clin Exp Res 2021 May;45(5):899-911 [FREE Full text] [doi: 10.1111/acer.14598] [Medline: 33970504]
76. Results from the 2019 national survey on drug use and health (NSDUH): key substance use and mental health indicators in the United States. Substance Abuse and Mental Health Services Administration. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2020 Sep. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm [accessed 2022-06-18]
77. Mental healthcare at our fingertips. YouTube. URL: https://www.youtube.com/watch?v=KZ8SbmXzFy8 [accessed 2022-06-18]
78. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 05;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
79. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54 [FREE Full text] [doi: 10.1177/0261927x09351676]
80. Tackman AM, Sbarra DA, Carey AL, Donnellan MB, Horn AB, Holtzman NS, et al. Depression, negative emotionality, and self-referential language: a multi-lab, multi-measure, and multi-language-task research synthesis. J Pers Soc Psychol 2019 May;116(5):817-834 [FREE Full text] [doi: 10.1037/pspp0000187] [Medline: 29504797]
81. Jacobucci R, Ammerman BA, Tyler Wilcox K. The use of text-based responses to improve our understanding and prediction of suicide risk. Suicide Life Threat Behav 2021 Feb;51(1):55-64 [FREE Full text] [doi: 10.1111/sltb.12668] [Medline: 33624877]
82. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
83. Belouali A, Gupta S, Sourirajan V, Yu J, Allen N, Alaoui A, et al. Acoustic and language analysis of speech for suicidal ideation among US veterans. BioData Min 2021 Feb 02;14(1):11 [FREE Full text] [doi: 10.1186/s13040-021-00245-y] [Medline: 33531048]
84. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6(7):e856 [FREE Full text] [doi: 10.1038/tp.2016.123] [Medline: 27434490]
85. Pavlou PA. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int J Electron Commer 2014 Dec 23;7(3):101-134. [doi: 10.1080/10864415.2003.11044275]
86. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
87. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WY, et al. Consumer attitudes and perceptions on mHealth privacy and security: findings from a mixed-methods study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
88. Parkinson B, Meacock R, Sutton M, Fichera E, Mills N, Shorter GW, et al. Designing and using incentives to support recruitment and retention in clinical trials: a scoping review and a checklist for design. Trials 2019 Nov 09;20(1):624 [FREE Full text] [doi: 10.1186/s13063-019-3710-z] [Medline: 31706324]
89. Prendergast M, Podus D, Finney J, Greenwell L, Roll J. Contingency management for treatment of substance use disorders: a meta-analysis. Addiction 2006 Nov;101(11):1546-1560. [doi: 10.1111/j.1360-0443.2006.01581.x] [Medline: 17034434]
90. Ginley MK, Pfund RA, Rash CJ, Zajac K. Long-term efficacy of contingency management treatment based on objective indicators of abstinence from illicit substance use up to 1 year following treatment: a meta-analysis. J Consult Clin Psychol 2021 Jan;89(1):58-71 [FREE Full text] [doi: 10.1037/ccp0000552] [Medline: 33507776]
91. Petry NM. Contingency management: what it is and why psychiatrists should want to use it. Psychiatrist 2011 May;35(5):161-163 [FREE Full text] [doi: 10.1192/pb.bp.110.031831] [Medline: 22558006]
92. Rendina HJ, Mustanski B. Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the United States. J Med Internet Res 2018 Jul 04;20(7):e233 [FREE Full text] [doi: 10.2196/jmir.9019] [Medline: 29973332]
93. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e12578 [FREE Full text] [doi: 10.2196/12578] [Medline: 30950799]

94. Prasad A, Sorber J, Stablein T, Anthony DL, Kotz D. Understanding user privacy preferences for mHealth data sharing. In: mHealth Multidisciplinary Verticals. Boca Raton, FL: CRC Press; 2014.
95. Ackerman MS, Mainwaring SD. Privacy issues and human-computer interaction. In: Cranor LF, Garfinkel S, editors. Security and Usability: Designing Secure Systems That People Can Use. Cambridge, MA: O'Reilly Media; 2005.
96. Schomerus G, Lucht M, Holzinger A, Matschinger H, Carta MG, Angermeyer MC. The stigma of alcohol dependence compared with other mental disorders: a review of population studies. Alcohol Alcohol 2011;46(2):105-112. [doi: 10.1093/alcalc/agq089] [Medline: 21169612]
97. Barry CL, McGinty EE, Pescosolido BA, Goldman HH. Stigma, discrimination, treatment effectiveness, and policy: public views about drug addiction and mental illness. Psychiatr Serv 2014 Oct;65(10):1269-1272 [FREE Full text] [doi: 10.1176/appi.ps.201400140] [Medline: 25270497]
98. Overton SL, Medina SL. The stigma of mental illness. J Couns Dev 2008;86(2):143-151 [FREE Full text] [doi: 10.1002/j.1556-6678.2008.tb00491.x]
99. Parcesepe AM, Cabassa LJ. Public stigma of mental illness in the United States: a systematic literature review. Adm Policy Ment Health 2013 Sep;40(5):384-399 [FREE Full text] [doi: 10.1007/s10488-012-0430-z] [Medline: 22833051]
100. Contextualized daily prediction of lapse risk in opioid use disorder by digital phenotyping. NIH RePORTER. URL: https://reporter.nih.gov/search/-uI6W9d_OEyhJadSGQU19g/project-details/10427354 [accessed 2023-07-26]
101. Marwick AE, Boyd D. Understanding privacy at the margins. Int J Commun 2018;12(2018):1157-1165 [FREE Full text]
102. Collins KM, Armenta RF, Cuevas-Mota J, Liu L, Strathdee SA, Garfein RS. Factors associated with patterns of mobile technology use among persons who inject drugs. Subst Abus 2016;37(4):606-612 [FREE Full text] [doi: 10.1080/08897077.2016.1176980] [Medline: 27092425]
103. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and Fitbit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
104. Shin G, Feng Y, Jarrahi MH, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]

Abbreviations

AUD: alcohol use disorder
EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
NIH: National Institutes of Health

Edited by L Buis; submitted 11.08.22; peer-reviewed by E O. Agu, B Curtis, H Hutton, M Johansson; comments to author 19.10.22; revised version received 14.03.23; accepted 25.07.23; published 28.08.23.

Please cite as:
Wyant K, Moshontz H, Ward SB, Fronk GE, Curtin JJ
Acceptability of Personal Sensing Among People With Alcohol Use Disorder: Observational Study
JMIR Mhealth Uhealth 2023;11:e41833
URL: https://mhealth.jmir.org/2023/1/e41833
doi:10.2196/41833
PMID:37639300

Adoption of Electronic Medical Records for Chronic Disease Care in Kenyan Refugee Camps: Quantitative and Qualitative Prospective Evaluation

Emily Lyles[1], MSPH, MPA, PhD; Kenneth Paik[2], MD; John Kiogora[3], MBChB; Husna Hussein[3], BS, CO; Alejandra Cordero Morales[1], MPH, MD; Lilian Kiapi[4], MPH, MD; Shannon Doocy[1], PhD

1

2

3

4

Corresponding Author:
Shannon Doocy, PhD

Abstract

Background: Noncommunicable disease (NCD) prevention and control in humanitarian emergencies is a well-recognized need, but there is little evidence to guide responses, leading to varying care delivery. The Sana.NCD mobile health (mHealth) app, initially developed in Lebanon, is the only known mHealth tool for NCD management designed to increase care quality and coverage for providers in humanitarian settings.

Objective: We evaluated a specialized mHealth app consisting of an abbreviated medical record for patients with hypertension or diabetes, adapted for a Kenyan refugee camp setting.

Methods: We tested an adapted version of the Sana.NCD app (diabetes and hypertension medical record) in an 11-month (May 2021 to March 2022) quantitative and qualitative prospective evaluation in Kenya's Hagadera refugee camp. Leveraging the rollout of a general electronic medical record (EMR) system in the Kakuma refugee camp, we compared a specialized NCD management app to a general EMR. We analyzed secondary data collected from the Sana.NCD app for 1539 patients, EMR data for 68 patients with NCD from Kakuma's surgical and outpatient departments, and key informant interviews that focused on Hagadera clinic staff perceptions of the Sana.NCD app.

Results: The Hagadera NCD clinic reported 18,801 consultations, 42.1% (n=7918) of which were reported in the NCD app. The Kakuma EMR reported 350,776 visits, of which 9385 (2.7%) were for NCDs (n=4264, 1.2% hypertension; n=2415, 0.7% diabetes). The completeness of reporting was used as a quality-of-care metric. Age, sex, prescribed medicines, random blood sugar, and smoking status were consistently reported in both the NCD app (>98%) and EMR (100%), whereas comorbidities, complications, hemoglobin A_{1c}, and diet were rarely reported in either platform (≤7% NCD app; 0% EMR). The number of visits, BMI, physical activity, and next visit were frequently reported in the NCD app (≥99%) but not in the EMR (≤15%). In the NCD app, the completeness of reporting was high across the implementation period, with little meaningful change. Although not significantly changed during the study, elevated blood sugar (P=.82) and blood pressure (P=.12) were reported for sizable proportions of patients in the first (302/481, 62.8%, and 599/1094, 54.8%, respectively) and last (374/602, 62.1%, and 720/1395, 51.6%, respectively) study quarters. Providers were satisfied with the app, as it standardized patient information and made consultations easier. Providers also indicated that access to historic patient information was easier, benefiting NCD control and follow-up.

Conclusions: A specialized record for NCDs outperformed a more general record intended for use in all patients in terms of reporting completeness. This CommCare-based NCD app can easily be rolled out in similar humanitarian settings with minimal adaptation. However, the adaptation of technologies to the local context and use case is critical for uptake and ensuring that workflows and time burden do not outweigh the benefits of EMRs.

(JMIR Mhealth Uhealth 2023;11:e43878) doi:10.2196/43878

KEYWORDS

mHealth; hypertension; diabetes; chronic disease; Kenya; refugees; mobile health; noncommunicable disease

Introduction

Access to and quality of health services are major concerns in conflict settings, where the destruction of health facilities, inadequate human resources for health, and displacement limit access to and quality of care. Innovative approaches to support facility and community care using digital solutions may help address these challenges and improve disease management in low-resource settings [1]. This is particularly important for noncommunicable diseases (NCDs), which require long-term care and are often complicated to manage. An estimated 41 million deaths due to NCDs occur annually, accounting for 71% of global mortality. Mortality disproportionately affects the poorest and most disenfranchised populations, with 85% of premature deaths due to NCDs among adults aged 30-69 years occurring in low- and middle-income countries [2].

Historically, humanitarian health interventions have not focused on NCDs; however, with demographic and epidemiologic transitions and increasing NCD burden, it is essential to address this gap. The need for NCD prevention and control in humanitarian emergencies is well recognized, but there is little evidence to guide the development of operational strategies or policies, resulting in varying delivery of NCD care [3,4]. Mobile health (mHealth) interventions are increasingly common and, with the widespread availability of devices, have the potential to overcome infrastructure, human resource, and access limitations in conflict settings [3,5]. Evidence for mHealth interventions, and electronic medical records (EMRs) specifically, is strong in high-income countries, where the need for context-specific modification of existing EMR models and the continued adaptability of these models to meet evolving clinical needs and environments is well established [6,7]. Although EMRs are increasingly common, little is empirically known about patient outcomes within these systems and the overall evidence base for EMRs, and mHealth interventions are generally less developed in low- and middle-income settings, particularly for decision support functions [8-10]. Decision support algorithms provide guidance to less-trained providers who are compelled to deliver care in the absence of physicians, increasing access to services and improving the quality of delivery. Enhanced recordkeeping is an important first step to improve the quality of care, and persistent digital health information can also improve the continuity of care over time and place in conflict and displacement contexts.

To date, the use of mHealth tools for NCDs in humanitarian settings has been limited and concentrated in the Middle East, specifically in Lebanon and Jordan. A 2012 pilot of cohort monitoring among Palestinian refugees with diabetes and hypertension at an UNRWA (United Nations Relief and Works Agency for Palestine Refugees in the Near East) primary health care clinic in Jordan showed that EMRs were a valuable monitoring mechanism, with potential to improve the quality and continuity of care for diabetes and hypertension [11,12]. The expansion of UNRWA's EMR system to additional health centers further demonstrated the system's ability to accurately capture service provision data and health outcomes [13,14]. In Iraq, a mobile diabetes self-management system that incorporates mobile glucose self-monitoring and a remote web interface and health management system was also found to lower hemoglobin A_{1c} (HbA_{1c}) in a case study of 12 patients with type 2 diabetes [15].

More recently, a randomized controlled trial was conducted in rural areas and refugee camps in Lebanon, assessing the effectiveness of an integrative NCD intervention that includes weekly SMS text messages containing medical information and appointment or follow-up reminders [16]. The study observed improved health outcomes, namely blood pressure and HbA_{1c}, among patients who received weekly SMS text messages for 1 year and concluded that the intervention was a simple, socially acceptable, and low-cost mechanism for improving NCD care. The SMS messaging intervention did not show as much promise in effecting behavior change, with fewer than one-third of participants reporting improvement in behaviors for managing diabetes or hypertension [17].

The Sana.NCD mHealth app (previously titled Sana.PCHR), developed by the Massachusetts Institute of Technology (MIT) Sana Mobile group, was initially designed for Syrian refugees in Lebanon. It is the only known mHealth tool for NCD management designed to increase the coverage and quality of care that is appropriate for a range of providers in primary and community settings in humanitarian contexts [18,19]. The app is an innovative and scalable approach that showed potential to improve adherence to guidelines and quality of care during development in Lebanon [18,19]. This study tested a modified version of the Sana mHealth app at the NCD clinic based in the Hagadera refugee camp hospital, which is supported by the International Rescue Committee (IRC). Established in 1992, the Hagadera camp is 1 of 3 camps in the Dadaab refugee camp complex, which collectively hosts more than 215,000 Somali refugees [20]. NCDs account for 24% and 27% of all deaths in Somalia and Kenya, respectively, and the camp context represented a stable location with high levels of need where the mHealth app could be tested [21].

This work aimed to evaluate a specialized mHealth app consisting of an abbreviated medical record for patients with hypertension or diabetes, adapted for a Kenyan refugee camp setting.

Methods

Study Design

A quantitative and qualitative prospective evaluation with an 11-month follow-up period (May 1, 2021, through March 31, 2022) was used to evaluate a version of the Sana.NCD mHealth app, adapted for the Hagadera refugee camp in Kenya.

Intervention Description and Study Implementation

The Sana.NCD mHealth app, as originally designed and tested in Lebanon, included a patient-controlled health record in addition to serving as a diabetes and hypertension EMR and a decision support tool for providers. In this pilot study, the app was adapted for the localized context based on input from clinicians who would be using the app and IRC health program staff. The mHealth app evaluated in this study was an abbreviated medical record and clinical protocol for hypertension and diabetes; a patient-controlled health record was not relevant in this context because patients returned to the same location for care. For this pilot study, the app was integrated with the CommCare platform (Dimagi, Inc) [22], which was already used by the IRC and advantageous for future scale-up. MIT Sana and Dimagi worked in partnership with the IRC to customize the app to the workflows and context of IRC-supported health facilities in Hagadera. Providers were trained by Dimagi on app use, which was followed by a 2-week pilot period wherein the IRC team continued to practice using the app prior to the formal data collection period. Technical support was provided to the health facility for the duration of the study implementation period. Images of the Sana.NCD app interface are provided in Multimedia Appendix 1.

The planned study involved a comparison of the patients and sites supported by the IRC through the Sana.NCD app (Hagadera refugee camp) to those without the Sana.NCD app (Kakuma refugee camp, which hosts predominantly South Sudanese refugees); however, an unanticipated change to recordkeeping at the comparison facilities necessitated a design change. The IRC began the rollout of an outpatient register (an individual EMR) for all patients in the Kakuma refugee camp that facilitated reporting to the Ministry of Health of Kenya and the Integrated Refugee Health Information System [23] on May 1, 2021. Given the adoption of the EMR at Kakuma at a similar time as the use of the Sana.NCD app at Hagadera, the evaluation was revised to leverage the opportunity to compare the use of a specialized app for NCD management to the use of a more general EMR.

The study's quantitative component consisted of secondary descriptive analysis of deidentified data collected in the Sana.NCD app for patients with hypertension and diabetes who received outpatient care at the Hagadera refugee camp hospital, in addition to EMR data provided by the IRC for patients visiting the IRC Kakuma hospital's surgical and medical outpatient clinic departments (where patients with NCD are seen). Data from a total of 1539 patients in Hagadera using the Sana.NCD mHealth app and 68 patients (identifiable as having diabetes or hypertension) in the Kakuma comparison facilities with the EMR were analyzed. As Kakuma's EMR system was newly implemented during the study period, reporting on several key indicators critical to diagnosis were frequently not captured or recorded in open-text "notes" fields, hindering the ability to quickly identify patients with specific health conditions. This limited the sample size from Kakuma and precluded more complex analyses comparing health outcomes between the intervention and comparison facilities, which was a planned study outcome.

The study's qualitative component consisted of key informant interviews with IRC health facility staff that focused on their professional opinions of the app. All 6 health facility staff involved in providing NCD care who had experience using the Sana.NCD app participated in the key informant interviews. Interviews were conducted remotely by study team members who had no previous relationships with the key informants. Notes were taken during each interview, but no recordings were taken. Because the interviewers had no relationship with the key informants prior to the interview and they were not associated with the health facility where the participants were employed, perceived influence or conflict of interest is believed to be minimal. A structured facilitation guide was used; however, interviews were iterative, allowing for deeper probing of topics where needed. Interview content focused on the challenges faced in the use and adoption of the app, changes in NCD care and service provision resulting from the use of the app, changes in patient data collection and management and if this informed changes to patient care, and recommendations for future modifications to the app.

Data Analysis

Quantitative data were analyzed using Stata 15 (StataCorp LLC). Descriptive statistics (means and proportions) were used to characterize the adoption of the app, compare reporting quality between the Hagadera (NCD app) and Kakuma (EMR) health facilities, and compare the completeness of reporting in Hagadera NCD app records between the first and last study quarters. Change from the first to last quarter of the study was also analyzed for key biometrics, namely BMI classification, elevated blood pressure, and elevated blood sugar. BMI was categorized as normal (<25 kg/m^2), overweight (25-29.9 kg/m^2), or obese (≥30 kg/m^2); blood pressure was classified as normal (systolic<140 mm Hg and diastolic<90 mm Hg) or elevated (systolic>140 mm Hg or diastolic>90 mm Hg); and blood sugar was similarly classified as having achieved target levels (HbA_{1c}<7.0% or random blood sugar<11.1 mmol/L, with preference given to HbA_{1c} where available) or elevated [24]. Linear probability models were used to estimate differences in biometric changes from the first to last study quarter between male and female patients and across patient age groups (<60 y, 60-69 y, and 70 y or older), with main terms for sex or age and study quarter and the interaction between sex or age and study quarter.

Qualitative key informant interview data (NCD app, Hagadera refugee camp only) were analyzed using thematic analysis, focusing on various process aspects of mHealth app use or adoption and user perceptions of changes in health service provision that may have resulted from the adoption of the app. Qualitative findings were used to contextualize trends in quantitative data, obtain a more nuanced understanding of the app's utility, and provide recommendations on possible adaptations for the mHealth app and deployment strategies for scale-up in other contexts.

Ethical Considerations

The Johns Hopkins School of Public Health Institutional Review Board reviewed this study and determined that it was not engaged in human research: only deidentified health program data were used for quantitative analysis, and interaction with human participants was limited to key informant interviews with health professionals and their experiences of using the technologies. Consent discussions for key informants occurred at the start of each interview, at which time key informants were read a consent statement and oral consent was obtained before proceeding with the interview.

Results

Facilities Comparison

The NCD clinic where the NCD app was implemented is situated at the outpatient department of the main Hagadera refugee camp hospital. The clinic operates from 8 AM to 4 PM, 6 days a week, with most patients seen in the morning and a few walk-ins in the afternoon. All new or suspected NCD cases are referred from other service points or by community health workers. All bookings are done at the clinic with a typical follow-up time of 1 month. The clinic is generally staffed by 1 clinician with additional support when needed. The comparison site covered by the EMR consists of 5 health facilities (including 1 camp hospital) in the Kakuma refugee camp. These clinics operate Monday through Friday and, as with the Hagadera NCD clinic, are staffed by 1 clinician per facility. Bookings are done independently in each facility on a daily basis, with specific "clinic days" for each condition. NCD services in Kakuma are decentralized and integrated with primary health care to enable patients to have easy access to NCD care in the facility closest to their residence. Care for chronic conditions is provided free of charge at both Hagadera and Kakuma health facilities.

During the 11-month study period, 1539 patients with diabetes or hypertension were seen on an outpatient basis in Hagadera and reported in the NCD app. Patients were predominantly female (n=928, 60.3%) and had an average age of 53 (range 6-100) years. The majority (n=898, 58.3%) were seen for hypertension, with diabetes and combined diabetes and hypertension diagnoses accounting for 23.2% (n=357) and 18.5% (n=284) of patients, respectively (Table 1). The identification of patients with hypertension or diabetes was a challenge in the EMR due to the absence of consistently used fields for reporting the diagnosis. Consequently, most patients with these conditions were not captured by this study; the 68 patients that are included were identified based on information provided in text fields. Demographic characteristics of the included Kakuma care seekers were similar in that the population was predominantly female (n=52, 77%) and had a similar age distribution (mean 53; range 15-95 y). A notable difference in diagnosis was observed, with 96% (n=65) of Kakuma patients having hypertension and small numbers of patients having only diabetes (n=2, 3%) or both diabetes and hypertension (n=1, 2%). There is, however, a substantial bias in this distribution that becomes evident when caseload is considered, which is difficult to quantify. Hagadera reported 18,801 visits to the NCD clinic during the study period, of which 42.1% (n=7918) were reported in the NCD app. In Kakuma, a total of 350,776 visits to the medical and surgical outpatient departments were reported for a broader range of conditions; of these consultations, 9385 (2.7%) were for NCDs, including 4264 (1.2%) for hypertension and 2415 (0.7%) for diabetes, suggesting that the EMR did not aptly characterize patient diagnoses.

Table . Demographics and diagnosed condition(s) of patients with hypertension and diabetes.

Characteristic		NCD[a] app in the Hagadera camp (n=1539)	EMR[b] in the Kakuma camp (n=68)
Age (years), mean (SD)[c]		52.6 (14.6)	52.6 (14.1)
Sex, n (%)			
	Male	610 (39.7)	16 (23.5)
	Female	928 (60.3)	52 (76.5)
Diagnosed condition(s) of interest, n (%)			
	Hypertension only	898 (58.3)	65 (95.6)
	Diabetes only	357 (23.2)	2 (2.9)
	Hypertension and diabetes	284 (18.5)	1 (1.5)

[a]NCD: noncommunicable disease.

[b]EMR: electronic medical record.

[c]NCD app: n=1501; EMR: n=68.

Reporting frequency for key information of interest (eg, relevant biometric data, comorbidities, and risk factors) was assessed as a measure of the quality of care and is summarized in Table 2. Patient age, sex, prescribed medicines, random blood sugar (among patients with diabetes), and smoking status were consistently reported in both the NCD app and the EMR. In contrast, reporting of comorbidities, complications, HbA_{1c} (the gold-standard metric of longer-term glycemic control), and diet were rarely reported in either platform. Indicators frequently reported in the NCD app but not the EMR included the number of health facility visits, BMI, blood pressure, smoking history, physical activity, and next follow-up visit. As previously noted, figures for the Kakuma EMR should be interpreted with caution given that a substantial number of patients with NCD were likely not captured by this study.

Table . Reporting quality: the proportion of records with key data reported (full study period).

Record	NCD[a] app in the Hagadera camp (n=7918), n (%)	EMR[b] in the Kakuma camp (n=68), n (%)
Patient demographics		
Age	7801 (98.5)	68 (100)
Sex	7908 (99.9)	68 (100)
Clinical history and presentation		
Comorbidities	219 (2.8)	0 (0)
Complications[c]	248 (3.1)	0 (0)
Number of visits to facility	7839 (99)	0 (0)
Any prescribed medicine(s)		
Among all patients	7834 (98.9)	68 (100)
Among patients with hypertension only[d]	6105 (100)	66 (100)
Among patients with diabetes only[e]	3432 (100)	3 (100)
Biometrics		
BMI	7868 (99.4)	0 (0)
Height	7868 (99.4)	0 (0)
Weight	7918 (100)	5 (7.4)
Blood pressure		
Both systolic and diastolic reported	7918 (100)	51 (75)
Systolic only	7918 (100)	0 (0)
Diastolic only	7918 (100)	0 (0)
Neither systolic nor diastolic reported	0 (0)	17 (25)
Blood sugar		
Random blood sugar (all patients)	3505 (44.3)	6 (8.8)
Random blood sugar (patients with diabetes)[e]	3368 (98.1)	3 (100)
Hemoglobin A_{1c} (patients with diabetes)[e]	10 (0.3)	0 (0)
Lifestyle risk factor		
Smoking		
Current smoking status	7839 (99)	68 (100)
History of tobacco intake	7839 (99)	0 (0)
Physical activity or sedentary lifestyle	7843 (99.1)	0 (0)
Diet		
Salt intake	522 (6.6)	0 (0)
Carbohydrate intake	81 (1)	0 (0)
Protein intake	10 (0.1)	0 (0)
Fat intake	2 (0)	0 (0)

Record	NCD[a] app in the Hagadera camp (n=7918), n (%)	EMR[b] in the Kakuma camp (n=68), n (%)
Follow-up: next care visit date	7884 (99.6)	10 (14.7)

[a]NCD: noncommunicable disease.

[b]EMR: electronic medical record.

[c]Complications include amputation, blindness, stroke, end-stage renal disease requiring dialysis, and other.

[d]NCD app: n=6105; EMR: n=66.

[e]NCD app: n=3432; EMR: n=3.

Reporting Quality Changes Over Time

The completeness of reporting was used as a metric of the quality of care, where having longitudinal information can be important for decision-making in chronic conditions care. Due to the limited Kakuma EMR sample size (first study quarter n=4; last study quarter n=28), reporting quality over time was evaluated only for the NCD app. A review of the paper record system that was in place in Hagadera prior to the adoption of the NCD app found numerous gaps in the information available in patient charts as well as inconsistencies in reporting across visits. Given that one of the primary aims of using the app was improved reporting, changes in reporting were tracked over time; reporting rates for the first and last quarters of the NCD app implementation period are presented in Table 3.

Completeness in reporting was consistently high across the implementation period, with little meaningful change over time. Nearly all fields were reported for ≥98% of patients each quarter. The exceptions were the reporting of complications (<4%), comorbidities (<3%), and diet (<6%), which were due to the data entry structure. Complications and comorbidities were entered through a multicheckbox question, with checkboxes for each complication or comorbidity observed but no checkbox for “none observed.” Diet was captured using an open-text question in a patient’s medical history, but according to information shared during qualitative interviews with the providers, most clinicians left the field blank because it did not include specific guidance on what to include as part of “diet and salt intake.”

Table . Completeness of reporting in the Hagadera noncommunicable disease app records in first and last study quarters.

Record			First quarter (May to July 2021; n=2380), n (%)	Last quarter (January to March 2022; n=1645), n (%)
Patient demographics				
	Age		2367 (99.5)	1604 (97.5)
	Sex		2380 (100)	1643 (99.9)
Clinical history and presentation				
	Comorbidities		63 (2.6)	56 (3.4)
	Complications[a]		60 (2.5)	54 (3.3)
	Number of visits to facility		2349 (98.7)	1633 (99.3)
Any prescribed medicine(s)				
	Among all patients		2349 (98.7)	1630 (99.1)
	Among patients with hypertension only[b]		1860 (100)	1242 (100)
	Among patients with diabetes only[c]		1052 (100)	692 (100)
Biometrics				
	BMI		2367 (99.5)	1623 (98.7)
	Height		2367 (99.5)	1623 (98.7)
	Weight		2380 (100)	1645 (100)
	Blood pressure			
		Systolic	2380 (100)	1645 (100)
		Diastolic	2380 (100)	1645 (100)
	Blood sugar (among patients with diabetes)			
		Random blood sugar[c]	1027 (97.6)	679 (98.1)
		Hemoglobin A_{1c}[c]	0 (0)	2 (0.3)
Lifestyle risk factor				
	Smoking			
		Current smoking status	2349 (98.7)	1633 (99.3)
		History of tobacco intake	2349 (98.7)	1633 (99.3)
	Physical activity or sedentary lifestyle		2352 (98.8)	1633 (99.3)
	Diet			
		Salt intake	130 (5.5)	104 (6.3)
		Carbohydrate intake	26 (1.1)	18 (1.1)
		Protein intake	3 (0.1)	2 (0.1)
		Fat intake	0 (0)	0 (0)
Follow-up: next care visit date			2365 (99.4)	1639 (99.6)

[a]Complications include amputation, blindness, stroke, end-stage renal disease, and other.

[b]First quarter: n=1860; last quarter: n=1242.

[c]First quarter: n=1052; last quarter: n=692.

Hypertension and Diabetes Management

Given the ultimate goal of improving health outcomes, several key biometric indicators were analyzed to examine change from the first to last quarter of NCD app implementation (Table 4). Differences in these indicators and their change over time between male and female patients and across patient age groups are also provided in Multimedia Appendix 2. Mean BMI was nearly identical in the first and last study quarters. Although the proportion of patients classified as overweight or obese decreased by 3.4% (95% CI −7.3% to 0.6%; *P*=.09), the change was not statistically significant. Approximately half of patients had elevated blood pressure in the first (599/1094, 54.8%) and last (720/1395, 51.6%) study quarters, with an overall decrease of 3.1% (95% CI −7.1% to 0.8%; *P*=.12). Elevated glucose or blood sugar levels were reported for 62.8% (302/481) of patients overall in the first study quarter and 62.1% (374/602) in the last quarter, equating to a 0.7% (95% CI −6.5% to 5.1%; *P*=.82) decrease during the study period.

Table . BMI, blood pressure, and blood sugar levels among patients with hypertension or diabetes in the Hagadera camp.

		First quarter, value (95% CI)	Last quarter, value (95% CI)	Change over time	
				Value (95% CI)	*P* value
BMI[a,b]					
	Mean BMI (kg/m^2)	25.2 (24.9 to 25.4)	25.1 (24.9 to 25.3)	N/A[c]	
	Normal (%)	54.5 (51.5 to 57.5)	57.9 (55.3 to 60.5)	N/A	
	Overweight (%)	34.4 (31.6 to 37.3)	31.1 (28.6 to 33.6)	N/A	
	Obese (%)	11.0 (9.2 to 12.9)	11.0 (9.3 to 12.7)	N/A	
	Overweight or obese (%)	N/A	N/A	−3.4 (−7.3 to 0.6)	.09
Blood pressure[d,e]					
	Normal (%)	45.2 (42.3 to 48.2)	48.4 (45.8 to 51.0)	N/A	
	Elevated (%)	54.8 (51.8 to 57.7)	51.6 (49.0 to 54.2)	−3.1 (−7.1 to 0.8)	.12
Blood sugar[f,g]					
	Target achieved (%)	37.2 (32.9 to 41.5)	37.9 (34.0 to 41.8)	N/A	
	Elevated (%)	62.8 (58.5 to 67.1)	62.1 (58.2 to 66.0)	−0.7 (−6.5 to 5.1)	.82

[a]Defined as <25 kg/m^2=normal, 25-29.9 kg/m^2=overweight, and ≥30 kg/m^2=obese.

[b]First quarter: n=1086; last quarter: n=1373.

[c]N/A: not applicable.

[d]Normal blood pressure defined as systolic<140 mm Hg and diastolic<90 mm Hg; elevated blood pressure defined as systolic>140 mm Hg or diastolic>90 mm Hg.

[e]First quarter: n=1094; last quarter: n=1395.

[f]Among patients with diabetes and based on results from either random blood sugar (RBS) or hemoglobin A_{1c} (HbA_{1c}; target blood sugar level defined as HbA_{1c}<7.0% or RBS<11.1 mmol/L), with preference given to HbA_{1c} results (when available) or most recent RBS test in cases of multiple measures.

[g]First quarter: n=481; last quarter: n=602.

Provider Perceptions

Overall, providers were satisfied with the app and indicated it standardized patient information and made consultations easier. There was consensus that access to information from previous consultations was easier when using the app and that this was beneficial for NCD control and follow-up. Prior to the implementation of the app, patient data were collected using a paper medical record, which was stored in file cabinets. Providers mentioned that these records frequently got lost and that the use of the app eliminated data loss and allowed for information to be stored securely. As one provider mentioned, “Most important aspect is the traceability of the data and the way the data is saved and stored. There is no problem if the patient loses their patient’s card or number.”

Although all providers were supportive of continued use of the app and felt that it was beneficial for multiple aspects of patient care, several challenges were observed. Duplication was a concern, where patients with NCD often have multiple diagnoses. For patients with multiple diagnoses, the app format required their information to be entered twice, which increased time requirements for providers. Another concern were field limits restricting minimum and maximum values for selected variables. These were intended to reduce data entry errors and improve data quality but proved problematic in the case of patients with outlying values. One provider shared the following: “When entering patients’ data, for example diabetic patients, with a weight of less than 50 kgs you can’t input that info…also, blood pressure has a cap, some patients with uncontrolled hypertension have high blood pressure which can’t be entered.” Another provider observation was the need for an updated monthly medication list within the app because medication availability and stock-outs can change monthly.

Discussion

Principal Findings

Thorough, consistent, and accessible reporting of patient data is essential for monitoring disease control and informing treatment decisions for patients with chronic NCD such as hypertension and diabetes. As the global NCD burden continues to rise and both the number of humanitarian crises and the scale of displacement increase, innovative solutions are needed to ensure access to and quality of care for NCDs, given the limitations and challenges in health service delivery in humanitarian contexts [4]. Although this study did not explicitly seek to assess the impact of the NCD app on clinical measures, it nevertheless provided important evidence of the successful implementation of an mHealth tool in the context of a refugee camp NCD clinic. Our findings demonstrate consistently higher reporting quality for patients with hypertension and diabetes in a specialized app for NCD management than a more general EMR. Additionally, positive provider perceptions further highlight several important facilitators of and barriers to effective uptake of an mHealth app such as the one evaluated.

EMR systems have meaningful potential to decrease care fragmentation and improve the continuity of care [25]. In conditions such as diabetes and hypertension, disease management, responsive treatment decisions, and the prevention of complications require continuous tracking of clinical measures such as blood sugar, blood pressure, and BMI, as well as timely access to historical patient data [26,27]. The NCD app tested at Hagadera provided valuable evidence of improved patient data tracking, follow-up, and continuity of care. Providers at Hagadera also indicated that by improving access to patients' data, the tool reduced distraction during consultations; this led to improved patient-provider interactions, which is another critical aspect of the effective management of chronic NCDs [7,26].

When implementing a digital tool such as the app in Hagadera, integration into the existing health care service environment is important to consider, as uptake requires providers to adopt and regularly use the tool and for patients to accept its use [6,27,28]. The time required to learn how to use an mHealth tool and the ease of use during appointments are decisive factors for the adoption of new technologies due to the overburden that health care providers often confront in their daily work [29,30]. In the case of the Hagadera experience, the app originally developed for use in Lebanon was substantially modified to meet the localized context in Kenya. Health providers' feedback on the app's utility and the consistently high reporting of patient data within the NCD app demonstrate the benefit of collaborating with users throughout the app's design and implementation, to create a tool that is not only acceptable but also meets users' specific needs and is feasible to implement in day-to-day clinical practice.

Successful deployment of an mHealth app, especially one specified for particular conditions, also requires integration into existing digital ecosystems, which is critical for both efficiency and long-term sustainability. Carefully designed systems integration is essential for uptake and reporting consistency in facilities with digital technologies such as EMRs or other health information systems previously in place [8]. When implemented correctly, this can reduce data redundancy and duplicated data entry, as well as improve reporting capabilities. Although this may be perceived as adding to providers' workload, feedback from users of the NCD app in Hagadera regarding the app's utility and usability suggest that by actively involving end users in the design of the app and using a platform already known to the IRC, the work required to integrate with existing systems was not an influential barrier to the app's use.

One advantage of most mHealth apps is that changes and improvements can be made incrementally. Although a common theme in the qualitative interviews was the Hagadera NCD app's ease of use, several relatively straightforward changes (eg, restrictions on permissible values for entering clinical measures, such as weight, height, and blood pressure, and regularly updating listed medication options to reflect availability) were identified and put in place at the end of the pilot period to improve patient information input and user experience. The ease with which the app can be modified also facilitates interoperability with other tools, applications, and reporting requirements, which can in turn facilitate its utility and benefits [7,27]. Being able to adapt the app to meet reporting requirements more efficiently is particularly valuable in humanitarian contexts, where those providing services often have to report various information to a range of bodies (eg, government, donor, and intraorganizational). With the increasing use of electronic health information systems, the ability to design and subsequently easily adapt mHealth tools to align with existing or new tools and applications is invaluable for ensuring consistency in reporting and the continuity of quality health care, especially in settings with multiple service providers and actors providing or supporting health services, such as in humanitarian crises.

Finally, the ethical aspects of the use of mHealth tools and other digital interventions must be considered due to the sensitive and personally identifiable information that is managed within these systems. Data confidentiality is a critical concern, and in the case of Hagadera, the app appeared to improve secure data collection and storage practices. Data storage within remote or local servers with password-protected access was perceived as an improvement over paper records.

Limitations

Despite the challenges to conducting rigorous research in humanitarian settings, evaluations such as this are nevertheless valuable in contributing to the limited evidence available to inform health services in such contexts. The change in study design necessitated by the unanticipated implementation of an EMR system at the Kakuma health facilities limited the planned comparison and subsequent analyses. The ideal comparison would have been a facility with a paper-based system; however, given increasing use of electronic systems for health information, comparing 2 electronic systems is nevertheless valuable. This evaluation was critically limited by the small number of patient records available for the comparison facilities at Kakuma. This hindered the ability to compare reporting quality over time in both sites and to examine health outcomes among patients with data in the Kakuma EMR. Another limitation was the length of the follow-up period with respect to examining changes in blood pressure and blood sugar control, where longer time periods and more robust metrics (eg, repeat blood pressure measurements, HbA_{1c}, or longitudinal fasting blood sugar readings for blood sugar) are necessary to analyze disease control. Finally, the evaluation approach used quantitative patient data and qualitative provider data but did not capture patients' perspectives on the app's use and if or how it affected interactions at health facilities or the quality of care.

Conclusions

Due demographic transitions and the increasing global NCD burden, more attention to the management of these conditions in complex settings with limited resources is critical. Given the growing burden of NCDs, strategies to increase the quality, continuity, and effectiveness of care that are relevant to primary care settings are urgently needed. In this study of EMR adoption in Kenyan refugee settings, a specialized record for NCDs outperformed a more general record intended for use in all patients in terms of reporting completeness. The demonstrated feasibility of high-quality, standardized routine data collection for key health parameters is a critical step toward improving the quality and continuity of care, but it was insufficient to improve blood pressure and blood sugar during the study period. Improving NCD care and outcomes in humanitarian and low-income settings will be a long-term and multifaceted undertaking. Enhanced recordkeeping is feasible, is valued by providers, and is an important first step to improving the quality of care. This CommCare-based NCD app can easily be rolled out in similar humanitarian settings with minimal adaptation; however, the adaptation of technologies to the local context and use case is critical for uptake and ensuring that workflows and time burden do not outweigh the benefits of EMRs.

Acknowledgments

We would like to acknowledge contributions from many individuals that made this work possible. Sarvesh Tewari and the team from Dimagi adapted the Sana Mobile health app for use with CommCare and led training and technical support for the app's use. This work would not have been possible without many colleagues from the International Rescue Committee, including Aston Benjamin Atwiine, Stephen Kinara, Winston Luseno, Mouna Mayoufi, and Felix Muloma. We also gratefully acknowledge Noah Baker from the Johns Hopkins Bloomberg School of Public Health for his data-cleaning support. This research was funded by Grand Challenges Canada. The funding body had no role in the design or implementation of the research and did not participate in the analysis or presentation of findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Images of the Sana.NCD mobile health app interface.

[PDF File, 806 KB - mhealth_v11i1e43878_app1.pdf]

Multimedia Appendix 2

BMI, blood pressure, and blood sugar levels among patients with hypertension or diabetes in the Hagadera camp by patient sex and age group.

[PDF File, 198 KB - mhealth_v11i1e43878_app2.pdf]

References

1. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387. [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
2. Noncommunicable diseases. World Health Organization. 2022 Sep 16. URL: www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-02-15]
3. Blanchet K, Sistenich V, Ramesh A, Frison S, Warren E, Smith J, et al. An evidence review of research on health interventions in humanitarian crises. ELRHA. 2015 Oct. URL: www.elrha.org/wp-content/uploads/2015/01/Evidence-Review-22.10.15.pdf [accessed 2022-10-24]
4. Demaio A, Jamieson J, Horn R, de Courten M, Tellier S. Non-communicable diseases in emergencies: a call to action. PLoS Curr 2013 Sep 6;5:ecurrents.dis.53e08b951d59ff913ab8b9bb51c4d0de. [doi: 10.1371/currents.dis.53e08b951d59ff913ab8b9bb51c4d0de] [Medline: 24056956]
5. Qiang CZ, Yamamichi M, Hausman V, Miller R, Altman D. Mobile applications for the health sector. World Bank. 2012 Apr. URL: documents.worldbank.org/en/publication/documents-reports/documentdetail/751411468157784302/mobile-applications-for-the-health-sector [accessed 2022-10-24]
6. Veinot TC, Zheng K, Lowery JC, Souden M, Keith R. Using electronic health record systems in diabetes care: emerging practices. IHI 2010;2010:240-249. [doi: 10.1145/1882992.1883026] [Medline: 25264545]

7. Azizi A, Aboutorabi R, Mazloum-Khorasani Z, Hoseini B, Tara M. Diabetic personal health record: a systematic review article. Iran J Public Health 2016 Nov;45(11):1388-1398. [Medline: 28032056]
8. Lessing SE, Hayman LL. Diabetes care and management using electronic medical records: a systematic review. J Diabetes Sci Technol 2019 Jul;13(4):774-782. [doi: 10.1177/1932296818815507] [Medline: 30556418]
9. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Jul 15;16:572. [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]
10. Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KSF. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 6;19(1):e13. [doi: 10.2196/jmir.5876] [Medline: 28062391]
11. Khader A, Farajallah L, Shahin Y, Hababeh M, Abu-Zayed I, Kochi A, et al. Cohort monitoring of persons with diabetes mellitus in a primary healthcare clinic for Palestine refugees in Jordan. Trop Med Int Health 2012 Dec;17(12):1569-1576. [doi: 10.1111/j.1365-3156.2012.03097.x] [Medline: 23051859]
12. Khader A, Farajallah L, Shahin Y, Hababeh M, Abu-Zayed I, Kochi A, et al. Cohort monitoring of persons with hypertension: an illustrated example from a primary healthcare clinic for Palestine refugees in Jordan. Trop Med Int Health 2012 Sep;17(9):1163-1170. [doi: 10.1111/j.1365-3156.2012.03048.x] [Medline: 22845700]
13. Khader A, Ballout G, Shahin Y, Hababeh M, Farajallah L, Zeidan W, et al. Diabetes mellitus and treatment outcomes in Palestine refugees in UNRWA primary health care clinics in Jordan. Public Health Action 2013 Dec 21;3(4):259-264. [doi: 10.5588/pha.13.0083] [Medline: 26393043]
14. Khader A, Ballout G, Shahin Y, Hababeh M, Farajallah L, Zeidan W, et al. Treatment outcomes in a cohort of Palestine refugees with diabetes mellitus followed through use of e-health over 3 years in Jordan. Trop Med Int Health 2014 Feb;19(2):219-223. [doi: 10.1111/tmi.12241] [Medline: 24341942]
15. Istepanian RSH, Mousa A, Haddad N, Sungoor A, Hammadan T, Soran H, et al. The potential of m-health systems for diabetes management in post conflict regions a case study from Iraq. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:3650-3653. [doi: 10.1109/EMBC.2014.6944414] [Medline: 25570782]
16. Saleh S, Farah A, Dimassi H, El Arnaout N, Constantin J, Osman M, et al. Using mobile health to enhance outcomes of noncommunicable diseases care in rural settings and refugee camps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 13;6(7):e137. [doi: 10.2196/mhealth.8146] [Medline: 30006326]
17. Saleh S, Farah A, El Arnaout N, Dimassi H, El Morr C, Muntaner C, et al. mHealth use for non-communicable diseases care in primary health: patients' perspective from rural settings and refugee camps. J Public Health (Oxf) 2018 Dec;40(suppl_2):ii52-ii63. [doi: 10.1093/pubmed/fdy172] [Medline: 30307516]
18. Doocy S, Paik KE, Lyles E, Tam HH, Fahed Z, Winkler E, et al. Guidelines and mHealth to improve quality of hypertension and type 2 diabetes care for vulnerable populations in Lebanon: longitudinal cohort study. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e158. [doi: 10.2196/mhealth.7745] [Medline: 29046266]
19. Doocy S, Paik K, Lyles E, Tam HH, Fahed Z, Winkler E, et al. Pilot testing and implementation of a mHealth tool for non-communicable diseases in a humanitarian setting. PLoS Curr 2017 Jun 5;9:ecurrents.dis.e98c648aac93797b1996a37de099be74. [doi: 10.1371/currents.dis.e98c648aac93797b1996a37de099be74] [Medline: 28744410]
20. Dadaab refugee complex. UNHCR Kenya. URL: www.unhcr.org/ke/dadaab-refugee-complex [accessed 2022-08-28]
21. Noncommunicable diseases progress monitor 2020. World Health Organization. 2020 Jun 12. URL: www.who.int/publications/i/item/9789240000490 [accessed 2022-02-15]
22. Commcare. Dimagi. URL: www.dimagi.com/commcare [accessed 2022-08-28]
23. Integrated refugee health information system (iRHIS). UNHCR. URL: his.unhcr.org/home [accessed 2022-09-27]
24. WHO package of essential Noncommunicable (PEN) disease interventions for primary health care. World Health Organization. 2020 Sep 7. URL: www.who.int/publications/i/item/who-package-of-essential-noncommunicable-(pen)-disease-interventions-for-primary-health-care [accessed 2022-09-13]
25. Williams F, Boren SA. The role of the electronic medical record (EMR) in care delivery development in developing countries: a systematic review. Inform Prim Care 2008;16(2):139-145. [doi: 10.14236/jhi.v16i2.685] [Medline: 18713530]
26. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]
27. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. Prev Chronic Dis 2013;10:E26. [doi: 10.5888/pcd10.120180] [Medline: 23428085]
28. Chaplin B, Meloni S, Eisen G, Jolayemi T, Banigbe B, Adeola J, et al. Scale-up of networked HIV treatment in Nigeria: creation of an integrated electronic medical records system. Int J Med Inform 2015 Jan;84(1):58-68. [doi: 10.1016/j.ijmedinf.2014.09.006] [Medline: 25301692]
29. Kavuma M. The usability of electronic medical record systems implemented in Sub-Saharan Africa: a literature review of the evidence. JMIR Hum Factors 2019;6(1):e9317. [doi: 10.2196/humanfactors.9317] [Medline: 30801251]
30. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010 Aug 6;10:231. [doi: 10.1186/1472-6963-10-231] [Medline: 20691097]

Abbreviations

EMR: electronic medical record
HbA_{1c}: hemoglobin A_{1c}
IRC: International Rescue Committee
mHealth: mobile health
MIT: Massachusetts Institute of Technology
NCD: noncommunicable disease
UNRWA: United Nations Relief and Works Agency for Palestine Refugees in the Near East

Edited by L Buis; submitted 27.10.22; peer-reviewed by K Morley, P Khorasani; revised version received 24.05.23; accepted 11.07.23; published 05.10.23.

Please cite as:
Lyles E, Paik K, Kiogora J, Hussein H, Cordero Morales A, Kiapi L, Doocy S
Adoption of Electronic Medical Records for Chronic Disease Care in Kenyan Refugee Camps: Quantitative and Qualitative Prospective Evaluation
JMIR Mhealth Uhealth 2023;11:e43878
URL: https://mhealth.jmir.org/2023/1/e43878
doi:10.2196/43878

Retention in an mHealth App Aiming to Promote the Use of HIV Pre-Exposure Prophylaxis Among Female Sex Workers in Dar es Salaam, Tanzania: Prospective Cohort Study

Christopher H Mbotwa[1,2], MA; Method R Kazaura[1], PhD; Kåre Moen[3], MD, PhD; Melkizedeck T Leshabari[4], PhD; Emmy Metta[4], PhD; Elia J Mmbaga[1,3], MD, PhD

1

2

3

4

Corresponding Author:
Christopher H Mbotwa, MA

Abstract

Background: Increasing access to smartphones in sub-Saharan Africa offers an opportunity to leverage mobile health (mHealth) technology to improve access to health care in underserved populations. In the domain of HIV prevention, mHealth interventions can potentially contribute to solving the challenges of suboptimal adherence to pre-exposure prophylaxis (PrEP) and low retention in PrEP services among populations most vulnerable to HIV acquisition. However, there is a gap in the knowledge about the use of such interventions in sub-Saharan Africa.

Objective: This study aims to evaluate the extent and predictors of retention in an mHealth app (*Jichunge*) that aims to promote adherence to PrEP and retention in PrEP care among female sex workers in Dar es Salaam, Tanzania.

Methods: A prospective cohort of female sex workers residing in Dar es Salaam were recruited, using respondent-driven sampling. All participants were provided with the *Jichunge* app as they started PrEP. A questionnaire was used to collect data on sociodemographics and other structural factors, while app use data for the 60-day period following the first 150 days of being in the intervention arm were extracted from the app's back end. A multivariable log-binomial model was used to determine predictors of 6-month retention in the *Jichunge* app.

Results: A total of 470 female sex workers were recruited. Nearly three-quarters of participants (206/284, 72.5%) who came to the 6-month follow-up interview no longer had the *Jichunge* app on their phones. The majority of these participants (193/206, 93.7%) no longer had access to the app because of issues related to their phones. Data extracted from the back end of the app showed that the use of the app declined over time, and only 13.4% (63/470) of the participants were retained (continued to use the app) after 6 months of intervention. At 6 months, women aged ≥35 years were >2 times more likely to use the app than women aged 18 to 24 years (adjusted risk ratio [aRR] 2.2, 95% CI 1.2-4.1; *P*=.01). Furthermore, retention in the app was higher among participants who demonstrated high PrEP awareness at baseline (aRR 1.8, 95% CI 1.1-3; *P*=.01) and among those who had experienced financial difficulties due to health care spending (aRR 1.9, 95% CI 1.2-3.2; *P*=.01).

Conclusions: Most female sex workers (206/284, 72.5%) who were enrolled in PrEP care in Tanzania no longer used the *Jichunge* app after 6 months. Retention in the app at 6 months was predicted by older age, high PrEP awareness, and financial difficulties due to health care spending. Strategies for the long-term retention of participants in mHealth apps, such as systems for reinstallations of apps, should be considered during the design phase.

Trial Registration: Pan African Clinical Trials Registry PACTR202003823226570; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=9781

(JMIR Mhealth Uhealth 2023;11:e46853) doi:10.2196/46853

KEYWORDS

mobile health; retention; engagement; mHealth; sex workers; pre-exposure prophylaxis; HIV; Africa; PrEP

Introduction

Mobile health (mHealth) apps have increasingly been reported to be effective in promoting access and adherence to treatment for both communicable diseases and noncommunicable diseases globally [1]. The increased use of smartphones in low- and middle-income countries, such as many of those in sub-Saharan Africa, offers an opportunity to leverage mHealth technology for health promotion purposes [2]. mHealth can be useful in minimizing some structural barriers to health care access (eg, by reducing the impact of long distance to health care facilities and reducing the likelihood of stigma experiences associated with certain conditions and treatments) and hence potentially contributes to the achievement of universal health coverage [3,4].

In the domain of biomedical HIV prevention, mHealth interventions have been shown to improve the effectiveness of pre-exposure prophylaxis (PrEP; eg, in the United States) through supporting adherence to medication and retention in care [5-7]. Information on the use of mHealth for such purposes is limited in sub-Saharan Africa. Furthermore, there is a scarcity of evidence on the extent and predictors of the longer-term use of mHealth apps for HIV prevention in populations most vulnerable to HIV acquisition and transmission. The longer-term use of many mHealth apps is fundamental if the whole spectrum of their intended effects is to be realized. Therefore, understanding the long-term use of various types of mHealth apps, as well as the association between use and different types of users and user characteristics, is vital if one aims to develop successful and innovative interventions tailored to the specific needs of the targeted population. The use of any app can be influenced by individual and structural factors, as well as by the condition for which its use is intended. For example, Crafoord et al [8] found that older age and higher education predicted higher engagement with an interactive mHealth app among patients with breast cancer or prostate cancer in Sweden. On the other hand, a study by Nelson et al [9] found that the use of a text message intervention to support self-care among patients with type 2 diabetes in the United States did not differ by sociodemographic characteristics.

In this study, we focused on retention in an mHealth app that aims to support PrEP use among female sex workers in Tanzania. Female sex workers are disproportionately affected by HIV [10,11]. According to the UNAIDS (Joint United Nations Programme on HIV/AIDS), the risk of acquiring HIV is 30 times higher among female sex workers than that among adult women in general [10]. In Tanzania, a recent study estimated the HIV prevalence among female sex workers in Dar es Salaam to be 15.3% [12], which is 3 times higher than the HIV prevalence in the general population [13]. In all African countries except Senegal, sex work is criminalized and not regulated by health policies [14], and this increases female sex workers' vulnerability to stigma, HIV infection, and a lack of adequate access to health services [15,16]. Thus, interventions that may do away with some barriers to health care service access among female sex workers could potentially make significant contributions toward achieving the global goal of ending HIV as a public health challenge by 2030.

As part of the Pragmatic Trial for HIV Pre-Exposure Prophylaxis Roll-Out in Tanzania (PREPTA), an interactive and free-of-charge smartphone app (called *Jichunge*) was developed with the aim of supporting the initiation and use of PrEP among the following two groups, who are at increased HIV risk: female sex workers and men who have sex with men. In this paper, we evaluate the extent and predictors of retention in the *Jichunge* app among female sex workers after a duration of 6 months.

Methods

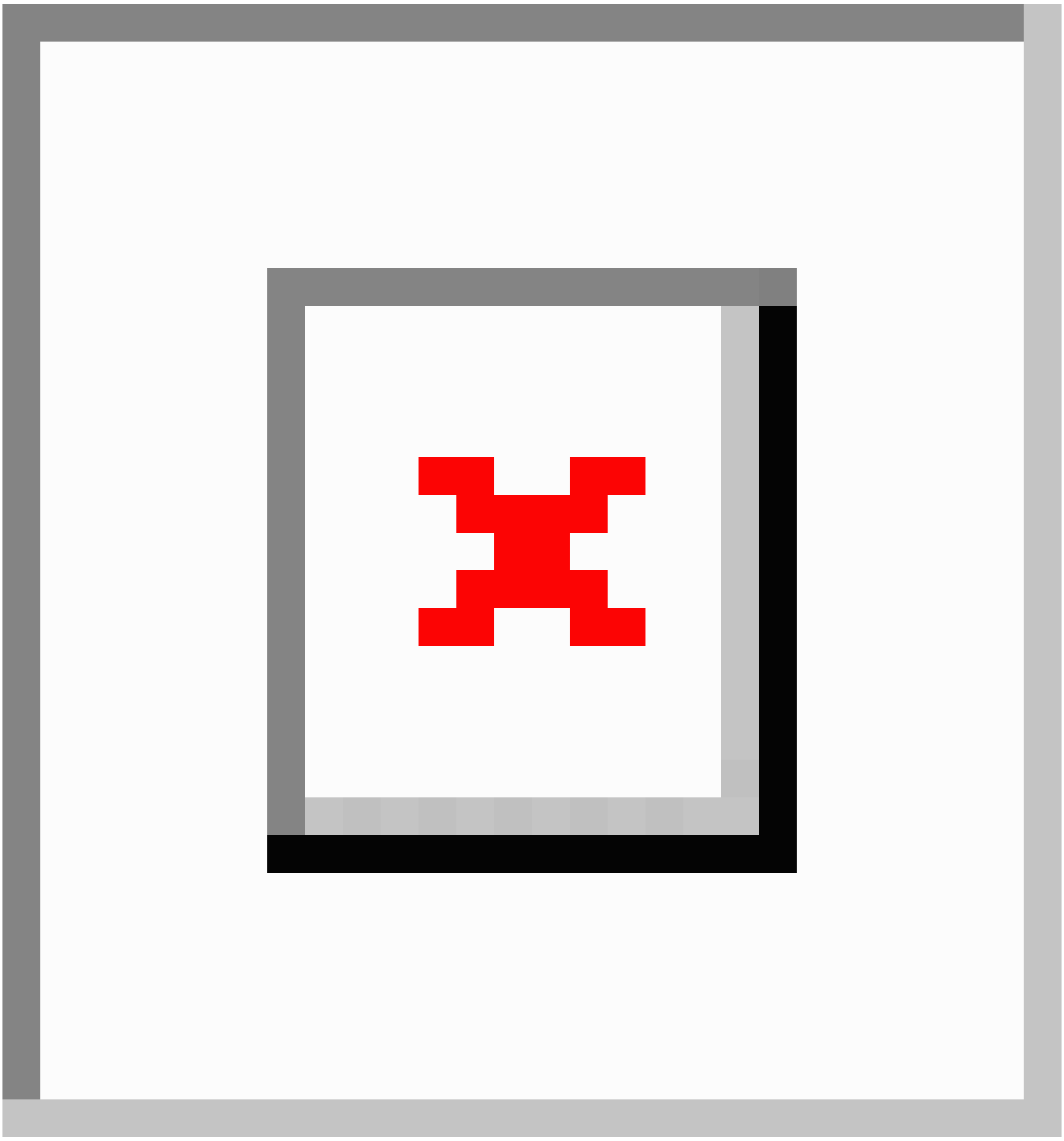
Study Design

This study was a prospective cohort analysis of participants who were recruited into the PREPTA study, which was partly described in our previous publications [17,18]. Briefly, the PREPTA is a research project with the overall aim of assessing the effectiveness of an mHealth intervention in promoting retention in PrEP care and adherence to PrEP among female sex workers and men who have sex with men in the following two cites of Tanzania: Dar es Salaam (intervention arm) and Tanga (control arm). The project is being implemented by the Muhimbili University of Health and Allied Sciences, Tanzania, and the University of Oslo, Norway. This study analyzed app usage after 6 months among female sex workers who were recruited in the intervention arm of the trial.

Intervention

A participatory design approach was adopted when developing the *Jichunge* app [17]. The app provides users with information about PrEP, reminds them to take their daily pill, provides an opportunity to consult a doctor and a peer educator via the web, and operates a web-based forum for anonymous discussions among PrEP users (Figure 1). The overall aim of the app is to improve adherence to PrEP and retention in the PrEP program. Further details about the *Jichunge* app can be found in our previous publications [17,18].

Figure 1. Main functionalities of the *Jichunge* app. PrEP: pre-exposure prophylaxis.



Study Population and Sample Size

We included women aged at least 18 years who had sold sex in the past 3 months. Participants were residents of Dar es Salaam (they had an address in the city and had lived there for the past 6 months), owned a smartphone, and were prepared to start PrEP treatment. The sample size for this analysis was estimated by using a formula for cohort studies [19,20]. Since the retention in PrEP services and adherence to PrEP were not known at the design stage of this study, we used a proportion (p) of 50%, which has been shown to provide an optimum sample size when the actual proportion is between 10% and 90% [21]. A precision of 5% and a design effect of 2 were used in estimating the sample size to achieve a statistical power of 80%. After adjusting for a 20% potential loss to follow-up, the minimum sample size was estimated to be 423 female sex workers.

Study Procedures

Participants were recruited between March and June 2021, using respondent-driven sampling (RDS). RDS is a method for sampling from populations for which there is no available sampling frame [19,22]. On the first day, 3 initial participants (referred to as "seeds") were recruited purposively by researchers and peer educators who had prior experience with biobehavioral surveys among female sex workers in Tanzania. The first seed was aged 20 years and was a street-based sex worker. The second seed was 32 years old and was a phone- and internet-based sex worker. The third seed was 28 years old and was a bar- and pub-based sex worker. Before the end of the recruitment phase, an additional 6 seeds were added to speed up recruitment, as individuals in the initial seeds' recruitment chains were facing difficulties in recruiting more participants. After having completed the study procedures, each seed was given 3 recruitment coupons and asked to pass them on to her peers and invite them to participate in the PREPTA study. Of the 9 seeds, only 1 did not recruit any other participant to the study. Every new study participant was asked to do the same until we had reached the desired sample size.

All participants were screened for PrEP eligibility by a health clinic, as per national guidelines [23]. PrEP eligibility criteria included a negative HIV test result, no signs of acute HIV infection, a serum creatinine clearance of >60 mL/min, and the willingness to start PrEP. Those who qualified were provided with PrEP pills for 30 days and thereafter invited to take part in the PREPTA study. After consenting to participate in the study, participants received the *Jichunge* app (free of charge), attended an app onboarding session, and thereafter participated in a face-to-face baseline interview with a trained interviewer. The *Jichunge* app was only available to study participants, and it was not available in mainstream app stores. If study participants needed to reinstall the app, they had to contact the project site managers and be given a link, through which they could download the app again. Upon completion of the interview, each participant was paid the modest amount of TZS 8000 (US $3.18) as compensation for transport costs and the time spent at the study site. In addition, participants received TZS 4000 (US $1.59) for each peer they referred to the study, as per the RDS protocol.

All participants were invited to participate in a 6-month follow-up interview regardless of whether they had used PrEP and the *Jichunge* app. For the 6-month follow-up data collection, we attempted to contact each participant by phone at least 3 times on different days before they were considered lost to follow-up.

Data Collection

Information on sociodemographics and other structural factors at baseline was collected by using a questionnaire, which was administered by trained research assistants. At 6 months following enrollment, participants were invited to participate in a follow-up survey consisting of an interviewer-administered questionnaire that focused on their experiences with PrEP and the *Jichunge* app. From the back end of the app, we also continuously collected data on participants’ use of the app’s different functionalities (ie, opening the app, registering medicine taking, reading editorial contents, accessing a web-based consultation, and entering the web-based discussion forum). Data from the back end of the app were collected from all participants regardless of whether they participated in the 6-month follow-up survey. In this study, we included user statistics pertaining to participants' use of the *Jichunge* app after 6 months (ie, the 60-day period following the first 150 days of being in the intervention arm).

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki. Details on how we worked to protect participants’ privacy were provided in a previous article [18]. Briefly, data from the questionnaires were stored on a secure server that was developed particularly for the handling of sensitive research data (a service for sensitive data called “TSD” [24]) at the University of Oslo, Norway. The *Jichunge* app does not store, transfer, or expose sensitive data. Further, no element in the user interface of the app refers to any information that can identify a participant. Written informed consent was obtained from all participants involved in this study for the collection and use of both questionnaire data and app data. The protocol for this study was reviewed and approved by the National Health Research Ethics Committee in Tanzania (protocol code: NIMR/HQ/R.8a/Vol. IX/3454) and by the Regional Committee for Medical and Health Research Ethics in Norway (protocol reference number: 33675).

Study Variables

Outcome Variable

The primary outcome variable for this study was 6-month retention in the *Jichunge* app. Retention in an mHealth app is one of the metrics for user engagement with mHealth apps [25-27]. In this study, participants were considered retained if they still used the app after 6 months. We assumed that after 6 months of intervention, a user would have established a routine and integrated the *Jichunge* app into their health care regimen. According to the PrEP implementation framework in Tanzania, a 6-month period is considered long enough to monitor a PrEP user for the effective use of PrEP and the presence of continuous HIV risk factors [23]. Accordingly, a 6-month period is long enough to evaluate user engagement with mHealth interventions that aim to promote the use of PrEP in the context of Tanzania. We defined *6-month retention* as app use within the 150 to 210 days following the first installation of the app, that is, within this 60-day period, one needed to open the app at least once to be categorized as “retained.”

Independent Variables

Independent variables were sociodemographics, sex work characteristics, and some sociostructural factors that were asked about at baseline and considered potentially associated with 6-month retention in the app. These included the age of the respondent, marital status, education level, social support, awareness of PrEP, experience with PrEP and sex work stigma, self-perceived HIV risk, and income from sex work. The measurement of these variables was done as described in the following paragraphs.

Social support was measured by using a Likert scale of 8 items, which was adapted from the Duke–UNC (University of North Carolina) Functional Social Support Questionnaire [28]. Participants were asked to respond to each question by choosing 1 of 5 possible responses (1=*Much less than I would like*; 2=*Less than I would like*; 3=*Some, but would like more*; 4=*Almost as much as I like*; 5=*As much as I like*). We computed the total score for all items, and a total score below 32 was considered to indicate "inadequate social support." The scale had a Cronbach α of .88, indicating high internal consistency.

Sex work stigma and perceived PrEP stigma were measured by using 13 and 10 scale items, respectively; each item had 5 response options (1=*Strongly disagree*; 2=*Disagree*; 3=*Neither disagree nor agree*; 4=*Agree*; 5=*Strongly agree*), and the scales had a Cronbach α of .84 and .88 respectively, signifying high reliability. Sex work stigma was thereafter categorized into the following three categories: "low" for scores of ≤26, "moderate" for scores between 27 and 38, and "high" for scores of ≥39. For PrEP stigma, a score above 30 was considered "high."

PrEP awareness was measured by using 8 true or false questions about PrEP. Participants who answered more than 6 questions correctly were categorized as being highly aware of PrEP.

Statistical Analyses

We started by generating the distribution of sociodemographic characteristics and descriptive statistics related to the use of the *Jichunge* app. Using the chi-square test, we assessed the association of 6-month retention in the app with sociodemographics, sex work characteristics, and other sociostructural factors. Since the outcome variable was binary and relatively common (proportion of retained participants: 63/470, 13.4%), we used a generalized linear model with a log link and binomial distribution (log-binomial regression) to examine independent predictors of 6-month retention in the app [29]. Log-binomial regression was chosen because it estimates the risk ratios directly, as opposed to logistic regression, which gives odds ratios that can overestimate the association if the outcome variable is common (proportion of >10%) [30]. We started with a bivariate analysis to obtain crude estimates of risk ratios. All variables with a *P* value of <.25 in the bivariate analysis were included in the multivariable model to obtain adjusted estimates of risk ratios. All analyses were performed using Stata version 17 (StataCorp LLC), and a *P* value of <.05 was considered statistically significant.

Results

Sociodemographic Characteristics

A total of 470 participants (age: median 26, IQR 22-30 years) were recruited at baseline. Of these, 277 (58.9%) reported to have had at least some secondary education, and 361 (76.8%) had never been married. Out of all 470 participants, we managed to reach 340 (72.3%) participants for the 6-month follow-up survey, and 284 (60.4%) participants were interviewed.

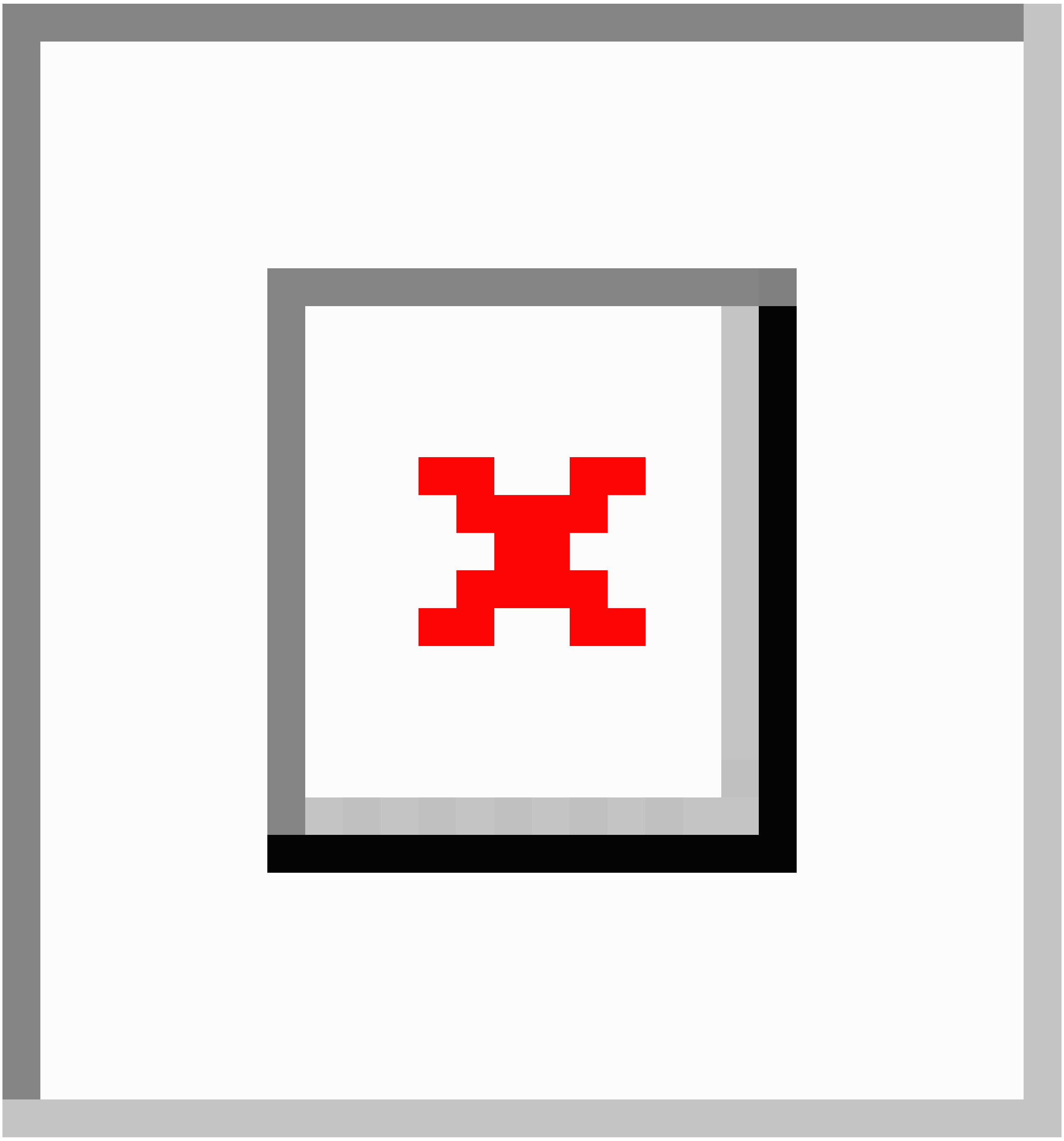
Participants' Possession of the *Jichunge* App After 6 Months

Participants who came to the 6-month follow-up interview (n=284) were asked if they still had the *Jichunge* app on their phones; nearly three-quarters (206/284, 72.5%) no longer had the app. When asked about the reasons for this, a large majority of these 206 study participants (n=193, 93.7%) said that they no longer had access to the app because of issues related to their phones, such as the phone having been stolen or lost (n=89, 43.2%), the phone having been sold or changed (n=53, 25.7%), or the phone having technical problems (n=51, 24.8%); only a few (n=13, 6.3%) reported to have uninstalled the app from their phones.

Pattern of the Use of the *Jichunge* App and Its Services

More than three-quarters of the participants (376/470, 80%) had accessed the app at least once after enrollment into the PREPTA study. Figure 2 shows the proportion of study participants who had opened the app and used different *Jichunge* services at least once per month during the 6-month observation period. App use was high in the first month and declined continually in the subsequent months. For all months, pill registration was the most used service, while web-based consultation was the least used service.

Figure 2. Pattern of use of the *Jichunge* app and its services.



Retention in the *Jichunge* App and Use of Its Services After 6 Months

Overall, of the 470 participants, 63 (13.4%) were retained in the sense that they had opened the app at least once after 6 months of enrollment. Specifically, of the 470 participants, 52 (11.1%) had registered their daily pill taking, 40 (8.5%) had accessed PrEP editorial contents, 35 (7.4%) had used the discussion forum, and 14 (3%) had used the *Jichunge* web-based consultation services.

Participants' Baseline Characteristics by 6-Month Retention in the *Jichunge* App

In Table 1, we present the distribution of participants' baseline characteristics by retention in the *Jichunge* app at 6 months. Retention increased significantly and linearly with age; 10.6% (21/199) of women aged 18 to 24 years were retained, whereas 12.8% (29/226) of women aged 25 to 34 years and 28.9% (13/45) of women aged ≥35 years were retained (χ^2_2=10.8; *P*=.01). The proportion of retained users was also higher among women with high PrEP awareness (41/232, 17.7%) than that among women with low PrEP awareness (22/238, 9.2%; χ^2_1=7.2; *P*=.01). Retention was higher among female sex workers who had experienced financial difficulties due to health care spending (43/244, 17.6%) than that among female sex workers who had not experienced such difficulties (19/224, 8.5%; χ^2_1=8.2; *P*=.004).

Table . Distribution of participants' baseline characteristics by 6-month retention in the *Jichunge* app (N=470).

Characteristics		All participants, n (%)	Participants retained, n (%)	Participants not retained, n (%)	Chi-square (*df*)	*P* value
Age (years)					10.8 (2)	.01[a]
	18-24	199 (42.3)	21 (10.6)	178 (89.4)		
	25-34	226 (48.1)	29 (12.8)	197 (87.2)		
	≥35	45 (9.6)	13 (28.9)	32 (71.1)		
Marital status					0.02 (1)	.90
	Never married	361 (76.8)	48 (13.3)	313 (86.7)		
	Married or previously married	109 (23.2)	15 (13.8)	94 (86.2)		
Education					2.3 (2)	.32
	Formal education	28 (6)	5 (17.9)	23 (82.1)		
	Primary education	165 (35.1)	17 (10.3)	148 (89.7)		
	Secondary education or higher	277 (58.9)	41 (14.8)	236 (85.2)		
Social support[b]					2.6 (1)	.11
	Inadequate	283 (60.2)	44 (15.6)	239 (84.4)		
	Adequate	184 (39.1)	19 (10.3)	165 (89.7)		
PrEP[c] **awareness**					7.2 (1)	.01
	Low	238 (50.6)	22 (9.2)	216 (90.8)		
	High	232 (49.4)	41 (17.7)	191 (82.3)		
Financial difficulties due to health care spending					8.2 (1)	.004
	Yes	244 (51.9)	43 (17.6)	201 (82.4)		
	No	224 (47.7)	19 (8.5)	205 (91.5)		
Sex work stigma[b]					0.5 (2)	.79
	Low	27 (5.7)	4 (14.8)	23 (85.2)		
	Moderate	398 (84.7)	53 (13.3)	345 (88.7)		
	High	32 (6.8)	3 (9.4)	29 (90.6)		
Perceived PrEP stigma[b]					0.3 (1)	.59
	Low	355 (75.5)	46 (13)	209 (87)		
	High	114 (24.3)	17 (14.9)	97 (85.1)		
Income from sex work (TZS [US $])[b]					2.4 (3)	.49
	≤150,000 (≤59.61)	116 (24.7)	15 (12.9)	101 (87.1)		
	150,001-299,999 (59.61-119.25)	108 (23)	13 (12)	87 (88)		
	300,000-449,999 (119.22-178.87)	144 (30.6)	25 (17.4)	119 (82.4)		
	≥450,000 (≥178.82)	90 (19.1)	10 (11.1)	80 (86.2)		

[a]*P* value based on the chi-square trend test.

[b]N is less than 470 due to missing observations.

[c]PrEP: pre-exposure prophylaxis.

Predictors of 6-Month Retention in the *Jichunge* App

Table 2 shows the results of the log-binomial regression of the independent predictors of retention in the *Jichunge* app. The findings revealed that older age, high PrEP awareness, and financial difficulties due to health care spending independently predicted retention in the app. Participants aged at least 35 years were more than 2 times more likely to be retained in the app than those aged 18 to 24 years (adjusted risk ratio [aRR] 2.3, 95% CI 1.2-4.1; *P*=.01). Similarly, participants who demonstrated high PrEP awareness at baseline were more likely to be retained in the app after 6 months than those who demonstrated low PrEP awareness (aRR 1.8, 95% CI 1.1-3; *P*=.01). Furthermore, female sex workers who had experienced financial difficulties due to health care spending were approximately 2 times more likely to be retained in the app after 6 months than those who had not experienced such difficulties (aRR 1.9, 95% CI 1.2-3.2; *P*=.01).

Table . Log-binomial regression model for independent predictors of 6-month retention in the *Jichunge* app.

Variable		Risk ratio (95% CI)	*P* value	Adjusted risk ratio (95% CI)	*P* value
Age (years)					
	18-24	1	Reference	1	Reference
	25-34	1.2 (0.7-2.1)	.47	1.1 (0.7-1.9)	.64
	≥35	2.74 (1.5-5)	.001	2.2 (1.2-4.1)	.01
Social support					
	Adequate	1	Reference	1	Reference
	Inadequate	1.5 (0.9-2.6)	.12	1.4 (0.8-2.3)	.23
Financial difficulties due to health care spending					
	Yes	2.1 (1.2-3.4)	.01	1.9 (1.2-3.2)	.01
	No	1	Reference	1	Reference
PrEP[a] awareness					
	High	1.9 (1.2-3.1)	.01	1.8 (1.1-3)	.01
	Low	1	Reference	1	Reference

[a]PrEP: pre-exposure prophylaxis.

Discussion

Principal Findings

Although several mHealth interventions have been designed to support the delivery of health care services, few studies have evaluated retention in those interventions. Understanding the retention in mHealth apps is crucial when planning and designing interventions. In this study, we examined the extent and predictors of 6-month retention in an interactive mHealth app (*Jichunge* app) designed to promote adherence to PrEP use and retention in PrEP care within Tanzania, a lower-middle–income country in sub-Saharan Africa.

The proportion of retained female sex workers (those who still used the *Jichunge* app after 6 months) was only 13.4% (63/470). This is much lower than the the optimal 46.4% (218/470) of users we found 1 month after the app had been installed [18]. The main reason for the lower retention in the app after 6 months did not appear to have much to do with the app itself but rather with the fact that a very large proportion of users had lost access to the app due to their phones having been lost, stolen, sold, changed, or damaged. Issues related to phones clearly represented a main challenge for implementing this mHealth intervention among female sex workers in Tanzania.

We have not found any previous study that evaluated the extent of retention in mHealth apps for HIV prevention. However, our results are in line with those of a study conducted in Poland, which found that less than one-quarter of participants used mHealth to monitor diet, weight, and physical activity in 2022 [31]. On the other hand, our estimate was lower than those reported in studies of the engagement with mHealth apps for self-management among patients with cancer in Sweden [8] and among adolescent and young adult cancer survivors in Pennsylvania [32]. It was also lower than the engagement with text message–delivered support among patients with type 2 diabetes in Tennessee [9]. The difference in the extent of use could be due to variations in contextual, population, and intended health outcomes for each of the interventions. Importantly, the studies mentioned for comparison entailed people diagnosed with chronic medical conditions, while PrEP is used for prevention among healthy people; therefore, the motivation to use the app might be different. Thus, when designing mHealth apps, strategies for attracting users and promoting long-term retention should take into consideration the contextual setting and the specific needs of the targeted population.

Understanding predictors of retention in mHealth apps is essential in tailoring the design of mHealth for its intended users. In this study, we found that retention in the *Jichunge* app was associated with older age, high PrEP awareness, and financial difficulties due to health care spending. Contrary to the evaluation of the use of the *Jichunge* app after 1 month [18], education level and social support were not associated with the retention in the app. This implies that the predictors of early mHealth use are not necessarily the same as those for retention in mHealth apps. Thus, continual evaluation of mHealth interventions is essential in understanding the users' early use and long-term retention behaviors.

In this study, older female sex workers were more likely to be retained in the *Jichunge* app than younger women after 6 months. Our findings are similar to those in a study by Chang et al [33], who found that older persons were more likely to use mobile phones for opioid use disorder telemedication during the COVID-19 pandemic in New York. Contrary to this, Bauer et al [34] found that younger age was associated with the use of mHealth tools among primary care patients in the Northwest United States. The use of the *Jichunge* app does not require sophisticated technical skills, and this might have contributed to easier use among older women. On the other hand, the app does not contain entertaining features, such as those found in many commercial social media apps, and this might have discouraged younger persons from engaging with the app over time. Therefore, future mHealth app development may consider the inclusion of features that may speak to the different needs of younger and older populations.

Participants who demonstrated high PrEP awareness at baseline were more likely to be retained in the *Jichunge* app after 6 months. A lack of prior awareness of medication is among factors that can affect the willingness to use PrEP and hence can affect the acceptability and use of a PrEP-related app [35]. Therefore, people with prior awareness of PrEP medication might have been motivated to use the *Jichunge* app to learn more about the medication. The primary goal of *Jichunge* was to promote adherence to PrEP and retention in PrEP care; hence, it is likely that participants who stopped using PrEP might have lost interest in the *Jichunge* app as well. The findings underline the need to create awareness about PrEP before launching innovative interventions to support PrEP adherence and retention.

One of the benefits of many mHealth interventions is that they can reduce the financial costs related to health care services [36,37]. In this study, we found that participants who reported to have experienced financial difficulties due to health care spending were more likely to be retained in the *Jichunge* app than those who did not have such difficulties. In Tanzania, PrEP is provided free of cost, and at the time of this study, only a few clinics were designated to prescribe PrEP. Participants with financial challenges could have been using *Jichunge* services, such as the web-based consultation that is provided free of cost, as an alternative to physical consultations for PrEP-related and other concerns or conditions. Our findings are contrary to those reported in a retrospective comparative study conducted in Boston by Xiong et al [38], who found that patients insured through Medicaid were less likely to use telemedicine to connect with orthopedic surgery services than privately insured patients. On the basis of our study, mHealth apps seem to have the potential to help improve universal health coverage for underserved populations by removing some of the financial barriers associated with access to health care services.

Strength and Limitations

First, among the strengths of this study is that it provides evidence on retention in mHealth apps within a real-world setting. Second, the large sample size ensured enough statistical power and stable estimates. Third, we evaluated the outcome variable (6-month retention) by using objective measures (electronic records of actual app use), thereby avoiding the bias that could have resulted from using subjective measures (such as self-reported use of the app). One limitation of this study is that we did not carry out any in-depth investigation of users' experiences with the *Jichunge* app, which limits our understanding of why our participants did or did not engage with the mHealth intervention. Furthermore, we only measured 1 aspect of user engagement with the app (retention), which limits our understanding of other aspects, such as time spent in the app, total sessions, and subjective measures.

Conclusion

Overall, retention in the *Jichunge* app after 6 months was low and was predicted by older age, high PrEP awareness, and financial difficulties due to health care spending. A large majority of the sex workers (193/206, 93.7%) who took part in the 6-month follow-up interview reported that they no longer had access to the *Jichunge* app because their phones had been lost, sold, stolen, or changed or were in disrepair. Clearly, female sex workers in Tanzania cannot be expected to use the same phone over long periods of time, and this may underscore some of the complexities associated with mHealth interventions in this population. Thus, the user patterns and "social lives" of smartphones should be taken into consideration when designing mHealth interventions. We recommend qualitative studies that could provide in-depth information on the users' experiences with the use of mHealth interventions for PrEP services.

Acknowledgments

This research was funded by the Research Council of Norway through the Global Health and Vaccination Programme (GLOBVAC; project number: 285361). The project is also part of the European & Developing Countries Clinical Trials Partnership (EDCTP2) program supported by the European Union. The technical development and operation of the *Jichunge* app was carried out by Sveve AS (Teiealleen 28, 2030 Nannestad, Norway), and we would like to acknowledge Ivar Moen and Ole Habberstad for their technical support throughout the research process. We extend our sincere thanks to Inga Haaland, Hanne Lichtwarck, and Willhellmuss Mauka for their efforts and support during fieldwork for this study. Finally, we thank all study participants, fieldwork assistants, and clinic staff for their contributions to this study.

Authors' Contributions

CHM conceptualized the study, designed the study, analyzed the data, and drafted the manuscript. MRK and EJM conceptualized the study, designed the study, interpreted the findings, and revised the manuscript. KM conceptualized the study, interpreted the findings, and revised the manuscript. MTL and EM conceptualized the study and reviewed the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016. URL: www.who.int/publications/i/item/9789241511780 [accessed 2023-03-18]
2. Measuring digital development: facts and figures 2022. International Telecommunication Union. 2022. URL: www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2023-03-18]
3. Nicosia FM, Kaul B, Totten AM, Silvestrini MC, Williams K, Whooley MA, et al. Leveraging telehealth to improve access to care: a qualitative evaluation of veterans' experience with the VA TeleSleep program. BMC Health Serv Res 2021 Jan 21;21(1):77. [doi: 10.1186/s12913-021-06080-5] [Medline: 33478497]
4. Stiles-Shields C, Reyes KM, Archer J, Lennan N, Zhang J, Julion WA, et al. mHealth uses and opportunities for teens from communities with high health disparities: a mixed-methods study. J Technol Behav Sci 2022 Sep 13:1-13. [doi: 10.1007/s41347-022-00278-y] [Medline: 36117748]
5. Weitzman PF, Zhou Y, Kogelman L, Rodarte S, Vicente SR, Levkoff SE. mHealth for pre-exposure prophylaxis adherence by young adult men who have sex with men. Mhealth 2021 Jul 20;7:44. [doi: 10.21037/mhealth-20-51] [Medline: 34345621]
6. Liu AY, Vittinghoff E, von Felten P, Amico KR, Anderson PL, Lester R, et al. Randomized controlled trial of a mobile health intervention to promote retention and adherence to preexposure prophylaxis among young people at risk for human immunodeficiency virus: the EPIC study. Clin Infect Dis 2019 May 30;68(12):2010-2017. [doi: 10.1093/cid/ciy810] [Medline: 30239620]
7. Fuchs JD, Stojanovski K, Vittinghoff E, McMahan VM, Hosek SG, Amico KR, et al. A mobile health strategy to support adherence to antiretroviral preexposure prophylaxis. AIDS Patient Care STDS 2018 Mar;32(3):104-111. [doi: 10.1089/apc.2017.0255] [Medline: 29565183]
8. Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2020 Aug 10;22(8):e17058. [doi: 10.2196/17058] [Medline: 32663140]
9. Nelson LA, Spieker A, Greevy R, LeStourgeon LM, Wallston KA, Mayberry LS. User engagement among diverse adults in a 12-month text message–delivered diabetes support intervention: results from a randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17534. [doi: 10.2196/17534] [Medline: 32706738]
10. UNAIDS data 2021. UNAIDS. 2021. URL: www.unaids.org/en/resources/documents/2021/2021_unaids_data [accessed 2023-03-18]
11. Shannon K, Crago AL, Baral SD, Bekker LG, Kerrigan D, Decker MR, et al. The global response and unmet actions for HIV and sex workers. Lancet 2018 Aug 25;392(10148):698-710. [doi: 10.1016/S0140-6736(18)31439-9] [Medline: 30037733]
12. Mizinduko MM, Moen K, Likindikoki S, Mwijage A, Leyna GH, Makyao N, et al. HIV prevalence and associated risk factors among female sex workers in Dar es Salaam, Tanzania: tracking the epidemic. Int J STD AIDS 2020 Sep;31(10):950-957. [doi: 10.1177/0956462420917848] [Medline: 32772690]
13. Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC). The Tanzania HIV Impact Survey 2016-2017 (THIS) - Final Report. Tanzania National Bureau of Statistics. 2018. URL: www.nbs.go.tz/index.php/en/census-surveys/health-statistics/hiv-and-malaria-survey/382-the-tanzania-hiv-impact-survey-2016-2017-this-final-report [accessed 2023-03-18]
14. Ito S, Lépine A, Treibich C. The effect of sex work regulation on health and well-being of sex workers: evidence from Senegal. Health Econ 2018 Nov;27(11):1627-1652. [doi: 10.1002/hec.3791] [Medline: 29978530]
15. HIV and sex work—human rights fact sheet series 2021. UNAIDS. 2021. URL: www.unaids.org/en/resources/documents/2021/05-hiv-human-rights-factsheet-sex-work [accessed 2023-03-18]
16. Platt L, Grenfell P, Meiksin R, Elmes J, Sherman SG, Sanders T, et al. Associations between sex work laws and sex workers' health: a systematic review and meta-analysis of quantitative and qualitative studies. PLoS Med 2018 Dec 11;15(12):e1002680. [doi: 10.1371/journal.pmed.1002680] [Medline: 30532209]
17. Mauka W, Mbotwa C, Moen K, Lichtwarck HO, Haaland I, Kazaura M, et al. Development of a mobile health application for HIV prevention among at-risk populations in urban settings in East Africa: a participatory design approach. JMIR Form Res 2021 Oct 7;5(10):e23204. [doi: 10.2196/23204] [Medline: 34617904]

18. Mbotwa C, Kazaura M, Moen K, Leshabari M, Metta E, Leyna G, et al. Predictors of mHealth use in promoting adherence to pre-exposure prophylaxis among female sex workers: an evaluation of the Jichunge intervention in Dar es Salaam, Tanzania. BMC Health Serv Res 2022 Jul 4;22(1):859. [doi: 10.1186/s12913-022-08245-2] [Medline: 35787285]
19. Introduction to HIV/AIDS and sexually transmitted infection surveillance: module 4: introduction to respondent-driven sampling. World Health Organization. 2013. URL: apps.who.int/iris/handle/10665/116864 [accessed 2023-03-18]
20. Kirkwood BR, Sterne JAC. Essential Medical Statistics 2nd ed: Wiley-Blackwell; 2003. ISBN: 978-0-865-42871-3.
21. Naing L, Nordin RB, Rahman HA, Naing YT. Sample size calculation for prevalence studies using Scalex and ScalaR calculators. BMC Med Res Methodol 2022 Jul 30;22(1):209. [doi: 10.1186/s12874-022-01694-7] [Medline: 35907796]
22. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl 1997 May 1;44(2):174-199. [doi: 10.2307/3096941]
23. United Republic of Tanzania. Ministry of Health, Community Development, Gender, Elderly and Children. National AIDS Control Program. Implementation framework for pre-exposure prophylaxis of HIV in Tanzania mainland. . 2021..
24. About TSD. University of Oslo. URL: www.uio.no/english/services/it/research/sensitive-data/about/ [accessed 2023-03-18]
25. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Sep 11;26:100454. [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
26. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
27. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong KL, et al. Assessing user engagement of an mHealth intervention: development and implementation of the Growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89. [doi: 10.2196/mhealth.7236] [Medline: 28663164]
28. Broadhead WE, Gehlbach SH, de Gruy FV, Kaplan BH. The Duke-UNC Functional Social Support Questionnaire: measurement of social support in family medicine patients. Med Care 1988 Jul;26(7):709-723. [doi: 10.1097/00005650-198807000-00006] [Medline: 3393031]
29. Cummings P. Methods for estimating adjusted risk ratios. Stata J 2009 Aug 1;9(2):175-196. [doi: 10.1177/1536867X0900900201]
30. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003 May 15;157(10):940-943. [doi: 10.1093/aje/kwg074] [Medline: 12746247]
31. Żarnowski A, Jankowski M, Gujski M. Use of mobile apps and wearables to monitor diet, weight, and physical activity: a cross-sectional survey of adults in Poland. Med Sci Monit 2022 Sep 9;28:e937948. [doi: 10.12659/MSM.937948] [Medline: 36081328]
32. Psihogios AM, King-Dowling S, O'Hagan B, Darabos K, Maurer L, Young J, et al. Contextual predictors of engagement in a tailored mHealth intervention for adolescent and young adult cancer survivors. Ann Behav Med 2021 Nov 18;55(12):1220-1230. [doi: 10.1093/abm/kaab008] [Medline: 33674863]
33. Chang JE, Lindenfeld Z, Thomas T, Waldman J, Griffin J. Patient characteristics associated with phone versus video telemedicine visits for substance use treatment during COVID-19. J Addict Med 2022;16(6):659-665. [doi: 10.1097/ADM.0000000000000985] [Medline: 35195540]
34. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin LM, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788. [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
35. Yi S, Tuot S, Mwai GW, Ngin C, Chhim K, Pal K, et al. Awareness and willingness to use HIV pre-exposure prophylaxis among men who have sex with men in low- and middle-income countries: a systematic review and meta-analysis. J Int AIDS Soc 2017 Jun 26;20(1):21580. [doi: 10.7448/IAS.20.1.21580] [Medline: 28691439]
36. Carrandi A, Hu Y, Karger S, Eddy KE, Vogel JP, Harrison CL, et al. Systematic review on the cost and cost-effectiveness of mHealth interventions supporting women during pregnancy. Women Birth 2023 Feb;36(1):3-10. [doi: 10.1016/j.wombi.2022.03.007] [Medline: 35339412]
37. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330. [doi: 10.2196/11330] [Medline: 30777845]
38. Xiong G, Greene NE, Lightsey HM, Crawford AM, Striano BM, Simpson AK, et al. Telemedicine use in orthopaedic surgery varies by race, ethnicity, primary language, and insurance status. Clin Orthop Relat Res 2021 Jul 1;479(7):1417-1425. [doi: 10.1097/CORR.0000000000001775] [Medline: 33982979]

Abbreviations

aRR: adjusted risk ratio
mHealth: mobile health
PrEP: pre-exposure prophylaxis
PREPTA: Pragmatic Trial for HIV Pre-Exposure Prophylaxis Roll-Out in Tanzania

RDS: respondent-driven sampling
UNAIDS: Joint United Nations Programme on HIV/AIDS
UNC: University of North Carolina

Edited by L Buis; submitted 21.03.23; peer-reviewed by J Stekler, V Cambiano; revised version received 17.06.23; accepted 04.08.23; published 17.10.23.

Please cite as:
Mbotwa CH, Kazaura MR, Moen K, Leshabari MT, Metta E, Mmbaga EJ
Retention in an mHealth App Aiming to Promote the Use of HIV Pre-Exposure Prophylaxis Among Female Sex Workers in Dar es Salaam, Tanzania: Prospective Cohort Study
JMIR Mhealth Uhealth 2023;11:e46853
URL: https://mhealth.jmir.org/2023/1/e46853
doi:10.2196/46853

Development and Pilot Implementation of Neotree, a Digital Quality Improvement Tool Designed to Improve Newborn Care and Survival in 3 Hospitals in Malawi and Zimbabwe: Cost Analysis Study

Hassan Haghparast-Bidgoli[1], PhD; Tim Hull-Bailey[2], MPhil; Deliwe Nkhoma[3], MSc; Tarisai Chiyaka[4,5], MSc; Emma Wilson[2], PhD; Felicity Fitzgerald[6], MD, PhD; Gwendoline Chimhini[7], MPH, MD; Nushrat Khan[2], PhD; Hannah Gannon[2], PhD; Rekha Batura[1], MSc, MD; Mario Cortina-Borja[2], PhD; Leyla Larsson[5], MSc; Msandeni Chiume[3], MD; Yali Sassoon[8], MPhil; Simbarashe Chimhuya[7,9], MD; Michelle Heys[2], MD

1
2
3
4
5
6
7
8
9

Corresponding Author:
Hassan Haghparast-Bidgoli, PhD

Abstract

Background: Two-thirds of the 2.4 million newborn deaths that occurred in 2020 within the first 28 days of life might have been avoided by implementing existing low-cost evidence-based interventions for all sick and small newborns. An open-source digital quality improvement tool (Neotree) combining data capture with education and clinical decision support is a promising solution for this implementation gap.

Objective: We present results from a cost analysis of a pilot implementation of Neotree in 3 hospitals in Malawi and Zimbabwe.

Methods: We combined activity-based costing and expenditure approaches to estimate the development and implementation cost of a Neotree pilot in 1 hospital in Malawi, Kamuzu Central Hospital (KCH), and 2 hospitals in Zimbabwe, Sally Mugabe Central Hospital (SMCH) and Chinhoyi Provincial Hospital (CPH). We estimated the costs from a provider perspective over 12 months. Data were collected through expenditure reports, monthly staff time-use surveys, and project staff interviews. Sensitivity and scenario analyses were conducted to assess the impact of uncertainties on the results or estimate potential costs at scale. A pilot time-motion survey was conducted at KCH and a comparable hospital where Neotree was not implemented.

Results: Total cost of pilot implementation of Neotree at KCH, SMCH, and CPH was US $37,748, US $52,331, and US $41,764, respectively. Average monthly cost per admitted child was US $15, US $15, and US $58, respectively. Staff costs were the main cost component (average 73% of total costs, ranging from 63% to 79%). The results from the sensitivity analysis showed that uncertainty around the number of admissions had a significant impact on the costs in all hospitals. In Malawi, replacing monthly web hosting with a server also had a significant impact on the costs. Under routine (nonresearch) conditions and at scale, total costs are estimated to fall substantially, up to 76%, reducing cost per admitted child to as low as US $5 in KCH, US $4 in SMCH, and US $14 in CPH. Median time to admit a baby was 27 (IQR 20-40) minutes using Neotree (n=250) compared to 26 (IQR 21-30) minutes using paper-based systems (n=34), and the median time to discharge a baby was 9 (IQR 7-13) minutes for Neotree (n=246) compared to 3 (IQR 2-4) minutes for paper-based systems (n=50).

Conclusion: Neotree is a time- and cost-efficient tool, comparable with the results from limited similar mHealth decision-support tools in low- and middle-income countries. Implementation costs of Neotree varied substantially between the hospitals, mainly due to hospital size. The implementation costs could be substantially reduced at scale due to economies of scale because of integration to the health systems and reductions in cost items such as staff and overhead. More studies assessing the impact and cost-effectiveness of large-scale mHealth decision-support tools are needed.

(JMIR Mhealth Uhealth 2023;11:e50467) doi:10.2196/50467

KEYWORDS

mHealth; clinical decision support; quality improvement tool; costs; cost; economic; economics; decision support; costing; expenditure; child; children; pediatric; pediatrics; paediatric; paediatrics; preterm; premature; baby; babies; newborn; newborns; maternal; neonatal; mobile health

Introduction

In 2020, around 2.4 million neonatal deaths occurred globally, with most of the deaths (75%) occurring during the first week of life [1]. Low coverage and poor quality of care are among the main factors contributing to the high burden of neonatal mortality in low- and middle-income countries (LMICs) [2-5], in particular in sub-Saharan Africa and South Asia, where the highest rates of neonatal mortality occur [1]. Shortages of trained health care professionals and inadequate access and adherence to clinical guidelines are the main barriers to high-quality newborn care in health care facilities in LMICs [6-9]. Most neonatal deaths can be prevented through implementing low-cost evidence-based interventions [3,10].

Computerized or electronic clinical decision-support systems (CDSSs) have shown to improve adherence to clinical guidelines and health outcomes [11], though evidence mainly comes from high-income settings. CDSSs are any type of electronic system designed to directly assist in clinical decision-making, utilizing patient-specific information to produce individualized assessments or suggestions that are subsequently presented to health care professionals for their deliberation [12]. Over the past decade, widespread adoption of smartphones and tablets has enabled clinical decision-support tools to be accessible to health care providers on mobile devices directly at the point of care [13]. This is especially significant in LMICs and among their remote and underserved populations, where resources are limited.

Despite the growing adoption of mobile health (mHealth) decision-support tools in LMICs, there is a lack of comprehensive evidence regarding their effectiveness. The limited existing evidence regarding use of these tools in LMICs suggests their potential to improve adherence to clinical guidelines and both the quality of maternal care and childcare at primary health care (PHC) centers [14-16] and hospitals [7]; these tools might also improve clinical outcomes [16]. There is, however, limited evidence on costs and cost-effectiveness of mHealth decision-support tools in LMICs. This evidence gap limits assessments of budget impact and value for money, deterring large-scale programmatic implementation and adoption. There is, however, some limited economic evidence on computerized CDSSs or mHealth decision-support tools addressing improvements in maternal care [4,17-20]. All these were implemented at a PHC level, that is, communities [17-19] and health centers [4,20], rather than in hospitals. Findings from these studies show that the costs vary by intervention type, and that these tools can be cost-effective, though these results are based on small-scale and short-term implementation of the interventions.

This paper presents the costs of a pilot implementation of Neotree, an mHealth clinical decision-support app [21]. The Neotree system is an Android-based, open-source, and fully integrated digital health intervention that enables immediate data capture by health care professionals (primarily nurses) at the bedside while simultaneously providing evidence-based clinical decision support and newborn care education [21,22]. The app operates on low-cost Android tablets or mobile phones at the hospital bedside and is used by health care professionals to support their care and treatment of small and sick newborns in 3 hospitals in Malawi and Zimbabwe. It aims to increase the performance of the health care professionals, and as a result enhance quality of care, by creating a platform to improve supervision, support, and motivation. Early pilot data from Zomba Central Hospital in Malawi demonstrated high usability, acceptability, and feasibility [23], with potential for electronic audits and feedback to drive quality improvement (eg, targets for hypothermia on admission) [24]. Subsequent pilot implementation evaluation in Malawi and Zimbabwe has shown similar high usability, acceptability, and feasibility, with both perceived and observed improvements in quality of care [24,25]. Neotree has been in use since implementation in November 2018 to support the care of over 30,000 babies by more than 1000 health care providers, with ongoing implementation evaluation [21,26]. Neotree has been designed as a holistic clinical decision-support tool to address the leading causes of neonatal mortality and morbidity. Clinical decision support is based on the best available evidence, clinical heuristics, and expert consensus [27] and currently includes the following pathways: resuscitation, thermoregulation, convulsions, low birth weight, prematurity, hypoglycemia, HIV, respiratory distress, neonatal encephalopathy, sepsis, syphilis, jaundice and congenital abnormalities, stabilization, and transfer. Neotree is currently being adopted for PHC settings in Malawi; in Zimbabwe, it is being integrated into the district and national health information systems.

This study aims to calculate the costs of Neotree in its first year of implementation, to estimate future costs at scale, and to estimate the time taken to deliver clinical care on admission and discharge when using Neotree compared to standard-of-care paper-based systems.

Methods

Study Design and Setting

Details on the wider development and pilot implementation of Neotree are presented elsewhere [21,26]. Below we explain the Neotree pilot implementation in 3 hospitals in Malawi and Zimbabwe: Kamuzu Central Hospital (KCH) in Malawi and Sally Mugabe Central Hospital (SMCH) and Chinhoyi Provincial Hospital (CPH) in Zimbabwe.

Neotree was implemented in SMCH in November 2018 as part of a quality improvement project [25] when neonatal mortality rates were 27 per 1000 births [28]. On average, 12,000 babies are delivered at SMCH annually, which is the largest of 3 tertiary neonatal units in Zimbabwe. In 2019, 2985 babies were admitted to the neonatal intensive care unit for whom a matched outcome was recorded, with a case fatality rate (CFR) of around 177 deaths per 1000 admissions (unpublished data).

Neotree was subsequently implemented in KCH as a pilot study in April 2019, when neonatal mortality rates in Malawi were reported at 20.2 per 1000 births [28]. KCH is 1 of 4 central hospitals in Malawi. In 2019, approximately 3000 babies were delivered at KCH; 2732 babies were admitted to the neonatal unit, where the CFR was around 204 per 1000 admitted babies [29].

In October 2019, a 3-year mixed methods implementation evaluation commenced. In December 2020, Neotree was implemented at CPH, where an estimated 4500 babies are delivered annually with a CFR of 180 per 1000 babies admitted to the neonatal unit. In 2021, around 700 babies were admitted to the neonatal unit.

Neotree implementation in Malawi and Zimbabwe was supported by two nongovernmental organizations (NGOs). In each site, a project manager and 2 incentivized or salaried Neotree ambassadors were employed by the NGOs. The project managers' role included both implementation and research (in their job plan, their time was equally split for these two roles). Neotree ambassadors were nursing staff who were paid a monthly salary supplement (100,000 Malawian kwacha [US $133], in Malawi) or full salary (US $1550, in Zimbabwe) to provide technical troubleshooting and implementation support. KCH (Malawi) had 2 Neotree ambassadors and SMCH and CPH (Zimbabwe) had 1 ambassador each but were supported by 5 incentivized nursing staff during weekends.

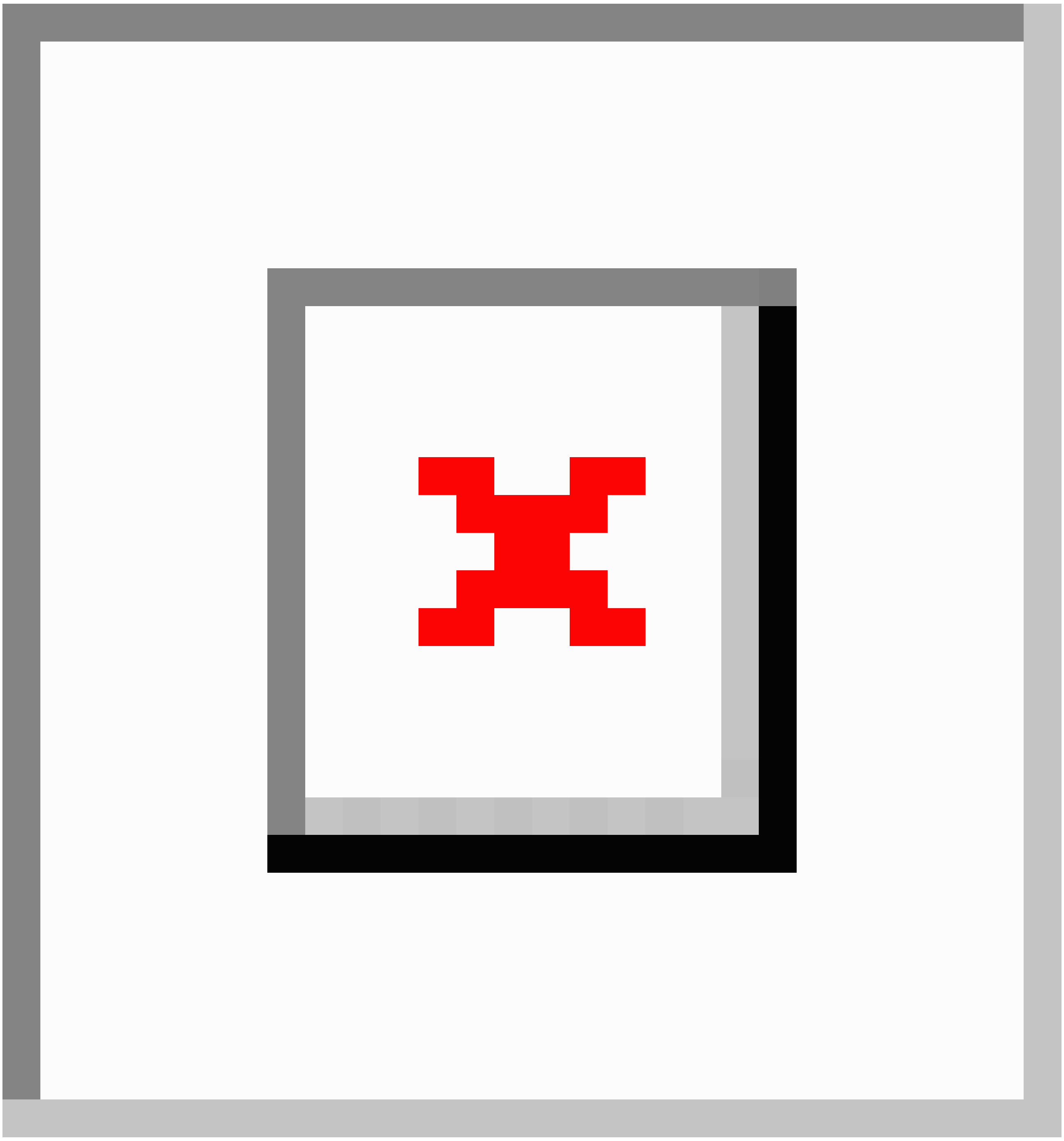
Costing Approach

We used a combination of activity-based costing [30] and expenditure (top-down) approaches [31] to estimate the cost of developing a pilot implementation of Neotree at newborn care units in KCH, SMCH, and CPH. We estimated the costs from a provider perspective, including the cost of setting up and implementing Neotree at the hospitals (programmatic costs) and the cost to the hospital as a result of implementing Neotree. Costs to the hospitals were measured as opportunity (indirect) costs of the hospital staff involvement in the implementation of Neotree.

In addition, to explore the impact of Neotree on time spent on procedures and activities in the delivery of newborn care, a pilot time-motion survey was conducted at KCH and a comparable hospital in Malawi (Bwaila District Hospital) where Neotree was not implemented. Since at the time of data collection in 2020 and 2021 Neotree was used for admission and discharge at the newborn care units, we recorded only admission and discharge time for a number of admissions and discharges in each hospital (34 admissions and 50 discharges at Bwaila hospital, 250 admissions and 264 discharges at KCH). A trained nurse recorded start and end time for each procedure at Bwaila District Hospital in November 2021 and Neotree data were extracted from the Neotree app for the same period. The Wilcoxon-Mann-Whitney test was used to compare differences in admission and discharge times between the 2 hospitals.

Figure 1 illustrates the conceptual framework used for the costing of Neotree. The initial time horizon for costing was 12 months (calendar year 2020 for KCH and SMCH, and 2021 for CPH). Programmatic cost data were collected through expenditure reports, monthly staff time-use surveys, and interviews with the project staff. Hospital-related costs were collected through staff time-use surveys, interviews with the project staff, and time-motion surveys. We estimated the economic cost of implementing Neotree, which included both financial costs (extracted from the expenditure reports) and opportunity costs of involving hospital staff in running Neotree.

Figure 1. The conceptual framework for costing of Neotree. HMIS: Health Management Information System.



An Excel-based costing tool was adapted from previous studies [32] to collect and analyze the cost data. The tool categorizes costs based on inputs or line items and activities. Main inputs or line items were staff, capital, material, web hosting, transportation, and overhead costs. Seven main activities identified through consultation with the implementation team were grouped into development and routine activities. Table 1 presents full descriptions of activities and line items/inputs. We excluded all research-related activities and costs, such as research meetings, as well as process and impact evaluations.

Table . Description of activities and inputs/line items.

Costs		Descriptions
Costs by line item		
	Staff	Value of staff time contributed in development and running of Neotree; staff costs mainly include incentives and salary paid to Neotree ambassadors (n=4), incentives paid to nurses for temporary weekend cover (in Zimbabwe), and the salary of the project manager at each site (n=2)
	Capital	Costs of tablets and other equipment used in running Neotree
	Materials—running Neotree	Costs of materials related to running Neotree
	Materials—other	Costs of materials such as office supplies, refreshments related to meetings, and training
	Transportation	Includes items such as costs of fuel and maintenance
	Travel	Includes cost items such as per diem and other allowances
	Overhead	Include cost items such as running costs and rent, utilities, communications, recruitment costs, and other overhead or joint costs
	Web hosting	Cost of monthly web hosting (in Malawi only)[a]
Costs by activity		
	Neotree data pipeline	Data pipeline testing and quality assurance
	Data backup	The process of manually backing up data, which entails exporting data from each tablet to AWS[b], then moving the Excel files from the tablets to a laptop, and then uploading them to the server
	Roll out and setting up	Includes training and setting up activities
	Admission and discharge or death data entry	The amount of time HCPs[c] take to do data entry during patient admission and outcomes; outcomes include being discharged alive, discharged on request, discharged on palliative care, death at less than or more than 24 hours of life, being transferred out, and other outcomes
	MM[d] dashboard preparation	Includes monthly data analysis, developing visuals from the monthly admission and outcome data, and preparing a PowerPoint presentation for the HCPs to use during the MM monthly meeting
	Neotree support	Includes providing support and troubleshooting with the tablets, data dashboards, and printers, as well as updating or editing content in the app
	Data quality checks/data audits	Includes running some SQL queries on the data to check anomalies, especially the patient unique identifier, and checking the physical files to verify any missing outcomes, especially death outcomes
	Joint activities/administration	Staff time spent on administration or joint activities, such as joint monthly meetings
	Monthly software maintenance	Cost of monthly software maintenance by a software developer

[a]Neotree pilot started at Kamuzu Central Hospital in 2019. Fom the beginning, data were stored on a cloud server (Amazon Web Services) with monthly charges. However, in 2 hospitals in Zimbabwe, we used a physical server to store data, which was less costly. These data storage solutions were developed to adhere to data regulations and preferences within each country.

[b]AWS: Amazon Web Services.
[c]HCP: health care professional.
[d]MM: morbidity and mortality.

The activities were defined as cost centers, and all costs were allocated to them. Staff costs were allocated to the activities using time-use data collected from all staff involved in running Neotree. The same allocation rule was used to allocate joint costs, such as web hosting and overhead costs of the activities.

In Zimbabwe, all costs were measured in US dollars, and in Malawi, all costs were converted to US dollars using the 2020 exchange rate of US $ 1=749.53 kwacha [33]. Capital costs were annualized using the expected life of each item and a discount rate of 5% [34].

We present the total economic and financial costs of developing and running Neotree, average monthly costs, and average monthly costs per admitted newborn at the 3 pilot hospitals.

Sensitivity and Scenario Analyses

We conducted a number of 1-way sensitivity analyses assessing the impact of uncertain variables or parameters on the results. These parameters are the discount rates for capital (3%, 5%, and 10%), exchange rate (±25%, only for Malawi), replacing web hosting with a server (in KCH, Malawi), uncertainty around implementation costs (±25%), and uncertainty around total number of admissions (min-max). We also analyzed a number of scenarios (described below and in Table 2) that reflect potential costs of Neotree at scale.

Table . Scenario analysis summary.

	Base case	Scenario 1 (routine costs)	Scenario 2	Scenario 3
Development activities	✓			
Neotree ambassadors	✓			
Project managers	NGO[a] salaries and 12% overhead	NGO salaries and 12% overhead	Hospital/MoH[b] salaries and 5% overhead	Hospital/MoH salaries and 5% overhead
Project management and server costs shared across multiple sites				✓

[a]NGO: nongovernmental organization.
[b]MoH: Ministry of Health.

Scenario 1 (Routine Conditions)

In costing the pilot implementation of Neotree, we have included monthly allowances or salary paid to the Neotree ambassadors (US $133 in KCH to US $1550 in SMCH and CPH) to provide technical troubleshooting and implementation support. Under routine conditions, or at Neotree's roll out, these routine activities will be conducted by staff without receiving an allowance. In addition, we considered that no development activities would be conducted under routine conditions, and as such these costs were removed under this scenario.

Scenario 2

The project managers in both settings were employed by the 2 NGOs, that is, they were paid at the NGO pay scale. In addition, a portion of the NGOs' overhead costs (12% in both settings, regardless of number of hospitals supported) were included in the implementation costs. At scale, it is likely that Neotree would be run by public hospitals, and therefore staff would be paid at the ministry of health (MoH) pay scale and overhead costs would be substantially lower. Therefore, in this scenario, we replaced the salary of the project managers with the salary of hospital staff with equivalent skills, and we assumed that the overhead costs would be reduced to 5%. Under this scenario, we used the estimates from scenario 1 (routine costs) and reduced overhead costs.

Scenario 3 (Potential Costs at Scale)

At scale, costs such as project management or support and servers can be shared across a number of hospitals. In Zimbabwe, the server was already used for both hospitals and the support costs were the same. We assumed these costs could be shared between 4 hospitals at scale.

Ethical Considerations

Ethics approvals for the Neotree pilot implementation and associated feasibility study data collection were obtained from the Malawi College of Medicine Research and Ethics Committee (P.01/20/2909; P.02/19/2613), University College London (17123/001, 6681/001, 5019/004), the Medical Research Council of Zimbabwe (MRCZ/A/2570), the Biomedical Research and Training Institute and Joint Research Ethics Committee for the University of Zimbabwe institutional review boards (AP155/2020; JREC/327/19), and the Sally Mugabe Hospital Ethics Committee (071119/64; 250418/48).

Results

Cost Data

Table 3 presents the total economic costs of developing and implementing a pilot of Neotree at 3 hospitals. Tables S1 and S2 in Multimedia Appendix 1 present the financial costs of implementing Neotree. The total economic cost of pilot implementation in the 3 hospitals ranged from US $37,748 in

KCH to US $52,331 in SMCH (Table 3). Considering average monthly admissions at newborn care units in the hospitals, the average monthly costs per admitted newborn ranged from US $15 in KCH and SMCH to US $58 in CPH, mainly reflecting the size of the hospitals.

Table . Total economic costs of the Neotree pilot implementation in Kamuzu Central Hospital (KCH), Sally Mugabe Central Hospital (SMCH), and Chinhoyi Provincial Hospital (CPH). Total cost represents the cost of all development and routine activities conducted in 12 months. Costs are estimated in 2020 US $ for KCH and SMCH and in 2021 US $ for CPH.

	KCH	SMCH	CPH
Total cost, US $	37,748	52,331	41,764
Average monthly costs, US $	3146	4361	3480
Average monthly admissions, n	210	298	60
Average monthly costs per admission, US $	15	15	58

Staff costs constituted on average 73% of total costs, ranging from 63% in KCH to 79% in CPH. The category "overhead costs" was the second highest input, with around 13% of total costs. However, in KCH, web hosting costs constituted around 17% of total costs (Table 4 and and Table S1 in Multimedia Appendix 1). In terms of activities, routine activities comprised around 72% of total costs, where support and data entry activities were on average the main cost drivers among these routine activities. However, the proportion of these activities vs the total routine costs varied substantially in each hospital (Table 5 and Table S2 in Multimedia Appendix 1).

Table . Total economic costs of Neotree pilot implementation in Kamuzu Central Hospital (KCH), Sally Mugabe Central Hospital (SMCH), and Chinhoyi Provincial Hospital (CPH) by line item/input. Total cost represents the cost of all development and routine activities conducted in 12 months. Costs are estimated in 2020 US $ for KCH and SMCH and in 2021 US $ for CPH.

Line item/input	Costs at KCH (total=US $37,748), US $ (%)	Costs at SMCH (total=US $52,331), US $ (%)	Costs at CPH (total=US $41,764), US $ (%)
Staff	23,696 (63)	40,524 (77)	32,995 (79)
Capital	941 (2)	1602 (3)	1408 (3)
Materials	1331 (4)	3600 (7)	2100 (5)
Transport/travel	161 (0)	192 (0)	153 (0)
Web hosting	6600 (17)	N/A[a]	N/A
Overhead	5019 (13)	6413 (12)	5109 (12)

[a]N/A: not applicable. Web-hosting costs were only incurred in KCH; SMCH and CPH used a physical server.

Table . Total economic costs of Neotree pilot implementation in Kamuzu Central Hospital (KCH), Sally Mugabe Central Hospital (SMCH), and Chinhoyi Provincial Hospital (CPH) by main activity. Total cost represents costs of all development and routine activities conducted in 12 months. Costs estimated in 2020 US $ for KCH and SMCH and in 2021 US $ for CPH.

Activities		Costs at KCH (total =US $37,748), US $ (%)	Costs at SMCH (total =US $52,331), US $ (%)	Costs at CPH (total =US $41,764), US $ (%)
Development activities				
	Neotree data pipeline	5003 (13)	0 (0)	4121 (10)
	Data backup	1714 (5)	10,524 (20)	6033 (14)
	Rollout	2460 (7)	4245 (8)	3456 (8)
Routine activities				
	Data entry	2355 (6)	13,014 (25)	12,624 (30)
	Morbidity and mortality dashboard preparation	1309 (3)	2654 (5)	7569 (18)
	Neotree support	14,828 (39)	10,909 (21)	2373 (6)
	Data quality checks/audits	3073 (8)	6810 (13)	4854 (12)
	Maintenance	7007 (19)	4175 (8)	734 (2)

Sensitivity and Scenario Analyses

The results from a 1-way sensitivity analysis showed that uncertainty around the number of admissions had a significant impact on costs in all hospitals, ranging from −38% to +239%. The varying exchange rate in Malawi also imposed a significant cost, ranging from −19% to +31% (Table S3 in Multimedia Appendix 1). Similarly, replacing monthly web hosting with an in-country physical server reduced the total cost by 16% in Malawi.

The results from the scenarios showed that, on average, compared to the base case scenario, the total costs would be reduced by 56%, 66%, and 70% under scenarios 1, 2, and 3, respectively (Table 6). Under scenario 1 (routine conditions), total costs in the hospitals would be reduced on average by 56%, ranging from 45% in KCH to 67% in CPH. For example, the cost per admitted child would be reduced from US $15 to US $8 in KCH, from US $15 to US $6 in SMCH, and from US $58 to US $19 in CPH.

Table . The results from scenario analyses of Neotree pilot implementation in Kamuzu Central Hospital (KCH), Sally Mugabe Central Hospital (SMCH), and Chinhoyi Provincial Hospital (CPH).

Hospitals and scenarios		Total costs, US $	Average monthly costs, US $	Total costs per admission, US $
KCH				
	Base case scenario	37,748	3146	15
	Scenario 1	20,745	1729	8
	Scenario 2	16,459	1372	7
	Scenario 3	13,711	1143	5
SMCH				
	Base case scenario	52,331	4361	15
	Scenario 1	22,801	1900	6
	Scenario 2	17,602	1467	5
	Scenario 3	14,612	1218	4
CPH				
	Base case scenario	41,764	3480	58
	Scenario 1	13,730	1144	19
	Scenario 2	10,746	895	15
	Scenario 3	10,150	846	14

Implementing Neotree with hospital or MoH staff (scenario 2), would further reduce costs by 22% on average (ranging from 21% in KCH to 23% in SMCH) in the hospitals compared with scenario 1. Similarly, under the potential scaled-up scenario (scenario 3), the total costs would be further reduced by 13% (ranging from 6% in CPH to 17% in KCH and SMCH) compared to scenario 2 (Table 6). Under this scenario, cost per admission would be $5, $4, and $14 in KCH, SMCH, and CPH, respectively.

Time-Motion Survey Results

The results from our time-motion study on these 2 procedures using Neotree (KCH) and a paper-based system (Bwaila District Hospital) showed that the median time to admit a baby was 27 (IQR 20-40) minutes (n=250) using Neotree and 26 (IQR 21-30) minutes (n=34) using the paper-based system, while the mean time to discharge a baby was 9 (IQR 7-13) minutes (n=246) using Neotree and 3 (IQR 2-4) minutes (n=50) using the paper-based system (Figure S1 in Multimedia Appendix 1).

There was no statistically significant difference between admission times between these two systems (P=.55) but there was a significant difference in discharge times (P=.001). Opportunity costs of the additional time spent on discharging a newborn using Neotree were around US $0.71, ranging from US $0.44 in KCH to US $1 in SMCH.

Discussion

Principal Findings

Our study estimates the costs of a pilot implementation of Neotree, an mHealth clinical decision-support app, in 3 hospitals in Malawi and Zimbabwe. It contributes to the limited cost and cost-effectiveness data for mHealth decision-support tools in LMICs. To our knowledge, this is the first study reporting cost data for a digital newborn care intervention in LMICs. Total cost of pilot implementation of Neotree ranged from US $37,748 to US $52,331. Taking into account average monthly newborn admissions in the hospitals, the average monthly cost per admitted child ranged from US $15 to US $58, which mainly reflects the size of the newborn care units. However, under routine conditions and at scale, these costs will be reduced substantially, up to 76%, reducing cost per admitted child to as low as US $5 in KCH (Malawi), US $4 in SMCH, and US $14 in CPH (the latter two both in Zimbabwe).

Findings Compared With Other mHealth Decision-Support Tools

Comparing costs of Neotree with other mHealth decision-support tools is challenging because of differences in type of intervention (eg, computerized CDSS or mHealth app), implementation site (eg, hospital, PHC center, or community), scale, and costing approach. To our knowledge, there are no published cost data on hospital-based implementations of mHealth decision-support tools in LMICs for any clinical cohort. However, the results from Neotree are comparable with 2 computer-assisted CDSSs piloted in PHC centers in Ghana [35] and Tanzania [36]. These 2 studies were part of the QUALMAT (Quality of prenatal and maternal care: bridging the know-do gap) project piloting computer-assisted CDSSs in PHC centers in a number of countries in sub-Saharan Africa. In Ghana, the tool was piloted in antenatal clinics and the labor wards of 6 PHC centers and implemented by trained nurses. During a 1-year pilot implementation, 22 nurses were trained on a CDSS, and 5595 antenatal consultations (44% of total consultations) and 872 labor patients (60% of total patients) were managed using the CDSS. The economic cost of the intervention was 2012 US $17,129 (or 2020 US $13,935), which included an approximately 2.5-year preintervention and a 1-year intervention implementation. Costs per antenatal consultation and labor care were 2012 US $3 (or 2020 US $3) and 2012 US $20 (or 2020 US $16), respectively [35]. In Tanzania, the tool was piloted for antenatal care in 6 PHC centers (5 public and 1 private). The economic cost of installing and piloting the intervention was 2013 US $127,506 (or 2020 US $121,914), including an approximately 2.5-year preintervention and a 1-year intervention implementation. During the 1-year implementation, 1665 antenatal contacts (70% of total contacts) and 754 childbirths (85% of total childbirths) were registered in the CDSS. Cost per total contact was 2013 US $53 (2020 US $50) [36].

Follow up cost-effectiveness analyses of these 2 interventions have shown that they were potentially cost-effective. In Ghana, the incremental cost-effectiveness ratio (ICER) of a computer-assisted CDSS compared to a paper-based system was estimated at 2012 US $1142 per pregnancy complication detected. Considering only additional costs implementing computer-assisted CDSSs, the cost per pregnancy complication detected was US $285 [20]. In Tanzania, ICERs were 2013 US $2469 and US $338 per 1% change in process quality for antenatal and childbirth care, respectively [4].

It should be taken into account that the cost evidence for Neotree and QUALMAT studies are from 1-year pilot studies and do not capture long-term economic and noneconomic impacts of the interventions. Economic evidence from long- and short-term implementation of CDSSs in high-income countries has shown promising results in reduction of health care expenditures, such as through reducing unnecessary laboratory testing and antibiotic prescriptions. However, the quality of these studies has been variable [37,38].

In terms of cost profile, staff costs constituted most of Neotree's implementation costs (on average 71%), followed by overhead costs (around 13%). This cost profile is similar to the QUALMAT study in Ghana [35], with staff constituting 39% and overhead 23% of total costs, followed by training (16%) and equipment (10%). In Tanzania, training costs constitute 48% of total costs, followed by staff costs with 22% [36].

Implications

The substantial differences in the average monthly cost of Neotree per admitted child, which varied from US $15 to US $58, mainly reflects the size of the newborn care units in the hospitals and thus the number of admitted babies. CPH, as a provincial hospital, has the smallest unit and the lowest number of admissions compared with KCH and SMCH (both are central, ie, tertiary hospitals). As we have shown in our scenario analyses, these unit costs can be significantly reduced when implemented as part of health system strengthening and at scale (ranging from US $4 to US $14). Implementation of an intervention as part of a health system has a distinct cost advantage. Implementation cost at scale can be substantially lower due to potential economies of scale on cost items such as maintenance and data hosting costs (in the case of KCH in Malawi). Some coordination and overhead costs will be reduced if Neotree is run directly by hospitals supported by the existing infrastructure without coordinating with an external agency, as was done in the Neotree pilot implementation. It should be acknowledged that the estimated costs might be generalizable to similar hospitals in Malawi and Zimbabwe but not necessarily to other facility types or settings other than these 2 countries.

In earlier studies, we noted a concern from HCPs that admitting and discharging a baby using Neotree might take an HCP (typically a nurse) too much time, thus distracting them from the care of other babies in the unit [23]. Human resources and capacity have been reported as key determinants of quality of newborn care [9]. Our pilot time-motion data suggest this has not been the case—at least in Malawi. Arguably, babies admitted to Bwaila District Hospital should be of lower clinical acuity (as it is a smaller hospital); however, our data show no statistically significant difference in time taken to admit a baby. Nevertheless, continual monitoring is required as the addition of more comprehensive clinical decision-support features may increase the time taken to complete clinical procedures (eg, to admit a baby).

Limitations

Our study has a number of potential limitations. We estimated costs from a provider perspective, including costs of developing and implementing Neotree and the opportunity cost of hospital staff contribution. However, we were not able to capture the full development cost of Neotree, as the first prototype for Neotree was developed in 2013, and it was further developed in 2016 and 2019. However, we have tried to reflect this in our sensitivity analysis. In addition, we were not able to measure the full impact of Neotree on the hospital or health system due to the short time horizon. For example, we did not measure the impact of Neotree on costs or savings due to changes in procedure times other than admission and discharge, nor did we measure savings due to reductions in the time needed for blood culture results (from 6 day to 3 days) or clinical auditing and quality review (from few days to few hours), nor the potential impact on tests and medication prescriptions. Lastly, due to the volatile nature of inflation in Zimbabwe, it was not possible to deflate the cost of CPH to 2020 to be comparable with KCH and SMCH.

Conclusion

Our findings show that Neotree is a time-efficient and cost-efficient tool, comparable with the results from limited mHealth clinical decision-support tools in LMICs. The implementation cost of Neotree varied substantially between the hospitals, mainly affected by the size of the hospitals. Our analysis showed that Neotree implementation costs can be substantially reduced at scale due to potential economies of scale as a result of integration to the health system and reductions in cost items such as staff and overhead. More studies assessing the impact and cost-effectiveness of larger-scale mHealth decision-support tools such as Neotree are needed.

Acknowledgments

This study was supported by the Wellcome Trust (grant 215742_Z_19_Z).

Authors' Contributions

HH-B contributed to conceptualization, design, analysis, and writing the first draft of the manuscript. TH-B and MH contributed to the design. DN, TC, and TH-B contributed to the data collection. MH, NK, EW, RB, MC-B and LL contributed to writing and editing of the manuscript. YS, TH-B and NK participated in Neotree software development. FF, HG, DN, TC, GC, SC, and MC facilitated pilot implementation of Neotree. All authors have approved the final version of the manuscript.

Conflicts of Interest

MH and FF are both trustees of the Neotree charity but receive no financial payment from this role.

Multimedia Appendix 1

Supplementary tables and figure.

[DOCX File, 62 KB - mhealth_v11i1e50467_app1.docx]

References

1. Levels and trends in child mortality. UNICEF. 2020. URL: https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2020 [accessed 2021-12-15]
2. The network for improving quality of care for maternal, newborn and child health: evolution, implementation and progress. World Health Organization. URL: www.who.int/publications/i/item/9789240023741 [accessed 2023-12-04]
3. Bhutta ZA, Das JK, Bahl R, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? Obstet Gynecol Surv 2014;69(11):641-643. [doi: 10.1097/OGX.0000000000000124]
4. Saronga HP, Duysburgh E, Massawe S, et al. Cost-effectiveness of an electronic clinical decision support system for improving quality of antenatal and childbirth care in rural Tanzania: an intervention study. BMC Health Serv Res 2017 Aug 7;17(1):537. [doi: 10.1186/s12913-017-2457-z] [Medline: 28784130]
5. Modi D, Desai S, Dave K, et al. Cluster randomized trial of a mHealth intervention "ImTeCHO" to improve delivery of proven maternal, neonatal, and child care interventions through community-based social health activists (ASHAs) by enhancing their motivation and strengthening supervision in tribal areas of Gujarat, India: study protocol for a randomized controlled trial. Trials 2017 Jun 9;18(1):270. [doi: 10.1186/s13063-017-1998-0] [Medline: 28599674]
6. Dickson KE, Simen-Kapeu A, Kinney MV, et al. Every newborn: health-systems bottlenecks and strategies to accelerate scale-up in countries. Lancet 2014 Aug 2;384(9941):438-454. [doi: 10.1016/S0140-6736(14)60582-1] [Medline: 24853600]
7. Muhindo M, Bress J, Kalanda R, et al. Implementation of a newborn clinical decision support software (NoviGuide) in a rural district hospital in Eastern Uganda: feasibility and acceptability study. JMIR Mhealth Uhealth 2021 Feb 19;9(2):e23737. [doi: 10.2196/23737] [Medline: 33605886]
8. Amoakoh HB, Klipstein-Grobusch K, Agyepong IA, et al. The effect of an mHealth clinical decision-making support system on neonatal mortality in a low resource setting: a cluster-randomized controlled trial. EClinicalMedicine 2019 Jul;12:31-42. [doi: 10.1016/j.eclinm.2019.05.010] [Medline: 31388661]
9. Kaur E, Heys M, Crehan C, et al. Persistent barriers to achieving quality neonatal care in low-resource settings. Perspectives from a unique panel of frontline neonatal health experts. J Glob Health Rep 2023;7:e2023004. [doi: 10.29392/001c.72089]
10. Knippenberg R, Lawn JE, Darmstadt GL, et al. Systematic scaling up of neonatal care in countries. Lancet 2005;365(9464):1087-1098. [doi: 10.1016/S0140-6736(05)71145-4] [Medline: 15781104]
11. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 3;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
12. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]

13. Agarwal S, Glenton C, Tamrat T, et al. Decision-support tools via mobile devices to improve quality of care in primary healthcare settings. Cochrane Database Syst Rev 2021 Jul 27;7(7):CD012944. [doi: 10.1002/14651858.CD012944.pub2] [Medline: 34314020]
14. Adepoju IO, Albersen BJA, De Brouwere V, van Roosmalen J, Zweekhorst M. mHealth for clinical decision-making in sub-Saharan Africa: a scoping review. JMIR Mhealth Uhealth 2017 Mar 23;5(3):e38. [doi: 10.2196/mhealth.7185] [Medline: 28336504]
15. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401. [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
16. Modi D, Dholakia N, Gopalan R, et al. mHealth intervention “ImTeCHO” to improve delivery of maternal, neonatal, and child care services-a cluster-randomized trial in tribal areas of Gujarat, India. PLoS Med 2019 Oct 24;16(10):e1002939. [doi: 10.1371/journal.pmed.1002939] [Medline: 31647821]
17. Modi D, Saha S, Vaghela P, et al. Costing and cost-effectiveness of a mobile health intervention (ImTeCHO) in improving infant mortality in tribal areas of Gujarat, India: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e17066. [doi: 10.2196/17066] [Medline: 33052122]
18. Prinja S, Gupta A, Bahuguna P, Nimesh R. Cost analysis of implementing mHealth intervention for maternal, newborn & child health care through community health workers: assessment of ReMIND program in Uttar Pradesh, India. BMC Pregnancy Childbirth 2018 Oct 3;18(1):390. [doi: 10.1186/s12884-018-2019-3] [Medline: 30285669]
19. Zurovac D, Larson BA, Sudoi RK, Snow RW. Costs and cost-effectiveness of a mobile phone text-message reminder programmes to improve health workers’ adherence to malaria guidelines in Kenya. PLoS One 2012;7(12):e52045. [doi: 10.1371/journal.pone.0052045] [Medline: 23272206]
20. Dalaba MA, Akweongo P, Aborigo RA, et al. Cost-effectiveness of clinical decision support system in improving maternal health care in Ghana. PLoS One 2015 May 14;10(5):e0125920. [doi: 10.1371/journal.pone.0125920] [Medline: 25974093]
21. Heys M, Kesler E, Sassoon Y, et al. Development and implementation experience of a learning healthcare system for facility based newborn care in low resource settings: the Neotree. Learn Health Syst 2022 Apr 6;7(1):e10310. [doi: 10.1002/lrh2.10310] [Medline: 36654803]
22. Khan N, Crehan C, Hull-Bailey T, et al. Software development process of Neotree - a data capture and decision support system to improve newborn healthcare in low-resource settings. Wellcome Open Res 2022;7:305. [doi: 10.12688/wellcomeopenres.18423.1]
23. Crehan C, Kesler E, Nambiar B, et al. The NeoTree application: developing an integrated mHealth solution to improve quality of newborn care and survival in a district hospital in Malawi. BMJ Glob Health 2019 Jan 16;4(1):e000860. [doi: 10.1136/bmjgh-2018-000860] [Medline: 30713745]
24. Crehan C, Chiume M, Mgusha Y, et al. Usability-focused development and usage of Neotree-beta, an app for newborn care in a low-resource neonatal unit, Malawi. Front Public Health 2022 Apr 28;10:793314. [doi: 10.3389/fpubh.2022.793314] [Medline: 35570891]
25. Gannon H, Chimhuya S, Chimhini G, et al. Electronic application to improve management of infections in low-income neonatal units: pilot implementation of the NeoTree beta app in a public sector hospital in Zimbabwe. BMJ Open Qual 2021 Jan;10(1):e001043. [doi: 10.1136/bmjoq-2020-001043] [Medline: 33472853]
26. Wilson E, Gannon H, Chimhini G, et al. Protocol for an intervention development and pilot implementation evaluation study of an e-health solution to improve newborn care quality and survival in two low-resource settings, Malawi and Zimbabwe: Neotree. BMJ Open 2022 Jul 5;12(7):e056605. [doi: 10.1136/bmjopen-2021-056605] [Medline: 35790332]
27. Evans M, Corden MH, Crehan C, Fitzgerald F, Heys M. Refining clinical algorithms for a neonatal digital platform for low-income countries: a modified Delphi technique. BMJ Open 2021 May 18;11(5):e042124. [doi: 10.1136/bmjopen-2020-042124] [Medline: 34006538]
28. Malawi: under-5 mortality rate. UN Interagency Group for Child Mortality Estimation. 2023. URL: https://childmortality.org/all-cause-mortality/data?refArea=MWI [accessed 2023-06-20]
29. Mgusha Y, Nkhoma DB, Chiume M, et al. Admissions to a low-resource neonatal unit in Malawi using a mobile app and dashboard: a 1-year digital perinatal outcome audit. Front Digit Health 2021 Dec 23;3:761128. [doi: 10.3389/fdgth.2021.761128] [Medline: 35005696]
30. Fiedler JL, Villalobos CA, De Mattos AC. An activity-based cost analysis of the Honduras community-based, integrated child care (AIN-C) programme. Health Policy Plan 2008 Nov;23(6):408-427. [doi: 10.1093/heapol/czn018] [Medline: 18755734]
31. Moreland S, Foley S, Morris L. A Guide to the Fundamentals of Economic Evaluation in Public Health: University of North Carolina; 2019.
32. Haghparast-Bidgoli H, Harris-Fry H, Kumar A, et al. Economic evaluation of nutrition-sensitive agricultural interventions to increase maternal and child dietary diversity and nutritional status in rural Odisha, India. J Nutr 2022 Oct 6;152(10):2255-2268. [doi: 10.1093/jn/nxac132] [Medline: 35687367]
33. Official exchange rate (LCU per US$, period average) - Malawi. World Bank. URL: https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=MW [accessed 2023-12-12]

34. Haacker M, Hallett TB, Atun R. On discount rates for economic evaluations in global health. Health Policy Plan 2020 Feb 1;35(1):107-114. [doi: 10.1093/heapol/czz127] [Medline: 31625564]
35. Dalaba MA, Akweongo P, Williams J, et al. Costs associated with implementation of computer-assisted clinical decision support system for antenatal and delivery care: case study of Kassena-Nankana district of northern Ghana. PLoS One 2014;9(9):e106416. [doi: 10.1371/journal.pone.0106416] [Medline: 25180831]
36. Saronga HP, Dalaba MA, Dong H, et al. Cost of installing and operating an electronic clinical decision support system for maternal health care: case of Tanzania rural primary health centres. BMC Health Serv Res 2015 Apr 2;15:132. [doi: 10.1186/s12913-015-0780-9] [Medline: 25888762]
37. Lewkowicz D, Wohlbrandt A, Boettinger E. Economic impact of clinical decision support interventions based on electronic health records. BMC Health Serv Res 2020 Sep 15;20(1):871. [doi: 10.1186/s12913-020-05688-3] [Medline: 32933513]
38. Fillmore CL, Bray BE, Kawamoto K. Systematic review of clinical decision support interventions with potential for inpatient cost reduction. BMC Med Inform Decis Mak 2013 Dec 17;13:135. [doi: 10.1186/1472-6947-13-135] [Medline: 24344752]

Abbreviations

AWS: Amazon Web Services
CDSS: clinical decision-support system
CFR: case fatality rate
CPH: Chinhoyi Provincial Hospital
HCP: health care professional
ICER: incremental cost-effectiveness ratio
KCH: Kamuzu Central Hospital
LMIC: low- and middle-income country
MCH: Sally Mugabe Central Hospital
mHealth: mobile health
MM: morbidity and mortality
NGO: nongovernmental organization
PHC: primary health care
QUALMAT: Quality of prenatal and maternal care: bridging the know-do gap

Edited by L Buis; submitted 02.07.23; peer-reviewed by A Carrandi, C Gissel; revised version received 21.10.23; accepted 07.11.23; published 22.12.23.

Please cite as:
Haghparast-Bidgoli H, Hull-Bailey T, Nkhoma D, Chiyaka T, Wilson E, Fitzgerald F, Chimhini G, Khan N, Gannon H, Batura R, Cortina-Borja M, Larsson L, Chiume M, Sassoon Y, Chimhuya S, Heys M
Development and Pilot Implementation of Neotree, a Digital Quality Improvement Tool Designed to Improve Newborn Care and Survival in 3 Hospitals in Malawi and Zimbabwe: Cost Analysis Study
JMIR Mhealth Uhealth 2023;11:e50467
URL: https://mhealth.jmir.org/2023/1/e50467
doi:10.2196/50467

Original Paper

The Implementation of a GPS-Based Location-Tracking Smartphone App in South Africa to Improve Engagement in HIV Care: Randomized Controlled Trial

Kate Clouse[1,2], PhD, MPH; Sandisiwe Noholoza[3], MPH; Sindiswa Madwayi[3]; Megan Mrubata[3]; Carol S Camlin[4], PhD, MPH; Landon Myer[3], MBChB, PhD; Tamsin K Phillips[3], PhD, MPH

[1]Vanderbilt University School of Nursing, Nashville, TN, United States

[2]Vanderbilt Institute for Global Health, Nashville, TN, United States

[3]Division of Epidemiology and Biostatistics, School of Public Health, University of Cape Town, Cape Town, South Africa

[4]Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Kate Clouse, PhD, MPH
Vanderbilt University School of Nursing
461 21st Avenue South
Nashville, TN, 37240
United States
Phone: 1 615 343 5351
Email: kate.clouse@vanderbilt.edu

Abstract

Background: Mobile health interventions are common in public health settings in Africa, and our preliminary work showed that smartphones are increasing in South Africa. We developed a novel smartphone app—CareConekta—that used GPS location data to characterize personal mobility to improve engagement in HIV care among pregnant and postpartum women living with HIV in South Africa. The app also used the user's location to map nearby clinics.

Objective: We aimed to describe the feasibility, acceptability, and initial efficacy of using the app in a real-world setting.

Methods: We conducted a prospective randomized controlled trial at a public sector clinic near Cape Town, South Africa. We enrolled 200 pregnant (third trimester) women living with HIV who owned a smartphone that met the required specifications. All participants installed the app, designed to collect 2 GPS heartbeats per day to geolocate the participant within a random 1-km fuzzy radius (for privacy). We randomized (1:1) participants to a control arm to receive the app with no additional support or an intervention arm to receive supportive phone calls, WhatsApp (Meta Platforms, Inc) messages, or both from the study team when traveling >50 km from the study area for >7 days. In addition to mobility data collected daily through the phone, participants completed questionnaires at enrollment and follow-up (approximately 6 months post partum).

Results: A total of 7 participants were withdrawn at enrollment or shortly after because of app installation failure (6/200, 3%) or changing to an unsuitable phone (1/200, 0.50%). During the study period, no participant's smartphone recorded at least 1 heartbeat per day, which was our primary feasibility measure. Of the 171 participants who completed follow-up, only half (91/171, 53.2%) reported using the same phone as that used at enrollment, with the CareConekta app still installed on the phone and GPS usually enabled. The top reasons reported for the lack of heartbeat data were not having mobile data, uninstalling the app, and no longer having a smartphone. Acceptability measures were positive, but participants at follow-up demonstrated a lack of understanding of the app's purpose and function. The clinic finder was a popular feature. Owing to the lack of consistent GPS heartbeats throughout the study, we were unable to assess the efficacy of the intervention.

Conclusions: Several key challenges impeded our study feasibility. Although the app was designed to reverse bill participants for any data use, the lack of mobile data was a substantial barrier to our study success. Participants reported purchasing WhatsApp data, which could not support the app. Problems with the web-based dashboard meant that we could not consistently monitor mobility. Our study provides important lessons about implementing an ambitious GPS-based study under real-world conditions in a limited-resource setting.

Trial Registration: ClinicalTrials.gov NCT03836625; https://clinicaltrials.gov/ct2/show/NCT03836625

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-020-4190-x

(JMIR Mhealth Uhealth 2023;11:e44945) doi:10.2196/44945

KEYWORDS

mobile health; mHealth; smartphone; mobile phone; HIV/AIDS; South Africa; pregnancy

Introduction

Background

There are an estimated 7.5 million people living with HIV in South Africa, more than the number of people living with HIV in any other country [1]. The country adopted a universal test-and-treat antiretroviral therapy (ART) policy in 2016, allowing for the initiation of lifelong ART regardless of clinical criteria [2]. Despite the widespread availability of free ART and the known efficacy of the treatment for prevention [3-5], South Africa still had >200,000 new HIV infections and 51,000 deaths from HIV in 2021 [1].

Continuous engagement in HIV care is a known challenge in South Africa [6,7]. Pregnant women living with HIV are at an especially high risk of dropping out of HIV care, particularly during the postpartum period [8-12]. Our earlier work has explored the potential of mobility, particularly long-distance travel to the mother's rural home, as a factor contributing to postpartum disengagement in care [13-15]. This work was limited in that it either required a retrospective analysis of existing data or relied on the self-reported mobility recall of participants still engaged in care.

Mobile health (mHealth) apps are frequently deployed in public health settings in Africa [16-19], and our preliminary research showed that smartphones are increasingly common among our target population of postpartum women living with HIV in South Africa [20]. Previous studies demonstrated the utility of aggregated cellular phone data in showing population mobility [21-24]. Furthermore, US-based studies demonstrated the feasibility of opt-in tracking of people at high risk of HIV [25,26]. Within this context, we developed a smartphone app that could track a participant's location—with their permission—throughout late pregnancy and the postpartum period to characterize mobility prospectively and offer a support intervention to those who traveled.

The CareConekta App

The CareConekta app was built through collaboration between the study team and Jembi Health Systems in Cape Town [27]. It followed an initial beta version developed and briefly tested in collaboration between the study team (KC, LM, and TKP) and Dr Martin Were of Vanderbilt University Medical Center's Department of Biomedical Informatics and was based on qualitative preliminary research that explored attitudes toward mHealth interventions and possible concerns regarding location tracking among potential users [28,29]. The app uses the phone's GPS signal to prospectively characterize mobility in real time, which is a real advancement in research that previously relied on retrospective analysis. Information on a participant's location would allow the study—and later, clinic—staff to intervene without delay to link traveling individuals to health facilities in the new area, with assistance through phone calls or WhatsApp (Meta Platforms, Inc) messages. In addition, loaded with a national list of HIV care facilities that could be located on a map, the app acted as a clinic finder. We previously published the study protocol [27] and descriptions of the cohort [30] and screening process [31]. In this paper, we describe our primary outcome of the feasibility of implementing this app and the operational lessons learned by conducting an mHealth study in a resource-limited setting in South Africa.

Methods

Study Design and Dates

We conducted a prospective, unblinded randomized controlled trial at the Gugulethu Midwife Obstetric Unit (MOU), a public sector clinic providing integrated HIV and peripartum care for pregnant women near Cape Town, South Africa. Full details of the study design can be found in our published protocol [27], and details of the participant characteristics and follow-up can be found in the cohort profile [30]. In addition to characterizing mobility during pregnancy and the postpartum period, we designed the app to serve as a tool for engagement in HIV care for mobile women living with HIV. Thus, we randomized (1:1) participants to a control arm to receive the app with no additional support or an intervention arm to receive the app and supportive phone calls, WhatsApp messages, or both from the study team when meeting our threshold for traveling: >50 km from the study area for >7 days. The enrollment goal was 200 participants, which we anticipated would be an attainable goal given our study period and our objective of describing the feasibility, acceptability, and initial efficacy of the intervention. Enrollment began in December 2019 and ended in February 2021. There was a 6-month pause in recruitment from March to September 2020 due to the COVID-19 pandemic. The participant follow-up ended in November 2021.

App Design Specifications

To characterize participant mobility during the study period, the CareConekta app was designed to collect 2 GPS location heartbeats per day. In addition, the app was built with a geographic list of health facilities in South Africa so that users could see a map of nearby facilities as they traveled. App connectivity and participant location (marked using anonymized study ID numbers) were viewable to the study team through a password-protected, web-based dashboard. To protect participant privacy, the location was made fuzzy by randomizing the location within a 1-km radius. The MOU was set as the home location from which to begin measuring movement. The mobility history was saved on an encrypted, password-protected server at a South African data center.

The app was available for free download through the Google Play Store (Google LLC) but required authentication and registration, so access was restricted to those enrolled in the study. CareConekta was designed such that it would cost the participant nothing: data costs associated with location tracking would be immediately reimbursed through reverse billing.

Therefore, the cellular service of 1 of the 4 major mobile providers in South Africa was a requirement for eligibility. In the event of disconnection, the app was designed such that location data would be stored on the phone and uploaded as soon as connectivity resumed. However, reverse billing did not apply for app installation or version updates, so participants installed the app at the clinic using free Wi-Fi, and small data bundles were provided by the study staff for reinstallation and updates, when needed.

Recruitment and Eligibility

Pregnant women were recruited during routine antenatal care at the MOU. Women were eligible if they were in the third trimester of pregnancy (≥28 weeks); aged ≥18 years; able to speak and understand isiXhosa (the predominant local language) or English; diagnosed with HIV at any time before enrollment; able to demonstrate basic smartphone-level literacy; and willing to participate in all aspects of the study, including randomization and mobility tracking. Eligible participants also needed to own a smartphone that met the technical requirements described in the subsequent section.

Smartphone Technical Requirements

For the purpose of this study, a smartphone was defined as a mobile phone device with a touchscreen interface and internet and GPS capabilities. CareConekta was designed for phones using the Android operating system, version 5.0 or later. In our preliminary work, nearly 90% of the smartphones of the women approached for study participation used the Android system [20]. Eligible participants were required to subscribe to service (prepaid or contract) from 1 of the main 4 cellular providers in South Africa: Vodacom (Vodacom Group Limited), Cell C (Cell C Limited), Telkom (Telkom SA SOC Limited), or MTN (MTN Group Limited). At recruitment, the eligible participant needed to demonstrate that the phone could use GPS by opening a map app, such as Google Maps (Google LLC), and finding the current location. Finally, the phone needed to be capable of holding battery charge; phones were ineligible if they needed to be charged more than twice per day on average (by self-report). Details of those found to be ineligible during the screening assessment can be found elsewhere [31].

App Installation and Operation

The app was installed at the study site by connecting the smartphone to the study's Wi-Fi source. Because the app was available on the Google Play Store, all participants first needed a Google email address to download the app. For the app to function properly after installation, the GPS needed to remain enabled (with location allowed), and the phone needed to have some data or airtime available for reverse billing to work. Participants were asked to contact the study team if they needed to reinstall the app because of changing devices or uninstalling the app.

Study Measures

Participant-reported data were collected at enrollment and follow-up—approximately 6 months post partum—directly into REDCap (Research Electronic Data Capture; Vanderbilt University) using tablet computers. REDCap is a secure, encrypted, and web-based software platform designed to support data capture for research studies [32]. Feedback on app installation experience was noted by the staff in REDCap at the end of the enrollment visit. App versions and participant contact attempts were maintained in study logs. Mobility data were exported from the CareConekta dashboard as CSV files.

Analysis

We report counts and proportions for categorical variables and medians and IQRs for continuous variables. Data analysis was performed using SAS (version 9.4; SAS Institute). Open-ended responses were reviewed to identify key topics or themes and illustrative quotes.

Ethics Approval

This study was approved by the institutional review boards of Vanderbilt University (reference 181640) and the University of California, San Francisco (237757), and the Human Services Research Committee of the University of Cape Town (659/2018). All participants signed a written informed consent form before enrollment, which included specific permission for location tracking.

Results

Installation Experience

App installation at enrollment was a highly variable experience. The most common reasons for slow experiences at installation were the need to delete items on the phone to make room for another app, the need to create a Google or Gmail account to use the Google Play Store, the need to update the Google Play Store app version, or problems sending and receiving heartbeats after installing the app. The following is a quote from the study staff's notes about a particularly troublesome installation experience:

> *Had to create a Google account for participant and update Play store for app installation. Downloading took a while still. Had to let participant go and she came back a day later for installation; phone still taking forever. The controls get frozen, cleaned phone for efficiency but it still freezes. Finally, app installed and registered participant. Heartbeat data transmitted after all necessary settings adjusted.*

Investigator Withdrawals

In total, 7 participants were withdrawn from the study soon after enrollment owing to technical issues. Most (6/200, 3%) were withdrawn at the end of the enrollment visit because of an app installation failure that could not be resolved. In addition, 1 (0.5%) other participant was withdrawn for changing their phone to an ineligible phone within 2 weeks of enrollment.

Participant-Reported Study Feasibility Measures

Figure 1 shows the feasibility measures of using the CareConekta app as designed, as reported by the participants at follow-up. Over one-third (64/171, 37.4%) of the participants reported that they were no longer using the smartphone in which the app was installed during enrollment. The top 3 reasons for changing phones were that the other phone stopped working

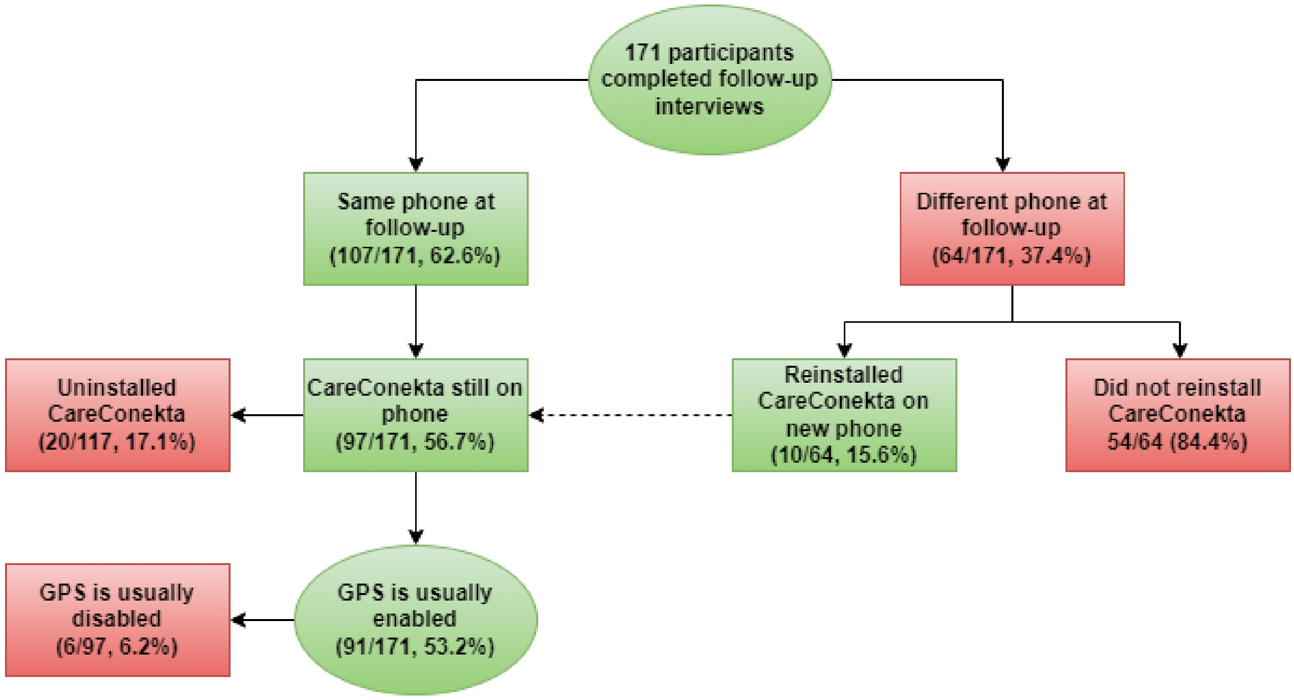
(27/64, 42%), the other phone was lost (18/64, 28%), and the other phone was stolen (12/64, 19%).

Of the 64 participants who reported changing devices during the study, 10 (16%) reported that the CareConekta app was reinstalled on the new device. Among the 107 participants with the same phone at follow-up and the 10 participants who had reinstalled the app, 17.1% (20/117) reported that the CareConekta app had been uninstalled.

Of the 97 who still had the app on their phone, 6 (6%) participants reported that GPS was usually disabled on their phones. The reasons mentioned by those who reported disabling the GPS setting were that their phone did this automatically or that they turned it off to conserve the battery.

Deducting those who changed phones, those who uninstalled the app, and those who disabled GPS, only 53.2% (91/171) of the participants at the time of follow-up reported the possession of a phone that would operate the app correctly.

Figure 1. Number of participants who reported being able to use the CareConekta app at study follow-up. A total of 173 participants completed the follow-up interviews, but the first 2 participants completed the study before the variables presented here were added to the questionnaire. In total, 10 participants who had uninstalled the CareConekta app reinstalled it with the assistance of the study team and continue through the flowchart, as shown.



Phone Sharing

Approximately one in seven (25/173, 14.4%) participants reported sharing their phone during the study. These participants most often shared their phone with one of their family members (12/25, 48%), their boyfriend or husband (11/25, 44%), or one of their friends who was not their boyfriend or husband (3/25, 12%). Of those who reported phone sharing, 80% (20/25) reported that they had their phone with them most of the day.

Version Updates

From December 2019 to December 2020, the CareConekta app moved from version 1 to version 7. At one point, a version change (version 5 to 6) meant that the heartbeat data submitted by older versions were no longer received. From July to August 2020, participants were phoned, notified of the new app version, and assisted with updating the app over the phone (Google Play > CareConekta > update app). In cases where participants struggled to follow the steps over the phone, we offered an option for them to come to the study clinic—especially in cases where they were scheduled for an upcoming visit—where we would update the app for them in person. On a few occasions, to those who experienced difficulty over the phone, we sent the steps to follow via SMS text message or WhatsApp. The participants were offered data (100 MB) to update the app over the phone.

Gaps in GPS Heartbeat Signals

A key feasibility measure was the successful transmission of at least 1 data location heartbeat per participant per day. Heartbeat data were observed daily by the study staff via the dashboard and periodically exported as a CSV file for analysis. During the study, we experienced periods when the dashboard was offline and data exports were unavailable, which resulted in missing data. Owing to difficulties with the dashboard and data exports resulting in substantial missing periods of data after March 19, 2021, our analysis of heartbeat data was conducted on all heartbeats received between the start of enrollment (December 2019) and March 2021. Table 1 shows the summary of gaps in heartbeats received during this period.

The daily GPS heartbeats of none of the participants were received without interruption. The range in heartbeat gaps was 2 to 273 days.

Among the 127 participants with gaps of ≥28 days, 55 (43.3%) were participants whose heartbeat transmission stopped altogether; the remaining 72 (57%) were women whose heartbeat transmission had long gaps but heartbeats resumed during the study period.

On the basis of 3302 heartbeats with GPS location, the median distance traveled from the study site was 2.4 (IQR 1.5-4.5) km. A total of 104 heartbeats (16 women) picked up a distance >50 km away. Only 3 (1.6%) out of 193 women were >50 km away for >7 days based on the heartbeat data, thus meeting our study definition of "travel." Of the 3 women, 2 (67%) were in the control arm and received no additional action, and 1 (33%) was in the intervention arm; the participant in the intervention arm had a break in GPS when she reached a rural area in November 2020, but her GPS heartbeats resumed in January 2021, and the intervention protocol was followed.

In comparison, self-reported mobility during the follow-up visit indicated 37 trips lasting ≥7 days during the study period. Most of these trips were missed by the CareConekta app data.

Table 1. Gaps in GPS heartbeats received during the CareConekta study.

	Total (n=193), n (%)	Intervention (n=98), n (%)	Control (n=95), n (%)
One heartbeat received per day	0 (0)	0 (0)	0 (0)
Maximum gap in heartbeats: <7 days	11 (5.7)	7 (7.1)	4 (4.2)
Maximum gap in heartbeats: 7-13 days	15 (7.8)	8 (8.2)	7 (7.4)
Maximum gap in heartbeats: 14-20 days	15 (7.8)	7 (7.1)	8 (8.4)
Maximum gap in heartbeats: 21-27 days	11 (5.7)	5 (5.1)	6 (6.3)
Maximum gap in heartbeats: ≥28 days	127 (65.8)	62 (63.3)	65 (68.4)
No heartbeat received during analysis period	14 (7.3)	9 (9.2)	5 (5.3)

Participant-Reported Reasons for GPS Gaps

From September 2020 to September 2021, our team attempted to contact all participants for whom a GPS heartbeat could not be detected. Contact attempts were made through phone call, followed by SMS text message, if needed. Over this 1-year period, participants could be contacted multiple times for a lack of heartbeat, and the result of the contact could differ each time. Therefore, we present the number of contact attempts and the primary reason reported for the GPS failure at the time of contact.

Out of the 454 contact attempts, 247 (54.4%) were unsuccessful, as the participants were unreachable or declined to talk. The reasons for GPS heartbeat loss gathered during the 207 (45.6%) successful contacts are listed in Table 2.

The most cited reason for why participants' phones failed to send GPS heartbeat signals was a lack of data. Overall, 40.6% (84/207) of the contacts reported this reason. Over one-third of them (32/84, 38%) reported that they had purchased mobile data only for WhatsApp. The second most cited reason for signal failure was that the CareConekta app was uninstalled from their phone (28/207, 13.5%). Phone change often occurred because of sharing phones with family members and friends, and children were often blamed for the deletion of apps from the phone. In addition, 9.7% (20/207) of the participants reported no longer having a smartphone, whereas 6.3% (13/207) changed to a different smartphone that did not have the app installed. Other reasons for GPS failure included phone malfunctions or broken phones (14/207, 6.8%), malfunctions related to the CareConekta app (14/207, 6.8%), and the disabling of GPS (13/207, 6.3%).

Table 2. Primary reason for GPS heartbeat gap reported through 207 participant contacts.

Primary reason for GPS heartbeat gap	Value (n=207), n (%)
Lack of mobile data	84[a] (40.6)
Uninstallation of the CareConekta app	28 (13.5)
Phone change: new phone is not a smartphone	20 (9.7)
Phone malfunction or broken phone	14 (6.8)
CareConekta app–related malfunction	14 (6.8)
Phone change: app not installed on new phone	13 (6.3)
Disabling of GPS	13 (6.3)
Unclear reason or failure of troubleshooting attempt	21 (10.1)

[a]32 (15.4%) participants mentioned having only WhatsApp data.

Raffle

In November 2020, to encourage participants to keep the CareConekta app installed on their phone and the GPS function enabled throughout the study period, we implemented an incentive: a weekly raffle of one 200 MB data bundle (worth approximately US $4). Participants were eligible to enter the raffle if their phone sent GPS coordinates at least once a day in the prior week. Among all eligible participants, 1 winner per week was randomly selected. The weekly raffle incentive only applied to the participants (n=86) enrolled from November 2020 onward who had signed version 5.0 or later of the informed consent document. There was no limit to the number of times a participant could win the raffle.

From the 41-week period of November 18, 2020, to September 15, 2021, the weekly raffle was drawn 32 times. Four drawings were missed because the CareConekta dashboard was down, and we could not see the GPS data. Two drawings were missed because of holidays. There were no winners for 3 weeks because no participants were eligible. From the 32 drawings, there were 23 unique winners. The same participants won the raffle repeatedly because of the small number of participants who met the eligibility criteria of having consistent heartbeats. In total, 3 participants won the raffle 3 times each during this period. We found that the raffle incentive made no difference to the consistency of GPS heartbeats.

Additional Technical Challenges

On multiple occasions, the staff-facing dashboard was either not accessible or not fully functional, which meant that the team was unable to view or download heartbeat data. During the early study period, these problems often resulted in app revisions and version updates. In some instances, when the dashboard was down, the app did not work either, and heartbeats were not transmitted or recorded. Although the app was originally specified to store heartbeats on the phone and transmit them to the server when the connection was restored, this did not happen. Similarly, the app was designed to be reverse billed so as to not cost participants data for using the app; however, some data or airtime was needed on the device for the app to initiate.

Data Expenditure

Overall, for all their cell phone needs, participants reported spending a median of R51 (IQR 30-100; US $2.75, IQR US $1.60-5.40) for data per month, with a similar response for monthly spend on airtime: median R50 (IQR 29-100; US $2.71, IQR US$1.57-5.40). Cellular data in South Africa cost approximately R10 (US $0.50) per 50 MB.

Participants' Understanding of the CareConekta App

At follow-up, participants were asked, "If you were to explain to a friend what the CareConekta app does, how would you explain it?" Nearly all participants (167/170, 98.2%) mentioned the clinic finder feature:

> *I'd say it's an app used to search for clinics when I travel to the Eastern Cape so I don't suffer when I run out of medication. I'd simply just search for a clinic on the app.* [Participant #103]
>
> *It is an app that can help you find clinics near you, so you don't say you did not go to the clinic because you did not know where it was.* [Participant #58]

Only 5 (2.9%) mentioned geolocation tracking or the app knowing the participant's location, and 2 (1.2%) participants said that they did not know the app's function. None of the participants mentioned the app notifications or staff contact via WhatsApp or phone.

Participants' Acceptability of the CareConekta App

Similar to the participants' responses regarding their understanding of the app, most of the responses regarding what participants liked about using the app were related to the clinic finder feature:

> *I found it useful because I don't have to look for clinics should I travel outside Cape Town. The app connects me to the clinics closest to me.* [Participant #19]
>
> *What I liked was that we had been looking for a pediatric clinic and got lost in a taxi, I then thought of this app, I used it and it showed me exactly where the clinic was.* [Participant #029]

Some responses indicated that the participants used the clinic finder for a general map too:

> *With this app I know for a fact I'd never get lost when I go somewhere. The app has a GPS function.* [Participant #88]

Initial Efficacy of the Intervention

Although the initial efficacy of the intervention was a secondary aim of the study, we were unable to assess this because the app did not function as designed. Without receiving regular GPS heartbeats, the study team did not know when a participant was traveling; therefore, the intervention could not be initiated. During the study period, only 1 participant in the intervention arm was flagged as traveling, as defined in our protocol, and received the additional notifications, so it was not possible to assess a statistically meaningful difference between the study arms.

Discussion

Principal Findings

This is one of the first GPS-based mHealth studies—if not the first GPS-based mHealth study—targeted at improving HIV care in South Africa, and we found that several key challenges impeded its implementation. This study was designed to test the feasibility and acceptability of the CareConekta app and the initial efficacy of using it as an intervention to improve engagement in care among mobile women living with HIV. Although we were able to accomplish our primary aim of assessing the feasibility and acceptability of the intervention, we were unable to assess the efficacy of the intervention because we did not receive consistent location-tracking data. It is important to note that the app missed picking up on travel that was reported at follow-up. Although this is disappointing, we feel that the lessons learned from the implementation of this ambitious mHealth study are important and will be useful to other researchers considering mHealth interventions for low-resource settings. In designing this study, we made a conscious choice to assess our app under real-world conditions. We briefly considered providing phones—particularly to avoid bias against those who did not own phones and to guarantee a consistent technical level of device—but decided that this would not allow us to interpret the real-world applicability of our results. Similar decisions were made against providing data to all participants. Thus, our results can be viewed as representing implementation in real-world conditions.

We developed an initial beta version of the CareConekta app and implemented it in a proof-of-concept trial in 2017. We enrolled 11 participants at the same study site. Among the 11 participants, app installation failed for 7 (64%) individuals. Because the app team was US-based, some requirements of the app did not align with the capabilities of many of the phones in use in South Africa, and we were also unable to offer real-time technological support in the event of installation difficulty. The importance of a local app development team, with available technology support, was one of the key lessons learned from this early work. We were also committed to collaborating with a local development company that works with and knows the mHealth agenda of the South African Department of Health; we wanted to be well poised for broader implementation if our app was successful. In the proof-of-concept trial, we were able to install the app on 4 participants' phones and deploy the app for 3 months. Of the 4 participants, 1 (25%) lost her phone after approximately 1 month, but the other 3 (75%) produced heartbeats at least weekly, often daily, during the 3-month period. This sufficiently proved the concept for us to proceed with this study.

In this study, during the 15 months of intensive data monitoring, no participant had GPS heartbeats every day without interruption, which is a key indicator of study feasibility. Despite specifically designing the app such that mobility data would be stored on the app in the event of interruptions in data, the app did not function correctly and did not provide the missing data. Our attempts to troubleshoot lost GPS signals unexpectedly required substantial staff time; indeed, at least 1 staff member phoned participants every week for a year to ask about missing heartbeats. Most participants did not respond, or if they did, they requested a later callback and then still did not respond. CareConekta was designed to reverse bill, which meant that even if the mobile device had no airtime and no data, the app would still be fully functional. However, through implementation, we found that if a participant had no data at all, the app would not work. That is, for reverse billing to work properly, a small amount of data was first required. This became a major stumbling block for implementation, as the top reason cited by participants for missing GPS heartbeats was a lack of data. In addition, we received numerous reports of purchasing only WhatsApp data, a product that was unfamiliar to the researchers at the time of study design but appears to have grown in popularity during the course of our study. Although the lack of electricity was not mentioned as a reason for lost GPS heartbeats, the study period coincided with regular periods of load shedding—scheduled electricity blackouts to conserve power in South Africa—which would have impacted participants' ability to keep their phones charged. Future mHealth studies will be wise to consider the high likelihood of the lack of data and electricity during study implementation.

Even as early as installation, difficulties arose in using the app. From our preliminary research, we knew that over 90% of our participant population used Android-based phones [20]. However, downloading and installing the app from the Google Play Store required a Google-authenticated email address, which not every participant had before enrollment. The availability of space on the phone for Google Play Store updates and the CareConekta app was also a challenge. Even after extensive troubleshooting and creating space, some phones that seemed to meet our technical specifications were still unable to successfully install CareConekta, and we had to withdraw 6 participants because of app installation failure. Despite being designed to collect heartbeats only twice per day, the app collected location data multiple times a day. This created a challenge for analysis and may have contributed to battery drain on devices.

Overall, we found that the participant acceptability of CareConekta was high, but we view this finding with caution because it does not align with the numbers of participants who reported losing phones and uninstalling the app or the high frequency of lost GPS heartbeats. It is possible that social desirability bias to report a positive experience to the interviewer influenced responses. In the follow-up responses, all participants understood the participant-facing function of the app—the clinic finder—but seemed to forget or not understand the passive geolocation tracking, despite the great efforts made to be explicit about location tracking at the time of informed consent. Future mHealth intervention developers should note that patient-facing features may be the ones that will be most understood and remembered among participants.

The proportion of phone sharing reported at follow-up (15%) is consistent with that reported at the time of enrollment (14%). This frequency of sharing is similar to another recent mHealth study in South Africa that found 11% phone sharing [33]. The possibility of phone sharing and potential lack of confidentiality should be considered when designing future mHealth studies in this setting. Relatedly, participants often reported that children who shared their phones were the ones responsible for uninstalling the app.

Ours is not the first smartphone-based study in South Africa to experience substantial feasibility challenges. A South African study conducted from 2015 to 2017 [34] was unable to meet its sample size requirements because 90.2% (n=3187) of the 3540 potential participants were ineligible at screening because of they did not have an Android phone (n=2100, 59.3%), their phone was not working (n=506, 14.3%), or they had a wrong Android version or inadequate RAM or mobile data (n=581, 16.4%). Screening for our study is described elsewhere [31] and did not encounter similar challenges, but the implementation of our study did encounter similar issues with the lack of data, and inadequate RAM may have caused the installation challenges we experienced, as apps and photos sometimes needed to be deleted to make space for the CareConekta app. A different South African study from 2019 found that 13% of the eligible participants were unable to download the study app on their Android-based phone, 6% were unable to scan a barcode, and 3% were unable to complete app registration [33]. During implementation, the authors noted reports of a lack of data access and lost or broken phones. MomConnect, a nationwide pregnancy registry service in South Africa, demonstrated broad success by sending simple SMS text messages but noted that network timeouts and failures were frequent, resulting in 1 in 4 users dropping out of the registration process prematurely, and recommended that platforms such as WhatsApp be adopted to encourage flexibility in messaging [35].

To our knowledge, this is the first mHealth intervention to use location-based participant tracking in a real-world setting in South Africa. Interventions such as this are becoming increasingly common in the United States and Europe. For example, in a study focused on travel health published in 2016, a total of 101 Swiss adult travel clients planning to travel to Thailand for <5 weeks were provided a smartphone equipped with an app to passively monitor their location and administer a daily questionnaire [36]. The authors found that the app was feasible and acceptable, but 10% (n=10) of the participants had technical difficulties, and 16% (n=16) dropped out during the brief follow-up period. In a US-based study of patients who were chronically ill published in 2018, a total of 27 participants were provided with a smartphone and location-tracking watch; 6 (22%) participants dropped out before the end of the 28-day study period, some owing to the inability to use the devices [37]. Of note, the investigators offered participants a financial incentive (up to US $100) to upload their data to the study. A US-based study published in 2021 provided smartphones to 30 individuals experiencing homelessness, who were followed for 4 months, with the goal of alerting their community-based team that the participants had entered hospital or emergency care [38]. The authors found that 6 (20%) participants were withdrawn after reporting their second study-provided smartphone stolen, and overall, only 19% of the GPS data aligned with hospital data, primarily owing to participants not having the smartphone with them during the visit, the smartphone being switched off, and gaps in GPS technology. Compared with our study, these studies had smaller sample sizes, shorter follow-up periods, very different settings, and provided smartphones or financial incentives; however, it is interesting to note that similar challenges were encountered in implementation. Despite the setbacks experienced in our study and in others, given the rise in off-the-shelf location-tracking apps in recent years and the clear importance of mobility in health, we anticipate that more mHealth interventions will focus on participant movement. Indeed, the massive *All of Us* campaign in the United States plans to incorporate location-based data from smartphones and wearable devices, including location, cardiac rate and rhythm, and respiratory rate [39].

Some of the problems experienced with app implementation may have been avoided through clearer communication between the research team and app development or technical team. Although we thought that we had very close communication, some aspects were lost in translation. It is critically important that the research team investigating any mobile app includes someone who fully understands the technical specifications and requirements and can liaise with or translate the research vision to the app development team, as well as caution the research team on potentially problematic areas of design.

Our study has some clear limitations, primarily that the app did not function as designed because of a combination of user actions and app malfunction. Another limitation is that we conducted the study at a single site, thus potentially limiting the generalizability of our results. However, given that we attempted to mimic real-world conditions, we anticipate that our study population will be similar to other adults attending public health clinics in South Africa.

However, we feel that many important lessons learned through this experience will be useful to other researchers and make the effort worthwhile. One strength of our study is that it was among the first to develop, implement, and clearly document experiences with a GPS-based location-tracking mHealth app in a low-resource setting, particularly in a real-world setting. Thus, our findings are meaningful, even if our intervention was not successful. In addition, we report on several technical factors traditionally unreported in the literature but critical to the feasibility of mHealth interventions.

Conclusions

In conclusion, we did not demonstrate the feasibility of using a GPS-based tracking app to characterize mobility and improve engagement in HIV care. Our most common problems that contributed to failure were a lack of mobile phone data, app uninstallations, phone changes, and missing heartbeat data. We are far less motivated to create a novel app in our future research endeavors and instead will use tools that people already are using, particularly WhatsApp, which all of our participants reported as their favorite app [31]. Although some of our participants had no data for other apps, they were buying data specifically for WhatsApp. By using apps that are already essential to participants' lives, researchers developing future studies can become better poised to ensure continuous engagement with the apps and decrease the likelihood of the apps being uninstalled. Careful translation of research aims into app design is essential in the development of future studies. In addition, the mHealth landscape is changing rapidly, and it is possible that Wi-Fi and internet availability will increase in the future, thus overcoming some of the challenges we experienced. We anticipate that mHealth interventions will continue to proliferate in resource-limited settings, such as our study setting, and our study results may offer guidance and words of caution for unanticipated challenges.

Acknowledgments

This work was funded by the US National Institute of Mental Health (R34 MH118028) and based on earlier work that was funded by the US National Institutes of Health (NIH) under grant P30 AI110527 provided to the Tennessee Center for AIDS Research. TKP was supported by a Collaborative Initiative for Pediatric HIV Education and Research grant from the International AIDS Society and by grant K43TW011943 jointly supported by the Fogarty International Center and National Institute of Mental Health of the NIH. The funders had no role in the study design, data collection and analysis, or decision to publish or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or other parties.

The authors are most grateful to the study participants, without whom this work would not have been possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 4263 KB - mhealth_v11i1e44945_app1.pdf]

References

1. Country factsheets South Africa. UNAIDS. URL: https://www.unaids.org/en/regionscountries/countries/southafrica [accessed 2023-03-10]
2. Implementation of the universal test and treat strategy for HIV positive patients and differentiated care for stable patients. Department of Health, Republic of South Africa. 2016. URL: https://tinyurl.com/4ds5pts3 [accessed 2023-03-10]
3. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
4. Rodger AJ, Cambiano V, Bruun T, Vernazza P, Collins S, Degen O, PARTNER Study Group. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. Lancet 2019 Jun 15;393(10189):2428-2438 [FREE Full text] [doi: 10.1016/S0140-6736(19)30418-0] [Medline: 31056293]
5. Bavinton BR, Pinto AN, Phanuphak N, Grinsztejn B, Prestage GP, Zablotska-Manos IB, Opposites Attract Study Group. Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. Lancet HIV 2018 Aug;5(8):e438-e447. [doi: 10.1016/S2352-3018(18)30132-2] [Medline: 30025681]
6. Fox MP, Rosen S. Retention of adult patients on antiretroviral therapy in low- and middle-income countries: systematic review and meta-analysis 2008-2013. J Acquir Immune Defic Syndr 2015 May 01;69(1):98-108 [FREE Full text] [doi: 10.1097/QAI.0000000000000553] [Medline: 25942461]
7. Retention in care a growing concern. UNAIDS. 2021 Nov 8. URL: https://www.unaids.org/en/resources/presscentre/featurestories/2021/november/20211108_retention-in-care-growing-concern [accessed 2023-10-03]
8. Kaplan R, Orrell C, Zwane E, Bekker L, Wood R. Loss to follow-up and mortality among pregnant women referred to a community clinic for antiretroviral treatment. AIDS 2008 Aug 20;22(13):1679-1681 [FREE Full text] [doi: 10.1097/QAD.0b013e32830ebcee] [Medline: 18670232]
9. Clouse K, Pettifor A, Maskew M, Bassett J, Van Rie A, Gay C, et al. Initiating antiretroviral therapy when presenting with higher CD4 cell counts results in reduced loss to follow-up in a resource-limited setting. AIDS 2013 Mar 20;27(4):645-650 [FREE Full text] [doi: 10.1097/QAD.0b013e32835c12f9] [Medline: 23169326]
10. Wang B, Losina E, Stark R, Munro A, Walensky RP, Wilke M, et al. Loss to follow-up in a community clinic in South Africa--roles of gender, pregnancy and CD4 count. S Afr Med J 2011 Apr;101(4):253-257 [FREE Full text] [doi: 10.7196/samj.4078] [Medline: 21786730]
11. Knettel BA, Cichowitz C, Ngocho JS, Knippler ET, Chumba LN, Mmbaga BT, et al. Retention in hiv care during pregnancy and the postpartum period in the option B+ era: systematic review and meta-analysis of studies in Africa. J Acquir Immune Defic Syndr 2018 Apr 15;77(5):427-438 [FREE Full text] [doi: 10.1097/QAI.0000000000001616] [Medline: 29287029]
12. Clouse K, Pettifor A, Shearer K, Maskew M, Bassett J, Larson B, et al. Loss to follow-up before and after delivery among women testing HIV positive during pregnancy in Johannesburg, South Africa. Trop Med Int Health 2013 Apr;18(4):451-460 [FREE Full text] [doi: 10.1111/tmi.12072] [Medline: 23374278]
13. Clouse K, Vermund SH, Maskew M, Lurie MN, MacLeod W, Malete G, et al. Mobility and clinic switching among postpartum women considered lost to HIV care in South Africa. J Acquir Immune Defic Syndr 2017 Apr 01;74(4):383-389 [FREE Full text] [doi: 10.1097/QAI.0000000000001284] [Medline: 28225717]
14. Clouse K, Fox MP, Mongwenyana C, Motlhatlhedi M, Buthelezi S, Bokaba D, et al. "I will leave the baby with my mother": long-distance travel and follow-up care among HIV-positive pregnant and postpartum women in South Africa. J Int AIDS Soc 2018 Jul;21 Suppl 4(Suppl Suppl 4):e25121 [FREE Full text] [doi: 10.1002/jia2.25121] [Medline: 30027665]
15. Phillips TK, Clouse K, Zerbe A, Orrell C, Abrams EJ, Myer L. Linkage to care, mobility and retention of HIV-positive postpartum women in antiretroviral therapy services in South Africa. J Int AIDS Soc 2018 Jul;21 Suppl 4(Suppl Suppl 4):e25114 [FREE Full text] [doi: 10.1002/jia2.25114] [Medline: 30027583]
16. Bervell B, Al-Samarraie H. A comparative review of mobile health and electronic health utilization in sub-Saharan African countries. Soc Sci Med 2019 Jul;232:1-16. [doi: 10.1016/j.socscimed.2019.04.024] [Medline: 31035241]

17. Manby L, Aicken C, Delgrange M, Bailey JV. Effectiveness of eHealth interventions for HIV prevention and management in Sub-Saharan Africa: systematic review and meta-analyses. AIDS Behav 2022 Mar;26(2):457-469 [FREE Full text] [doi: 10.1007/s10461-021-03402-w] [Medline: 34427813]
18. Onukwugha FI, Smith L, Kaseje D, Wafula C, Kaseje M, Orton B, et al. The effectiveness and characteristics of mHealth interventions to increase adolescent's use of Sexual and Reproductive Health services in Sub-Saharan Africa: a systematic review. PLoS One 2022;17(1):e0261973 [FREE Full text] [doi: 10.1371/journal.pone.0261973] [Medline: 35061757]
19. Pugh LE, Roberts JS, Viswasam N, Hahn E, Ryan S, Turpin G, et al. Systematic review of interventions aimed at improving HIV adherence to care in low- and middle-income countries in Sub-Saharan Africa. J Infect Public Health 2022 Aug 28;15(10):1053-1060 [FREE Full text] [doi: 10.1016/j.jiph.2022.08.012] [Medline: 36063721]
20. Mogoba P, Phillips TK, Myer L, Ndlovu L, Were MC, Clouse K. Smartphone usage and preferences among postpartum HIV-positive women in South Africa. AIDS Care 2019 Jun 31;31(6):723-729 [FREE Full text] [doi: 10.1080/09540121.2018.1563283] [Medline: 30596261]
21. Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW. Mobile phones and malaria: modeling human and parasite travel. Travel Med Infect Dis 2013;11(1):15-22 [FREE Full text] [doi: 10.1016/j.tmaid.2012.12.003] [Medline: 23478045]
22. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science 2012 Oct 12;338(6104):267-270 [FREE Full text] [doi: 10.1126/science.1223467] [Medline: 23066082]
23. Blumenstock J. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. Inform Technol Develop 2012 Feb 03;18(2):107-125 [FREE Full text] [doi: 10.1080/02681102.2011.643209]
24. Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med 2011 Aug;8(8):e1001083 [FREE Full text] [doi: 10.1371/journal.pmed.1001083] [Medline: 21918643]
25. Duncan DT, Kapadia F, Regan SD, Goedel WC, Levy MD, Barton SC, et al. Feasibility and acceptability of global positioning system (gps) methods to study the spatial contexts of substance use and sexual risk behaviors among young men who have sex with men in New York City: a P18 cohort sub-study. PLoS One 2016;11(2):e0147520 [FREE Full text] [doi: 10.1371/journal.pone.0147520] [Medline: 26918766]
26. Goedel WC, Reisner SL, Janssen AC, Poteat TC, Regan SD, Kreski NT, et al. Acceptability and feasibility of using a novel geospatial method to measure neighborhood contexts and mobility among transgender women in New York City. Transgend Health 2017;2(1):96-106 [FREE Full text] [doi: 10.1089/trgh.2017.0003] [Medline: 29082330]
27. Clouse K, Phillips TK, Camlin C, Noholoza S, Mogoba P, Naidoo J, et al. CareConekta: study protocol for a randomized controlled trial of a mobile health intervention to improve engagement in postpartum HIV care in South Africa. Trials 2020 Mar 12;21(1):258 [FREE Full text] [doi: 10.1186/s13063-020-4190-x] [Medline: 32164771]
28. Clouse K, Phillips TK, Mogoba P, Ndlovu L, Bassett J, Myer L. Attitudes toward a proposed GPS-based location tracking smartphone app for improving engagement in HIV care among pregnant and postpartum women in South Africa: focus group and interview study. JMIR Form Res 2021 Mar 08;5(2):e19243 [FREE Full text] [doi: 10.2196/19243] [Medline: 33555261]
29. Phillips TK, Bonnet K, Myer L, Buthelezi S, Rini Z, Bassett J, et al. Acceptability of interventions to improve engagement in HIV care among pregnant and postpartum women at two urban clinics in South Africa. Matern Child Health J 2019 Sep;23(9):1260-1270 [FREE Full text] [doi: 10.1007/s10995-019-02766-9] [Medline: 31218606]
30. Clouse K, Noholoza S, Ngcobo N, Madwayi S, Mrubata M, Camlin CS, et al. Cohort profile: CareConekta: a pilot study of a smartphone application to improve engagement in postpartum HIV care in South Africa. BMJ Open 2022 Nov 22;12(11):e064946 [FREE Full text] [doi: 10.1136/bmjopen-2022-064946] [Medline: 36414286]
31. Noholoza S, Phillips T, Madwayi S, Mrubata M, Camlin CS, Myer L, et al. Smartphone ownership and use among pregnant women with HIV in South Africa. MedRxiv 2022 Sep (forthcoming) [FREE Full text] [doi: 10.1101/2022.09.29.22280417]
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
33. Lalla-Edward ST, Mashabane N, Stewart-Isherwood L, Scott L, Fyvie K, Duncan D, et al. Implementation of an mHealth app to promote engagement during HIV care and viral load suppression in Johannesburg, South Africa (iThemba Life): pilot technical feasibility and acceptability study. JMIR Form Res 2022 Mar 02;6(2):e26033 [FREE Full text] [doi: 10.2196/26033] [Medline: 35107427]
34. Venter W, Coleman J, Chan VL, Shubber Z, Phatsoane M, Gorgens M, et al. Improving linkage to HIV care through mobile phone apps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e155 [FREE Full text] [doi: 10.2196/mhealth.8376] [Medline: 30021706]
35. Peter J, Benjamin P, LeFevre AE, Barron P, Pillay Y. Taking digital health innovation to scale in South Africa: ten lessons from MomConnect. BMJ Glob Health 2018;3(Suppl 2):e000592 [FREE Full text] [doi: 10.1136/bmjgh-2017-000592] [Medline: 29713511]
36. Farnham A, Blanke U, Stone E, Puhan MA, Hatz C. Travel medicine and mHealth technology: a study using smartphones to collect health data during travel. J Travel Med 2016 Jun;23(6):taw056. [doi: 10.1093/jtm/taw056] [Medline: 27592821]

37. Hardy J, Veinot TC, Yan X, Berrocal VJ, Clarke P, Goodspeed R, et al. User acceptance of location-tracking technologies in health research: implications for study design and data quality. J Biomed Inform 2018 Mar;79:7-19 [FREE Full text] [doi: 10.1016/j.jbi.2018.01.003] [Medline: 29355784]
38. Moczygemba LR, Thurman W, Tormey K, Hudzik A, Welton-Arndt L, Kim E. GPS mobile health intervention among people experiencing homelessness: pre-post study. JMIR Mhealth Uhealth 2021 Nov 03;9(11):e25553 [FREE Full text] [doi: 10.2196/25553] [Medline: 34730550]
39. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "all of us" research program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health
MOU: Midwife Obstetric Unit
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 09.12.22; peer-reviewed by M Rabbi2, S Badawy; comments to author 21.02.23; revised version received 10.03.23; accepted 21.04.23; published 19.05.23.

Please cite as:
Clouse K, Noholoza S, Madwayi S, Mrubata M, Camlin CS, Myer L, Phillips TK
The Implementation of a GPS-Based Location-Tracking Smartphone App in South Africa to Improve Engagement in HIV Care: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e44945
URL: https://mhealth.jmir.org/2023/1/e44945
doi:10.2196/44945
PMID:37204838

Original Paper

Using Chatbot Technology to Improve Brazilian Adolescents' Body Image and Mental Health at Scale: Randomized Controlled Trial

Emily L Matheson[1], PhD; Harriet G Smith[1], MSc; Ana C S Amaral[2], PhD; Juliana F F Meireles[3], PhD; Mireille C Almeida[4], MD, PhD; Jake Linardon[5], PhD; Matthew Fuller-Tyszkiewicz[6], PhD; Phillippa C Diedrichs[1], PhD

[1]Centre for Appearance Research, University of the West of England, Bristol, United Kingdom

[2]Federal Institute of Education, Science and Technology of Southeast of Minas Gerais, Barbacena, Brazil

[3]Department of Family and Community Medicine, School of Community Medicine, University of Oklahoma Health Science Center, Tulsa, OK, United States

[4]Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil

[5]School of Psychology, Deakin University, Geelong, Victoria, Australia

[6]Center for Social and Early Emotional Development, Deakin University, Burwood, Victoria, Australia

Corresponding Author:
Emily L Matheson, PhD
Centre for Appearance Research
University of the West of England
Coldharbour Ln
Bristol, BS16 1QY
United Kingdom
Phone: 44 1173284398
Email: emily.matheson@uwe.ac.uk

Abstract

Background: Accessible, cost-effective, and scalable mental health interventions are limited, particularly in low- and middle-income countries, where disparities between mental health needs and services are greatest. Microinterventions (ie, brief, stand-alone, or digital approaches) aim to provide immediate reprieve and enhancements in mental health states and offer a novel and scalable framework for embedding evidence-based mental health promotion techniques into digital environments. Body image is a global public health issue that increases young peoples' risk of developing more severe mental and physical health issues. Embedding body image microinterventions into digital environments is one avenue for providing young people with immediate and short-term reprieve and protection from the negative exposure effects associated with social media.

Objective: This 2-armed, fully remote, and preregistered randomized controlled trial assessed the impact of a body image chatbot containing microinterventions on Brazilian adolescents' state and trait body image and associated well-being outcomes.

Methods: Geographically diverse Brazilian adolescents aged 13-18 years (901/1715, 52.54% girls) were randomized into the chatbot or an assessment-only control condition and completed web-based self-assessments at baseline, immediately after the intervention time frame, and at 1-week and 1-month follow-ups. The primary outcomes were mean change in state (at chatbot entry and at the completion of a microintervention technique) and trait body image (before and after the intervention), with the secondary outcomes being mean change in affect (state and trait) and body image self-efficacy between the assessment time points.

Results: Most participants who entered the chatbot (258/327, 78.9%) completed ≥1 microintervention technique, with participants completing an average of 5 techniques over the 72-hour intervention period. Chatbot users experienced small significant improvements in primary (state: $P<.001$, Cohen d=0.30, 95% CI 0.25-0.34; and trait body image: P=.02, Cohen d range=0.10, 95% CI 0.01-0.18, to 0.26, 95% CI 0.13-0.32) and secondary outcomes across various time points (state: $P<.001$, Cohen d=0.28, 95% CI 0.22-0.33; trait positive affect: P=.02, Cohen d range=0.15, 95% CI 0.03-0.27, to 0.23, 95% CI 0.08-0.37; negative affect: P=.03, Cohen d range=−0.16, 95% CI −0.30 to −0.02, to −0.18, 95% CI −0.33 to −0.03; and self-efficacy: P=.02, Cohen d range=0.14, 95% CI 0.03-0.25, to 0.19, 95% CI 0.08-0.32) relative to the control condition. Intervention benefits were moderated by baseline levels of concerns but not by gender.

Conclusions: This is the first large-scale randomized controlled trial assessing a body image chatbot among Brazilian adolescents. Intervention attrition was high (531/858, 61.9%) and reflected the broader digital intervention literature; barriers to engagement were discussed. Meanwhile, the findings support the emerging literature that indicates microinterventions and chatbot technology are acceptable and effective web-based service provisions. This study also offers a blueprint for accessible, cost-effective, and scalable digital approaches that address disparities between health care needs and provisions in low- and middle-income countries.

Trial Registration: Clinicaltrials.gov NCT04825184; http://clinicaltrials.gov/ct2/show/NCT04825184

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-021-12129-1

(JMIR Mhealth Uhealth 2023;11:e39934) doi:10.2196/39934

KEYWORDS

adolescent; Brazil; body image; chatbot; microintervention; randomized controlled trial; mobile phone

Introduction

Background

Accessible, cost-effective, and scalable mental health interventions are limited, particularly in low- and middle-income countries (LMICs), where disparities between mental health needs and services are greatest [1,2]. Digital interventions offer a unique opportunity to address this gap, particularly among adolescents. For instance, in Brazil, 91% of those aged 9 to 17 years have internet access, with 88% of internet users having a social media profile [3]. Brazilian adolescents use social media for an average of 4 hours per day [4], with most (53.2%) reporting problematic smartphone behaviors including overuse, preoccupation, and withdrawal [5]. Despite this high level of exposure to potentially harmful social media content, digital environments have been underused to connect with and provide Brazilian adolescents with evidence-based mental health and well-being resources, particularly regarding body image.

Body image concerns are a potent risk factor for eating disorders, with this group of mental health concerns incurring a yearly economic cost of US $149 billion in the United States alone [6,7]. Costs are also likely to rise owing to the impact of the global COVID-19 pandemic on body and eating concerns [8]; thus, accessible, cost-effective, and scalable prevention and intervention approaches are critical, particularly for already underserved countries. For instance, body image concerns are a global phenomenon and are reported across various countries and cultures [9]. However, advancements in body image research and the availability of evidence-based prevention and intervention vary greatly between countries. In some countries, body image is an unexplored construct; in others, there are a handful of prevalence studies; or in the case of high-income, English-speaking countries with majority White population, there is an abundance of research on prevalence, causation, prevention, and intervention [10]. This disparity in research does not reflect the prevalence or severity, with body image concerns in LMICs comparable with those in high-income countries [8,10]. Furthermore, owing to political and economic disparities and competing priorities, research and early intervention efforts for body image in LMICs are largely unfunded, unexplored, and unaddressed.

Previous Work

In Brazil, 8 in 10 young people report body image concerns, with 1 in 5 reporting engagement in disordered eating and unhealthy weight control behaviors [11,12]. However, to date, only 2 body image interventions have been evaluated among Brazilian populations and neither are widely available [13,14]. This is, in part, owing to the intervention modality and the historical and current funding restrictions experienced by Brazilian researchers. First, both interventions involve in-person implementation in group settings; therefore, their sustainability is reliant on human and infrastructural resources. Second, in 2017, the Brazilian government reduced its health budget by US $210 million, with spending cuts of 15% for public universities and 45% for scientific research [15]. Therefore, developing innovative and sustainable evidence-based body image interventions is a challenge for Brazilian researchers. One way to overcome these barriers is to use international partnerships among academia, industry, and community. Such partnerships afford the creation of accessible, cost-effective, and scalable interventions that reach those in need and, in turn, reduce health care disparities.

Chatbot technology offers an innovative pathway for engaging young people with evidence-based resources. In a recent review, 11 mental health chatbots were identified across 12 studies, with the majority targeting affective disorders, including anxiety and depression (eg, *Woebot*) [1]. The chatbots used predefined rules or decision trees (8 out of 12 studies), machine learning, or artificial intelligence (4 out of 12 studies), and half were personified with an avatar. Although the chatbots showed promise as an effective and safe intervention modality, the review heeded caution about their clinical significance relative to treatment as usual and, therefore, concluded that greater research is needed to draw solid conclusions about this emerging service provision. Furthermore, all studies were conducted in high-income countries (eg, Australia, China, Sweden, and the United States), and therefore, the authors called for concerted research efforts in LMICs, stating that there is a greater need for this technology in countries where the shortage of mental health professionals is the highest. To date, no chatbot has been developed for or tested among Brazilians.

Topity, a new Brazilian body image chatbot hosted on *Facebook Messenger*, comprises 8 microintervention techniques that address risk and protective factors for body image [16-18]. Microinterventions are generally designed as brief, digital, and self-guided approaches that use in-the-moment techniques to provide immediate symptom relief or enhancement [19]. To date, this intervention model has been predominantly developed for and tested among adult samples with body image and mood concerns, with techniques including brief web-based written tasks and instructional audio and video clips [20,21]. More recently, microinterventions have been developed and applied to young people (eg, short films and web-based games), with these approaches proving acceptable and effective at eliciting immediate and short-term improvements in body image and mood [22,23].

During the *Topity* experience, users are given the choice to interact with either an avatar of a young woman (Dandara) or a man (Gabriel). The chatbot uses predefined rules and gamification to guide users through the completion of 8 microinterventions, which are clustered into 3 themes (*Family, Friends and Body Image* [2 techniques]; *Social Media and Body Image* [4 techniques]; and *Body Appreciation and Functionality* [2 techniques]). The microinterventions were informed by and adapted from existing body image interventions that have been traditionally delivered in hard copy (eg, self-help books) or in-person settings (eg, individual therapy or group-based programs). These techniques teach users how to critically analyze and evaluate social media content to reduce vulnerability to negative influences (ie, media literacy) [16], identify and challenge unhelpful thinking styles and behaviors that perpetuate body image distress (ie, cognitive behavior theory for body image) [17], and appreciate the features and functions of the body beyond appearance (ie, positive body image and embodiment theory) [18]. Each technique takes between 5 and 10 minutes to complete and has a distinct beginning and end. Access to microintervention techniques is gamified, with users needing to complete an initial technique within a thematic cluster before "unlocking" more activities. For example, in the *Family, Friends and Body Image* cluster, users need to complete the *Banish Body Talk* microintervention before being able to access and complete the *Dealing with Provocative People* microintervention. Gamification is a key feature of digital interventions because of its effect on users' motivation, engagement, and skill mastery [24]. An overview of the cocreation process for *Topity* and a summary of the microintervention techniques are reported in Tables 1 and 2 of the protocol, respectively [25].

Topity is 1 of the 4 body- and eating-related chatbots that have been developed in recent years. The other 3 bots, including *KIT* [26], *Tessa* [27], and *Alex* [28], target English-speaking populations and offer different service provisions. Furthermore, each chatbot is at a different phase of the development or testing process (eg, user experience vs effectiveness). First, *KIT* addresses body and eating concerns by providing users with psychoeducation, help seeking, and coping strategies. It was highly acceptable among users in Australia, and its effectiveness has been reported in gray literature [26]. Second, *Tessa* features the *Body Positive* program, which is a distilled version of the cognitive-behavioral therapy web-based program, *StudentBodies*, and comprises 8- × 10-minute conversations with the chatbot [27]. *Body Positive* has proven effective at reducing weight and shape concerns among at-risk American women (mean age 21, SD 3.09 years) and shows potential for reducing the onset of eating disorders. Finally, and most recently, *Alex* was developed to increase treatment motivation and the use of mental health services among individuals who screen positive for an eating disorder and are not yet in treatment [28]. *Alex* is currently undergoing efficacy testing [29].

Our Study

This 2-armed, fully remote randomized controlled trial (RCT) addresses several gaps within the limited but emerging field of mental health chatbots. It is the first body image chatbot for non–English-speaking users and one of the first studies to assess a chatbot in an LMIC [1]. Specifically, this trial evaluated the effectiveness of *Topity* in eliciting immediate and short-term improvements in adolescents' state- and trait-based body image, affect, and self-efficacy in managing body image concerns. Research hypotheses were prespecified on page 3 of the protocol [25] and were formulated for overall chatbot efficacy (hypotheses 1-2), moderation effects within subsamples (hypothesis 3), and intervention engagement and adherence (hypothesis 4):

1. *Hypothesis 1:* the chatbot was designed to provide immediate benefits to users. Therefore, it is anticipated that adolescents will experience improvements in state-based body satisfaction and affect at the time of engaging with the chatbot.
2. *Hypothesis 2:* adolescents randomized into the intervention condition will experience greater improvements in trait-based body esteem, affect, and body image self-efficacy immediately after the intervention time frame (eg, postintervention assessment) and at the 1-week and 1-month follow-ups relative to the assessment-only control condition.
3. *Hypothesis 3:* on the basis of previous research on the moderating effects of gender [30] and trait psychopathology [23,31] on body image intervention effectiveness, it was hypothesized that intervention effects will be moderated by gender and baseline levels of body esteem, affect, and body image self-efficacy. Specifically, intervention effects will be greatest among girls, girls and boys with lower levels of trait body esteem and body image self-efficacy, and girls and boys with higher levels of trait negative affect.
4. *Hypothesis 4:* with regard to user engagement and adherence, given the novelty of this intervention, analyses will be exploratory. However, it is anticipated that greater engagement and adherence with chatbot interventions will lead to greater improvements in trait-based outcomes.

Methods

Study Design

The study was a 2-armed, fully remote, and preregistered RCT conducted between April 7 and August 8, 2021, in Brazil. Details of the study rationale and protocol have been published elsewhere [25]. The research was conducted in accordance with ethical standards and guidelines for conducting research on young people in Brazil. Before obtaining consent and assent, parents and participants were provided with an information sheet outlining the research procedures and associated risks, respectively. Consent withdrawal was possible at any time without cause for justification. At study completion, participants were provided with a debrief form disclosing research aims, ancillary mental health resources and an electronic voucher of R$100 (approximately US $20) and R$80 (approximately US $15) for the intervention and control conditions, respectively.

Data were anonymized and accessed only by the authoring research team using masked and password-protected data files.

Ethics Approval and Trial Registration

This study received ethics approval from the *Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais* (4.232.804); *Comissão Nacional de Ética em Pesquisa* (4.582.466); and the University of the West of England (HAS 19.12.090). The study was registered with ClinicalTrials.gov (NCT04825184) [32].

Participants

The participants were recruited from diverse ethnic, geographic, and socioeconomic backgrounds across Brazil. Recruitment was conducted via a Brazilian research agency (ie, via email to their participant databases) and United Nations International Children's Emergency Fund's web-based communication platforms (eg, U-Report; a free tool for community participation). Eligible participants were adolescents aged 13 to 18 years who spoke Brazilian Portuguese, were a Brazilian resident, and had access to *Facebook Messenger*.

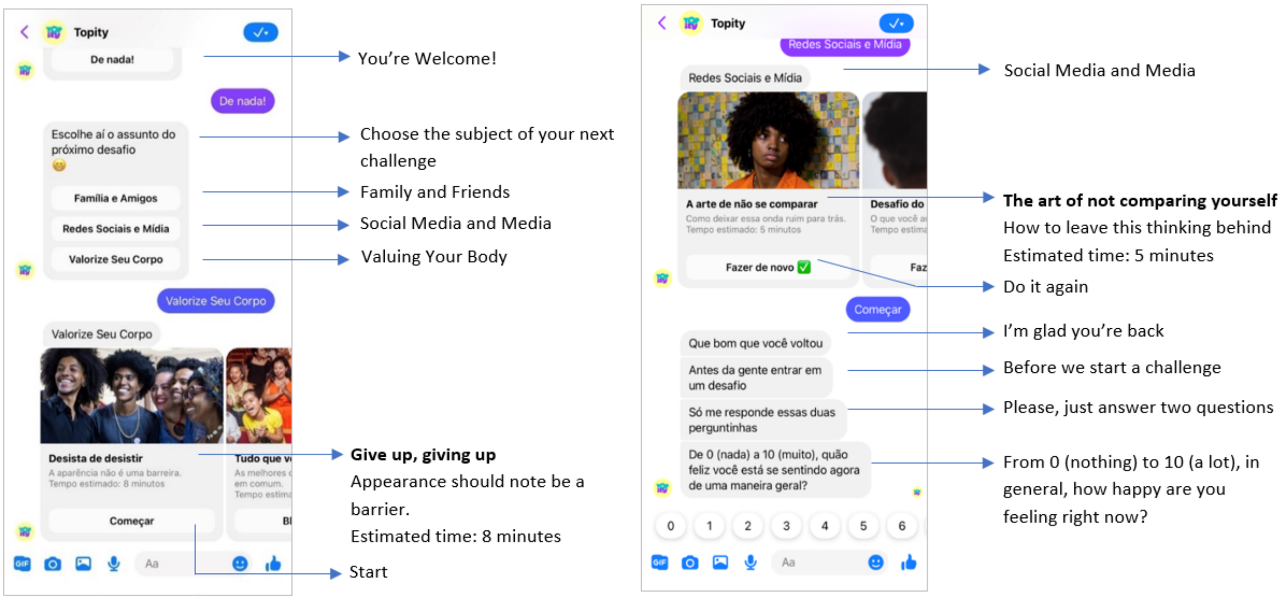
Randomization and Masking

Participants were randomly assigned 1:1 to receive either the chatbot intervention or an assessment-only control. The randomization scheme was generated by a research agency using a validated computer software. Blinding of the participants was not possible because of the nature of the intervention. The risk of bias from researchers was minimized because of having no contact with participants during the trial (ie, recruitment; randomization, survey dissemination, and compensation was completed by the research agency). Data analysts were blinded when conducting analyses of primary and secondary trait outcomes; however, this was not possible for primary and secondary state measures because of the within-group design.

Procedure

Following parental consent and participant assent, eligible participants completed preintervention self-assessments via a closed web-based survey (received via email), after which they were randomized into the intervention or control conditions. The web-based survey underwent user testing before the trial began and the research agency monitored the response rate and completion. Those in the intervention condition received a unique access link to the chatbot, 24 hours after randomization, and were encouraged to interact with the intervention as much as possible over the next 72-hour period (Figure 1). The unique access code was only compatible with the assigned user's *Facebook Messenger*, thus ensuring confidential interactions between the user and chatbot and prohibiting nonparticipants from entering the chatbot during the trial period. The control condition received "standard online care" for body image concerns among Brazilian adolescents, which at present is no intervention. Initially, *Topity* users had access to 3 of the 8 microintervention techniques (eg, the first technique from each thematic cluster), and once completed, they gained access to the next technique in the cluster. Therefore, once a technique was unlocked, there were no limits to the number of completions. The ordering of techniques was informed by the difficulty of the microintervention concept and skill, whereby users mastered easier concepts and skills (eg, psychoeducation) before progressing to those that were more challenging (eg, behavior and cognition change). Upon entering the chatbot and after completing a microintervention technique, users were assessed for state body satisfaction and affect. In the event that participants did not engage with the bot after 12, 16, and 23.5 hours, they received a *Facebook Messenger* notification encouraging participation. At the completion of the 72-hour intervention period, all participants were sent postintervention surveys and again at the 1-week and 1-month time points. During the 1-month follow-up period, all participants were provided with contact details for accessible mental health support services in Brazil, and the control condition was invited to interact with *Topity*; however, their engagement was not monitored or assessed for effectiveness. The participants received an electronic voucher of R$100 (approximately US $20) and R$80 (approximately US $15) for the intervention and control condition, respectively. Although participants were aware that they would receive compensation at study commencement, the compensation amount was not disclosed until study completion.

Figure 1. Screenshots and translations of a Topity conversation on mobile view.



Outcomes

A comprehensive overview of the outcome measures is provided in Table 3 of the protocol [25]. The primary outcome measures were the mean change in state (ie, a single 11-point scale) and trait (ie, Body Esteem Scale for Adolescents and Adults Brazil [33]) body image—specifically, the mean change in state body satisfaction between chatbot entry and at the completion of a microintervention technique and the mean change in trait body esteem between preintervention and postintervention assessments. Secondary outcomes included the mean change in state affect (ie, a single 11-point scale), trait positive and negative affect (ie, the Positive and Negative Affect Scale for Children 8-item [34]), and self-efficacy in addressing body image concerns (ie, The Body Image Self-Efficacy Scale [35-37]). Treatment adherence (eg, ≥1 microintervention technique) and user acceptability were also assessed after the intervention time frame.

Statistical Analysis

Statistical analyses were prespecified on pages 10 and 11 of the protocol [25]. Specifically, missing data were handled using multiple imputations (m=50) with chained equations, and participants were retained in the group they were assigned to at baseline, consistent with the principles of intention-to-treat. Linear mixed models were used to test our hypotheses. For hypothesis 1, scores on state-based outcomes were regressed onto a time variable (coded 0=precontent; 1=postcontent) for the intervention arm only. For hypotheses 2 to 4, trait-level outcome variables were regressed onto a time variable (dummy coded as baseline vs after the intervention, 1-week follow-up, and 1-month follow-up), group (0=control; 1=intervention), and an interaction between time and group to evaluate efficacy. Random effects were included for time and an unstructured covariance matrix was used to estimate the covariance among these random effects. Hypotheses 3 and 4 included the moderators of these random effects over time.

Protocol Amendments

Owing to the recruitment processes being heavily reliant on parents' literacy levels to provide informed consent, a video communicating the study information was embedded into the recruitment materials. This amendment was implemented 2 weeks into recruitment, following feedback from stakeholders that the original recruitment methods were restrictive. Furthermore, the recruitment phase was extended from 2 to 12 weeks because of slow uptake. Extensions were incrementally increased as part of the review process (eg, every 2 weeks). This rate was somewhat accelerated by adapting the recruitment materials; however, given that the trial was conducted during the peak of the COVID-19 pandemic in Brazil, there are likely several contextual factors outside the researchers' control that led to lower than anticipated uptake (eg, screen and social media fatigue) [38].

Results

Baseline Characteristics

The baseline characteristics of the participants are presented in Table 1. There was an approximately equal number of girls and boys and a nationally representative distribution across ethnicity and region. There were no notable differences and negligible effect sizes on any baseline variables between the intervention and control groups, indicating that the randomization process was successful.

Table 1. Baseline characteristics of the sample.

Variable[a]	Total sample (N=1715)	Intervention (n=858)	Control (n=857)	*P* value	ES[b]
Age (years), n (%)				.29	0.06
13	276 (16.1)	135 (15.7)	141 (16.5)		
14	237 (13.8)	108 (12.6)	129 (15.1)		
15	230 (13.4)	124 (14.5)	106 (12.4)		
16	255 (14.9)	140 (16.3)	115 (13.4)		
17	220 (12.8)	109 (12.7)	11 (1.3)		
18	497 (29)	242 (28.2)	255 (29.8)		
Gender, n (%)				.64	0.02
Boy	801 (46.7)	406 (47.3)	395 (46.1)		
Girl	901 (52.5)	447 (52.1)	454 (53)		
Gender diverse	13 (0.8)	5 (0.6)	8 (0.9)		
Ethnicity, n (%)				.91	0.03
Asian	40 (2.3)	23 (2.7)	17 (2)		
Black	181 (10.6)	92 (10.7)	89 (10.4)		
Indigenous	14 (0.8)	8 (0.9)	6 (0.7)		
Mixed race	630 (36.7)	314 (36.6)	316 (36.9)		
White	845 (49.3)	419 (48.8)	426 (49.7)		
Other	5 (0.3)	2 (0.2)	3 (0.4)		
Region, n (%)				.68	0.03
North	92 (5.4)	50 (5.8)	42 (4.9)		
Northeast	348 (20.3)	176 (20.5)	172 (20.1)		
Central west	134 (7.8)	60 (7)	74 (8.6)		
Southeast	906 (52.8)	452 (52.7)	454 (53)		
South	235 (13.7)	120 (14)	115 (13.4)		
Baseline variables, mean (SD)					
Appearance positive	3.33 (0.83)	3.36 (0.84)	3.31 (0.83)	.24	0.05
Appearance negative	3.34 (1.08)	3.36 (1.08)	3.33 (1.08)	.64	0.02
Weight	3.11 (1.18)	3.10 (1.19)	3.12 (1.17)	.70	0.01
Positive affect	3.55 (0.96)	3.56 (0.97)	3.54 (0.96)	.58	0.02
Negative affect	2.52 (0.98)	2.50 (0.96)	2.54 (0.99)	.37	0.04
Body image self-efficacy	64.73 (22.18)	65.71 (22.04)	63.75 (22.28)	.07	0.08

[a]Test statistic: chi-square for nominal variables and 2-tailed *t* tests for continuous variables.

[b]ES: effect size (phi coefficient for nominal variables and Cohen *d* for continuous variables).

Attrition, Adherence, and Acceptability

The participant flow diagram is shown in Figure 2. Of the 1715 participants, 798 (46.53%) provided postintervention data, 580 (33.82%) provided 1-week follow-up data, and 459 (26.76%) provided 1-month follow-up data. The intervention group had significantly higher drop-out rates at the postintervention assessment point, relative to the control condition (503/858, 58.6% vs 414/857, 48.3%; *P*<.001; However, the study condition was not significantly associated with dropout at 1 week (586/858, 68.3% vs 579/857, 64.1%) or 1 month (611/858, 71.2% vs 645/857, 75.3%) follow-up (*P* values >.05). Participants who dropped out at the primary time point (after the intervention time frame) were compared with those who completed the baseline variables (Table 2). Dropouts reported significantly lower weight esteem scores at baseline than completers. Notable differences also existed in age (those aged 18 years were the most likely to drop out), ethnicity (Indigenous participants were the most likely to drop out), and region (those from the central west region were the least likely to drop out).

Of the 858 participants randomized into the intervention group, 327 (38.1%) entered the chatbot, and of those 327 participants, 258 (78.9%) completed 1 or more microintervention techniques, thus meeting treatment adherence. On average, participants completed 5 techniques over the 72-hour intervention period, with a minimum of 1 and a maximum of 17 completed across participants. Finally, most participants (251/327, 77%) selected Dandara as their avatar, and 155 of the postintervention responses included acceptability data, with *Topity* receiving an overall score of 6.07 out of 7.

Figure 2. Research design and participant flow using CONSORT (Consolidated Standards of Reporting Trials) eHealth guidelines.

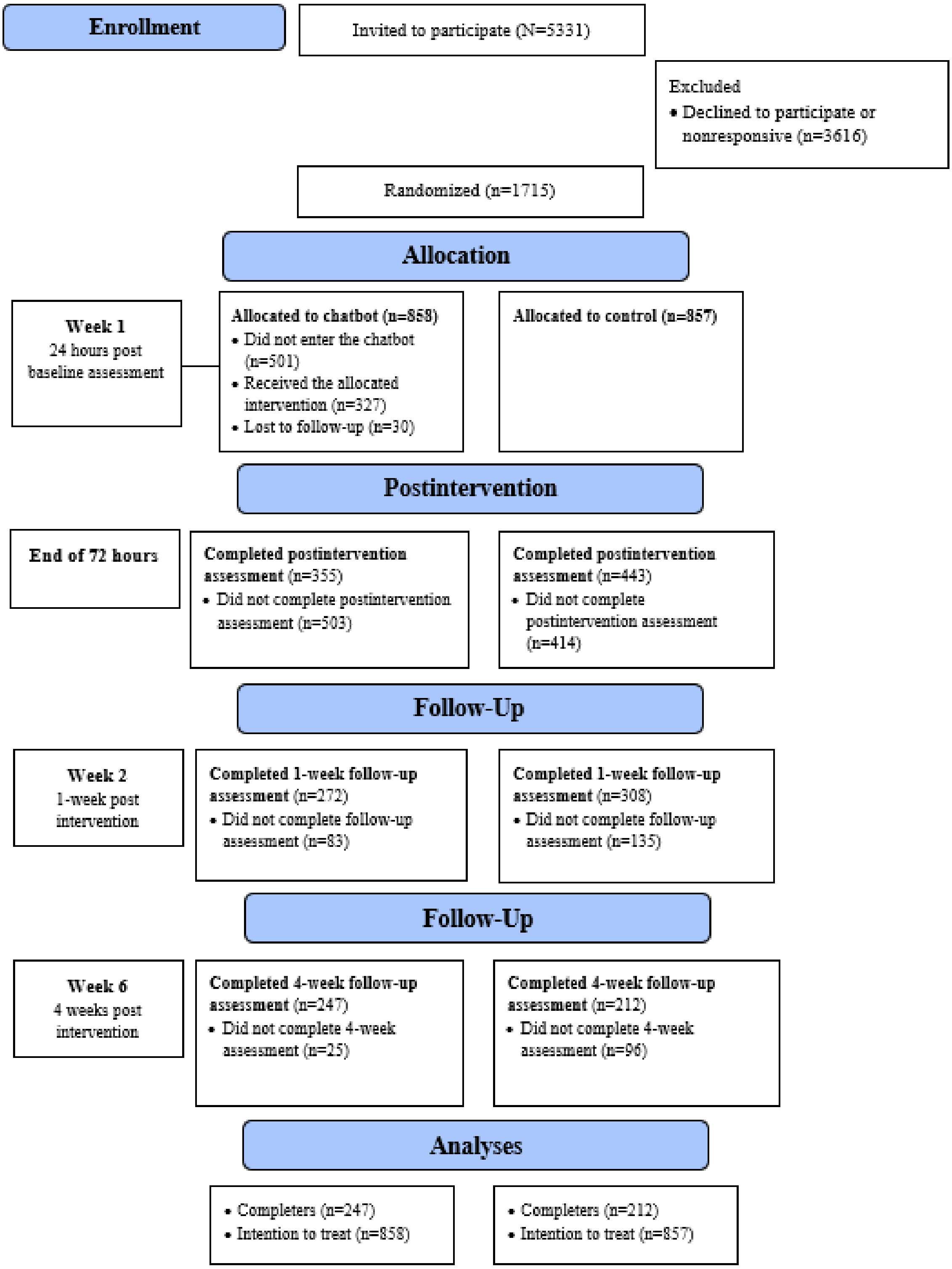


Table 2. Comparisons of dropouts and completers after the intervention time frame on baseline variables.

Variables[a]	Dropout T2[b]	Completer T2	*P* value	ES[c]
Gender, n (%)			.32	0.03
Boy (n=801)	442 (55.2)	359 (44.8)		
Girl (n=901)	467 (51.8)	434 (48.2)		
Gender diverse (n=13)	8 (61.5)	5 (38.5)		
Age (years), n (%)			<.001	0.30
13 (n=276)	118[d] (42.8)	158[d] (57.2)		
14 (n=237)	107[d] (45.1)	130[d] (54.9)		
15 (n=230)	93[d] (40.4)	137[d] (59.6)		
16 (n=255)	115[d] (45.1)	140[d] (54.9)		
17 (n=220)	102[d] (46.4)	118[d] (53.6)		
18 (n=497)	382[d] (76.9)	115[d] (23.1)		
Ethnicity, n (%)			.04	0.08
Asian (n=40)	23 (57.5)	17 (42.5)		
Black (n=181)	99 (54.7)	82 (45.3)		
Indigenous (n=14)	12[d] (85.7)	2[d] (14.3)		
Mixed race (n=630)	355 (56.3)	275 (43.7)		
White (n=845)	426[d] (50.4)	419[d] (49.6)		
Other (n=5)	2 (40)	3 (60)		
Region, n (%)			.005	0.09
North (n=92)	57 (62)	35 (38)		
Northeast (n=348)	202 (58)	146 (42)		
Central west (n=134)	58[d] (43.3)	76[d] (56.7)		
Southeast (n=906)	464[d] (51.2)	442[d] (48.8)		
South (n=235)	136 (57.9)	99 (42.1)		
Baseline, mean (SD)				
Appearance positive	3.31 (0.81)	3.36 (0.86)	.22	0.05
Appearance negative	3.37 (1.07)	3.31 (1.08)	.20	0.05
Weight	3.03 (1.17)	3.20 (1.18)	.003	0.14
Positive affect	3.57 (0.98)	3.54 (0.95)	.52	0.03
Negative affect	2.54 (0.98)	2.50 (0.98)	.41	0.04
Body image self-efficacy	64.32 (22.15)	65.21 (22.21)	.41	0.04

[a]Test statistic: chi-square for nominal variables and 2-tailed *t* tests for continuous variables.

[b]T2 refers to the postintervention time point.

[c]ES: effect size (phi coefficient for nominal variables and Cohen *d* for continuous variables).

[d]When groups significantly differed in a pairwise Bonferroni corrected significance test.

Main Statistical Analyses

Hypothesis 1

Hypothesis one tested whether engagement with the chatbot produced immediate increases in the state of body satisfaction and positive affect. The results revealed significant main effects of time on satisfaction (unstandardized coefficient, *b*=0.60, 95% CI 0.50-0.70; *P*<.001; Cohen *d*=0.30) and affect (*b*=0.51, 95% CI 0.41-0.61; *P*<.001; Cohen *d*=0.28), indicating that participants reported momentary improvements in these outcomes immediately following exposure to chatbot microintervention techniques.

Hypothesis 2

Overview

Hypothesis 2 tested the differences in primary and secondary trait outcomes immediately after the intervention and at the 2 follow-up points. The results of these analyses are presented in Table 3.

Table 3. Means, SDs, and change scores on outcomes between study groups.

Outcomes	Study group				Difference in change score	Cohen *d*	*P* value (2-tailed)
	Control		Experimental		Experimental-control[a]		
	Values, n	Values, mean (SD)[b]	Values, n	Values, mean (SD)	Values, mean (95% CI)		
Primary outcome							
Appearance positive							
Baseline	857	3.31 (0.83)	858	3.36 (0.85)	N/A[c]	N/A	N/A
Postintervention	443	3.32 (0.86)	355	3.52 (0.84)	0.11 (0.02 to 0.19)	0.13	.02
1-week follow-up	308	3.39 (0.88)	272	3.67 (0.85)	0.09 (0.01 to 0.18)	0.1	.03
1-month follow-up	246	3.35 (0.88)	213	3.74 (0.81)	0.22 (0.13 to 0.32)	0.26	<.001
Appearance negative							
Baseline	857	3.34 (1.09)	858	3.36 (1.08)	N/A	N/A	N/A
Postintervention	443	3.29 (1.09)	355	3.46 (1.07)	0.12 (0.00 to 0.24)	0.11	.047
1-week follow-up	308	3.31 (1.11)	272	3.58 (1.12)	0.15 (0.04 to 0.27)	0.13	.01
1-month follow-up	246	3.32 (1.11)	213	3.6 (1.03)	0.15 (0.02 to 0.28)	0.13	.02
Weight							
Baseline	857	3.12 (1.17)	858	3.1 (1.19)	N/A	N/A	N/A
Postintervention	443	3.23 (1.12)	355	3.35 (1.1)	0.02 (−0.10 to 0.13)	0.01	.70
1-week follow-up	308	3.28 (1.1)	272	3.54 (1.12)	0.1 (−0.02 to 0.22)	0.08	.10
1-month follow-up	246	3.33 (1.1)	213	3.65 (1.02)	0.19 (0.07 to 0.31)	0.16	.003
Secondary outcome							
Positive affect							
Baseline	857	3.54 (0.96)	858	3.57 (0.98)	N/A	N/A	N/A
Postintervention	443	3.55 (0.94)	355	3.72 (0.93)	0.09 (−0.03 to 0.19)	0.09	.17
1-week follow-up	308	3.56 (0.98)	272	3.82 (1.01)	0.15 (0.03 to 0.27)	0.15	.02
1-month follow-up	246	3.6 (0.96)	213	3.93 (0.91)	0.23 (0.08 to 0.37)	0.23	.002
Negative affect							
Baseline	857	2.54 (0.99)	858	2.5 (0.97)	N/A	N/A	N/A
Postintervention	443	2.45 (1.01)	355	2.34 (1.03)	−0.07 (−0.20 to 0.07)	−0.07	.30
1-week follow-up	308	2.42 (1.06)	272	2.23 (1.05)	−0.16 (−0.30 to −0.02)	−0.16	.03
1-month follow-up	246	2.37 (1.02)	213	2.15 (1.01)	−0.18 (−0.33 to −0.03)	−0.18	.02
Body image self-efficacy							
Baseline	857	63.76 (22.28)	858	65.72 (22.05)	N/A	N/A	N/A
Postintervention	443	65.92 (22.87)	355	70.44 (21.7)	2.18 (−0.54 to 4.90)	0.09	.12
1-week follow-up	308	67.27 (23.36)	272	73.37 (21.41)	3.13 (0.63 to 5.63)	0.14	.02
1-month follow-up	246	67.83 (24.29)	213	76.43 (20.21)	4.37 (1.70 to 7.04)	0.19	.002

[a]Experimental-control: calculation of the mean difference between groups by subtracting the control group's mean from the experimental group's mean.

[b]Mean (SD) values are based on nonimputed data; mean differences and effect sizes were derived from the intention-to-treat analysis.

[c]N/A: not applicable.

After the Intervention Time Frame

For the primary outcome of body esteem, significant mean differences were observed for the appearance positive (Cohen d=0.13; P=.02) and negative (Cohen d=0.11; P=.047) subscales, in favor of the intervention group experiencing improvements. A significant mean difference was not observed for the weight subscale (Cohen d=0.01; P=.75). Nonsignificant mean differences between the 2 groups immediately after the intervention were also observed for the secondary outcomes of positive affect (Cohen d=0.09), negative affect (Cohen d=−0.07), and body image self-efficacy (Cohen d=0.14).

Follow-Up

As seen in Table 3, significant mean differences were observed at both the 1-week and 1-month follow-up on all primary and secondary outcomes, except for the weight subscale at the 1-week follow-up. In all cases, the intervention group experienced greater improvements in these constructs than the control group. Effect sizes were small, ranging from Cohen d=0.10 to Cohen d=0.26.

Hypothesis 3

Hypothesis 3 tested whether mean differences in trait outcomes at each time point were moderated by participant gender and baseline severity. The results of these analyses are shown in Tables S1 and S2 in Multimedia Appendix 1.

Participant gender was not a notable moderator for any outcome variable immediately after the intervention time frame or at any of the 2 follow-up periods. However, there is some evidence to suggest that baseline levels of a particular outcome significantly affect responsiveness to the intervention. In particular, those with lower body esteem (on all 3 subscales), lower positive affect, and lower body image self-efficacy experienced greater intervention benefits than those with higher levels on these outcome variables. This occurred at all 3 time points for the appearance positive subscale, at the 1-week and 1-month follow-ups for the appearance negative and weight subscales and body image self-efficacy, and only at the 1-month follow-up for positive affect.

Hypothesis 4

Hypothesis 4 tested whether there were any relationships between the 4 indices of chatbot use and the level of improvement in the primary and secondary outcomes. The results of these analyses are presented in Table S3 in Multimedia Appendix 1. We did not observe any notable relationships between the levels of chatbot use and the primary and secondary outcomes immediately after the intervention time frame and follow-up periods.

Discussion

Principal Findings

This 2-armed, fully remote RCT was the first to assess body image chatbot among Brazilian adolescents. These findings indicate that a chatbot containing microinterventions is an effective approach for eliciting small notable improvements in state- and trait-based outcomes for body image and associated well-being constructs. Girls and boys aged 13 to 18 years experienced small but substantial improvements in state body satisfaction and positive affect immediately following engagement with *Topity* techniques. Users also reported small notable improvements in trait-based body esteem, positive and negative affect, and body image self-efficacy relative to the assessment-only control condition. These group differences were observed at all 3 time points for the appearance positive and negative subscales of the Body Esteem Scale for Adolescents and Adults, and at 1-week and 1-month follow-ups for positive and negative affect and body image self-efficacy. Group differences in the weight subscale of the Body Esteem Scale for Adolescents and Adults were only observed at 1-month, in favor of the intervention group experiencing improvements. The intervention effects were comparable across girls and boys; therefore, they were not moderated by gender. However, the effects were moderated by baseline concerns, with those reporting lower baseline body esteem, positive affect, and self-efficacy experiencing greater intervention benefits than those with lower levels of concern. Finally, no dose effects were observed (eg, greater engagement was not associated with greater improvement).

Comparison With Previous Work

This is the third RCT to report small immediate and sustained improvements in young people's state- and trait-based body image following engagement with a microintervention [22,23]. These small but notable effects across all 3 trials are likely owing to both the brevity of the intervention phases (ie, single sessions; 72 hours) and the use of universal samples (ie, varying degrees of body image concerns). Larger and more sustained effects have been observed in microintervention [39] and chatbot [27] studies with longer intervention phases (ie, 3 weeks [39]; 1 month [27]) and selected samples (eg, women with heighted weight and shape concerns), which is likely owing to the greater scope for symptom change. The current moderator analyses mirror these selected sample effects, with intervention benefits being the largest among young people with poor baseline levels of body image, affect, and self-efficacy. Notably, improvements in secondary outcomes (ie, positive and negative affect and body image self-efficacy) were not observed immediately after the intervention but emerged at 1-week and 1-month follow-ups. This pattern may be attributed to the specificity of the intervention content and mediating effect of improved body image on affect and self-efficacy, that is, body image has shown to be predictive of mood states [40], and participants may need to experience improved body image before thinking and feeling that they are self-efficacious in managing their body image concerns.

With respect to intervention adherence and user engagement, less than half (327/858, 44% of participants) of those randomized into the intervention group used the chatbot. Of these, the majority (258/327, 79% of participants) completed the minimum intervention dose of one microintervention technique, with an average of 5 techniques completed over the 72-hour intervention period. There was also a preference for the woman avatar, Dandara, with most girls and boys opting to receive guidance from her, relative to her counterpart, Gabriel. These engagement levels [27,39] and gender preferences [41] mirror previous body image research.

First, with respect to engagement levels, in 2 comparable microintervention [39] and chatbot [27] trials, women completed an average of 4 techniques over 21-day and 1-month periods, respectively. Although the number of techniques is comparable across the 3 trials, the current participants completed more techniques over a shorter period. More research is needed on the naturalistic and longitudinal use of microinterventions and chatbots to better understand how and when users engage with the content, particularly when an intervention time frame is not prescribed. Relatedly, this trial did not find support for dose response, with intervention effects unaffected by the number of techniques completed. To our knowledge, only one other microintervention study has considered dose response and found no support for this relationship [42].

Second, with respect to gender preferences of the avatar, a small body of research has examined the role of gender in body image interventions, particularly the gender of the intervention facilitator [41]. Specifically, adult women preferred when interventions were delivered by a woman, whereas men did not have a preference and were content with either a man or a woman. The current findings reflect this pattern among adolescents, with most girls (152/174, 87%) selecting Dandara and boys selecting a combination of Dandara (94/147, 64%) and Gabriel (53/147, 36%). Overall, there is limited research on adolescents' chatbot intervention preferences, including avatar demographics (eg, age, gender, sexuality, and ethnicity), hosting platforms (eg, Facebook and WhatsApp), and chat functions (eg, predefined rules, machine learning, or artificial intelligence). These features require exploration with the end user during intervention development, to ensure an acceptable, feasible, and safe intervention is created.

The current attrition rates and patterns are also comparable with the abovementioned trials [27,39] and broader digital mental health research [43]. First, with respect to condition-specific dropouts, rates were higher in the current intervention condition relative to the control, a pattern that is likely owing to higher participant burden among this group. Second, with respect to overall attrition, rates were higher but still reflective of digital mental health interventions, including body- and eating-related interventions [44] and chatbots [45]. The current rates are likely explained by several methodological features known to exacerbate attrition in digital trials, including a burdensome recruitment and onboarding process, no contact or follow-up calls with a researcher, and the masking of the compensation amount [43]. These features are discussed in this section.

First, Brazil passed the General Data Protection Law in 2018, with policies coming into effect in February 2020. These laws require parents to read, sign, and upload information and consent forms to a secure portal for identity verification before their child's participation. This process requires parents to have good literacy skills, and without literacy skills, they are unable to engage with and comprehend the research materials, thus restricting adolescents' participation. The recruitment methods were adjusted based on stakeholder feedback, with an information video provided alongside the written content. However, parents were still required to sign and upload consent forms. Second, to streamline the recruitment and onboarding processes, communication with participants was conducted primarily through automated processes (eg, email), with little to no contact between participants and the researchers or a recruitment agency. Third, during the informed consent phase, participants were advised that they would receive compensation at study completion; however, according to Brazilian ethics, the amount was not disclosed until study completion (eg, at the 1-month follow-up). Collectively, these methodological features and the general nature of digital interventions likely explain the current attrition rates.

Overall, previous and present findings indicate that microinterventions are an effective intervention for young people; however, they may serve different purposes depending on the severity of a participant's concern. From a stepped-care approach, microinterventions may serve as a stand-alone approach for young people with milder concerns (eg, offered to young people during a moment of body image distress) or as an adjunctive approach to those participating in longer traditional body image interventions (eg, offered to young people ahead of talking therapy to increase motivation and self-efficacy). Unfortunately, to date, a stepped-care model for the eHealth of body and eating concerns has not been formally conceptualized and warrants consideration. Finally, given the growing literature on microinterventions, a systematic review and meta-analysis of this intervention model are both timely and necessary. Specifically, the review should provide a comprehensive overview of extant mental health microinterventions and analyze different intervention and trial features to determine those associated with greater engagement, adherence, and effectiveness.

Limitations, Strengths, and Future Impact

This trial is not without limitations, most of which speak to the digital and remote nature of the trial, which were compounded by the global COVID-19 pandemic. As noted, this trial comprised several methodological features known to impact attrition (eg, a burdensome recruitment and onboarding process, no contact or follow-up calls with a researcher; the masking of the compensation amount). Furthermore, the trial was conducted during the peak of the COVID-19 pandemic in Brazil (ie, 92,625 cases per day) [46]. Notably, a high level of screen and social media fatigue was reported among adolescents during the pandemic [38], which is largely because of the exponential reliance on technology during this socially restrictive period (ie, web-based schooling and entertainment [eg, social media and streaming services] and video calls with friends and family). Moreover, numerous web-based mental health resources were developed, tested, or made available to young people during this time. Therefore, although body and eating concerns were equally problematic during the pandemic [8], it is possible that this intervention and research opportunity were dismissed because of the overwhelming digital demands of adolescents during this time. Finally, dropout was affected by participants' age (those aged 18 years were the most likely to drop out), ethnicity (Indigenous participants were the most likely to drop out), and region (those from the central west region were the least likely to drop out). This suggests poor acceptance among these demographics and warrants further exploration.

Despite these limitations, the strengths and findings associated with this trial provide avenues for advancing this research field beyond those previously discussed. First, *Topity* has reached >40,000 young people since the RCT began in April 2021. Second, the digital infrastructure developed for this chatbot can be implemented and disseminated in other countries, with the content adapted for suitability in these contexts. This may include translating the content into different languages and incorporating intervention stimuli that are appropriate and salient to a particular country and culture. Next, this trial provides further support for the adaptation of traditional prevention and intervention techniques (eg, hard copy and in-person) for use in digital environments [47]. More broadly, mental health researchers are encouraged to examine existing evidence-based approaches and identify techniques that could be adapted for stand-alone use in digital environments. Isolating these techniques for self-guided use in digital settings that are already frequented by the consumer is likely to increase the accessibility, acceptability, and scalability of mental health resources and, in turn, lead to impactful and sustainable change.

Conclusions

This trial supports the use of chatbot technology to deliver mental health microinterventions within digital environments frequented by young people (eg, social media platforms and messaging apps). Microinterventions are proving to be an effective method for providing adolescents with immediate and short-term symptom relief and reducing imbalances in the ratio of harmful and helpful body image content on social media platforms. Although microinterventions are a promising intervention model, more research is needed to conceptualize how this model can be integrated into and enhance a stepped-care model for digital approaches targeting body and eating concerns.

Acknowledgments

The authors would like to thank our project partners at the Dove Self-Esteem Project, United Nations International Children’s Emergency Fund (UNICEF) Brazil, UNICEF UK, Talk2U, and Toluna for their expertise and collaborative partnership. The views and opinions expressed in this manuscript are those of all the authors and do not necessarily represent these partners. This study was funded by a research grant from the Dove Self-Esteem Project (Unilever). The funders had no role in data analysis, decision to publish, or manuscript preparation. The funders were permitted to review the manuscript, but the authors exclusively retained the final decision on content. Upon review, the funders did not suggest changes. The views expressed are those of the authors and not necessarily those of the funder.

Data Availability

The body image chatbot, *Topity*, is freely available to those with access to *Facebook Messenger*. Anonymized survey data will be available on reasonable request for noncommercial purposes.

Authors' Contributions

ELM, HGS, ACSA, JFFM, MCA, JL, MFT, and PCD were responsible for the formulation or evolution of overarching research goals and aims. ELM and PCD were responsible for the development and design of methodology. ELM and HGS were responsible for the provision of study materials. HGS, JL, and MFT accessed and verified underlying data. ELM was responsible for the preparation, creation, and presentation of the published work, specifically writing the initial draft, visualization, and data presentation. HGS, ACSA, JFFM, MCA, JL, MFT, and PCD were responsible for preparation, creation, and presentation of the published work by those from the original research group, specifically critical review, commentary, or revision including the pre- or postpublication stage. ELM and HGS were responsible for management and coordination responsibility for the research activity planning and execution. JL and MFT were responsible for application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data. ELM and PCD were responsible for oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team, and the acquisition of the financial support for the project leading to this publication.

Conflicts of Interest

JL was supported by the National Health and Medical Research Council Investigator Grant (APP1196948). PCD is an independent consultant to the mental health policy and programing team on Instagram (owned by Meta, the parent company of Facebook Messenger) and Dove (Unilever). All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplementary tables pertaining to results of hypotheses 3 and 4.
[DOCX File , 41 KB - mhealth_v11i1e39934_app1.docx]

Multimedia Appendix 2
CONSORT E-HEALTH Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 98 KB - mhealth_v11i1e39934_app2.pdf]

References

1. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
2. FUKUDA CC, PENSO MA, AMPARO DM, ALMEIDA BC, MORAIS CD. Mental health of young Brazilians: barriers to professional help-seeking. Estud psicol (Campinas) 2016 Jun;33(2):355-365. [doi: 10.1590/1982-02752016000200017]
3. Cetic B. Executive Summary ICT Kids Online Brazil Survey. Centro Regional para o Desenvolvimento da Sociedade da Informação. URL: https://cetic.br/media/docs/publicacoes/2/20221121121039/executive_summary_ict_kids_online_2021.pdf [accessed 2021-12-15]
4. Malheiros LE, da Costa BG, Lopes MV, Chaput J, Silva KS. Association between physical activity, screen time activities, diet patterns and daytime sleepiness in a sample of Brazilian adolescents. Sleep Med 2021 Feb;78:1-6. [doi: 10.1016/j.sleep.2020.12.004] [Medline: 33370617]
5. Andrade AL, Scatena A, Bedendo A, Enumo SR, Dellazzana-Zanon LL, Prebianchi HB, et al. Findings on the relationship between internet addiction and psychological symptoms in Brazilian adults. Int J Psychol 2020 Dec 07;55(6):941-950. [doi: 10.1002/ijop.12670] [Medline: 32266726]
6. Deloitte Access Economics. The social and economic cost of eating disorders in the United States of America: a report for the strategic training initiative for the prevention of eating disorders and the academy for eating disorders. Deloitte Access Economics. 2020 Jun. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1267/2020/07/Social-Economic-Cost-of-Eating-Disorders-in-US.pdf [accessed 2021-09-21]
7. The real cost of beauty ideals the staggering economic and social cost of body dissatisfaction and appearance-based discrimination in the United States. Deloitte Access Economics. 2022 Oct. URL: https://www.hsph.harvard.edu/striped/wp-content/uploads/sites/1267/2022/10/Real-Cost-of-Beauty-Report-10-4-22.pdf [accessed 2022-12-16]
8. Rodgers RF, Lombardo C, Cerolini S, Franko DL, Omori M, Fuller-Tyszkiewicz M, et al. The impact of the COVID-19 pandemic on eating disorder risk and symptoms. Int J Eat Disord 2020 Jul;53(7):1166-1170 [FREE Full text] [doi: 10.1002/eat.23318] [Medline: 32476175]
9. Swami V. Cultural influences on body size ideals. Eur Psychologist 2015 Jan 01;20(1):44-51. [doi: 10.1027/1016-9040/a000150]
10. Andersen N, Swami V. Science mapping research on body image: a bibliometric review of publications in Body Image, 2004-2020. Body Image 2021 Sep;38:106-119. [doi: 10.1016/j.bodyim.2021.03.015] [Medline: 33838539]
11. Amaral AC, Ferreira ME. Body dissatisfaction and associated factors among Brazilian adolescents: a longitudinal study. Body Image 2017 Sep;22:32-38. [doi: 10.1016/j.bodyim.2017.04.006] [Medline: 28570920]
12. Leal GV, Philippi ST, Alvarenga MD. Unhealthy weight control behaviors, disordered eating, and body image dissatisfaction in adolescents from São Paulo, Brazil. Braz J Psychiatry 2020 Jun;42(3):264-270 [FREE Full text] [doi: 10.1590/1516-4446-2019-0437] [Medline: 32022158]
13. Amaral AC, Stice E, Ferreira ME. A controlled trial of a dissonance-based eating disorders prevention program with Brazilian girls. Psicol Reflex Crit 2019 Jun 17;32(1):13 [FREE Full text] [doi: 10.1186/s41155-019-0126-3] [Medline: 32026167]
14. Dunker KL, Claudino AM. Preventing weight-related problems among adolescent girls: a cluster randomized trial comparing the Brazilian 'New Moves' program versus observation. Obes Res Clin Pract 2018 Jan;12(1):102-115. [doi: 10.1016/j.orcp.2017.07.004] [Medline: 28797704]
15. Doniec K, Dall'Alba R, King L. Brazil's health catastrophe in the making. Lancet 2018 Sep;392(10149):731-732 [FREE Full text] [doi: 10.1016/s0140-6736(18)30853-5]
16. Aufderheide P. Media Literacy. A Report of the National Leadership Conference on Media Literacy. Queenstown, Maryland: The Aspen Institute; 1993.
17. Cash T. Body Image: A Handbook of Theory, Research, and Clinical Practice. New York: Guildford Press; 2002.
18. Menzel J, Levine M. Embodying experiences and the promotion of positive body image: the example of competitive athletics. In: Self-objectification in Women: Causes, Consequences, and Counteractions. Washington, D.C., United States: American Psychological Association; 2011.
19. Baumel A, Fleming T, Schueller SM. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. J Med Internet Res 2020 Oct 29;22(10):e20631 [FREE Full text] [doi: 10.2196/20631] [Medline: 33118946]
20. Atkinson MJ, Diedrichs PC. Examining the efficacy of video-based microinterventions for improving risk and protective factors for disordered eating among young adult women. Int J Eat Disord 2021 May 07;54(5):708-720. [doi: 10.1002/eat.23460] [Medline: 33415790]
21. Elefant AB, Contreras O, Muñoz RF, Bunge EL, Leykin Y. Microinterventions produce immediate but not lasting benefits in mood and distress. Internet Interv 2017 Dec;10:17-22 [FREE Full text] [doi: 10.1016/j.invent.2017.08.004] [Medline: 29270366]

22. Matheson EL, Lewis-Smith H, Diedrichs PC. The effectiveness of brief animated films as a scalable micro-intervention to improve children's body image: a randomised controlled trial. Body Image 2020 Dec;35:142-153. [doi: 10.1016/j.bodyim.2020.08.015] [Medline: 33049455]
23. Matheson EL, Smith HG, Lewis-Smith H, Arbon RE, Diedrichs PC. Game on! A randomised controlled trial evaluation of playable technology in improving body satisfaction and negative affect among adolescents. New Media Soc 2021 Mar 09;24(12):2635-2658. [doi: 10.1177/1461444821993800]
24. Yim SH, Schmidt U. Experiences of computer-based and conventional self-help interventions for eating disorders: a systematic review and meta-synthesis of qualitative research. Int J Eat Disord 2019 Oct 25;52(10):1108-1124. [doi: 10.1002/eat.23142] [Medline: 31343088]
25. Matheson EL, Smith HG, Amaral AC, Meireles JF, Almeida MC, Mora G, et al. Improving body image at scale among Brazilian adolescents: study protocol for the co-creation and randomised trial evaluation of a chatbot intervention. BMC Public Health 2021 Nov 20;21(1):2135 [FREE Full text] [doi: 10.1186/s12889-021-12129-1] [Medline: 34801002]
26. Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G. Development of a positive body image chatbot (KIT) with young people and parents/carers: qualitative focus group study. J Med Internet Res 2021 Jun 16;23(6):e27807 [FREE Full text] [doi: 10.2196/27807] [Medline: 34132644]
27. Fitzsimmons-Craft EE, Chan WW, Smith AC, Firebaugh M, Fowler LA, Topooco N, et al. Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. Int J Eat Disord 2022 Mar;55(3):343-353. [doi: 10.1002/eat.23662] [Medline: 35274362]
28. Shah J, DePietro B, D'Adamo L, Firebaugh M, Laing O, Fowler LA, et al. Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. Int J Eat Disord 2022 Sep 18;55(9):1229-1244 [FREE Full text] [doi: 10.1002/eat.23798] [Medline: 36056648]
29. Fitzsimmons-Craft EE. Developing an optimized conversational agent or 'chatbot' to facilitate mental health services use in individuals with eating disorders. National Institutes of Health. URL: https://grantome.com/grant/NIH/K08-MH120341-01 [accessed 2022-12-15]
30. Diedrichs PC, Atkinson MJ, Steer RJ, Garbett KM, Rumsey N, Halliwell E. Effectiveness of a brief school-based body image intervention 'Dove Confident Me: Single Session' when delivered by teachers and researchers: results from a cluster randomised controlled trial. Behav Res Ther 2015 Nov;74:94-104 [FREE Full text] [doi: 10.1016/j.brat.2015.09.004] [Medline: 26469131]
31. Alleva JM, Sheeran P, Webb TL, Martijn C, Miles E. A meta-analytic review of stand-alone interventions to improve body image. PLoS One 2015;10(9):e0139177 [FREE Full text] [doi: 10.1371/journal.pone.0139177] [Medline: 26418470]
32. The impact of an artificial intelligence chatbot on Brazilian adolescents' body image. University of the West of England. 2022. URL: https://clinicaltrials.gov/ct2/show/NCT04825184 [accessed 2022-12-15]
33. Smith HG, Garbett KM, Matheson EL, Amaral AC, Meireles JF, Almeida MC, et al. The body esteem scale for adults and adolescents: translation, adaptation and psychometric validation among Brazilian adolescents. Body Image 2022 Sep;42:213-221. [doi: 10.1016/j.bodyim.2022.05.012] [Medline: 35779360]
34. Damásio B, Pacico J, Poletto M, Koller S. Refinement and psychometric properties of the eight-item Brazilian positive and negative affective schedule for children (PANAS-C8). J Happiness Stud 2012 Sep 17;14(4):1363-1378 [FREE Full text] [doi: 10.1007/s10902-012-9383-x]
35. Bandura A. Self-efficacy Beliefs of Adolescents. Charlotte, North Carolina, United States: Information Age Publishing, Incorporated; 2006.
36. McCusker J, Lambert S, Cole M, Ciampi A, Strumpf E, Freeman E, et al. Activation and self-efficacy in a randomized trial of a depression self-care intervention. Health Educ Behav 2016 Dec;43(6):716-725 [FREE Full text] [doi: 10.1177/1090198116637601] [Medline: 27179288]
37. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct 01;21(10):1644-1651. [doi: 10.2337/diacare.21.10.1644] [Medline: 9773724]
38. Pandya A, Lodha P. Social connectedness, excessive screen time during COVID-19 and mental health: a review of current evidence. Front Hum Dyn 2021 Jul 22;3. [doi: 10.3389/fhumd.2021.684137]
39. Fuller-Tyszkiewicz M, Richardson B, Lewis V, Linardon J, Mills J, Juknaitis K, et al. A randomized trial exploring mindfulness and gratitude exercises as eHealth-based micro-interventions for improving body satisfaction. Comput Human Behav 2019 Jun;95:58-65. [doi: 10.1016/j.chb.2019.01.028]
40. Bornioli A, Lewis-Smith H, Slater A, Bray I. Body dissatisfaction predicts the onset of depression among adolescent females and males: a prospective study. J Epidemiol Community Health 2020 Dec 07;75(4):343-348. [doi: 10.1136/jech-2019-213033] [Medline: 33288655]
41. Yager Z, Diedrichs P, Drummond M. Understanding the role of gender in body image research settings: participant gender preferences for researchers and co-participants in interviews, focus groups and interventions. Body Image 2013 Sep;10(4):574-582 [FREE Full text] [doi: 10.1016/j.bodyim.2013.06.004] [Medline: 23876877]
42. Lokman S, Leone S, Sommers-Spijkerman M, van der Poel A, Smit F, Boon B. Complaint-directed mini-interventions for depressive complaints: a randomized controlled trial of unguided web-based self-help interventions. J Med Internet Res 2017 Jan 04;19(1):e4 [FREE Full text] [doi: 10.2196/jmir.6581] [Medline: 28052840]

43. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
44. Linardon J, Shatte A, Messer M, Firth J, Fuller-Tyszkiewicz M. E-mental health interventions for the treatment and prevention of eating disorders: an updated systematic review and meta-analysis. J Consult Clin Psychol 2020 Nov;88(11):994-1007. [doi: 10.1037/ccp0000575] [Medline: 32852971]
45. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
46. Brazil: the latest coronavirus counts, charts and maps. Reuters COVID-19 Tracker. URL: https://graphics.reuters.com/world-coronavirus-tracker-and-maps/countries-and-territories/brazil/ [accessed 2021-09-21]
47. Wade T, Wilksch S. Internet eating disorder prevention. Curr Opin Psychiatry 2018 Nov;31(6):456-461. [doi: 10.1097/YCO.0000000000000450] [Medline: 30234524]

Abbreviations

LMIC: low- and middle-income country
RCT: randomized controlled trial

Edited by L Buis; submitted 02.06.22; peer-reviewed by J Pry, Anonymous, K Burnell; comments to author 15.12.22; revised version received 21.12.22; accepted 10.04.23; published 19.06.23.

Please cite as:
Matheson EL, Smith HG, Amaral ACS, Meireles JFF, Almeida MC, Linardon J, Fuller-Tyszkiewicz M, Diedrichs PC
Using Chatbot Technology to Improve Brazilian Adolescents' Body Image and Mental Health at Scale: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e39934
URL: https://mhealth.jmir.org/2023/1/e39934
doi:10.2196/39934
PMID:37335604

Original Paper

Recommendations for the Quality Management of Patient-Generated Health Data in Remote Patient Monitoring: Mixed Methods Study

Robab Abdolkhani[1,2], PhD; Kathleen Gray[1], PhD; Ann Borda[3], PhD; Ruth DeSouza[4], PhD

[1]Centre for Digital Transformation of Health, The University of Melbourne, Melbourne, Australia

[2]Department of General Practice, Melbourne Medical School, The University of Melbourne, Melbourne, Australia

[3]Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

[4]School of Art, Royal Melbourne Institue of Technology University, Melbourne, Australia

Corresponding Author:
Robab Abdolkhani, PhD
Centre for Digital Transformation of Health
The University of Melbourne
Level 13
305 Grattan Street
Melbourne, 3000
Australia
Phone: 61 90358703 ext 03
Email: Robab.abdolkhani@unimelb.edu.au

Abstract

Background: Patient-generated health data (PGHD) collected from innovative wearables are enabling health care to shift to outside clinical settings through remote patient monitoring (RPM) initiatives. However, PGHD are collected continuously under the patient's responsibility in rapidly changing circumstances during the patient's daily life. This poses risks to the quality of PGHD and, in turn, reduces their trustworthiness and fitness for use in clinical practice.

Objective: Using a sociotechnical health informatics lens, we developed a data quality management (DQM) guideline for PGHD captured from wearable devices used in RPM with the objective of investigating how DQM principles can be applied to ensure that PGHD can reliably inform clinical decision-making in RPM.

Methods: First, clinicians, health information specialists, and MedTech industry representatives with experience in RPM were interviewed to identify DQM challenges. Second, these stakeholder groups were joined by patient representatives in a workshop to co-design potential solutions to meet the expectations of all the stakeholders. Third, the findings, along with the literature and policy review results, were interpreted to construct a guideline. Finally, we validated the guideline through a Delphi survey of international health informatics and health information management experts.

Results: The guideline constructed in this study comprised 19 recommendations across 7 aspects of DQM. It explicitly addressed the needs of patients and clinicians but implied that there must be collaboration among all stakeholders to meet these needs.

Conclusions: The increasing proliferation of PGHD from wearables in RPM requires a systematic approach to DQM so that these data can be reliably used in clinical care. The developed guideline is an important next step toward safe RPM.

(JMIR Mhealth Uhealth 2023;11:e35917) doi:10.2196/35917

KEYWORDS

data quality management; patient-generated health data; remote patient monitoring; wearable electronic devices; remote sensing technology; telemedicine; big data

Introduction

Remote Patient Monitoring

The use of remote patient monitoring (RPM) solutions and production of patient-generated health data (PGHD) to enable continuous monitoring of patients outside clinical settings are increasing with the growing availability of health wearable devices and the connected mobile apps and web portals [1]. The COVID-19 pandemic has accelerated the use of RPM to monitor mild cases of the disease remotely, given the limited capacity of acute care facilities [2].

As the pandemic is not yet over, RPM will likely contribute more to health care delivery owing to the availability of various affordable technologies and the need for remote treatment and monitoring. However, despite the urgent need and rapid implementation and use of RPM, investigation on how quality PGHD can best be collected and managed to lead to accurate decision-making is still lacking.

Ensuring the Quality of PGHD

Patients may collect some data as instructed by clinicians, mainly from medical wearables. Patients may also collect data, on their own accord or on advice from clinicians, from consumer wearables. There are fundamental similarities between the data collected upon patient initiation and those collected upon clinician initiation, whether from consumer or medical wearables, that erode the regulators' distinctions: the data are generated outside the controlled environment of the clinic; the data collection is the responsibility of the individual wearer; and the data are shared electronically with parties who operate outside a controlled clinical setting, namely wearable companies. Thus, RPM data collected from wearables, whether upon patient initiation or upon clinician initiation, are covered by the broad concept of PGHD.

Outside the clinic, consumer and medical wearable technologies used in RPM capture a large amount of data continuously in rapidly changing circumstances during a patient's daily life under the patient's or caregiver's supervision [3]. The wearable platform includes sensors that capture data automatically and a mobile app and web portal where the person enters data manually. Inside the clinic, RPM solutions are not integrated well into patient records or clinical workflows, and various digital health devices and platforms are used for different RPM purposes [4]. The quality of PGHD collected from disparate devices is compromised by various technical, behavioral, or operational issues that occur during data capture by the patient or caregiver, during the transmission of the data from the patient to the clinician, and during the clinician's review of the data for decision-making [5].

Health data quality plays a vital role in health care systems. Clinicians need to trust the available data to make accurate decisions and provide efficient and timely care for their patients. Data are of good quality when they are fit for their intended use [6], that is, when they are accurate, accessible, consistent, complete, interpretable, timely, relevant, and compliant with the standards defined by health care organizations [7]. Any quality issue with data can affect patient safety, the reimbursement of health services, and the quality of clinical outcomes and other aspects of health care delivery [8].

Data quality management (DQM) refers to the processes of ensuring data quality when data are collected, stored, analyzed, reviewed, and used in clinical decision-making [9]. The core outcome of DQM is establishing the fitness of data for its intended use. National and international health care and health information–related organizations provide guidelines for the quality management of patient data that are generated within clinical settings [8-12]. However, in RPM, data are collected outside the clinical setting, and different stakeholders are involved at different stages of PGHD management both outside and inside the health care settings.

This paper describes the DQM recommendations provided to ensure that data from wearables are fit for use in clinical care.

Methods

Overview

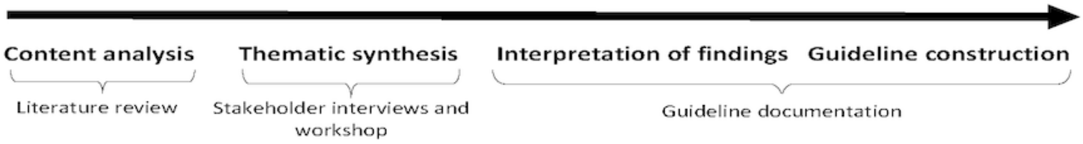
Recommendations for the quality management of PGHD arose from a mixed methods study on the quality management of PGHD from wearables and were constructed following a guideline development convention in health care [13-18] through the stages listed in Textbox 1.

Most of the data collection in this research was done before the COVID-19 pandemic. However, the rapid deployment of RPM during the pandemic emphasizes the need for guidelines, such as the one constructed in this study, to improve the use of RPM initiatives and efficiently integrate them into the routine care.

Textbox 1. Stages involved the construction of the recommendations.

- Evidence reviews: a comprehensive literature review focused on original research within a 10-year time frame that discussed the barriers to and concerns of using patient-generated health data (PGHD) in clinical practice was conducted and published previously [19].
- Stakeholder involvement: in-depth interviews were conducted with PGHD stakeholders directly involved in the remote monitoring of patients with chronic diseases, including those with diabetes, those with cardiac arrythmia, and those sleep disorders, in primary care, secondary care, and tertiary care settings to identify the challenges related to the quality management of PGHD. The interview participants were from Australia, the United Kingdom, and the United States. The interview results were published previously [5]. Then, a participatory workshop was held in Australia with stakeholders, including patients, clinicians, health information professionals, wearable developer companies, PGHD integration service providers and remote patient monitoring (RPM) consultants, to discuss the identified challenges and address potential solutions and stakeholders' needs and expectations. The results of this study were published elsewhere [20].
- Documentation of recommendations: we used the approach of integrating multiple types of qualitative evidence to produce new knowledge [21], as shown in Figure 1, to construct a set of recommendations that cover all aspects of the quality management of PGHD during data flow from the patient to the clinician. We synthesized the findings of the aforementioned 2 stages along with supporting evidence from an updated review of scientific literature and new policies related to PGHD. The interpretation stage aimed to draw connections between the data points, themes, and findings. Then, the construction stage expressed new meaning and uncovered ways of understanding the realities regarding the research topic for the stakeholders. It acknowledged the importance of synthesized and interpreted elements in terms of stakeholders, context, and influencing issues. Moving from evidence to recommendations, the construction stage sought to examine how the synthesized findings related to the broader context in the past, present, and future. Details of the findings and the emerged themes applied in the construction of the guideline are provided in Multimedia Appendix 1. This step produced a guideline containing 19 separate recommendations. One of the researchers interpreted the findings and constructed the guideline, and then the guideline content was reviewed separately by 3 researchers and discussed in multiple meetings.
- Validation of recommendations: a 1-round Delphi method [22] using a web-based survey and 5-point Likert scale was adopted, 14 Australian international health informatics and health information management experts participated in the survey. The interview and workshop studies involved several groups of PGHD stakeholders, who shared their experience and perspectives related to the quality management of PGHD. However, except for health information professionals, the other stakeholders were experts in their own clinical or technical field but not in managing and governing patient health data. The data quality management (DQM) elements in the guideline still required validation by experts with a high-level understanding of and experience with health information management and health informatics principles and practices. These experts were purposefully selected based on their professional reputation and their known interest in PGHD. None of them had been involved in the previous studies of this project, so they could form an independent view of the resulting recommendations. The survey had 19 items in total, representing the 19 recommendations in the guideline. It used a 5-point Likert scale (not important, slightly important, moderately important, important, and very important) to capture the participants' expert opinions about the extent to which each DQM recommendation is potentially important in contributing to the safety and the quality of RPM. Each survey item also included a free-text comment option so that the participants could further explain their response. Consensus on each recommendation for each DQM aspect of PGHD was deemed to be achieved by having 60% of votes fall within 2 adjacent categories of the 5-point scale. A method to group the responses for analysis was determined: if the participants reached at least an aggregated 60% agreement that a recommendation is "important" or "very important," it was deemed to have been rated as essential; if the participants reached at least an aggregated 60% agreement that a recommendation is "slightly important" or "moderately important," it was deemed to have been rated as desirable; and if the participants reached at least an aggregated 60% agreement that a recommendation is "not important," it was deemed to have been rated as unnecessary.

Figure 1. Continuum of integrating multiple qualitative findings to create new evidence.



Ethics Approval

The stakeholder involvement studies received approval from the Human Ethics Advisory Group at the Department of General Practice at the University of Melbourne [5,20]. The ethics approval number for the validation study from the same group is 1955682.1.

Results

Recommendations and Key Themes

The ensuing guideline encompasses 19 recommendations. These recommendations were grouped according to 7 overarching DQM aspects. Table 1 lists these 7 aspects; their adapted definition for this research; and the key themes identified from the literature review, interviews, and workshop studies. The sociotechnical issues to be considered in relation to each DQM aspect have been discussed in the corresponding recommendations. Through this style of presentation, PGHD stakeholders can understand what actions they and others need to take to collect, manage, and use trustworthy PGHD in RPM.

Table 1. Data quality management (DQM) recommendations for patient-generated health data (PGHD) in remote patient monitoring.

DQM aspects and key themes	Recommendations
PGHD accessibility: authorized users of PGHD access t hem across all data management stages	
Patients' and clinicians' access to PGHD	Both raw and processed PGHD from wearables should be accessible to the patient and clinician.
Patients' and clinicians' awareness of PGHD access by others	A mechanism should be available to the patient and clinician to set up notice recurrence on where, when, how, and by whom PGHD from wearables are accessed.
Patients' consent to PGHD access by different clinicians	A mechanism should be available to the patient to change permissions for clinicians to access PGHD from wearables.
PGHD accuracy: error-free data	
Automatic and manual PGHD collection	PGHD should be collected automatically by the wearable device, with as little as possible manual intervention.
PGHD annotation	Annotation function for manually and automatically entered PGHD should be available to the patient and clinician in order to comment on inaccurate data.
Wearable calibration	The wearable should be calibrated automatically as required by the clinical standard of care of diseases.
PGHD completeness: no PGHD are missing	
No active data collection	A protocol should be available to the patient and clinician that defines PGHD "downtime," that is, the time range during which it is acceptable if the wearable is not collecting data.
Resuming PGHD collection after downtime	A protocol should be available to the patient and clinician for resuming PGHD collection when the acceptable downtime period is exceeded.
Context for incomplete PGHD	Annotation function should be available to the patient in order to provide context for any period of missing PGHD.
PGHD consistency: data convey the same meaning no matter whether they are collected from one or different brands of wearables	
PGHD definitions and formats	PGHD from wearables should be collected based on clinically accepted and structured data definitions and standard formats.
PGHD integration with electronic medical records	PGHD from wearables should be integrated into the patient's clinical care record.
PGHD exchange within and outside care settings	PGHD from wearables should be consistently exchanged inside and between clinical settings.
PGHD interoperability: data presentation highlights the key message that is understood by PGHD stakeholders	
PGHD contextualization	PGHD from wearables should be accompanied by contextual data that are clinically important to patient management.
Dynamic and static PGHD visualization	Dynamic visual representation as well as a static snapshot (such as in PDF format) of PGHD from wearables should be available to the patient and clinician.
The patient's understanding of PGHD	Alerts should be sent to the patient during PGHD collection by the wearable when data are outside the acceptable range, accompanied by clinical advice on action to take.
PGHD relevancy: data are pertinent to the standard of care for the condition being monitored	
PGHD relevancy to the standards of care	There should be a shared understanding between the patient and clinician of relevant data for the disease based on the standards of care and make sure that all the relevant data are collected.
PGHD timeliness: availability of up-to-date PGHD for patients and clinicians when needed	
PGHD availability to patients when needed	PGHD from wearables should be available to the patient within a timeframe (continuously to periodically) according to the standards of care of diseases.
PGHD availability to clinicians when needed	PGHD from wearables should be available to the clinician within a timeframe (continuously to periodically) according to the standards of care of diseases.
Time frame for PGHD sharing between patients and clinicians	A timeframe for sharing PGHD from wearables should be available to the patient and clinician.

PGHD Accessibility

PGHD accessibility was characterized by data access methods, privacy protection, and data ownership issues to be explored in RPM.

Recommendation 1: Both Raw and Processed PGHD From Wearables Should Be Accessible to the Patient and Clinician

The extent to which patients and clinicians currently have access to all the recorded PGHD is questionable. PGHD accessibility largely depends on who owns the data to have complete access to them. PGHD have not yet been fully incorporated into clinical workflows; therefore, these data are neither controlled nor owned by health care organizations. Rather, the raw and processed PGHD from each wearable platform are accessed and controlled by the device company outside the health care setting.

Access to raw and processed PGHD during data collection may increase patients' awareness of their health status and whether they are required to take action or change their behavior and improve self-care. Now, the trend in wearable design is shifting toward data visibility to patients [23,24]. Nevertheless, clinicians may intentionally disable the access of raw data to patients during data collection, as it would lead to patient behavior change that might conflict with the purpose of the RPM program. The ability to access all raw and processed PGHD could also be limited by wearable companies. Medical device manufacturers should share comprehensive and contemporary health information with patients upon request [25]. Therefore, patients are within their rights to request health information that is captured, stored, and analyzed by and retrieved from a legally marketed medical device. Different policies suggest that wearable developers, regardless of the wearable type, should provide patients complimentary access to PGHD [26-28].

Considering these policies, PGHD ownership has not yet been defined clearly enough to determine who owns part or all of the data, affecting patients' access to PGHD [29,30].

In terms of clinicians' access to raw and processed PGHD, the necessity to access all the collected raw and processed data depends on which data are needed for decision-making. Our findings showed that it would be difficult for a clinician to find the log-in details of a patient's wearable portal if the patient has changed their portal account information or the device without informing the clinician. Clinicians' access to PGHD might also be prevented by patients, which might reveal that they have not followed their care plans [31].

Collecting various types of PGHD from different wearables outside the clinical environment means that data are stored across different platforms. Ideally, PGHD should be accessible to the people who collect them, and access methods should be transparent. The purpose of giving patients and clinicians access to PGHD is to enable them to have a clear picture of the former's health status.

Recommendation 2: A Mechanism Should Be Available to the Patient and Clinician to Set Up Notice Recurrence on Where, When, How, and by Whom PGHD From Wearables Are Accessed

It is important that patients and clinicians be aware of who else has access to PGHD during data management from outside the health care setting to inside it. Patients and clinicians in this project had little understanding of how and by whom PGHD are accessed during data management stages in RPM [32]. Clinicians placed responsibility on the wearable developer for informing patients about who can access their data. Also, the installation terms and conditions of a large number of health wearables' apps indicate that the wearable developers are the owners of PGHD and have authority to grant data access to others [26]. A review study of 4 known wearable products showed that the privacy policy of only 1 platform asserted PGHD as users' sole and exclusive property [33]. However, these companies' statements were not accompanied by strategies to support patients' awareness of the accessibility of their data to others. Patients should be informed about what PGHD are collected and accessed, including possible lawful access by third parties; whether these data are identifiable or depersonalized; and how they are accessible for clinical decision-making [27,28,34,35]. Patients need transparency about PGHD access not only before data collection but also throughout all the PGHD management stages. Not knowing who has access to their data can deter or inhibit PGHD collection [23].

Initiatives such as the privacy notice checklist developed by the US Office of National Coordinator for Health Information Technology are to be used by wearable developer companies to disclose their privacy and security policies to patients and inform them about what happens to their PGHD once they purchase and use the device [36]. However, this notice appears to be more applicable to wearables for self-management than to those for RPM. In addition, one-off use of the privacy notice checklist cannot ensure the notification of all PGHD accesses during all data management stages. For example, PGHD might be transferred from the patient to the clinician through communication networks that might be hacked. Many RPM programs lack robust cybersecurity mechanisms [37].

Using PGHD in clinical practice means that clinicians might also need to be notified about PGHD flow to be able to track patient monitoring instructions from other clinicians if necessary. Patients and clinicians should be able to set up notification recurrence of PGHD access based on their preference.

Recommendation 3: A Mechanism Should Be Available to the Patient to Change Permissions That Clinicians Have to Access PGHD From Wearables

The patients' and clinicians' awareness of circumstances under which PGHD are accessed does not give patients the authority to consent to PGHD access by others.

It is unclear to PGHD stakeholders how patient's consent to PGHD access should look [38]. Patients in the RPM of our 3 use cases, diabetes, cardiac arrhythmia, and sleep disorders, sign a consent form at the beginning of the program. However, the continuous nature of data collection and access in RPM might require constant PGHD access authorization when different clinicians need to access the data for different purposes of patient care [29,39]. There are concerns that the clinicians may access PGHD at a stage where the data have not been granted access to by the patient. This might not be ethical even if done to benefit the patient [40].

Patients themselves may have little awareness of PGHD consent, and their attention may be confined to the terms and conditions statement before pressing the consent button for installing the wearable components. However, the wearable developers' privacy policies and terms of service are often difficult to read and understand [41,42].

Appropriate consent management mechanisms enable patients to manage their consent preferences. Nevertheless, there is not yet a well-established consent mechanism for continuous data collection and use [28]. As various types of PGHD might be collected through different wearable platforms, sensitive data might be released when using one consent at the beginning of the RPM program. For example, patients might not want to provide details about their behaviors or lifestyles to clinicians in a certain time frame if it would lead to judgment or being shamed for perceived unhealthy choices during data collection. Thus, a process of dynamic consent might be more feasible to give patients control over the level of access to their data for different purposes and the choice of whether these data are anonymized or identifiable [43,44]. It could provide more personalized approaches and improve the continuous patient-clinician communication. Also, it gives patients the ability to understand and decide to what extent they are willing to share their data. Moreover, defining different levels of permission enables patients to review consent over a period to update or withdraw data at any time without affecting previously collected data [28].

Validation Results

Each of these 3 recommendations about PGHD accessibility was rated as "essential" to the to the safety and quality of care in RPM (reached an aggregated 60% agreement as being important to very important).

PGHD Accuracy

PGHD accuracy is compromised by a patient's errors or other error sources during data management, as well as uncontrolled possibilities for data revision.

This aspect depends on the technical features of the wearable and its components and the behaviors of the patient or caregiver at the point of data collection. Clinicians' trust of PGHD accuracy is significantly impacted by the differentiation between medical grade and consumer wearables. Clinicians trust the level of accuracy in PGHD captured by medical wearables owing to their preassessment and approval from regulatory bodies. Our findings showed that consumer wearables were not used in RPM because of not being regulated for clinical use. However, even a medical wearable may not work accurately in some instances, as identified in our interviews and workshop studies. Moreover, a study showed that the inaccuracy of continuous glucose monitoring (CGM) wearables was the most critical impediment (53%) to the use of these devices by diabetic adults [45]. Nevertheless, as the wearables collect data longitudinally, clinicians may trust the overall trends rather than doubting whether a single data point was captured correctly.

Recommendation 4: PGHD Should Be Collected Automatically by the Wearable Device, With as Little as Possible Manual Intervention

The way PGHD are collected can pose risks for data accuracy. Automated sensing via algorithms embedded into wearables can provide persistent collection and analysis, providing a comprehensive picture of a patient's status over time. Automation can lower the tracking burden, improve PGHD accuracy, and accelerate data filtering for timely access [39]. For a patient with low digital health literacy, automated data collection can reduce the level of disengagement with the device.

In addition to automatic data collection, some wearables require types of PGHD such as meal, activity, and mood data to be entered manually in the wearable platform on a daily basis. This can place a burden on patients and result in inaccurate and inconsistent recordings [46]. Yet, there has been no innovation to change the manual collection of these types of PGHD into a seamless automatic process; however, the extent of engagement in manual PGHD collection and documentation might depend on the patients' level of understanding of the data and the message that PGHD could convey to the patients [47]. From clinicians' perspectives [48,49], automated data collection and transmission to the associated app is a more accurate mechanism to evaluate peak flow variability than a patient's difficult and time-consuming manual calculations.

Some PGHD types that patients were required to record manually—such as activity data in the remote monitoring of patients with diabetes—could be automatically captured via consumer wearables. Synchronization shortages between different types of medical and consumer wearables and a lack of adoption of consumer wearables in RPM are barriers to increasing automation. Although it might not yet be possible for some data elements to be captured automatically, there could be strategies to limit free-text entries. For example, wearable developers can reduce the possibility of errors in manual data entry in the associated apps by requiring the user to choose from a list of options instead of entering free-text [50].

However, having all PGHD collected automatically may lead to less control by patients over their health status and reduce their engagement in their self-care [51]. In addition, behavioral factors such as improper application of a sensor on the body or changing the device settings can have adverse impacts on PGHD accuracy. Automation can provide more accurate data if it does not negatively impact patients' engagement in self-care.

Recommendation 5: Annotation Function for Manually and Automatically Entered PGHD Should Be Available to the Patient and Clinician in Order to Comment on Inaccurate Data

Whether PGHD are collected automatically or manually, the ability to annotate them during data collection is a critical contribution to their accuracy [31]. In addition to the annotation of manual entries, the annotation of automatically collected data can help patients prevent errors in them and mark questionable data to discuss with clinicians [52].

The rapidly changing environment surrounding the patients may contribute to inaccuracies in manual and automatic captures [53]. Sometimes, the wearable works inappropriately or the patient makes mistakes in wearing the device or entering data; however, the feature of annotating both data collected manually and those collected automatically is not designed in many wearable platforms and is often overlooked in the testing of wearables for use in RPM [47].

According to our findings, patients can add notes on inaccuracies only through their diaries to discuss them with clinicians during the clinical consultations. However, the annotation feature could be embedded in the wearable design to reflect on data inaccuracy in real time instead of writing a diary note that might be forgotten. Patients could be notified of the incorrect values to annotate data or redo data collection instead of sending incorrect data to the clinician [30]. The wearable developers could also enable passive data annotation upon patients' request [28]. Nonetheless, it is uncertain, if the patients themselves do not notice the errors, how they could annotate PGHD given that wearables often lack feedback mechanisms to alert the wearer about inaccuracies [54].

There is a concern that patients may use this functionality to override real actions. Therefore, the patients and caregivers need to be educated on PGHD annotation and build trust upon this functionality to enhance patient-clinician interaction and shared decision-making [46,55].

Clinicians should also be able to annotate the processed PGHD to understand data collection barriers and provide more efficient personalized care plans [29]. However, as the processed PGHD are usually represented as static snapshots to the clinician, it would be difficult for a clinician to annotate the reports and highlight the problematic areas of PGHD [56,57].

Recommendation 6: The Wearable Should Be Calibrated Automatically as Required by the Clinical Standard of Care of Diseases

Both medical and consumer wearables may collect inaccurate data. Therefore, it is important to ensure that wearables are calibrated to guarantee accurate sensing [30,39,58].

Some wearables need one-off calibration by the clinicians before initiating remote monitoring, whereas for other types such as CGM devices, the patient should frequently calibrate the device via a glucometer to ensure PGHD accuracy. Nonetheless, a patient's responsibility in terms of how often and when they should calibrate the wearable device in RPM is often unclear [31]. The need to calibrate wearables not only is a burden on patients or their caregivers but also increases the likelihood of inaccuracies. Patients should be taught the importance of calibration and when it should be done. For example, from a clinician's point of view, the 12-hour calibration for CGM wearables prescribed by most wearable developer companies [59] is not clinically acceptable; rather, calibration should be done 3 times a day when the blood glucose is not rapidly changing. Similarly, the best time to calibrate the device is when the glucose level is stable [60]. A study showed that nearly half of the participants reported calibrating CGM at more intervals than recommended by the wearable developer to ensure data accuracy [45].

Although regular calibration might be a burden, understanding its value would encourage patients to do it correctly [55]. However, considering the possibility for error to arise from the manual calibration of CGM devices with glucometers, an automatic calibration mechanism could be preferable.

Moreover, the wearable developers could conduct dynamic testing of the products. Clinicians want more collaboration with wearable developers to define strategies for continuous wearable assessment that can be achieved through various RPM interventions. Ideally, there should be a consensus among clinicians and wearable developers regarding guiding patients on the frequency of and providing instructions on calibration based on clinical principles.

Validation Results

Each of the 3 recommendations about PGHD accuracy was rated as "essential" (each one reached an aggregated 60% agreement as being important to very important).

PGHD Completeness

Incomplete PGHD may be a result of technical or behavioral issues. Battery failure, wearable dysfunction, lack of synchronization in different time zones, internet disconnection, or patient's neglect are among the accidental causes of insufficient PGHD. This may compel clinicians to reorder data collection. Moreover, there might be deliberate data omissions for both manual and automatic entries because of demotivation, lack of digital literacy, body pain, or perception of having no changes owing to seeing similar trends over time [61].

Lack of continuous follow-up may also result in incomplete data. Different health care settings define data sharing time frames differently in the remote monitoring of the same health condition; this can create confusion in patient and clinician communication. Lack of continuous interaction with patients during remote monitoring could result in a lack of engagement in self-care and motivation to collect data [62]. None of the published RPM studies have reported an approach to identify the exact reason for data incompleteness.

Recommendation 7: A Protocol Should Be Available to the Patient and Clinician That Defines PGHD Downtime, That Is, the Time Range During Which It Is Acceptable if the Wearable Is Not Collecting Data

Given the constant automated sensing capabilities of wearables, it is unclear whether patients are required to wear the devices continuously in different RPM programs to provide sufficient data for their care planning. There was lack of awareness among PGHD stakeholders in our studies on standardizing "downtime" when patients can stop data collection in different RPM programs. From clinicians' point of view, a CGM wearable should be worn for at least 80% of the RPM period so that it can provide complete data for interpretation and decision-making. However, it is thought to be burdensome for patients to have to wear the device day and night and calibrate it frequently [55]. New generations of wearables seem to address "downtime" by letting patients turn off the device. Alerts could be designed in wearables to help clinicians discuss the reasons for incomplete data with patients.

As mentioned in the *PGHD Accuracy* section, clinicians would prefer focusing on data trends over time rather than single data points; therefore, some degree of missing data is acceptable [51]. Some clinicians do not see PGHD completeness as fundamental for sound decision-making [63]. However, it is important to know the extent of the impact of incompleteness on data interpretation and decisions made for patient care in the remote monitoring of different diseases [51,64]. Having a predefined and transparent downtime protocol on which the patient and clinician agree could clarify the completeness of PGHD [65].

Recommendation 8: A Protocol Should Be Available to the Patient and Clinician for Resuming PGHD Collection When the Acceptable Downtime Period Is Exceeded

Applying the acceptable time frame in which patients can stop collecting PGHD cannot be thoroughly understood unless patients are aware of when to resume data collection.

However, this might not happen if patients forget to do so. Moreover, owing to a lack of technical infrastructure for the real-time transmission of data from outside to inside the health care setting, clinicians are not aware of the missing data during data collection and thus are unable to alert patients to resume data capture [66,67]. This could be considered in the wearable design; for instance, wearables can provide patients the ability to set an alarm based on the predefined acceptable downtime schedule.

Recommendation 9: Annotation Function Should Be Available to the Patient in Order to Provide Context for Any Period of Missing PGHD

Any missing data need to be supplemented by contextual information to help clinicians identify the causes and discuss them with patients [63]. Contextual information about the missing data can help clinicians understand whether the problem was technical, behavioral, or related to the process of data transmission. Incomplete data in themselves do not explain the circumstances that led to their incompleteness [51].

Although some wearables were reported to provide notification of missing data, they still lack contextual information. This could place a burden on patients to be constantly attentive to record the causes of incompleteness. Innovative mechanisms could be designed to increase the interaction between the device and the wearer to record contexts for the incomplete data in a real-time or on a daily basis. The annotation feature that was mentioned in the *PGHD Accuracy* section is equally important to enable patients to enter information regarding missing data [56].

Validation Results

Each of the 3 recommendations about PGHD completeness was rated as "essential" (all reached an aggregated 60% agreement as being important to very important).

PGHD Consistency

PGHD consistency is characterized by the ability to compare PGHD of one measurement from different devices as well as the ability to relate PGHD to the corresponding conventional clinical measurement.

Various wearables and associated mobile apps and web portals used in RPM programs may not represent data in a consistent manner. PGHD inconsistency can happen during data collection, transmission, and review, which immensely impacts data presentation that may have not been thoroughly recognized by PGHD stakeholders.

Recommendation 10: PGHD From Wearables Should Be Collected Based on Clinically Accepted and Structured Data Definitions and Standard Formats

Both consumer and medical wearable platforms may fail to represent data in clinically standardized formats [30,68,69]. Nonstandard presentation can result in confusion in data interpretation and inability to discern whether PGHD reports show normal or abnormal trends [51]. The standardization of health data elements is intended to define what data are to be collected, decide on how the collected data should be represented, and specify how the data should be encoded for transmission [70].

Collecting PGHD from different types and brands of wearables, each with its own data presentation format, could result in inconsistent reports [49,51]. Most of the recent PGHD-related policies advise developing standardized formats for PGHD collection that align with the clinical data standards, which are defined as protocols, terminologies, and specifications that are used during data management stages [23,50,54,71,72].

To ensure consistent definitions and formats for PGHD, 2 approaches should be considered: PGHD consistency at data collection and PGHD consistency at the data processing stage. Patients may need to be advised to collect PGHD from one type or brand of wearables for the remote monitoring of a particular health condition to provide consistent reports. Clinicians in our studies and others [62] preferred to give patients autonomy over device selection and stated that patients should have the right to select a convenient and easy-to-use wearable device. Nonetheless, because PGHD cannot be further filtered by information systems within a health care setting to fix the inconsistencies that emerge from collecting data in different formats, data that are presented for review might be difficult to interpret. Inconsistent reports at the data review stage are a consequence of collecting data from disparate wearable platforms. The second approach is to standardize PGHD at the data processing stage regardless of the wearable used in data collection. In this case, robust technical infrastructure needs to be in place to allow gathering PGHD from different wearables and their apps and portals in one database to filter and process data and present standardized reports that are similar to clinical data presentation formats.

Universally accepted data definitions and data exchange formats are required to facilitate effectual data transfer. Data should be codified according to the known clinical standards. In addition, ontologies that could aggregate and enrich PGHD with definitions, synonyms, and term relationships can be developed to provide standardized formats and make data semantically exchangeable [73].

Recommendation 11: PGHD From Wearables Should Be Integrated Into the Patient's Clinical Care Record

Lack of PGHD integration with electronic medical record (EMR) systems is another barrier to PGHD consistency [30,39,74-76]. Current RPM programs are project oriented and not embraced in routine clinical practices. Moreover, most current EMR systems are not designed to seamlessly gather various types of data from outside the clinical setting in a straightforward manner [77]. PGHD should be combined with the patient's clinical record to identify potential correlation with past conditions and be used in future interventions [30,46,54].

Despite the clinicians' preference for the patients to choose their own brand of wearable, PGHD integration with EMRs constrains the selection of wearable. Patients should only use wearables in RPM that follow the interoperability standards used in the health care setting.

Most policies addressed the necessity of integrating PGHD with EMRs [23,26,35,54], but few provided specific suggestions. For example, the American Medical Association's best practices for digital health implementation recommend that standard communication templates be designed before implementing RPM intervention to ensure consistency in data documentation during the whole process [72]. Therefore, PGHD integration with EMRs might be facilitated by modifying the EMRs, developing external dashboards, or limiting the choices of the brand of wearable used for data collection.

Recommendation 12: PGHD From Wearables Should Be Consistently Exchanged Inside and Between Clinical Settings

In addition to the need for standardized formats and integration with EMRs, PGHD exchange inside and outside the health care setting needs to be consistent by following health data exchange protocols.

If >1 health care setting is actively represented in an RPM program for one disease cohort or different departments implement RPM in one health care setting, data should be exchanged consistently to be understandable by different clinicians. This requires standardized formats for various types of PGHD.

Interoperability initiatives developed by the Australian Digital Health Agency [78] defined standards to facilitate data and information exchange and provided compliance mechanisms in connected health programs. These standards are broad and cover both PGHD and clinical data collected from outside and within health care settings. More specific interoperability standards were introduced by the Personal Connected Health Alliance with its Continua guidelines for data interoperability in personal connected health devices [79]. These initiatives need to be tested in various RPM programs to assess the consistency in data exchange.

Validation Results

recommendations 10 and 11 about PGHD consistency were rated as "essential" (they reached an aggregated 60% agreement as being important to very important), whereas recommendation 12 was rated by 9 (90%) of the 10 participants and reached less than 60% agreement for inclusion in the defined categories.

PGHD Interpretability

Interpretability is affected by the way in which PGHD are presented as well as by the availability of contextual information regarding PGHD.

Not understanding the presented data can reduce patients' and clinicians' motivation for data collection and review [31,80,81]. Challenges of PGHD interpretation can occur at any stage of data management. This aspect was mentioned as the most challenging feature in our studies.

Recommendation 13: PGHD From Wearables Should Be Accompanied by Contextual Data That Are Clinically Important to Patient Management

The increasingly high volume and dynamically changing nature of PGHD make it difficult and time-consuming to gain a holistic view of a patient's status from the data alone [31,74,82].

Most PGHD are not supplemented by contextual information about the circumstances in which the data were collected. Lack of context can lead to misunderstanding; misinterpretation; and, consequently, unsound decisions [80]. For example, when a clinician tries to discern a pattern in a processed PGHD report, it may be unclear whether a graph showing a lack of activity reflects the patient's demotivation, a problem in the wearable's function, or a medication interruption [56]. In this situation, relying on the patient's verbal expression without recorded contextual information is not sufficient to draw an understanding of the patient's situation per trend.

Similar to its application in PGHD completeness, context for PGHD is important for the data review stage so that clinicians can understand what the patient was doing at the point of data collection [31,52,74,83]. Among the wearables studied in this project, only CGM devices allowed the manual capture of limited contextual data—such as those about mood and exercise—in cases where the automatically captured data were reported to be erroneous or incomplete. PGHD from medical wearables can be contextualized by data that are automatically captured from consumer wearables [80,84]. However, no mechanism exists to integrate these 2 types of wearables in RPM, and there is uncertainty about which contextual data are more relevant to patient care. In addition, the ability to understand and interpret contextual information is still beyond clinicians' expertise [85].

More collaboration between patients, clinicians, and wearable developers is needed to identify what contextual data need to be collected for each health condition and whether these data elements should be incorporated within the wearable design or require wearable integration.

Recommendation 14: Dynamic Visual Representation as Well as a Static Snapshot (Such as in PDF Format) of PGHD From Wearables Should Be Available to the Patient and Clinician

Clinicians review the processed PGHD reports either from the patient's or clinician's portal that the wearable developers design for them with static visualization of data that can be downloaded in a PDF format.

Having a snapshot of all the data collected over time provides a summary of the patient's status, but as the amount of PGHD increases, such a report could be progressively more complex for their clinician to interpret [86]. Designing interactive visualization tools based on clinicians' needs can result in easy PGHD interpretation [52,76,82,87].

Interactive visualizations could enable clinicians to highlight the most concerning areas and customize the reports based on different variables [87,88]. Interactive visualization supported by annotation capability can facilitate the cointerpretation of PGHD report, such as the ability to add highlights in a graph to detect changes. It is beneficial to patients to have a saved version of points and notes of what they and their clinicians identified and discussed for use in the next consultations. Likewise, in the subsequent clinic visit, the clinician could readily recollect what the previous consultation was focused on, which helps recognize patterns and set efficient care plans [56]. A dynamic and interactive visualization could also layer PGHD displays based on clinicians' preferences. Studies have shown that different layers of data presentation, such as a holistic summary, an individual data summary, and detailed individual data, support the comprehensive interpretation of PGHD [52,76,82].

Notwithstanding, it is a challenge for wearable developers to design data presentation formats that please all clinicians with varying levels of digital health literacy. Collaboration among PGHD stakeholders is needed to determine who should design interactive dashboards for PGHD presentation, whether the dashboards should be implemented within EMRs or somewhere else in the health care setting, and how the reports should be presented to the patient and clinician to inform shared understanding and decision-making.

Recommendation 15: Alerts Should Be Sent to the Patient During PGHD Collection by the Wearable When Data Are Outside the Acceptable Range, Accompanied by Clinical Advice on Action to Take

In addition to clinicians' interpretation of PGHD, it is important to ensure that patients can also interpret the data correctly. Not understanding PGHD can reduce the motivation to continue data collection. Efforts toward changing the patients' roles from passive participants to active players in RPM require patients to understand PGHD and make necessary changes during data collection [89].

However, wearables do not provide understandable contextual information on PGHD. Most wearables display PGHD without further explanations of their meaning, the normal range, what will happen to the patient's health status if their measurements are out of range, or what actions the patient could take if their measurements are out of range. This is problematic if the patient cannot immediately communicate with their clinician when they see significant changes occur in their data trends and do not know what action to take.

There are alarms embedded in some types of consumer wearables that notify out of range measurements [85]. Although PGHD from these tools may not require urgent actions, such features could improve patients' interpretations of the data and better inform and influence behavior change.

Some medical wearables are equipped with a feature that alarms when the raw data go outside the normal range. Devices without this feature could be dangerous to a patient's health, as immediate medical action may not be undertaken when needed. Nevertheless, the questions of to what extent patients could interpret PGHD augmented with contextual information during RPM without a clinician's intervention and how sound a patient's decision would be based on the interpretation are largely unexplored.

Cointerpretation of PGHD improves the shared understanding of data reports and generates an additional layer of meaning for PGHD in patient care plans [90]. These strategies particularly depend on patients' and clinicians' training and collaboration with the wearable developers to improve PGHD presentation design and interpretation.

Validation Results

Each of the 3 recommendations about PGHD interpretability was rated as "essential" (reached an aggregated 60% agreement as being important to very important).

PGHD Relevancy

PGHD relevancy is characterized in various manners depending on the scope and coverage of data for each health condition. The values of conventional clinical data collected inside the health care setting are defined based on the standards of care. However, PGHD include a wide range of heterogeneous and new types of data whose relevancy to the monitored health condition might be unclear. Only 1 common theme was found in the previous studies of this project for PGHD relevancy, which resulted in the recommendation discussed next.

Recommendation 16: There Should Be a Shared Understanding Between the Patient and Clinician of Relevant Data for the Disease Based on the Standards of Care, and Make Sure That All the Relevant Data Are Collected

PGHD relevancy was perceived as the most distinguishing DQM factor in using PGHD from consumer wearables versus those from medical wearables in RPM, and its lack was perceived as the most predominant barrier to the adoption of PGHD in clinical practice [91]. Patients and clinicians might have different perspectives on which types of PGHD are relevant to patient care [31]. Patients' enthusiasm to use a wide range of consumer wearables and collect new types of PGHD that have not been collected easily before (eg, heart rate, sleep quality, and activity level) increases their expectation from clinicians to review the data. By contrast, clinicians might not be convinced of the extent to which the data are relevant to the health condition and supplement the clinical data collected from medical wearables to provide a better picture of patients' status.

Clinicians involved in this project along with other studies indicated that PGHD from consumer wearables have not yet been proven to correlate with most health conditions and that they are different from other clinical data in terms of clinical value [39,76,82].

Even if PGHD are collected from medical wearables, it would still be challenging to identify whether all the data are relevant to the specific health condition. Conversely, some wearables cannot capture all the relevant PGHD; therefore, important relevant data might be missed, which might lead to incorrect decisions about patient care [76,83]. The need for the collection and analysis of relevant data was addressed by recent PGHD-related policies [23,54,72]. Only 2 clinical guidelines developed to address the details and level of relevancy of PGHD collected from wearable devices for the remote monitoring of patients with diabetes and those with cardiac arrhythmia were identified [60,65]. More guidelines are needed to determine relevant PGHD for the remote monitoring of each health condition.

Validation Results

Half of the participants rated PGHD relevancy recommendation as very important (40%) to important (10%), whereas 40% addressed it as moderately (30%) to slightly important (10%).

PGHD Timeliness

PGHD timeliness is characterized by the timing and frequency of PGHD availability to patients and clinicians.

Recommendation 17: PGHD From Wearables Should Be Available to the Patient Within a Timeframe (Continuously to Periodically) According to the Standards of Care of Diseases

Our findings showed that the timing of PGHD availability to patients was overlooked. Although accessing data during data collection is critical when a decision needs to be made, some wearables do not provide real-time PGHD access to patients during data collection. As discussed in the *PGHD Accessibility* section, depending on the health condition and the clinical purpose of RPM, PGHD presentation to patients in real time might be deliberately disabled by clinicians. However, studies have shown that accessing real-time data from the wearable increased patients' awareness of the wearable's function, further engaged them in self-care, and enhanced shared decision-making [92-94].

As PGHD collection in RPM are led by clinicians, patients may not be fully aware of their rights in accessing PGHD at data collection and how it might impact their safety. PGHD access in real-time or periodic mode needs to be defined according to the standards of care of the health condition [65]. RPM interventions could be designed based on patient-centered care models where time frames could be established so that patients can access their data during data collection to make a proper decision, change their behavior, or immediately contact the clinician.

Recommendation 18: PGHD From Wearables Should Be Available to the Clinician Within a Timeframe (Continuously to Periodically) According to the Standards of Care of Diseases

The most challenging issue reported about PGHD timeliness is the lack of clinicians' access to data in real time [39,77,83]. As PGHD are not yet integrated with EMRs, it is difficult and time-consuming to frequently receive the data and follow-up with patients. PGHD integration with EMRs would provide possibilities for generating alerts on newly added PGHD in the EMR system in a real-time or near real–time basis so that clinicians can be updated on a patient's status and provide prompt feedback [83]. Notwithstanding, the technical integration by itself is not the ultimate solution. PGHD need to be fully incorporated into clinical workflows such that clinicians could receive data based on predefined protocols and be able to provide timely advice to patients [95]. Timely access to PGHD without immediate feedback to patients would lead to patient demotivation on data sharing [96]. However, RPM interventions may have different protocols for PGHD availability to clinicians. In some cases of remote monitoring of patients with diabetes, clinicians remotely obtain PGHD reports from patients during data collection, whereas in others, they see the report after data collection during the clinic consultation. Having predefined protocols might facilitate clinicians' access to PGHD within a specific time frame.

Recommendation 19: A Timeframe for Sharing PGHD From Wearables Should Be Available to the Patient and Clinician

As noted earlier, RPM programs apply disparate time frames for PGHD sharing. This way of accessing data can be challenging. Patients who access data in real time may also need to receive a clinical advice immediately, whereas data are not available to clinician in the same time frame. Frequent data sharing during data collection could help recognize some behaviors that might not be identified when the collection period is finished.

PGHD need to be available when there is an urgent need for clinical advice so that the patient can change the way of data collection or their behavior accordingly [30,97]. However, findings showed that this depends on the health condition; for example, the guideline on using wearables in cardiac RPM emphasized that these services should not be mistaken with acute care; therefore, there is no urgent need for real-time feedback [65]. Hence, based on the health context, having transparent protocols on data sharing could help clinicians review PGHD and set patients' expectations for data transmission and feedback [76].

As different RPM programs may need different approaches on data timeliness based on the standards of care, there should be a single time frame defined for the remote monitoring of each health condition to ensure consistency among the programs.

Validation Results

All of the recommendations about PGHD timeliness were rated as “essential” to the safety of and quality of care in RPM (reached an aggregated 60% agreement as being important to very important).

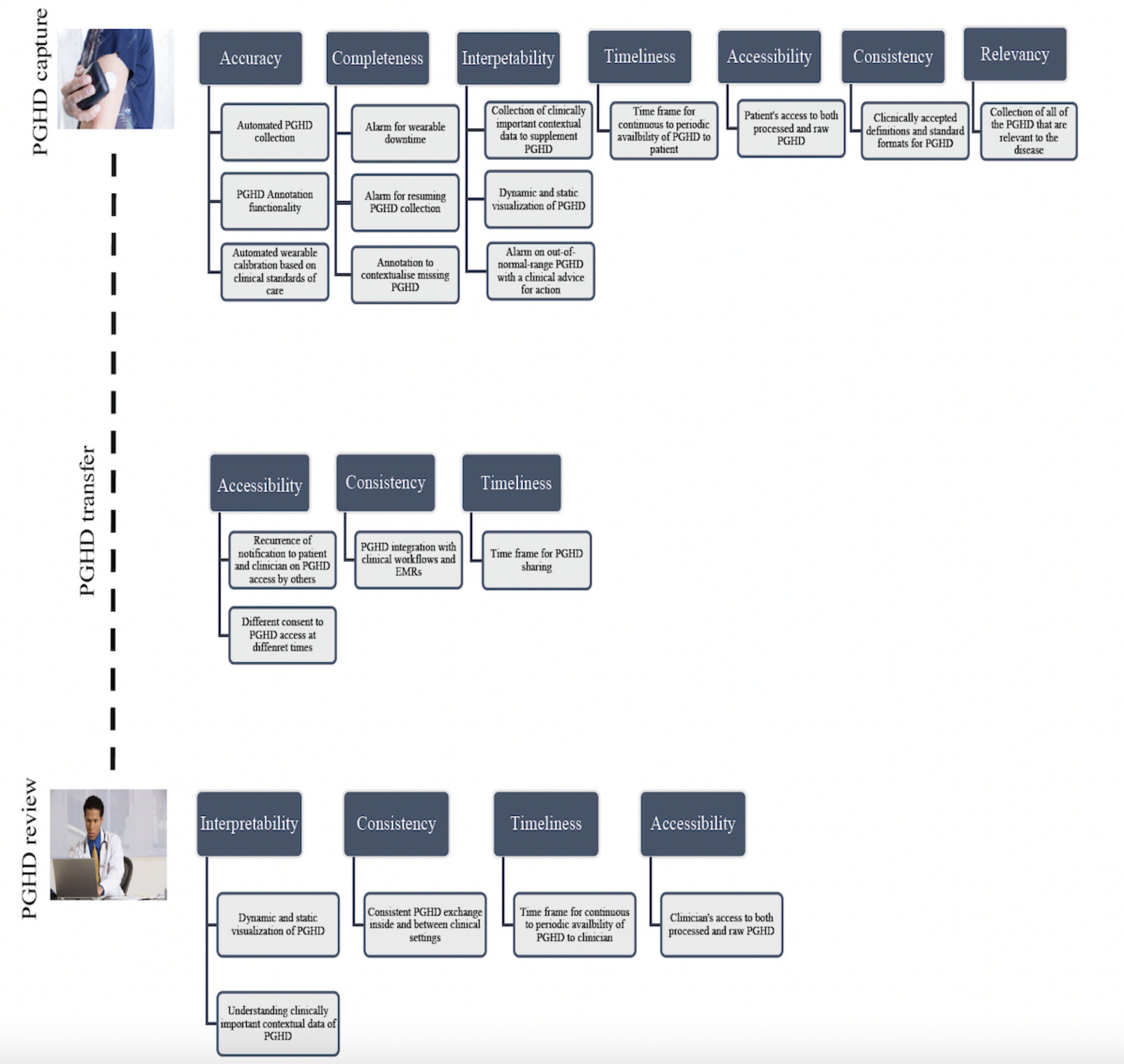
A Staged Model of Quality Management of PGHD in RPM

Figure 2 illustrates the recommendations according to the importance of their consideration at different stages of data management. This model can assist PGHD stakeholders in understanding what DQM actions need to be taken to efficiently collect, manage, and use PGHD in RPM.

As shown in Figure 2, all DQM aspects of PGHD require attention at the data collection stage. It indicates that the quality management of PGHD is critical when data are collected outside the clinical environment under patients’ or their caregivers’ supervision. Data access, consistency, and timeliness were the most critical DQM aspects to be considered during PGHD transmission from the patient to clinician. These 3 aspects along with PGHD interpretability require emphasis when clinicians review the data reports for shared decision-making and creating patient care plans.

As there are interconnections among DQM aspects, this model indicates that collaborative actions need to be undertaken by different PGHD stakeholders to practice DQM and ensure high-quality PGHD in RPM.

Figure 2. Recommendations for the quality management of patient-generated health data (PGHD) at the 3 stages of PGHD management. EMR: electronic medical record.



Discussion

Principal Findings

This paper presented the development of 19 recommendations for 7 DQM aspects of PGHD collected from wearable devices in RPM programs. The guideline aims to assure that high-quality data are collected, managed, and used in RPM programs to improve the safety and quality of these programs and enhance PGHD fitness for use in routine clinical practice.

The guideline was constructed by following 4 steps of guideline development process through 5 qualitative studies. The guideline was then conceptualized to address 3 main concepts: PGHD management process, DQM aspects of PGHD, and sociotechnical issues that influence the quality management of PGHD during the data management process.

The DQM guideline for PGHD is distinguished from conventional DQM guidelines for clinical data in several ways: (1) it emphasizes the need for action corresponding to each DQM aspect at each stage of PGHD management; (2) it considers both external sociotechnical factors and internal organizational factors that impact the quality management of PGHD in RPM; (3) it recognizes patients' and clinicians' needs for each DQM aspect of PGHD, as the key PGHD stakeholders in RPM. This guideline is intended mainly for using PGHD for patient care. It is anticipated that the guidelines can also be used alongside conventional DQM guidelines for clinical data to assure PGHD quality and when these data are integrated into EMR systems.

To effectively apply the guideline in the remote monitoring of various health conditions, wearable devices should not be considered as stand-alone tools that work in isolation. Instead, they should be looked at as one component of a bigger ecosystem where different stakeholders interact with each other, with the devices, data, technical infrastructure of the health care setting, and standards to ensure that high-quality PGHD are collected, managed, and used for patient care [3]. The guideline can be best applied when RPM is implemented for >1 health condition across the health care system and when PGHD are collected from >1 type of wearable device and system interconnections are facilitated [98]. Also, realizing the value of high-quality PGHD for patient care can potentially blur the reliability distinctions between the 2 types of wearables, consumer and medical wearables. Being approved by regulatory agencies as a medical grade wearable does not ensure that the PGHD from it achieve a satisfactory level of quality. PGHD from consumer wearables are rarely used in current RPM services, and the research findings mainly included PGHD from medical wearables, so unseen challenges might exist to the quality management of PGHD from consumer devices. Advances in the capabilities of consumer devices and patients' and clinicians' accessibility to them are likely to see greater crossover between medical and consumer wearables in the future.

The DQM guideline for PGHD in RPM cannot be successfully implemented and used if the health system does not address the factors listed in Textbox 2.

The implementation of PGHD quality management in RPM can benefit the health care system and those who are considered the stakeholders of PGHD and who might advantage from the incorporation of these data into clinical practice, including the groups listed in Textbox 3.

Textbox 2. Considerations for the implementation of the data quality management (DQM) guideline.

- Policies: clinical, technical, and organizational policies need to be in place in parallel with the guideline for the quality management of patient-generated health data (PGHD) to increase the likelihood that PGHD will be trustworthy for use in clinical care.
- Technical infrastructure: PGHD from wearables used in remote patient monitoring (RPM) programs are not yet integrated routinely into electronic medical record systems or able to flow securely across the health care system; both factors are key barriers to using PGHD in clinical care. The guideline can be best applied when a technical infrastructure is established to follow the recommendations for the systematic management, interactive and standardized presentation, and consistent exchange of PGHD and standardized and timely access of PGHD reports to patients and clinicians.
- Digital health literacy: understanding the quality management of PGHD requires sufficient digital health literacy among all PGHD stakeholders. The conceptual model shows that all DQM aspects need action in the stage when patients collect PGHD, emphasizing patients' need for literacy to understand DQM. Training delivered to patients and caregivers could enhance their engagement in the collection and management of high-quality data. Moreover, RPM teams could be expanded to include professionals who could provide DQM advice and support to clinical stakeholders.
- Collaboration: without collaboration among all PGHD stakeholder groups, the guideline recommendations cannot be implemented effectively in RPM. In addition to the stakeholders that were involved in this project, other stakeholders such as payers, policy makers, and health care administrative need to collaborate with the RPM team to understand the implementation requirements of the DQM guideline of PGHD. Continuous collaborative efforts to evaluate wearable devices, PGHD, and the data management processes could provide the health system with high-quality data that are fit for clinical care, population health management, and secondary uses.

Textbox 3. Stakeholders who can benefit from patient-generated health data (PGHD) quality management in remote patient monitoring (RPM)

- Patients and carers: patients will be able to collect high-quality data to manage their conditions. They will learn to correctly use digital health devices and collect high-quality PGHD that could be clinically valuable and used optimally in clinical care. Moreover, by collecting relevant and quality-assured PGHD and sharing them with a single RPM system, patients' role in RPM could be changed from passive to active participants, strengthening their interactions with clinicians, improving shared decision-making, and better engaging them in their health self-management.
- Clinicians: the pandemic has increased clinicians' awareness of the potential uses of RPM and PGHD. However, they need a reliable and convenient way to determine the utility of PGHD from patients, based on how and when these data are collected and reported and how and when they and others can access and interact with the data. The use of PGHD quality management recommendations enable clinicians to assess the quality of available data to support a patient consultation and how these data can form a valuable basis for efficient shared decision-making. Through this, they could optimize their focus on PGHD during and between patient visits.
- Health information professionals: health information professionals are nonclinical staff, including health informaticians, health information managers, and other experts who monitor data and information management within the health care system. The implementation of the guideline could bring new responsibilities and roles for these professionals. For example, these experts can play a critical role in defining new approaches to manage PGHD and use their skills to work collaboratively on data integration and management. In addition, they could act as gatekeepers before PGHD become available to clinicians and filter and analyze the data to provide the most meaningful information for clinicians.
- Health care organizations: for maintaining RPM after the pandemic, health care organizations need to be sure that PGHD from different digital health tools will support safer, higher-quality, and faster decision-making, based on more persuasive patient-clinician communication, leading to more effective and efficient health outcomes. The implementation of the RPM system driven by the PGHD quality management guideline could assist health care organizations in taking a standard approach to data integration, quality assurance, and risk management of these data to increase their trustworthiness for use in patient care. It may also provide health care organizations with strategies to think about the required infrastructure, policies, human resources, and potential collaborations with other parties to enhance the use of PGHD in clinical practice.
- Digital health technology companies: medical device companies may be alerted to the existing unrecognized problems in ensuring the quality of data from their proprietary devices and offer solutions to overcome these. This may address the need for better synergies with the existing health data standards and health information system architectures to enable data sharing from various devices. Consumer device companies may also realize the gaps in ensuring the quality of data from their devices. This could assist in developing higher-quality devices with user-friendly and validated data handling solutions that are capable of being integrated more readily into the existing clinical information systems.
- Other beneficiaries: findings from this research may also have indirect benefits, including providing insights into PGHD features and functions to the developers of electronic clinical information systems, such as patient records and point-of-care decision-support. These insights may inform the development of health policies and regulation of PGHD, including their use in research and public health, and could also provide more research opportunities in this area considering other kinds of PGHD and solutions for the further use of such data in broader contexts.

Limitations

PGHD collection for self-management purposes without clinical use was out of the scope of this research. Moreover, this study did not concentrate on the concept of data quality as used in the biomedical engineering domain, such as the accuracy of the formula or algorithms embedded in wearables. We also limited the exploration of PGHD to their use in direct patient care, engaging with the stakeholders in this kind of use, and excluded the secondary uses of data, such as in outcomes research, surveillance, reimbursement strategies, and purposes other than patient care.

The recommendations of the DQM guideline of PGHD were defined at a high level. They would benefit from the addition of details that specify the roles and responsibilities of different stakeholder groups. This would require the guideline to be investigated more deeply with participation from different stakeholder groups to identify further considerations in different contexts.

The guideline might be questioned as not being specific to one health condition when it is known that RPM initiatives are distinctive in different contexts of care. However, it was extracted from the RPM initiatives for 3 chronic conditions that showed similarities and commonalities in the quality management of PGHD. It is worth noting that digital health implementation is moving toward focusing on the patient as a whole rather than the disease. Therefore, RPM initiatives, as well as the data they collect, could also shift their focus from a specific disease and wearable to services for the integrated management of all the health conditions that a patient might have [72]. Nevertheless, for further exploration, the guideline can be implemented in each disease-based RPM to provide more specific recommendations based on particular needs. Understanding what makes PGHD more reliable for shared decision-making can motivate PGHD stakeholders to have a shared understanding of the value of these data and use them more efficiently to achieve better health outcomes.

Comparison With Prior Work

Research on the adoption, integration, and evaluation of RPM, wearables, and PGHD in clinical practice is rapidly growing [99-106], particularly during the COVID-19 pandemic, when many RPM initiatives were implemented around the world.

However, a few studies focused on PGHD quality [51,62,69,107] had aims and scopes that were different from those of our research. This is the first study of its kind that adapted 7 common aspects of DQM and investigated them in PGHD context during PGHD management stages. It also involved various groups of international PGHD stakeholders to share their experiences, concerns, and expectations regarding the quality management of PGHD and constructed and validated a set of recommendations as a novel guideline. This process helped reach a consensus among the participants on the recommendations they could follow to effectively collaborate for better patient care.

Conclusions

Although the quality of PGHD is addressed as a vital factor in increasing their reliability in clinical decision-making, this research is the first of its kind to explore the quality management of PGHD through 7 aspects during data management stages. The guideline developed in this research provides a major step forward in this regard. It gives PGHD stakeholders a framework for improving the quality management of PGHD collected and used in RPM underpinned by collaboration.

Acknowledgments

This project is funded through a University of Melbourne PhD scholarship. The corresponding author has made a substantial contribution to the conception and design of the work; the acquisition, analysis, and interpretation of data; and drafting the manuscript and revising it critically for important intellectual content. The other authors contributed to the conception and design of the work, revising the manuscript critically for important intellectual content, and final approval of the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The synthesis of the findings of this project's previous studies and the resulting themes that constructed the recommendations.
[DOCX File , 42 KB - mhealth_v11i1e35917_app1.docx]

References

1. Field MJ, Grigsby J. Telemedicine and remote patient monitoring. JAMA 2002 Jul 24;288(4):423-425. [doi: 10.1001/jama.288.4.423] [Medline: 12132953]
2. Taiwo O, Ezugwu AE. Smart healthcare support for remote patient monitoring during covid-19 quarantine. Inform Med Unlocked 2020;20:100428 [FREE Full text] [doi: 10.1016/j.imu.2020.100428] [Medline: 32953970]
3. Azodo I, Williams R, Sheikh A, Cresswell K. Opportunities and challenges surrounding the use of data from wearable sensor devices in health care: qualitative interview study. J Med Internet Res 2020 Oct 22;22(10):e19542 [FREE Full text] [doi: 10.2196/19542] [Medline: 33090107]
4. Kraef C, van der Meirschen M, Free C. Digital telemedicine interventions for patients with multimorbidity: a systematic review and meta-analysis. BMJ Open 2020 Oct 13;10(10):e036904 [FREE Full text] [doi: 10.1136/bmjopen-2020-036904] [Medline: 33051232]
5. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]
6. Tayi GK, Ballou DP. Examining data quality. Commun ACM 1998 Feb;41(2):54-57. [doi: 10.1145/269012.269021]
7. WHO Regional Office for the Western Pacific, World Health Organization. Regional Office for the Western Pacific. Improving Data Quality A Guide for Developing Countries. Geneva: World Health Organization; 2003.
8. Data duality frameworknterprise business intelligence solutions; 2013. The Australian Capital Territory Health. https://health.act.gov.au/sites/default/files/Policy_and_Plan/Data%20Quality%20Framework.pdf.
9. Data Quality Management Model (2015 Update) - Retired. The American Health Information Management Association. 2015 Oct 14. URL: https://library.ahima.org/PB/DataQualityModel#.Y77GNnZBzIU [accessed 2023-01-11]
10. Andrews TN. Guide to the health facility data quality report card. SILO. 2017 Jun 4. URL: https://silo.tips/download/guide-to-the-health-facility-data-quality-report-card [accessed 2023-02-13]
11. Data quality assessment and improvement strategies. The European Institute for Innovation through Health Data. 2016. URL: https://www.i-hd.eu/wp-content/uploads/2020/11/iHD-Annual-Report-2016.pdf
12. Canadian Institute for Health Information. The CIHI Data Quality Framework. Ottawa, Ontario: Canadian Institute for Health Information; Mar 22, 2009.
13. Woolf S. Practice guidelines, a new reality in medicine. Arch Intern Med 1992 May 01;152(5):946-952. [doi: 10.1001/archinte.1992.00400170036007]
14. The guidelines manual Process and methods. National Institute for Health and Care Excellence. 2012 Nov 30. URL: https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-pdf-2007970804933 [accessed 2023-02-13]
15. Hanson D, Hoss B, Wesorick B. Evaluating the evidence: guidelines. AORN J 2008 Aug;88(2):184-196. [doi: 10.1016/j.aorn.2008.04.017] [Medline: 18722835]
16. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003 Feb;12(1):18-23 [FREE Full text] [doi: 10.1136/qhc.12.1.18] [Medline: 12571340]

17. Committee on Clinical Practice Guidelines, Institute of Medicine, Field M, Lohr K. Guidelines for Clinical Practice From Development to Use. Washington, D.C., United States: National Academies Press; 1992.
18. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A data quality assessment guideline for electronic health record data reuse. EGEMS (Wash DC) 2017 Sep 04;5(1):14 [FREE Full text] [doi: 10.5334/egems.218] [Medline: 29881734]
19. Abdolkhani R, Borda A, Gray K. Quality management of patient generated health data in remote patient monitoring using medical wearables - a systematic review. Stud Health Technol Inform 2018;252:1-7. [Medline: 30040674]
20. Abdolkhani R, Gray K, Borda A, DeSouza R. Quality assurance of health wearables data: participatory workshop on barriers, solutions, and expectations. JMIR Mhealth Uhealth 2020 Jan 22;8(1):e15329 [FREE Full text] [doi: 10.2196/15329] [Medline: 32012090]
21. Major C, Savin-Baden M. Integration of qualitative evidence: towards construction of academic knowledge in social science and professional fields. Qual Res 2011 Dec 02;11(6):645-663. [doi: 10.1177/1468794111413367]
22. Hsu C, Sandford B. The delphi technique: making sense of consensus. Pract Asses Res Eval 2007;12(12).
23. Our data-driven future in healthcare. Academy of Medical Sciences. 2018 Nov. URL: https://acmedsci.ac.uk/file-download/74634438 [accessed 2023-02-13]
24. Turakhia M, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
25. Manufacturers sharing patient-specific information from medical devices with patients upon request 2017. U.S. Food & Drug Administration. 2017 Oct. URL: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm505756.pdf [accessed 2023-02-13]
26. Redefining our picture of health: towards a person-centered integrated care, research, wellness, and community ecosystem. American Medical Informatics Association. 2017. URL: https://www.amia.org/sites/default/files/API-2017-White-Paper-Redefining-our-Picture-of-Health.pdf [accessed 2023-02-13]
27. Art. 13 GDPR Information to be provided where personal data are collected from the data subject. Intersoft Consulting. URL: https://gdpr-info.eu/art-13-gdpr/ [accessed 2023-02-13]
28. Best practices for consumer wearables and wellness apps and devices 2016. Robert Wood Johnson Foundation. 2016 Aug 17. URL: https://fpf.org/wp-content/uploads/2016/08/FPF-Best-Practices-for-Wearables-and-Wellness-Apps-and-Devices-Final.pdf [accessed 2023-02-13]
29. Chung C, Dew K, Cole A, Zia J, Fogarty J, Kientz J. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16: Computer Supported Cooperative Work and Social Computing; Feb 27-Mar 2, 2016; San Francisco California USA URL: https://doi.org/10.1145/2818048.2819926 [doi: 10.1145/2818048.2819926]
30. Rickardsson I. Patient-generated health data: professionals' opinions and standardized data transfer. Digitala Vetenskapliga Arkivet. 2016. URL: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A941216&dswid=7788 [accessed 2023-02-13]
31. West P, Giordano R, Kleek M, Shadbolt N. The quantified patient in the doctor's office:challenges and opportunities. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA URL: https://doi.org/10.1145/2858036.2858445 [doi: 10.1145/2858036.2858445]
32. Bellekens X, Nieradzinska K, Bellekens A, Seeam P, Hamilton A, Seeam A. A study on situational awareness security and privacy of wearable health monitoring devices. Int J Cyber Situational Awareness 2016 Nov 7;1(1):74-96. [doi: 10.22619/IJCSA]
33. Paul G, Irvine J. Privacy implications of wearable health devices. In: Proceedings of the 7th International Conference on Security of Information and Networks. 2014 Presented at: SIN '14: The 7th International Conference on Security of Information and Networks; Sep 9 - 11, 2014; Glasgow Scotland UK. [doi: 10.1145/2659651.2659683]
34. Castell S, Roinson L, Ashford H. Future data-driven technologies and the implications for use of patient data. Ipsos MORI. 2018 Nov. URL: https://acmedsci.ac.uk/file-download/6616969 [accessed 2023-02-13]
35. Using remote patient monitoring technologies for better cardiovascular disease outcomes guidance. American Heart Association. 2019. URL: https://www.heart.org/-/media/files/about-us/policy-research/policy-positions/clinical-care/remote-patient-monitoring-guidance-2019.pdf?la=en [accessed 2023-02-13]
36. 2018 Model privacy notice. The U.S. Office of National Coordinator for Health IT. 2018. URL: https://www.healthit.gov/sites/default/files/2018modelprivacynotice.pdf [accessed 2023-02-13]
37. Medical device cyber security Draft guidance and information for consultation. Australian Government Department of Health. 2018 Dec. URL: https://www.tga.gov.au/sites/default/files/consultation-medical-device-cyber-security.pdf [accessed 2023-02-13]
38. Addonizio G. The privacy risks surrounding consumer health and fitness apps, associated wearable devices, and HIPAA's limitations. Law School Student Scholarship. 2016. URL: https://tinyurl.com/4fbh8zp3 [accessed 2023-02-13]
39. Zhu H, Colgan J, Reddy M, Choe E. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]

40. Hernández N, Castro L, Favela J, Michán L, Arnrich B. Data quality in mobile sensing datasets for pervasive healthcare. In: Handbook of Large-Scale Distributed Computing in Smart Healthcare. Scalable Computing and Communications. Cham: Springer; 2017:217-238.
41. Ostherr K, Borodina S, Bracken RC, Lotterman C, Storer E, Williams B. Trust and privacy in the context of user-generated health data. Big Data Soc 2017 Apr 17;4(1):205395171770467. [doi: 10.1177/2053951717704673]
42. U.S. Department of Health and Human Services. 2016. URL: https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf [accessed 2023-02-13]
43. New technologies that use patient data. The Academy of Medical Sciences. 2018 May. URL: https://acmedsci.ac.uk/file-download/77418765 [accessed 2023-02-13]
44. Sharko M, Wilcox L, Hong M, Ancker J. Variability in adolescent portal privacy features: how the unique privacy needs of the adolescent patient create a complex decision-making process. J Am Med Inform Assoc 2018 Aug 01;25(8):1008-1017 [FREE Full text] [doi: 10.1093/jamia/ocy042] [Medline: 29788423]
45. Engler R, Routh T, Lucisano J. Adoption barriers for continuous glucose monitoring and their potential reduction with a fully implanted system: results from patient preference surveys. Clin Diabetes 2018 Jan;36(1):50-58 [FREE Full text] [doi: 10.2337/cd17-0053] [Medline: 29382979]
46. Kakkanatt C, Benigno M, Jackson VM, Huang PL, Ng K. Curating and integrating user-generated health data from multiple sources to support healthcare analytics. IBM J Res Dev 2018 Jan 1;62(1):2:1-2:7. [doi: 10.1147/jrd.2017.2756742]
47. Badawy R, Raykov Y, Evers L, Bloem B, Faber M, Zhan A, et al. Automated quality control for sensor based symptom measurement performed outside the lab. Sensors (Basel) 2018 Apr 16;18(4):1215 [FREE Full text] [doi: 10.3390/s18041215] [Medline: 29659528]
48. Genes N, Violante S, Cetrangol C, Rogers L, Schadt E, Chan YF. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018;1:23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
49. Lewinski A, Drake C, Shaw R, Jackson G, Bosworth H, Oakes M, et al. Bridging the integration gap between patient-generated blood glucose data and electronic health records. J Am Med Inform Assoc 2019 Jul 01;26(7):667-672 [FREE Full text] [doi: 10.1093/jamia/ocz039] [Medline: 31192360]
50. Robeznieks A. Xcertia releases updated mobile health-app guidelines for comment. American Medical Association. 2019 Feb 15. URL: https://www.ama-assn.org/practice-management/digital/xcertia-releases-updated-mobile-health-app-guidelines-comment [accessed 2023-02-13]
51. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada URL: https://doi.org/10.1145/3173574.3174058 [doi: 10.1145/3173574.3174058]
52. Ryokai K, Michahelles F, Kritzler M, Syed S. Communicating and interpreting wearable sensor data with health coaches. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul URL: http://doi.org/10.4108/icst.pervasivehealth.2015.259055 [doi: 10.4108/icst.pervasivehealth.2015.259055]
53. King RC, Villeneuve E, White RJ, Sherratt RS, Holderbaum W, Harwin WS. Application of data fusion techniques and technologies for wearable health monitoring. Med Eng Phys 2017 Apr;42:1-12 [FREE Full text] [doi: 10.1016/j.medengphy.2016.12.011] [Medline: 28237714]
54. Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024. Accenture Federal Services for the Office of the National Coordinator for Health Information Technology. 2018 Jan. URL: https://www.healthit.gov/sites/default/files/onc_pghd_practical_guide.pdf [accessed 2023-02-13]
55. Lawton J, Blackburn M, Allen J, Campbell F, Elleri D, Leelarathna L, et al. Patients' and caregivers' experiences of using continuous glucose monitoring to support diabetes self-management: qualitative study. BMC Endocr Disord 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12902-018-0239-1] [Medline: 29458348]
56. Mentis H, Komlodi A, Schrader K, Phipps M, Gruber-Baldini A, Yarbrough K, et al. Crafting a view of self-tracking data in the clinical visit. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA URL: https://doi.org/10.1145/3025453.3025589 [doi: 10.1145/3025453.3025589]
57. Davidson E, Simpson C, Demiris G, Sheikh A, McKinstry B. Integrating telehealth care-generated data with the family practice electronic medical record: qualitative exploration of the views of primary care staff. Interact J Med Res 2013 Nov 26;2(2):e29 [FREE Full text] [doi: 10.2196/ijmr.2820] [Medline: 24280631]
58. Shin M. Secure remote health monitoring with unreliable mobile devices. J Biomed Biotechnol 2012;2012:546021-546025 [FREE Full text] [doi: 10.1155/2012/546021] [Medline: 22910449]
59. Cappon G, Acciaroli G, Vettoretti M, Facchinetti A, Sparacino G. Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment. Electronics 2017 Sep 05;6(3):65. [doi: 10.3390/electronics6030065]

60. Danne T, Nimri R, Battelino T, Bergenstal R, Close K, DeVries J, et al. International consensus on use of continuous glucose monitoring. Diabetes Care 2017 Dec;40(12):1631-1640 [FREE Full text] [doi: 10.2337/dc17-1600] [Medline: 29162583]
61. Shin G, Feng Y, Jarrahi M, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]
62. Codella J, Partovian C, Chang H, Chen C. Data quality challenges for person-generated health and wellness data. IBM J Res Dev 2018 Jan 1;62(1):3:1-3:8. [doi: 10.1147/jrd.2017.2762218]
63. Rutjes H, Kersten-van DE, Willemsen M, IJsselsteijn W. Tell-tale data: the value of self-tracked data for healthcare professionals. In: Proceedings of the 2018 Short Workshop on Next Steps Towards Long Term Self Tracking". 2018 Presented at: 2018 Short Workshop on Next Steps Towards Long Term Self Tracking"; Apr 21, 2018; Montreal, QC, Canada URL: https://research.tue.nl/en/publications/tell-tale-data-the-value-of-self-tracked-data-for-healthcare-prof
64. Cai L, Zhu Y. The challenges of data quality and data quality assessment in the big data era. Data Sci J 2015 May 22;14:2. [doi: 10.5334/dsj-2015-002]
65. Slotwiner D, Varma N, Akar JG, Annas G, Beardsall M, Fogel RI, et al. HRS Expert Consensus Statement on remote interrogation and monitoring for cardiovascular implantable electronic devices. Heart Rhythm 2015 Jul;12(7):e69-100. [doi: 10.1016/j.hrthm.2015.05.008] [Medline: 25981148]
66. Wang E, Zhou L, Parmanto B, Watzlaf V, Abdelhak M. Clinician's perceptions and expectations on a mHealth platform for supporting patient data integration and clinical service delivery: a case study in evidence-based communication rehabilitation. In: Proceedings of the Hawaii International Conference on System Sciences 2018. 2018 Presented at: Hawaii International Conference on System Sciences 2018; Jan 3-6, 2018; Hawaii, USA URL: https://aisel.aisnet.org/hicss-51/hc/it_adoption_in_healthcare/9/ [doi: 10.24251/hicss.2018.386]
67. Jung S, Kim J, Hwang H, Lee K, Baek RM, Lee HY, et al. Development of comprehensive personal health records integrating patient-generated health data directly from Samsung s-health and apple health apps: retrospective cross-sectional observational study. JMIR Mhealth Uhealth 2019 May 28;7(5):e12691 [FREE Full text] [doi: 10.2196/12691] [Medline: 31140446]
68. Botha M, Botha A, Herselman M. Data quality challenges: a content analysis in the e-health domain. In: Proceedings of the 2014 4th World Congress on Information and Communication Technologies (WICT 2014). 2014 Presented at: 2014 4th World Congress on Information and Communication Technologies (WICT 2014); Dec 08-11, 2014; Melaka, Malaysia URL: https://doi.org/10.1109/WICT.2014.7077311 [doi: 10.1109/wict.2014.7077311]
69. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. Front Public Health 2017 Nov 01;5:284 [FREE Full text] [doi: 10.3389/fpubh.2017.00284] [Medline: 29209601]
70. Patient Safety: Achieving a New Standard for Care. Washington (DC): National Academies Press (US); 2004.
71. Health information technology standards 2013. Public Health Data Standards Consortium. URL: https://phdsc.org/standards/health-informa-tion/d_standards.asp [accessed 2023-02-13]
72. Remote Patient Monitoring Playbook. American Medical Association. 2018. URL: https://www.ama-assn.org/system/files/ama-remote-patient-monitoring-playbook.pdf [accessed 2023-02-13]
73. Kim H, Lee S, Baik S, Kim J. MELLO: medical lifelog ontology for data terms from self-tracking and lifelog devices. Int J Med Inform 2015 Dec;84(12):1099-1110. [doi: 10.1016/j.ijmedinf.2015.08.005] [Medline: 26383495]
74. Chung C, Cook J, Bales E, Zia J, Munson S. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]
75. Giordanengo A, Arsand E, Grottland A, Bradway M, Hartvigsen G. Acceptance barriers of using patients' self-collected health data during medical consultation. In: Proceedings of the 17th Scandinavian Conference on Health Informatics. 2019 Presented at: 17th Scandinavian Conference on Health Informatics; Nov 12 -13, 2019; Oslo Norway.
76. Cohen D, Keller S, Hayes G, Dorr D, Ash J, Sittig D. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign. JMIR Hum Factors 2016 Oct 19;3(2):e26 [FREE Full text] [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
77. Kumar R, Goren N, Stark D, Wall D, Longhurst C. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537 [FREE Full text] [doi: 10.1093/jamia/ocv206] [Medline: 27018263]
78. Interoperability and connected healthcare in Australia early engagement paper. Australian Government Australian Digital Health Agency. 2018 Feb 27. URL: http://files-au.clickdimensions.com/digitalhealthgovau-a5xdx/files/earlyengagementpaperv1.0.pdf [accessed 2023-02-13]
79. Personal Connected Health Alliance.Interoperability design guidelines for personal connected health systems. Continua Design Guidelines®. 2017. URL: https://www.pchalliance.org/resources [accessed 2023-02-13]
80. Islind AS, Lindroth T, Lundin J, Steineck G. From narratives to numbers: data work and patient-generated health data in consultations. Stud Health Technol Inform 2018;247:491-495. [doi: 10.3233/978-1-61499-852-5-491]

81. Backonja U, Haynes S, Kim K. Data visualizations to support health practitioners' provision of personalized care for patients with cancer and multiple chronic conditions: user-centered design study. JMIR Hum Factors 2018 Oct 16;5(4):e11826 [FREE Full text] [doi: 10.2196/11826] [Medline: 30327290]
82. Kim Y, Heo E, Lee H, Ji S, Choi J, Kim J, et al. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver, Colorado, USA. [doi: 10.1145/3025453.3025570]
83. Mishuris RG, Yoder J, Wilson D, Mann D. Integrating data from an online diabetes prevention program into an electronic health record and clinical workflow, a design phase usability study. BMC Med Inform Decis Mak 2016 Jul 11;16(1):88 [FREE Full text] [doi: 10.1186/s12911-016-0328-x] [Medline: 27401606]
84. Holt J, Cusatis R, Asan O, Williams J, Nukuna S, Flynn K, et al. Incorporating patient-generated contextual data into care: clinician perspectives using the consolidated framework for implementation science. Healthc (Amst) 2020 Mar;8(1):100369. [doi: 10.1016/j.hjdsi.2019.100369] [Medline: 31445878]
85. Raj S, Lee JM, Garrity A, Newman MW. Clinical Data in Context. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-20. [doi: 10.1145/3314409]
86. Austin R. Picturing patterns in whole-person health: leveraging visualization techniques with structured consumer-generated mHealth data. Libraries Digital Conservancy. 2018 Dec. URL: https://conservancy.umn.edu/handle/11299/202216 [accessed 2023-02-13]
87. Feller D, Burgermaster M, Levine M, Smaldone A, Davidson P, Albers D, et al. A visual analytics approach for pattern-recognition in patient-generated data. J Am Med Inform Assoc 2018 Oct 01;25(10):1366-1374 [FREE Full text] [doi: 10.1093/jamia/ocy054] [Medline: 29905826]
88. Visualising health. Visualising Health Homepage. URL: http://www.vizhealth.org/about/ [accessed 2023-02-13]
89. Petersen C. Patient informaticians: turning patient voice into patient action. JAMIA Open 2018 Oct;1(2):130-135 [FREE Full text] [doi: 10.1093/jamiaopen/ooy014] [Medline: 31984326]
90. Van KJ, Bogers S, Rutjes H, Deckers E, Frens J, Hummels C. Exploring the value of parent tracked baby data in interactions with healthcare professionals: a data-enabled design exploration. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3173871]
91. Harle C, Listhaus A, Covarrubias C, Schmidt S, Mackey S, Carek P, et al. Overcoming barriers to implementing patient-reported outcomes in an electronic health record: a case report. J Am Med Inform Assoc 2016 Jan;23(1):74-79 [FREE Full text] [doi: 10.1093/jamia/ocv085] [Medline: 26159464]
92. Daley C, Chen E, Roebuck A, Ghahari R, Sami A, Skaggs C, et al. Providing patients with implantable cardiac device data through a personal health record: a qualitative study. Appl Clin Inform 2017 Dec 14;08(04):1106-1116. [doi: 10.4338/aci-2017-06-ra-0090]
93. Mirro M, Daley C, Wagner S, Rohani Ghahari R, Drouin M, Toscos T. Delivering remote monitoring data to patients with implantable cardioverter-defibrillators: does medium matter? Pacing Clin Electrophysiol 2018 Nov;41(11):1526-1535. [doi: 10.1111/pace.13505] [Medline: 30225880]
94. Daley C, Toscos T, Mirro M. Data integration and interoperability for patient-centered remote monitoring of cardiovascular implantable electronic devices. Bioengineering (Basel) 2019 Mar 17;6(1):25 [FREE Full text] [doi: 10.3390/bioengineering6010025] [Medline: 30884894]
95. Fajingbesi FE, Olanrewaju RF, Rasool Pampori B, Khan S, Yacoob M. Real time telemedical health care systems with wearable sensors. Asian J Pharmaceutical Res Health Care 2017 Jul 10;9(3):138. [doi: 10.18311/ajprhc/2017/14971]
96. Healthcare coaching: multiplying the value of wearables and patient-generated health data 2018. fitbit Health Solutions. URL: https://2nwchq3a3ags2kj7bq20e3qv-wpengine.netdnassl.com/wpcontent/uploads/FINAL_HIMSS_FitBit_WP_10.01.20181.pdf [accessed 2023-02-13]
97. Ding H, Fatehi F, Russell A, Karunanithi M, Menon A, Bird D, et al. User experience of an innovative mobile health program to assist in insulin dose adjustment: outcomes of a proof-of-concept trial. Telemed J E Health 2018 Jul;24(7):536-543. [doi: 10.1089/tmj.2017.0190] [Medline: 29261476]
98. Williams R, Will C, Weiner K, Henwood F. Navigating standards, encouraging interconnections: infrastructuring digital health platforms. Inform Commun Society 2020 Jul 13;23(8):1170-1186. [doi: 10.1080/1369118x.2019.1702709]
99. Smuck M, Odonkor C, Wilt J, Schmidt N, Swiernik M. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45-48 [FREE Full text] [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
100. Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1051-1056 [FREE Full text] [doi: 10.1093/jamia/ocab017] [Medline: 33822095]
101. Turner K, Jo A, Wei G, Tabriz A, Clary A, Jim H. Sharing patient-generated data with healthcare providers: findings from a 2019 national survey. J Am Med Inform Assoc 2021 Feb 15;28(2):371-376 [FREE Full text] [doi: 10.1093/jamia/ocaa272] [Medline: 33180896]

102. Petersen C. Use of patient-generated health data for shared decision-making in the clinical environment: ready for prime time. Mhealth 2021;7:39 [FREE Full text] [doi: 10.21037/mhealth.2020.03.05] [Medline: 34345616]
103. Rosner B, Kvedar J, Adler-Milstein J. Patient-generated health data earn a seat at the table: clinical adoption during the COVID-19 transition to telemedicine. JAMIA Open 2021 Oct;4(4):ooab097 [FREE Full text] [doi: 10.1093/jamiaopen/ooab097] [Medline: 35155996]
104. Winter JS, Davidson E. Harmonizing regulatory regimes for the governance of patient-generated health data. Telecommun Policy 2022 Jun;46(5):102285. [doi: 10.1016/j.telpol.2021.102285]
105. Melstrom L, Rodin A, Rossi L, Fu JP, Fong Y, Sun V. Patient generated health data and electronic health record integration in oncologic surgery: a call for artificial intelligence and machine learning. J Surg Oncol 2021 Jan;123(1):52-60 [FREE Full text] [doi: 10.1002/jso.26232] [Medline: 32974930]
106. Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. NPJ Digit Med 2020 Mar 13;3(1):37 [FREE Full text] [doi: 10.1038/s41746-020-0237-3] [Medline: 32195372]
107. Cho S, Ensari I, Weng C, Kahn M, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid systematic review. JMIR Mhealth Uhealth 2021 Mar 19;9(3):e20738 [FREE Full text] [doi: 10.2196/20738] [Medline: 33739294]

Abbreviations

CGM: continuous glucose monitoring
DQM: data quality management
EMR: electronic medical record
PGHD: patient-generated health data
RPM: remote patient monitoring

Edited by L Buis; submitted 23.12.21; peer-reviewed by E Krupinski, B Johnson, D Snider, M Mars; comments to author 23.02.22; revised version received 01.04.22; accepted 20.12.22; published 24.02.23.

Please cite as:
Abdolkhani R, Gray K, Borda A, DeSouza R
Recommendations for the Quality Management of Patient-Generated Health Data in Remote Patient Monitoring: Mixed Methods Study
JMIR Mhealth Uhealth 2023;11:e35917
URL: https://mhealth.jmir.org/2023/1/e35917
doi:10.2196/35917
PMID:36826986

Review

Interventions Aimed at Enhancing Health Care Providers' Behavior Toward the Prescription of Mobile Health Apps: Systematic Review

Ohoud Alkhaldi[1,2], BSc, MSc; Brian McMillan[3], BSSc, MBChB, PhD; Noha Maddah[1,4], BSc, MSc; John Ainsworth[1], MSc, PhD

[1]Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[2]Health Information Management and Technology Department, College of Public Health, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

[3]Centre for Primary Care and Health Services Research, The University of Manchester, Manchester, United Kingdom

[4]Department of Health Services and Hospitals Administration, Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author:
Ohoud Alkhaldi, BSc, MSc
Division of Informatics, Imaging and Data Sciences, School of Health Sciences
Faculty of Biology, Medicine and Health
The University of Manchester
Vaughan House
Portsmouth St
Manchester, M13 9GB
United Kingdom
Phone: 44 7415085174
Email: ohoud.alkhaldi@postgrad.manchester.ac.uk

Abstract

Background: Mobile health (mHealth) apps have great potential to support the management of chronic conditions. Despite widespread acceptance of mHealth apps by the public, health care providers (HCPs) are reluctant to prescribe or recommend such apps to their patients.

Objective: This study aimed to classify and evaluate interventions aimed at encouraging HCPs to prescribe mHealth apps.

Methods: A systematic literature search was conducted to identify studies published from January 1, 2008, to August 5, 2022, using 4 electronic databases: MEDLINE, Scopus, CINAHL, and PsycINFO. We included studies that evaluated interventions encouraging HCPs to prescribe mHealth apps. Two review authors independently assessed the eligibility of the studies. The "National Institute of Health's quality assessment tool for before-and-after (pretest-posttest design) studies with no control group" and "the mixed methods appraisal tool (MMAT)" were used to assess the methodological quality. Owing to high levels of heterogeneity between interventions, measures of practice change, specialties of HCPs, and modes of delivery, we conducted a qualitative analysis. We adopted the behavior change wheel as a framework for classifying the included interventions according to intervention functions.

Results: In total, 11 studies were included in this review. Most of the studies reported positive findings, with improvements in a number of outcomes, including increased knowledge of mHealth apps among clinicians, improved self-efficacy or confidence in prescribing, and an increased number of mHealth app prescriptions. On the basis of the behavior change wheel, 9 studies reported elements of environmental restructuring such as providing HCPs with lists of apps, technological systems, time, and resources. Furthermore, 9 studies included elements of education, particularly workshops, class lectures, individual sessions with HCPs, videos, or toolkits. Furthermore, training was incorporated in 8 studies using case studies or scenarios or app appraisal tools. Coercion and restriction were not reported in any of the interventions included. The quality of the studies was high in relation to the clarity of aims, interventions, and outcomes but weaker in terms of sample size, power calculations, and duration of follow-up.

Conclusions: This study identified interventions to encourage app prescriptions by HCPs. Recommendations for future research should consider previously unexplored intervention functions such as restrictions and coercion. The findings of this review can help inform mHealth providers and policy makers regarding the key intervention strategies impacting mHealth prescriptions and assist them in making informed decisions to encourage this adoption.

(JMIR Mhealth Uhealth 2023;11:e43561) doi:10.2196/43561

KEYWORDS

mHealth; mobile apps; prescription; behavioral change; mobile phone

Introduction

Background

The number of patients living with chronic conditions continues to increase worldwide [1], and empowering these patients to manage their diseases is vital. Mobile health (mHealth) provides digital solutions to patients to help them track and manage their diseases. With the increased number of available mHealth apps to download and use [2], it is expected that the number of consumers, whether they are members of the general public, patients, or health care providers (HCPs), will continue to grow. The purpose of different types of mHealth apps vary from well-being, prevention, management, and monitoring to follow up with HCPs. Some of these apps may be potentially suitable to be prescribed to patients for the diagnosis or treatment of medical conditions. The concept of "prescribable" health apps is recently introduced to refer to health apps that are currently available and have demonstrated effectiveness [3].

Studies evaluating the effectiveness of mobile apps on health outcomes are increasing in number. Several systematic reviews have concluded that mobile apps have the potential to improve patients' health conditions, such as diabetes [4], mental health [5], and cardiovascular diseases [6]. Governments in several countries have acknowledged the benefits of mHealth and have endeavored to meet the urgent need to accelerate the adoption of mHealth apps. Germany became the first country in the world to prescribe mobile apps. HCPs can prescribe mHealth apps, which can be reimbursed by health insurance companies [7]. In the United Kingdom, the National Institute for Health and Care Excellence has published guidance about "Sleepio," a digital therapeutic to treat insomnia, and has recommended the use of Sleepio as a cost-saving option in comparison with sleep hygiene or sleeping pills [8]. On the basis of the results of 28 studies, it has been concluded that Sleepio is more effective than the usual treatment in reducing symptoms of insomnia in adults [9].

In a survey study, HCPs from the United Kingdom were more likely to prescribe apps if they were tagged with National Health Service approval or recommended by work colleagues [10]. The same study reported that National Health Service–approved mHealth apps were more influential than evidence-based research. In Germany, a study of physicians' attitudes toward prescribable mHealth apps found that only one-third of the physicians intended to prescribe apps, and the rate of HCPs who had already prescribed them was lower than expected, despite the existence of regulations and facilitators from the government accelerating the mHealth app adoption among HCPs [11]. The study authors suggested that a range of factors influenced app prescribing, including gender, age, the lack of intention to prescribe, and limited apps for some specialties.

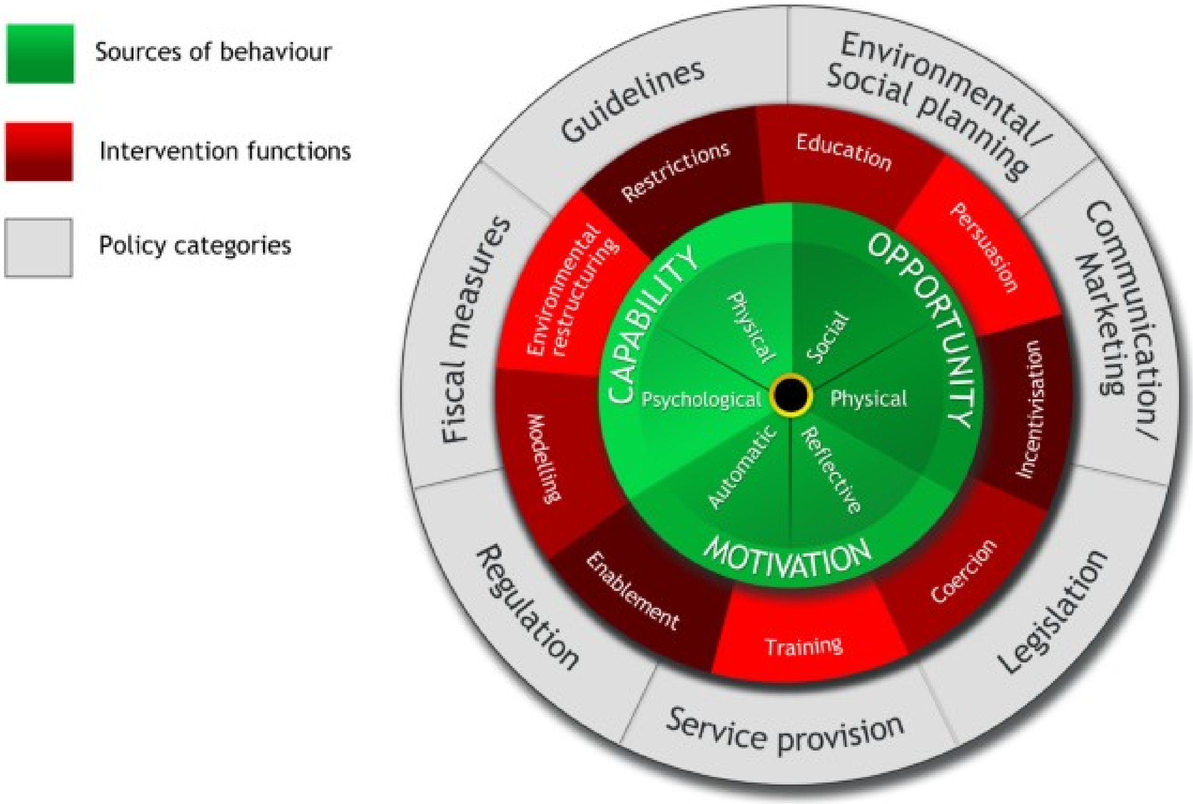
These studies shed light on various barriers to mHealth app adoption in clinical care and provide opportunities to design future behavior change interventions to improve HCPs' evidence-based app prescription behaviors. To date, there have been no systematic or comprehensive literature reviews that compile evidence of interventions for enhancing HCPs' app prescription behaviors. Bringing together the findings from such interventions could potentially provide policy makers and stakeholders with a better understanding of valuable strategies that can be implemented to enhance HCPs' uptake of mHealth apps. In this review, we address this gap by identifying interventions that aim to redirect HCPs' behavior toward prescribing or recommending apps to patients.

Behavior Change Framework

Several approaches are available to guide behavior change intervention designs. Among these approaches are the person-based approach [12], the British Medical Research Council's framework on the development and evaluation of complex interventions [13], and intervention mapping [14]. Although each of these approaches offers considerable value to researchers, each concentrates on a different component of intervention development or has been criticized for lacking comprehensiveness and coherence [15].

The behavior change wheel (BCW) is used to characterize and evaluate behavior change interventions [15]. This framework provides a comprehensive approach to identifying sources of behavior and classifying them into the capability, opportunity, motivation, and behavior (COM-B) model, which represents the wheel's hub (Figure 1). These components interact with each other to produce a change in behavior. Surrounding this is a layer of 9 intervention functions that can be selected depending on the behavioral analysis reached with the COM-B. The final layer contains 7 types of policies that one can use to deliver these intervention functions. The intervention functions are connected to behavior change techniques, which are the smallest active elements of an intervention (eg, self-monitoring, goal setting, action planning etc) [16]. Behavior change techniques used in interventions can be categorized using a taxonomy comprising 93 different techniques.

Figure 1. The behaviour change wheel (reproduced from Michie et al et al[15], which is published under Creative Commons Attribution 4.0 International License [17]).



Understanding the target behavior is essential before designing an intervention. However, the BCW can also be applied retrospectively to intervention studies to identify and describe the behavior change strategies that have been used. It can also be used to improve current interventions or to introduce and evaluate an intervention that looks promising. Therefore, the interventions included in this review are classified into intervention functions of the BCW.

Objectives

The study objectives were (1) to summarize and evaluate interventions aimed at encouraging HCPs to prescribe mHealth apps to patients and (2) to classify and map intervention strategies with the intervention functions of the BCW.

Methods

Research Question

This study was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [18]. The research question was created based on the PICO framework (population, intervention, comparator, and outcome) and was defined as follows: "How do interventions designed to encourage mobile health app prescription change the practices, knowledge or self-confidence of healthcare providers?"

Eligibility Criteria

The population of interest was HCPs or trainees (eg, health personnel, general practitioners [GPs], physicians, clinicians, dietitians, and students of health specialties). Intervention studies to encourage mHealth app prescriptions, regardless of the design, were considered eligible. The primary outcome of interest reflected any measures of practice or behavior changes such as number of app prescriptions or self-reported or objectively measured changes in knowledge or confidence. The included studies had to be conducted in primary care settings. Studies were excluded if they were about patients' adoption of mHealth apps or interventions to improve patients' health outcomes. Other mHealth technologies such as wearables, mobile phone messaging, video consultations, or electronic health records (EHRs) were excluded. Studies of apps for HCPs' medical education and training or decision support systems via mobile devices were also excluded.

Sources of Information and Search

A search of 4 electronic databases (MEDLINE, Scopus, CINAHL, and PsycINFO) was conducted to identify studies published between January 1, 2008, and August 5, 2022. The official start of mobile apps was chosen as 2008, as the iPhone App Store was launched that year [19]. Studies published in English and peer-reviewed papers were included. A manual search of the reference lists of eligible studies was conducted. Medical Subject Headings terms were used wherever possible to locate the relevant studies. The Boolean operators AND and OR were used to enhance the search strategy. The search string used in 2 databases is shown in Multimedia Appendix 1.

Study Selection

The search results were exported to the EndNote Web software (Clarivate Analytics) for screening and removing duplicates. After eliminating duplicates, screening of all titles and abstracts was independently conducted by 2 reviewers (OA and NM). The same 2 researchers reviewed the full texts of the papers identified as relevant to the objectives. In cases of disagreement, the research team discussed them and made the final decision.

Risk of Bias Assessment

The risk of bias was assessed using 2 quality appraisal tools based on the study design. The National Institute of Health's quality assessment tool for before-and-after (pretest-posttest design) studies with no control group was used for uncontrolled before-and-after studies [20]. This tool is composed of 12 items with response options "yes," "no," "not reported," and "not applicable," and the overall quality of each study can be classified as "good," "fair," or "poor." The grading was decided by the total score: 0 to 4 (poor), 5 to 9 (fair), and 10 to 12 (good).

The mixed methods appraisal tool (MMAT) was used to assess the quality of the remaining studies. The MMAT is a comprehensive tool for assessing the quality of quantitative, qualitative, and mixed methods study designs [21]. This tool begins with 2 screening questions to determine whether a research objective is clear and whether the collected data allow a research question to be answered. The remaining 5 questions assess the methodologies. A score of 0- 2 was considered low; a score of 3-4 was considered moderate; and a score of 5 was considered high.

Data Collection and Synthesis

Data Extraction

After the study selection, data were extracted from eligible studies. The following data were extracted: study characteristics (author, year of publication, country, aim, types of mHealth apps used, mode of delivery, length of study/number of sessions, study design, and sample size); outcomes of each study; and main findings related to the research question of this systematic review. One reviewer (OA) performed the data extraction, and the research team checked the accuracy of the extracted data.

Data Synthesis

The diversity of measures and outcomes identified in the eligible studies did not allow for quantitative data synthesis; therefore, a narrative synthesis was conducted. The 9 intervention functions of the BCW [15] were used to classify intervention strategies to help inform future attempts to design interventions.

Each intervention was categorized as performing one or more of the 9 functions [22]. Definitions of the intervention functions are listed in Textbox 1.

Textbox 1. Definitions of intervention functions in the behavior change wheel.

Education

- Increasing knowledge or understanding

Persuasion

- Using communication to induce positive or negative feelings or stimulate action

Incentivization

- Creating an expectation of reward

Coercion

- Creating an expectation of punishment or cost

Training

- Imparting skills

Restriction

- Using rules to reduce the opportunity to engage in the target behavior (or to increase the target behavior by reducing the opportunity to engage in competing behaviors)

Environmental restructuring

- Changing the physical or social context

Modeling

- Providing an example for people to aspire to or imitate

Enablement

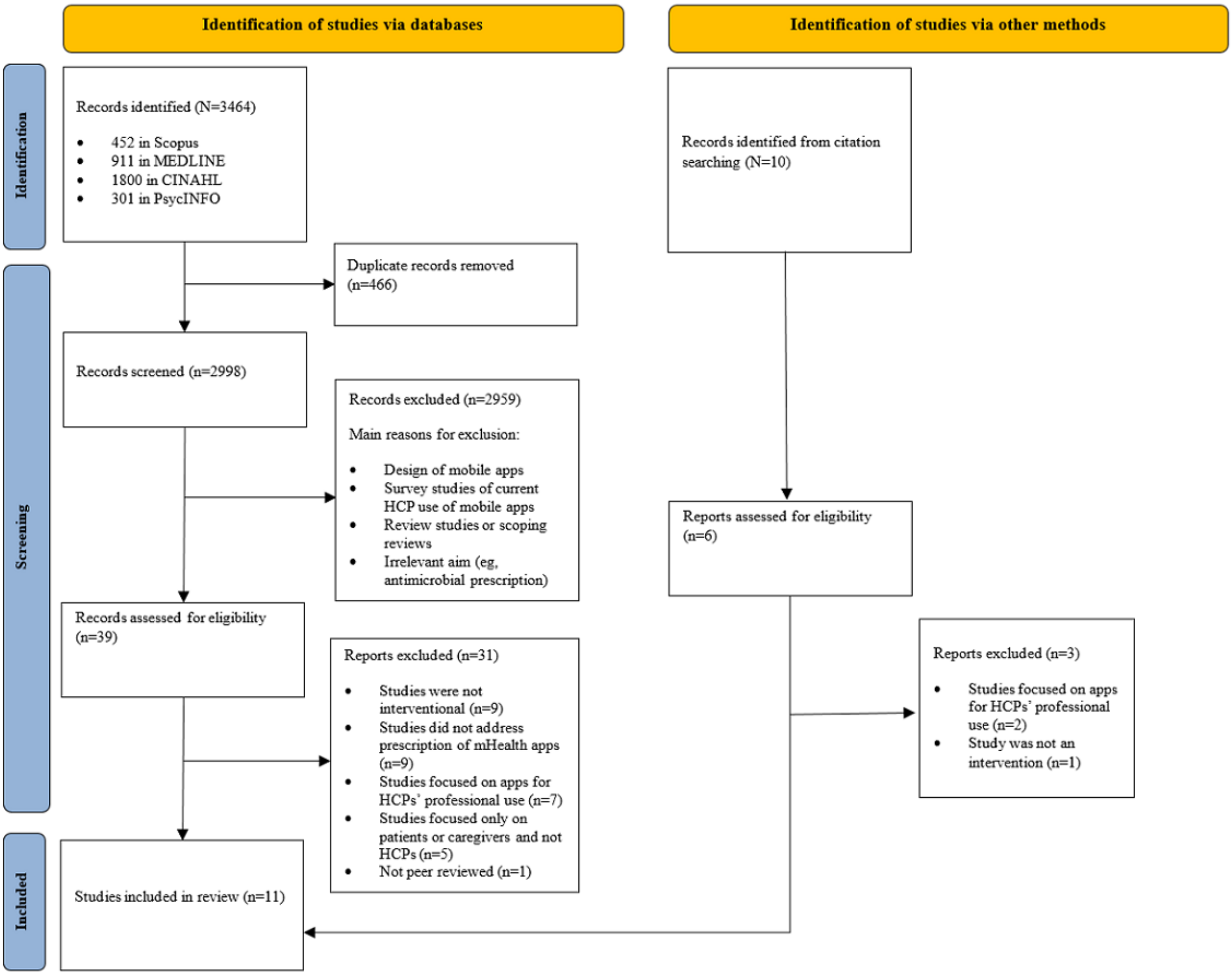
- Increasing means or reducing barriers to increase capability or opportunity beyond environmental restructuring

Results

Study Selection

The search strategy retrieved 3464 records. Of these, 466 (13.6%) studies were duplicates and were removed, leaving 2998 (86.5%) studies for screening. The screening of titles and abstracts excluded 2959 (98.7%) studies. Therefore, 39 (1.3%) studies were eligible for full-text screening. Another 10 records were identified through citation searching, and 6 (60%) of these were included for full-text screening. Full-text screening of the 45 studies yielded a total of 11 (24%) studies in the final review. The study selection process and the reasons for exclusion are shown in Figure 2.

Figure 2. Study selection flow diagram based on the PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) guidelines. HCP: health care professional; mHealth: mobile health.



Characteristics of the Included Studies

The studies were published between 2015 and 2020 and were performed in the United States (6/11, 54%), Australia (2/11, 18%), the Netherlands (1/11, 9%), Catalonia (1/11, 9%), and the United Kingdom (1/11, 9%). All the included studies contained an interventional component. Of the total 11 studies, 6 (54%) were pretest-posttest design studies with no control group; 3 (27%) were mixed methods studies; 1 (9%) was a usability study; and 1 (9%) was a qualitative description study.

The studies varied regarding study participants; some were focused on HCPs in primary care settings [23-25], whereas others targeted specific specialties, such as dietitians [26,27], behavioral health providers [28,29], providers in weight management clinics [30], clinical nutrition and physician assistant students [31], or interdisciplinary groups [32,33]. Most studies (8/11, 73%) had a sample size ranging from 5 to 40 HCPs, apart from 3 (27.3%) studies that reported the results of interventions conducted over multiple years or many training sessions in which the sample size ranged from 78 to 760 [28,29,31]. The functions of the mHealth apps used in these studies included weight management [27,30] or diet and activity tracking [26,31]. A total of 45% (5/11) of the studies used a list of approved apps for a range of health conditions [23,25,28,29,32], and 9% (1/11) used an app for chronic obstructive pulmonary disease (COPD) [33]. An overview of the characteristics of the individual studies included in the systematic review is provided in Multimedia Appendix 2 [23-33].

The most common changes resulting from the intervention were self-reported changes in knowledge, confidence, or self-efficacy [28,30-33]. Only 1 (1/11, 9%) study reported HCPs' intention to use or recommend the app [33], and 1 (1/11, 9%) study evaluated app acceptance and use in dietetic care [27]. However, the outcomes were objectively estimated using the change in the number of apps prescribed in 36% (4/11) studies [23-25,32].

Quality of the Evidence and Risk of Bias in the Included Studies

Using the National Institute of Health quality assessment tool, the studies by Armstrong et al [28,29], Byambasuren et al [25], Chen et al [26], and Rodder et al [31] were assessed as having a moderate risk of bias. The study by Al-Lami et al [30] was assessed as having a high risk of bias owing to a lack of clarity around several categories that were not reported, particularly regarding the selection and eligibility of participants. In all studies (6/11, 54%), it was difficult to determine whether the researchers were blinded to the intervention. An additional limitation of many of the included studies was their small sample size. Of the 11 studies, this was particularly the case for 2 studies (18%): one with 5 participants [26] and the other with 6 [30]. Most studies (5/6, 83%) lacked power calculations to accommodate the consequences of participants dropping out, thus resulting in missing values in the postintervention measurements, with the exception of Byambasuren et al [25]. The studies by Armstrong et al [28,29] and Byambasuren et al [25] did not present *P* values to compare outcome measures at pretest and posttest. The details of the quality assessment are presented in Multimedia Appendix 3 [23-33].

Using the MMAT, the qualitative description study was deemed high quality [27]. Out of the 3 mixed methods studies, 1 (33%) was judged to be of low quality [27] and 2 (67%) were of moderate quality [24,33]. The independent assessment of their quantitative and qualitative components lowered the methodological quality of all mixed methods studies. The quantitative study was deemed to be of low quality [32] owing to concerns regarding the poor description of patient selection and the high nonresponse rate. However, all studies clearly stated the objectives, interventions, and outcomes. The follow-up periods were generally short (immediately after the intervention up to 4 months after the intervention). More details on the quality assessments for each study are shown in Multimedia Appendix 3.

Effectiveness of Interventions

The results reported by the interventions included were generally positive, with improvements being seen in several outcomes such as changes in current practices, increased knowledge, or improved self-efficacy. Although the levels of significance for the included outcomes varied, only 1 (9.1%) study found that the interventions had nonsignificant results [27].

Changes in Practice

A total of 54% (6/11) of studies reported changes in HCP practice following the intervention. Byambasuren et al [25] reported effective changes in GPs' prescriptions of mHealth apps. Although the study did not report a *P* value for the number of prescriptions at pretest and posttest, the number of apps recommended per GP per fortnight increased from 1.7 to 4.1. The use of videos had no significant impact on the number of app prescriptions. Similarly, Makhni et al [32] reported the results of an 8-week trial; users at the 5 clinical sites prescribed more than 2000 apps; this exceeded the adoption targets, which was set at 100 health apps. In the study by Segui et al [23], the use of the AppSault platform to prescribe apps was reported. A total of 32 doctors made 79 app recommendations to patients, which represented 79% of the recommendations compared with what was expected during the pilot design [23]. This increase in the percentage of platform use was seen as a successful change in current practice. The staff use of apps was reported in the study by Hoffman et al [24], but it was self-reported through questionnaires. Clinical staff were receptive to apps, with 83% (19/23) incorporating behavioral health apps into their clinical work, and 25% (5/20) introducing apps to patients up to 50% of the time.

In total, 2 (18%) studies measured changes in practice using qualitative methods. First, the intervention by Korpershoek et al [33] measured the feasibility of using the Copilot app and reported HCPs' high satisfaction and high levels of interest in the app. They also believed that the app was user-friendly and relevant to daily practice and that it fit well within the organizational culture. Second, Barnett et al [27] reported the myPace app's acceptance among HCPs who were positive about the app. However, the uptake and recommendation of the myPace app were lower than expected.

Change in Knowledge

Of the 11 studies, 3 (27%) studies reported changes in the knowledge of HCPs about mHealth apps. The core competency training designed and delivered by Armstrong et al [28,29] was successful in transferring the knowledge of the enrolled clinicians. One of these studies reported the results of 3 years of the training program [28]. There was a 28.96% increase in HCPs' self-reported knowledge of the use of mHealth in clinical care when comparing pretraining measurements (mean 2.97, SD 1.07; n=537) with posttraining measurements (mean 4.31, SD 0.76; n=537). The other study reported the results of 1 year of training and showed that the number of HCPs who rated their overall knowledge of the use of mHealth in clinical care as good or excellent before training increased from 34% (67/199) to 93% (185/199) after the training [29]. The intervention of Al-Lami et al [30] consisted of providing HCPs with a list of evaluated apps to make recommendations from, educating them about the efficacy of using mHealth apps in weight management and training them in the use of apps' critical appraisal tools. A significant knowledge increase was reported (P=.02).

Changes in Confidence and Self-efficacy

In the remaining 2 studies, the curriculum expansion to enable physician assistant (PA) and clinical nutrition (CN) students to use mHealth apps in clinical care yielded increased self-reported confidence in their skills from pretest to posttest ($P \leq .001$) [31]. The findings were supported by students' Objective Structured Clinical Examination (OSCE) scores, which showed that both PA and CN students effectively taught standardized patients to use mobile apps for disease management. In the intervention by Chen et al [26], dietitians rated their self-efficacy before and after completing an educational and skill-training session on apps and after receiving 12 weeks of real-world experience using mHealth apps in their practice. A significant improvement in dietitians' overall self-efficacy with mHealth apps was reported (ANOVA $F_{2, 12}$=7.0; P=.01).

Intervention Functions

The included studies were analyzed using the BCW framework. By doing so, the framework allowed an examination of which intervention functions are most commonly applied in the context of mHealth app prescriptions. Given that some strategies can be classified as falling into >1 category, that is, multiple intervention functions, it was difficult to link outcomes to a single intervention function. Therefore, the following sections report how the intervention strategies fit within the BCW's 9 intervention functions. The intervention functions, outcomes, and main findings are reported separately in Table 1.

Table 1. Intervention functions adopted in each of the reviewed studies and main findings.

Author and year	Education	Persuasion	Incentivization	Coercion	Training	Restriction	Environmental restructuring	Modeling	Enablement	Main findings
Armstrong et al [28], 2018 and Armstrong et al [29], 2019[a]	7-hour CE[b] workshop	Interactive material to allow learners to engage with the material	N/A[c]	N/A	1. mHealth[d] apps used as examples for hands-on experience 2. Interpersonal skills on how to discuss mobile apps with patients	N/A	N/A	Site champion to offer additional training	Site champion to offer support	• In total, 93.7% reported that the information and skills learned from the training would be used in their clinical care [28]. • In total, 95.8% reported that the information and skills learned from the training would be used in their clinical care [29].
Byambasuren et al [25], 2020	A letter with a brief description of each app and the study time frames and procedures was mailed.	1. Having prescription pads worked as a visual reminder or cue to prescribe apps 2. Videos after 2 months demonstrating the content, features, and function of each app	N/A	N/A	N/A	N/A	Prescription pads were developed and given to GPs[e] with apps that are relevant in general practice.	N/A	N/A	• 1324 app prescriptions were dispensed over 4 months • The GPs' confidence in prescribing apps doubled from a mean of 2 (not so confident) before the study to 4 (very confident) at the end of the study

Author and year	Education	Persuasion	Incentivization	Coercion	Training	Restriction	Environmental restructuring	Modeling	Enablement	Main findings
Chen et al [26], 2019	Workshop to educate dietitians about a range of mHealth apps	Verbal persuasion about capabilities to master app use even in difficult situations	N/A	N/A	1. Case studies 2. Dietitians were trained to appraise the quality of these apps	N/A	Easy Diet Diary Connect platform	1. Workshop facilitator is a dietitian modeling and working with apps 2. Working with colleagues enabled social comparisons to be made	Ongoing support during 12 weeks of the intervention	• A significant improvement in overall self-efficacy with using mHealth apps (ANOVA $F_{2,12}$=7.0; P=.01)
Rodder et al [31], 2018	Curriculum expansion to educate students on the use of mHealth	N/A	Pass the OSCE[f] assessment	N/A	1. Students trained in how to evaluate apps using the SAAT[g] appraisal tool 2. Students trained in interpersonal skills, such as how to educate patients 3. Students trained to download and use recommended apps using case studies	N/A	MyNetDiary and Withings Health Mate apps	Peer comparison	N/A	• Confidence levels improved significantly for all survey measures, for both PA[h] and CN[i] students (P≤.001) • OSCE results showed that both PA and CN students were able to download MyNetDiary (96.4%), enter food into the app (98.4%), and discuss the advantages of using the app for food tracking with patients (90.3%).
Al-Lami et al [30], 2019		N/A	N/A	N/A	Training in how to use the Ped-WHAT App appraisal tool before making app recommendations		Providing HCPs[j] with a list of evaluated apps from which to make recommendations	N/A	N/A	

Author and year	Education	Persuasion	Incentivization	Coercion	Training	Restriction	Environmental restructuring	Modeling	Enablement	Main findings
	Workshop to educate clinical staff about the efficacy of using mHealth apps in weight management									• Provider knowledge of the use of apps significantly increased after the training (mean 1.00, SD 1.00 vs mean 1.67, SD 0.52; *t*=3.16; *P*=.025). • Provider confidence in recommending apps to patients increased significantly after the training (mean 1.00, SD 0 vs mean 1.67, SD 0.52; *t*=3.16; *P*=.025)
Makhni et al [32], 2017	N/A	N/A	N/A	N/A	Personal training in the use of the RxUniverse platform	N/A	RXuniverse app-prescribing system	Demonstrating a trial process for prescribing an app	Considerations of the workflow to minimize disruption and time burdens of participants	• During an 8-week trial, over 2000 apps were prescribed to all users in the 5 clinical sites • Users felt that RxUniverse performed well. The group mean for the overall SUS[k] score was 84.2, an “excellent” rating.
Hoffman et al [24], 2019	Series of staff meetings on best practices for using mental health apps within clinical care	N/A	N/A	N/A	N/A	N/A		N/A	N/A	

Author and year	Education	Persuasion	Incentivization	Coercion	Training	Restriction	Environmental restructuring	Modeling	Enablement	Main findings
							1. A list of recommended apps for self-management 2. Two EHR[l] standardized "smart phrases" to facilitate the use of apps 3. A guide to staff was created as to when to introduce and discuss apps with patients			• In total,–82.6% incorporated BH[m] apps into their clinical work; 25% introduced apps to patients 25% to-50% of the timeIn total, 42% expressed a need for more practice and training in using each tool within the CHA's[n] mobile app toolkit.
Korperhoek et al [33], 2020	HCPs were introduced to the Copilot app and the scenario for use of the app in daily practice, and the intended role of both HCPs and patients	N/A	N/A	N/A	A fictional patient scenario was given to HCPs, who were asked to conduct several tasks	N/A	Copilot app for COPD[o] self-management	N/A	N/A	• Main themes: high satisfaction, user-friendly, relevant for daily practice, app fit well within the organizational culture, high level of interest • An average score of 83.8 (SD 15.1) on the SUS, indicating good usability of the app
Barnett et al [27], 2015	N/A	N/A	N/A	N/A	Personal training in the use of myPace software and how to make app recommendations to patients	N/A	MyPace app for weight loss designed to fit daily dietetic practice	N/A	N/A	• The dietitians were positive and enthusiastic about the app; however, their enthusiasm did not translate into actual uptake, use, and recommendation
Segui et al [23], 2018		Follow-up and monitoring	N/A	N/A	N/A	N/A	AppSalut platform to prescribe apps	N/A		

Author and year	Education	Persuasion	Incentivization	Coercion	Training	Restriction	Environmental restructuring	Modeling	Enablement	Main findings
	Instructions to train HCPs in prescribing and using the platform								Support by periodic follow-ups accompanied by training for doctors and solving any technical problems	• A total of 32 doctors made 79 app recommendations to patients, representing 160% of doctors and 79% of recommendations compared with what was expected

[a]Studies are combined owing to similar intervention components.
[b]CE: continuing education.
[c]N/A: not applicable.
[d]mHealth: mobile health.
[e]GP: general practitioner.
[f]OSCE: Objective Structured Clinical Examination.
[g]SAAT: smartphone application appraisal tool.
[h]PA: physician assistant.
[i]CN: clinical nutrition.
[j]HCP: health care practitioner.
[k]SUS: system usability score.
[l]EHR: electronic health record.
[m]BH: behavioral health.
[n]CHA: Cambridge Health Alliance.
[o]COPD: chronic obstructive pulmonary disease.

Education

In total, of the 11, 9 (81.8%) studies included elements of education. Education came in the form of workshops in 4 studies [26,28-30]. The workshops covered the best practices for using mHealth apps in patient care and considerations of privacy, security, and ethical and cultural issues of using mHealth apps with patients [28,29]. The workshop in the study by Chen et al [26] educated dietitians about the range of commercially available apps. Similarly, the workshop in the study by Al-Lami et al [30] provided background information on mobile apps' efficacy and validity in weight management therapy. It introduced a list of apps to use when making recommendations to patients. However, Hoffman et al [24] used a series of staff meetings as the mode of delivery to educate behavioral health staff about the best practices for using mHealth apps with patients. One study, which was conducted in an academic health center, also involved education through curriculum expansion and focused on mHealth apps [31].

In the study by Byambasuren et al [25], education was part of the intervention in 2 ways. A letter was mailed to each participant describing the study and containing instructions and guides on prescribing the app . The second education element was also delivered via videos showing the app's content, features, and functions. In this study, the authors aimed to assess the impact of videos on the number of app prescriptions.

Creating guides or providing HCPs with instructions on when to introduce and discuss apps with patients were carried out in 2 studies as an educational form [23,24]. Another study used education; however, the content was tailored to the platform under testing. Koreospek et al [33] introduced the Copilot app to study participants and explained the intended use of the app for the self-monitoring of symptoms by patients with COPD. Moreover, a tutorial was given on how to use the app and perform essential functions, such as registering patients and customizing an action plan.

These educational elements of the interventions aimed to improve HCPs' knowledge. In addition, of the 11 studies, 3 (27%) captured and reported changes in self-reported knowledge before and after the intervention [28-30]. The remaining studies measured different outcomes, such as a change in self-confidence [25,31], self-efficacy [26], or an increased number of prescriptions [25]. However, these measures were reported as the results of the intervention as a whole and were not specific to education.

Persuasion

A total of 45% (5/11) of the studies reported elements of persuasion. Periodic follow-ups with study participants served as a method of reminding and motivating HCPs to modify their prescribing habits [23]. Visual reminders were used in another study in 2 forms. First, videos were sent to study participants after month 2 of the intervention [25]. These videos not only were educational but also worked as a tool to remind study participants of the study. Second, the design and dissemination of prescription pads involved reminders or cues to prescribe apps.

Persuasion was reflected in both studies by Armstrong et al [28,29]. The development of educational materials involved adopting evidence-based interactive educational experiences to allow learners to engage in the material and ensure promising results regarding behavior change. In a study by Chen et al [26], persuasion was sought verbally to convince dietitians of their capabilities to prescribe mHealth apps, even in difficult situations such as short consultations.

Incentivization

Incentivization can be social, such as a promotion in status, or fiscal. Students' desires to pass the OSCE exam by demonstrating their capabilities to recommend and use mHealth apps in nutrition care worked as a social incentive [31]. According to the OSCE results, PA and CN students successfully taught patients how to use mobile apps to track food intake and test blood pressure. None of the remaining interventions used strategies that fall into the incentivization category.

Training

A total of 8 (72.7%) interventions reported elements of training to improve HCPs' skills when using apps. The focus and range of skills offered to HCPs varied in each study. Three interventions used case studies or scenarios to help HCPs develop and master the basic skills of using and recommending apps [26,31,33]. Training on how to use Ped-WHAT, an app appraisal tool, when making recommendations to patients and their families was critical in the intervention by Al-Lami et al [30]. Similarly, Rodder et al [31] trained students to evaluate mobile apps using a smartphone app appraisal tool.

Participants received personal training in using myPace software and were allowed to practice with it [27]. They were also allowed to provide feedback on their first impression of the software during these training sessions and were encouraged to use or recommend the app to make it a standard tool to support everyday practice. The training in the intervention by Makhni et al [32] intervention was also individualized. HCPs were instructed on how to use the RxUniverse system to make app prescriptions and were then observed to ensure successful app prescription.

The term "hands-on experience" was used in the study by Armstrong et al [28] to refer to practical experience; however, if training was involved, it was not detailed enough. Interpersonal skills were targeted in 2 interventions [28,29]. Competencies in discussing mobile apps with patients and showing an understanding of patients' concerns about privacy and security were at the core of these interventions. Rodder et al [31] also mentioned measuring students' communication skills in OSCE examinations. Students were asked to demonstrate skills in discussing the benefits of using apps for food tracking and blood pressure measurement.

Environmental Restructuring

To promote app prescription behavior among HCPs, 9 interventions contained methods of environmental restructuring. Most included studies offered technical resources to facilitate app prescriptions, such as developing an electronic platform. The study by Segui et al [23] tested the feasibility of using an app catalog named AppSault for recommending apps to patients. The apps were free to download and passed the quality control process, guaranteeing a safe and reliable environment for their use. Makhni et al [32] used RxUniverse, a digital medicine-focused care delivery platform with a library of apps chosen based on published evidence-based reviews of their efficacy and usability. Some studies provided HCPs with specialized apps and measured how they fit daily practice. The MyNetDiary and Withings Health Mate apps were provided for PA and CN students in a study by Rodder et al [31], myPace for dietitians in a study by Barnett et al [27], and Copilot for COPD self-management in a study by Korpershoek et al [33]. The Easy Diet Diary Connect platform was used to allow dietitians to track patients in a study by Chen et al [26]. Other studies have used lists of apps given to HCPs to make recommendations [24,25,30].

Environmental restructuring came in the form of EHRs' standardized "smart phrases" to facilitate the prescription process [24].

Modeling

A total of 5 (45.5%) studies reported on methods of modeling [26,28,29,31,32]. The intervention by Armstrong et al [28,29] used "site champions" as onsite facilitators who offered additional training to local staff. Modeling was presented in different forms in the intervention by Chen et al [26]. First, the workshop moderator, who was also a dietitian, like the other study participants, modeled and used the app. Second, during the intervention, working with colleagues who had participated in the study and were successful in using the app enabled a social comparison to be made. Another study used modeling by comparison and competition among students or peers of each class [31]. Makhni et al [32] used a similar modeling method by demonstrating the functionality of the platform to participants who were thereafter asked to prescribe an mHealth app.

Enablement

Enablement was used in 5 (45.5%) studies. In 1 study, enablement came in the form of ongoing support throughout the study to reduce barriers associated with prescribing mHealth apps [26]. Support was also provided to study participants by consulting office managers and study participants about the specific operational workflows of each clinical site and the optimal implementation plan for RxUniverse at each pilot site to minimize the time burden [32]. The periodic follow-ups in the study by Segui et al [23] were accompanied by solving any technical problems, which enabled the implementation of corrective actions and extra training [23]. Recognizing site champions in Armstrong et al [28,29] offered support to sustain behavior change after the training.

Coercion and Restriction

None of the interventions included in this review used the intervention functions of coercion or restriction.

Linking Interventions With the COM-B Model

We looked at the most often used intervention functions in the studies with successful outcomes to gain a better understanding of why the included intervention functions provided substantial changes. As seen in Table 1, successful interventions used a variety of intervention functions (environmental restructuring, education, and training). These intervention functions cover most components of the COM-B model, which suggests that interventions are more likely to produce effective results and make changes in current behavior if they target a wide spectrum of the COM-B model's components. For example, the study by Chen et al [26] used all COM-B components to increase dietitians' app use behaviors through dietitians' education and skill training as well as environmental restructuring by providing physical app-based infrastructure. Modeling, coaching, and improving self-efficacy were also addressed. Therefore, the reported change in ratings of dietitians' self-efficacy when using mHealth apps was significant (P=.01). The Tukey post hoc test revealed significantly higher post–workshop mHealth app self-efficacy ratings compared with the baseline (P=.02), and the ratings were sustained at 12 weeks (P=.01).

In addition, the intervention functions used in the ineffective study were linked to only 2 COM-B model components (physical capability and physical opportunity), suggesting a need to understand the target behavior by collecting data from multiple sources to ensure successful behavior change.

Discussion

Principal Findings

This review included 11 studies investigating the effects of interventions aimed at encouraging app prescription behavior among HCPs. Most studies demonstrated positive findings for outcomes such as self-confidence, knowledge of mHealth apps, and number of app prescriptions. The studies included differed widely in terms of interventions, measures of practice change, types of mHealth apps used, modes of delivery, and study settings.

A broad range of interventions, all related to methods for enhancing mHealth app adoption, specifically mHealth prescribing, was covered in this review, including education (workshops, class lectures, individual sessions with a dietitian or research team, videos, or toolkits), persuasion (reminders or verbally about capabilities), incentivization (expectation to pass the course), training (case studies or scenarios or app appraisal tools), modeling (site champions or observing peers), enablement (support), and environmental restructuring (lists of apps, technological systems, time, and resources).

Comparisons With Other Works

More than half of the interventions included in this review had training (8/11, 72.7%) or educational (9/11, 81.8%) functions that targeted HCPs' capabilities (physical or psychological). A lack of knowledge and awareness of available apps are major barriers reported in the literature [34-36]. A study reported that HCPs consider the lack of knowledge of available apps that have proven their effectiveness in improving patient health outcomes an important barrier to prescribing [35]. This confirms the urgent need to provide training programs or educational sessions regarding the available mHealth apps that support patient self-management of long-term conditions.

However, the terms training and education are often used interchangeably. Some studies reported training as part of the intervention, but these studies included only elements of imparting knowledge and understanding, not skill development [24,28]. One possible explanation is the lack of attention to different types of training in behavior change interventions. Hence, there is difficulty in perceiving what training might entail. Therefore, it is vital to distinguish between the 2 terms. This overlap has also been reported between other intervention functions. Another study reported the existence of some overlap when providing education to participants because it serves a persuasion function at the same time [37]. This is mainly because education may induce positive feelings toward app prescription.

The lack of coercion (defined as "creating expectation of punishment or cost") may be because it was not deemed a suitable approach for improving self-efficacy or confidence in app prescriptions or it may have been deemed counterproductive when trying to create positive attitudes or encourage app prescriptions. Furthermore, it is impractical to penalize HCPs for not prescribing mHealth apps.

None of the interventions adopted restrictions such as rules to increase HCP app prescriptions or decrease any competing behavior. The absence of regulations that ensure apps' highest quality and accuracy and the lack of data validity and reliability of existing apps keep HCPs from prescribing apps to their patients [38,39]. To compensate, some interventions have provided study participants with a list of trustworthy apps to use when making app recommendations to patients. Other interventions adopted digital solutions such as building software or electronic systems containing approved apps. Removing such barriers by providing practical resources is a form of environmental restructuring.

Minimizing the disruption of HCPs' time is a form of enablement reported in only one intervention [32]. HCPs' concerns about time to discuss and instruct patients on how to use apps were reported as barriers to app prescription. In a pilot study, participants were concerned that recommending apps to patients would lengthen the duration of consultations [23]. In France, a qualitative study presented a theme after interviews with GPs about "Doctor Protection," which mainly introduced concerns about increased workload and prescriptions of apps as an additional task [40]. One significant distinction between apps and medications is that many drugs can be prescribed with simple directions, whereas an app may require more specific instructions.

One way to minimize physicians' workloads is to integrate and synchronize health information produced from mHealth apps to patients' EHRs. By doing this, the physician's ability to access patient data is centralized. In an acceptability and feasibility study that examined integrating patient data generated from smartphones into EHRs [41], clinicians reported that by using the graph feature, they could evaluate longitudinal data during consultations, which was quick and easy. When compared with retrieving information by recording histories, this was thought to be a possible time saver. Furthermore, this approach provided an accurate reflection of disease changes and treatment responses.

Strengths and Limitations of the Study

This review has several strengths and limitations. This is the first review that addresses interventions to improve HCPs' confidence and capabilities to prescribe or recommend mHealth apps to patients. We consider the included studies to be a complete set of studies from 2008 to 2022. The studies were sourced from a variety of electronic databases, with the reference lists of the included papers checked for potentially relevant studies. This systematic review used a robust methodology that included screening all the studies for relevance by 2 independent reviewers.

However, the number of studies included in this review was limited, and the findings depended on the quality of the included studies. Of the 11 studies, 3 (27.3%) studies were found to be at high risk of bias, 7 (63.6%) at moderate risk of bias, and only 1 (9.1%) at low risk of bias (see Appendix 3). A total of 6 (54.5%) studies were pretest-posttest design interventions; these are known for their methodological issues such as selection bias and short durations, which do not make it possible to determine whether the intervention is effective and sustainable [42]. Furthermore, 7 (63.6%) studies used self-reported data to reflect possible changes in behaviors. With self-reported outcomes, it is impossible to tell whether the reported change in knowledge or practice is owing to response-shift bias or an actual adoption of the targeted behaviors. This emphasizes the importance of using objective measures instead. Objective measures such as the number of app prescriptions and OSCE scores were reported in 4 (36.4%) studies. However, the assessment of behavior changes or intentions to change app prescription behaviors cannot always be performed using objective measures. This is the case, in particular with interventions that lack technical systems to track the changes in the number of prescriptions before and after the intervention. Combining objective and self-reported measures could bring more insight and different perspectives to the study findings.

Unanswered Questions and Future Research

Several gaps in the research were identified in this review. Coercion and restriction were not reported in any of the interventions included; however, this may be because they were not deemed appropriate approaches for changing app prescription behaviors in this population. The impact of other forms of intervention functions not used in any of the interventions reviewed could be explored, such as the use of incentives (financial or nonfinancial) that, if appropriately applied and supported with other intervention functions, could potentially make an impact and encourage HCPs to prescribe apps [43]. Evidence on the acceptability and impact of such programs in the context of mHealth is lacking. This can be answered with future studies.

High-quality studies with adequate sample sizes and longer study periods are now essential for detecting differences in app prescription behaviors. Future interventions could adopt theoretical frameworks and behavior change frameworks to systematically understand HCPs behavior toward the prescription of mHealth apps.

The COM-B model, as part of the BCW, is a useful tool to make behavioral diagnoses and identify what needs to change [22]. Evidence from successful interventions, interviews, or surveys with HCPs about what motivates or limits their mHealth app prescription behaviors can provide sufficient information to understand the sources of those behaviors. Therefore, future interventions can address the target behavior (app prescriptions) and its influencing factors.

Conclusions

This study identified interventions aimed at improving HCPs' app-prescribing behaviors. On the basis of the BCW, environmental restructuring and education were the most frequently used intervention functions in the included studies, followed by training. The findings of this study provide evidence that combining elements of training, education, and environmental restructuring is more likely to produce effective changes in HCPs' behavior toward app prescribing.

Acknowledgments

This research is part of a PhD study sponsored by the Ministry of Education in Saudi Arabia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results of the search strategies used in the MEDLINE and CINAHL databases.

[DOCX File , 16 KB - mhealth_v11i1e43561_app1.docx]

Multimedia Appendix 2

Characteristics of individual studies included in the systematic review.

[DOCX File , 18 KB - mhealth_v11i1e43561_app2.docx]

Multimedia Appendix 3
Risk of bias assessment of included studies.
[DOCX File , 16 KB - mhealth_v11i1e43561_app3.docx]

References

1. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization. 2013. URL: https://apps.who.int/iris/bitstream/handle/10665/94384/?sequence=1 [accessed 2022-05-05]
2. Statistics & Facts. statista. 2021. URL: https://www.statista.com/topics/2263/mhealth/ [accessed 2022-05-05]
3. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018 May 9;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
4. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017 Mar 1;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
5. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
6. Zhang H, Jiang Y, Nguyen HD, Poo DC, Wang W. The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: a pilot randomised controlled trial. Health Qual Life Outcomes 2017 Mar 14;15(1):49 [FREE Full text] [doi: 10.1186/s12955-017-0623-y] [Medline: 28288636]
7. Olesch A. DiGA: A Year With Apps on Prescription in Germany. SidekickHealth. 2021. URL: https://sidekickhealth.com/news/a-year-with-apps-on-prescription-in-germany/ [accessed 2022-03-15]
8. Sleepio to treat insomnia and insomnia symptoms. National Institute for Health and Care Excellence. 2022 May 20. URL: https://www.nice.org.uk/guidance/mtg70 [accessed 2022-07-09]
9. Sleepio is the first-ever digital therapeutic to receive NICE guidance, confirming clinical and cost effectiveness. BigHealth. 2022. URL: https://www.bighealth.co.uk/news/sleepio-is-the-first-ever-digital-therapeutic-to-receive-nice-guidance-confirming-clinical-and-cost-effectiveness/ [accessed 2022-07-25]
10. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17704 [FREE Full text] [doi: 10.2196/17704] [Medline: 32628118]
11. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
12. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
13. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
14. Fernandez ME, Ruiter RA, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019 Aug 14;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
15. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/2.0/ [accessed 2023-02-21]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Strain M. 1983 to today: a history of mobile apps. The Guardian. 2015 Feb 13. URL: https://www.theguardian.com/media-network/2015/feb/13/history-mobile-apps-future-interactive-timeline [accessed 2022-05-08]
20. Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group. National Heart, Lung, and Blood Institute. 2014. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2022-07-27]
21. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]

22. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, UK: Silverback Publishing; 2014.
23. Lopez Segui F, Pratdepadua Bufill C, Abdon Gimenez N, Martinez Roldan J, Garcia Cuyas F. The prescription of mobile apps by primary care teams: a pilot project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701 [FREE Full text] [doi: 10.2196/10701] [Medline: 29929948]
24. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting mental health in primary care: a 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Front Psychiatry 2019 Feb 28;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]
25. Byambasuren O, Beller E, Hoffmann T, Glasziou P. mHealth app prescription in Australian general practice: pre-post study. JMIR Mhealth Uhealth 2020 Jun 01;8(6):e16497 [FREE Full text] [doi: 10.2196/16497] [Medline: 32478660]
26. Chen J, Allman-Farinelli M. Impact of training and integration of apps into dietetic practice on dietitians' self-efficacy with using mobile health apps and patient satisfaction. JMIR Mhealth Uhealth 2019 Mar 04;7(3):e12349 [FREE Full text] [doi: 10.2196/12349] [Medline: 30829577]
27. Barnett J, Harricharan M, Fletcher D, Gilchrist B, Coughlan J. myPace: an integrative health platform for supporting weight loss and maintenance behaviors. IEEE J Biomed Health Inform 2015 Jan;19(1):109-116. [doi: 10.1109/JBHI.2014.2366832] [Medline: 25438329]
28. Armstrong CM, Ciulla RP, Edwards-Stewart A, Hoyt T, Bush N. Best practices of mobile health in clinical care: the development and evaluation of a competency-based provider training program. Prof Psychol Res Pr 2018 Oct;49(5-6):355-363. [doi: 10.1037/pro0000194]
29. Armstrong CM. Mobile health provider training: results and lessons learned from year four of training on core competencies for mobile health in clinical care. J Technol Behav Sci 2019 Mar 26;4(2):86-92. [doi: 10.1007/s41347-019-00089-8]
30. Al-Lami N, Sear K, Dai W, Chen JL. Obesity applications in smartphones: development and use of an evaluation tool. J Pediatr Health Care 2020;34(4):377-382. [doi: 10.1016/j.pedhc.2020.03.001] [Medline: 32414553]
31. Rodder SG, Kindratt TB, Xiao C, Orcutt V, Koch C, Mcilvaine K, et al. Teaching and evaluating smartphone applications: the effectiveness of a curriculum expansion. Educ Health (Abingdon) 2018;31(2):95-102 [FREE Full text] [doi: 10.4103/efh.EfH_322_16] [Medline: 30531051]
32. Makhni S, Zlatopolsky R, Fasihuddin F, Aponte M, Rogers J, Atreja A. Usability and learnability of RxUniverse, an enterprise-wide app prescribing platform used in an academic tertiary care hospital. AMIA Annu Symp Proc 2018 Apr 16;2017:1225-1232 [FREE Full text] [Medline: 29854191]
33. Korpershoek YJ, Holtrop T, Vervoort SC, Schoonhoven L, Schuurmans MJ, Trappenburg JC. Early-stage feasibility of a mobile health intervention (Copilot) to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease: multimethods approach. JMIR Form Res 2020 Nov 19;4(11):e21577 [FREE Full text] [doi: 10.2196/21577] [Medline: 33211013]
34. Zhang Y, Li X, Luo S, Liu C, Xie Y, Guo J, et al. Use, perspectives, and attitudes regarding diabetes management mobile apps among diabetes patients and diabetologists in China: national web-based survey. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12658 [FREE Full text] [doi: 10.2196/12658] [Medline: 30735147]
35. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
36. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and use of mHealth apps: a study from England. Pharmacy (Basel) 2017 Jun 14;5(2):33 [FREE Full text] [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
37. Staddon SC, Cycil C, Goulden M, Leygue C, Spence A. Intervening to change behaviour and save energy in the workplace: a systematic review of available evidence. Energy Res Soc Sci 2016 Jul;17:30-51 [FREE Full text] [doi: 10.1016/j.erss.2016.03.027]
38. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
39. Bellicha A, Macé S, Oppert JM. Prescribing of electronic activity monitors in cardiometabolic diseases: qualitative interview-based study. J Med Internet Res 2017 Sep 23;19(9):e328 [FREE Full text] [doi: 10.2196/jmir.8107] [Medline: 28947415]
40. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
41. Austin L, Sharp CA, van der Veer SN, Machin M, Humphreys J, Mellor P, et al. Providing 'the bigger picture': benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. Rheumatology (Oxford) 2020 Feb 01;59(2):367-378 [FREE Full text] [doi: 10.1093/rheumatology/kez207] [Medline: 31335942]
42. Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. Emerg Med J 2015 Jul;32(7):507-508 [FREE Full text] [doi: 10.1136/emermed-2015-204761] [Medline: 25820301]
43. Lavergne MR. Financial incentives for physicians to improve health care. CMAJ 2017 Dec 11;189(49):E1505-E1506 [FREE Full text] [doi: 10.1503/cmaj.171126] [Medline: 29229710]

Abbreviations

BCW: behavior change wheel
CN: clinical nutrition
COM-B: capability, opportunity, motivation, and behavior
COPD: Chronic Obstructive Pulmonary Disease
EHRs: electronic health records
GPs: general practitioners
MMAT: mixed methods appraisal tool
OSCE: Objective Structured Clinical Examination
PA: physician assistant
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by L Buis; submitted 16.10.22; peer-reviewed by M Kapsetaki, O Byambasuren, I Shubina; comments to author 29.11.22; revised version received 12.12.22; accepted 15.12.22; published 27.02.23.

Please cite as:
Alkhaldi O, McMillan B, Maddah N, Ainsworth J
Interventions Aimed at Enhancing Health Care Providers' Behavior Toward the Prescription of Mobile Health Apps: Systematic Review
JMIR Mhealth Uhealth 2023;11:e43561
URL: https://mhealth.jmir.org/2023/1/e43561
doi:10.2196/43561
PMID:36848202

Original Paper

Effectiveness of a Mindfulness Meditation App Based on an Electroencephalography-Based Brain-Computer Interface in Radiofrequency Catheter Ablation for Patients With Atrial Fibrillation: Pilot Randomized Controlled Trial

Ying He[1*], BSc; Zhijie Tang[2*], MSc; Guozhen Sun[1], MSc; Cheng Cai[1], MD; Yao Wang[1], MD; Gang Yang[1], MD; ZhiPeng Bao[1], MSc

[1]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

[2]School of Nursing, Nanjing Medical University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
ZhiPeng Bao, MSc
Department of Cardiology
The First Affiliated Hospital of Nanjing Medical University
300 Guangzhou Road, Gulou District
Nanjing, 210029
China
Phone: 86 15895903958
Fax: 86 025 68303041
Email: baozhipeng1219@163.com

Abstract

Background: Radiofrequency catheter ablation (RFCA) for patients with atrial fibrillation (AF) can generate considerable physical and psychological discomfort under conscious sedation. App-based mindfulness meditation combined with an electroencephalography (EEG)-based brain-computer interface (BCI) shows promise as effective and accessible adjuncts in medical practice.

Objective: This study aimed to investigate the effectiveness of a BCI-based mindfulness meditation app in improving the experience of patients with AF during RFCA.

Methods: This single-center pilot randomized controlled trial involved 84 eligible patients with AF scheduled for RFCA, who were randomized 1:1 to the intervention and control groups. Both groups received a standardized RFCA procedure and a conscious sedative regimen. Patients in the control group were administered conventional care, while those in the intervention group received BCI-based app–delivered mindfulness meditation from a research nurse. The primary outcomes were the changes in the numeric rating scale, State Anxiety Inventory, and Brief Fatigue Inventory scores. Secondary outcomes were the differences in hemodynamic parameters (heart rate, blood pressure, and peripheral oxygen saturation), adverse events, patient-reported pain, and the doses of sedative drugs used in ablation.

Results: BCI-based app–delivered mindfulness meditation, compared to conventional care, resulted in a significantly lower mean numeric rating scale (mean 4.6, SD 1.7 [app-based mindfulness meditation] vs mean 5.7, SD 2.1 [conventional care]; *P*=.008), State Anxiety Inventory (mean 36.7, SD 5.5 vs mean 42.3, SD 7.2; *P*<.001), and Brief Fatigue Inventory (mean 3.4, SD 2.3 vs mean 4.7, SD 2.2; *P*=.01) scores. No significant differences were observed in hemodynamic parameters or the amounts of parecoxib and dexmedetomidine used in RFCA between the 2 groups. The intervention group exhibited a significant decrease in fentanyl use compared to the control group, with a mean dose of 3.96 (SD 1.37) mcg/kg versus 4.85 (SD 1.25) mcg/kg in the control group (*P*=.003).The incidence of adverse events was lower in the intervention group (5/40) than in the control group (10/40), though this difference was not significant (*P*=.15).

Conclusions: BCI-based app–delivered mindfulness meditation effectively relieved physical and psychological discomfort and may reduce the doses of sedative medication used in RFCA for patients with AF.

Trial Registration: ClinicalTrials.gov NCT05306015; https://clinicaltrials.gov/ct2/show/NCT05306015

(JMIR Mhealth Uhealth 2023;11:e44855) doi:10.2196/44855

KEYWORDS

atrial fibrillation; radiofrequency catheter ablation; mindfulness meditation; brain computer interface; mHealth; smartphone app; randomized controlled trial

Introduction

Atrial fibrillation (AF) is recognized as the most common cardiac arrhythmia worldwide, with an estimated prevalence of 2%-4% in adults [1]. Epidemiological studies indicate that AF increases the risk of stroke by 5-fold and the risk of overall mortality by 3.5-fold [2,3]. AF is becoming an increasingly extensive public health problem and causes substantial health economic burden [4-6]. Radiofrequency catheter ablation (RFCA) has become a first-line therapy for AF to improve symptoms, cardiac function, and quality of life, and has been shown to be cost-effective [7-9]. In China, the number of patients with AF is estimated to be approximately 20 million, and the number of RFCAs for AF exceeds 30,000 every year [10,11].

Considering the longer general anesthesia preparation time, higher economic costs, and potential complications, conscious sedation is used for RFCA at most centers in China [12-14]. However, even under well-tolerated doses of sedative drugs, sedation-related side effects such as nausea, vomiting, and oversedation are common [15]. Furthermore, patients are required to remain motionless and endure radiofrequency energy burning their myocardia for hours during the complex RFCA procedure. Consequently, even under deep sedation, patients may still experience considerable pain, anxiety, fatigue, and other discomforts, which may be associated with poor outcomes [16-18]. Several studies have indicated that nonpharmacological interventions could be ideal adjuncts to sedative drugs, effectively reducing patients' physical or psychological discomfort and the required doses of sedative drugs during medical invasive procedures [19,20].

Mindfulness meditation originates from Buddhist teachings and refers to a category of techniques used to pay attention to the present moment and accept all that arises without judgment [21]. Numerous studies suggest that mindfulness meditation may hold potential for alleviating pain, fatigue, and negative emotions [22,23]. However, the effects of mindfulness meditation during AF ablation remain uncertain. In recent years, with the rapid development of digital medicine, app-based mindfulness interventions have been preliminarily shown to be effective and accessible [24].

A brain-computer interface (BCI) is defined as a technology for establishing external information communication and control pathways between the human brain and computers or other electronic devices [25]. Electroencephalography (EEG) is a conventional form of brain signal acquisition, which can be recognized and reflected (usually through visual or auditory signals) by BCI [26]. Studies have shown that EEG-based BCI devices can sense and classify human psychological states, which may facilitate mindfulness meditation practice [27,28].

This study aimed to determine the effects of a BCI-based mindfulness meditation app on RFCA for patients with AF. The primary hypothesis was that the intervention group would experience significant improvements in perceived pain, anxiety, and fatigue compared to the control group. We also hypothesized that the intervention might decrease the use of sedative drugs and the incidence of adverse events.

Methods

Study Design

This was a single-center, 2-arm, parallel-group, prospective, pilot randomized controlled trial. Patients were randomized 1:1 to the intervention and control groups using a computer-generated randomization list. Due to the nature of the study design, neither program implementers nor patients could be blinded to the intervention. However, the investigators performing the outcome assessments and data analysis were blinded to the group allocation.

Participants

Overview

Patients were eligible for the study if they were (1) diagnosed with AF, (2) at least 18 years old, (3) undergoing their initial RFCA procedure, and (4) willing to participate in the study. Patients were excluded if they had (1) severe systemic diseases such as malignant tumors, (2) a history of mental illness and cognitive complaints, and (3) difficulty understanding the questionnaire and the study aims. They were also excluded if they experienced drastic changes in their condition during RFCA.

Sample size calculations were conducted using PASS 2021 (NCSS LLC) software and based on previous studies [29,30]. Power analysis showed that a sample size of 70 participants was sufficient to have 90% statistical power at a 2-sided α of .05 for significance. To account for a 20% loss rate, the planned sample size was 84 participants.

All patients underwent a standardized RFCA procedure for AF with a 3D mapping system (CARTO 3, Biosense Webster) and were provided with standardized information about the study and the potential benefits and risks of the interventions. Both groups received the same sedative regimen during RFCA, which was adjusted by the interventional physician based on the patient's response to the medication and reported pain levels. The regimen included a single dose of parecoxib (40 mg), fentanyl (1 mcg/kg/hour), and dexmedetomidine as necessary, with dosing adjustments made in accordance with standard pain management procedures at our institution. The fentanyl maintenance infusion rate ranged from 0 to 2 mcg/kg/hour, while the dexmedetomidine maintenance infusion rate ranged from 0 to 1 mcg/kg/hour.

Intervention Group

Patients in the intervention group received mindfulness meditation guidance delivered through a Chinese-language interface and voice app (Focus Zen, version 2.1.1) along with a BCI-based headband. A mobile phone and a Samsung tablet device with the preinstalled app were prepared in the cardiac catheterization laboratory. Before ablation, study staff briefly introduced the method and meaning of mindfulness meditation to help patients understand the intervention content. Mindfulness meditation represents a practice of awareness in which the person gradually and purposefully focuses on the present without judgement to achieve a state of deep relaxation [31]. The app's developers designed a 35-minute mindfulness meditation course specifically for patients with AF to help them relax during ablation without affecting the ablation procedure (eg, the course instructed patients to breathe evenly rather than deeply). We provided patients with Bluetooth earphones and set the background sound within the app in accordance with the patients' preferences, such as forest, beach, or rain sounds. During the mindfulness meditation practice, patients were guided by a female voice through the app to relax muscles, regulate breathing, and practice visualization and body scanning. Simultaneously, the app collected EEG information using a headband device that was synchronized with the app through Bluetooth technology (Figure 1). An artificial intelligence algorithm included in the app was used to analyze the EEG data and classify the patient's brain state as active, calm, relaxed, or meditative. The app interface and headband light color were adjusted in accordance with the patients' state. Additionally, the app prompted the patients' current brain state through background sound effects and guided the patient to maintain the state or make adjustments through the app voice.

Figure 1. Screenshots of the study app. (A) Device connection, (B) background sound setting, and (C) mindfulness meditation course).

Control Group

The control group received routine care for their ablation procedure and was informed about the procedure of ablation and characteristics of impending pain in ablation, as in the intervention group. Psychological and supportive care were provided in accordance with the patients' needs. However, patients in the control group wore the headband device without using the earphones and did not receive mindfulness meditation guidance provided by the app.

Outcome Measurements

Both the intervention and control groups were administered 2 surveys, one 30 minutes before ablation and one within 30 minutes after ablation, to assess pain intensity, fatigue, and anxiety using specific paper questionnaires. Demographic information and characteristics of participants were collected at baseline. The study staff recorded patients' hemodynamic parameters (heart rate, blood pressure, and peripheral oxygen saturation [SpO_2]), spontaneously reported pain, the doses of sedative drugs used, and adverse events during ablation.

The primary outcomes were pain and anxiety levels during ablation and fatigue severity after ablation. The intensity of pain was measured using the numeric rating scale, with scores that ranged from 0 (no pain) to 10 (the worst possible pain) [32,33]. The State Anxiety Inventory (A-State) is a subscale of the State-Trait Anxiety Inventory [34], which is mainly used to assess the anxiety state in a specific situation. The A-State score was used in this study to explore the anxiety level of patients during ablation. We evaluated the patients' fatigue after ablation using the Brief Fatigue Inventory (BFI) [35], a 10-item validated scale. A higher score indicates a greater level of fatigue.

Secondary outcomes included mean heart rate, blood pressure, and SpO_2 during ablation. Adverse events were defined as excessive fluctuations in blood pressure (fluctuations of >50 mm Hg in systolic blood pressure), nausea and vomiting, and vasovagal reaction. The study staff also recorded the number of times of spontaneously reported pain and the doses of sedative drugs used during ablation.

Statistical Analysis

Statistical analysis was performed using SPSS (version 22.0; IBM Corp) and based on the intention-to-treat principle with a 2-sided significance level of .05. Data were analyzed using descriptive statistics and checked for the normality of their distribution. Descriptive continuous variables are presented as mean (SD) values and categorical variables as frequency and percentage values. Differences between study groups were analyzed using an independent 2-sample *t* test for numerical variables and the Mann-Whitney *U* test, chi-square test, or the Fisher exact test for categorical variables.

Ethical Considerations

This study was conducted at the cardiac catheterization laboratory of the First Affiliated Hospital of Nanjing Medical University, Nanjing, China, from April to September 2022. All study patients provided oral or written informed consent. This study was approved by the ethics committee of the First Affiliated Hospital of Nanjing Medical University (2022-SR-086) and registered at ClinicalTrials.gov (NCT05306015). Procedures were conducted in accordance with the tenets of the Declaration of Helsinki.

Results

Baseline Characteristics

A total of 84 patients (42 patients each in the intervention and control groups) were enrolled and completed baseline measures in this study. A total of 4 patients (2 each from the intervention and control groups) were excluded from the study. Figure 2 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram for this clinical trial.

No significant differences were found between the intervention group and the control group in baseline characteristics (Table 1). Neither group had previous experience with practicing mindfulness meditation.

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

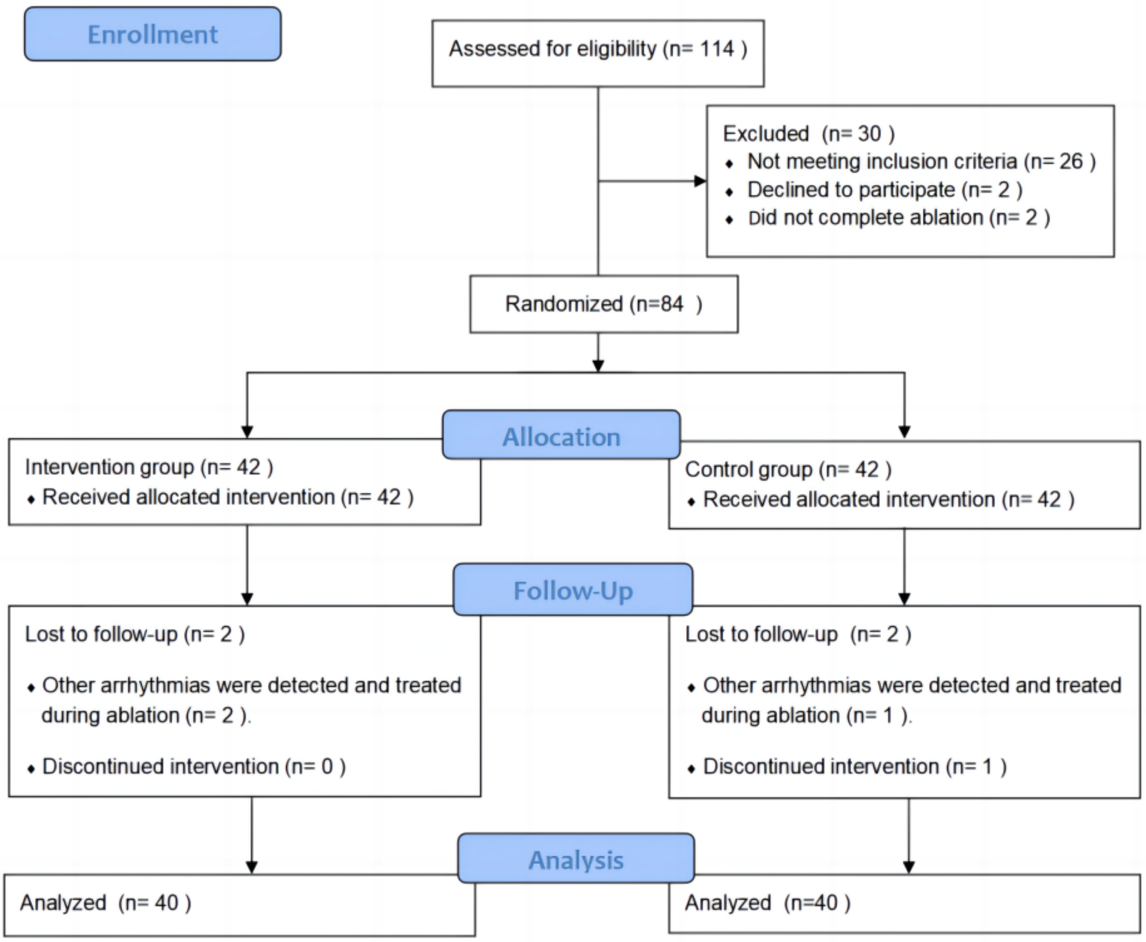


Table 1. Baseline characteristics of the study participants (N=80).

Characteristics	Intervention group (n=40)	Control group (n=40)	*P* value
Age (years), mean (SD)	58.1 (10.4)	60.0 (11.2)	.44
Gender (female), n (%)	15 (38)	16 (40)	.82
Weight (kg), mean (SD)	71.3(13.6)	69.5(11.2)	.53
BMI (kg/m²), mean (SD)	24.9 (3.0)	24.9 (3.3)	.97
LAD[a] (mm), mean (SD)	40.6 (6.5)	40.8 (4.2)	.89
LVEF[b] (%), mean (SD)	62.0 (5.9)	60.9 (6.4)	.40
Hypertension, n (%)	14 (35)	16 (40)	.64
DM[c], n (%)	6 (15)	4 (10)	.50
CHD[d], n (%)	3 (8)	9 (23)	.06
Type of AF[e], n (%)			.37
Paroxysmal	24 (60)	20 (50)	
Persistent	16 (40)	20 (50)	
NYHA[f] class, n (%)			.59
Class I	32 (80)	30 (75)	
Class II	8 (20)	10 (25)	
RFCA[g] time (minutes), mean (SD)	40.1 (14.3)	42.6 (15.0)	.44
RFCA energy (Watts), mean (SD)	42.2 (3.8)	42.1 (3.6)	.81
RFCA temperature (°C), mean (SD)	28.8 (3.3)	29.5 (4.3)	.39

[a]LAD: left atrium diameter.
[b]LVEF: left ventricular ejection fraction.
[c]DM: diabetes mellitus.
[d]CHD: coronary heart disease.
[e]AF: atrial fibrillation.
[f]NYHA: New York Heart Association.
[g]RFCA: radiofrequency catheter ablation.

Primary Outcomes

We found no significant difference in the baseline pain, anxiety, and fatigue scores between the intervention and control groups (Table 2). After the intervention, compared to the control group, there were significant differences in numeric rating scale (mean 4.6, SD 1.7 [intervention group] vs mean 5.7, SD 2.1 [control group]; *P*=.008) and A-State (mean 36.7, SD 5.5 vs mean 42.3, SD 7.2; *P*<.001) scores after ablation. The BFI score after ablation was significantly lower in the intervention group than in the control group (mean 3.4, SD 2.3 vs mean 4.7, SD 2.2; *P*=.01).

Table 2. NRS[a], A-State[b], and BFI[c] scores of the intervention and control groups.

Variable	Baseline			Post intervention		
	Intervention group, mean (SD)	Control group, mean (SD)	*P* value	Intervention group, mean (SD)	Control group, mean (SD)	*P* value
NRS score	0.3 (0.5)	0.4 (0.5)	.66	4.6 (1.7)	5.7 (2.1)	.008
A-State score	30.7 (4.4)	31.8 (6.2)	.39	36.7 (5.5)	42.3 (7.2)	<.001
BFI score	1.4 (1.7)	1.2 (1.5)	.49	3.4 (2.3)	4.7 (2.2)	.01

[a]NRS: numerical rating scale.
[b]A-State: State Anxiety Inventory.
[c]BFI: Brief Fatigue Inventory.

Secondary Outcomes

Between the intervention and control groups in ablation, there were no significant differences in the mean heart rate (mean 87.4, SD 15.7 [intervention group] vs mean 91.1, SD 16.4 [control group] beats per minute; *P*=.31), systolic blood pressure (mean 127.2, SD 15.7 vs mean 131.9, SD 17.4 mm Hg; *P*=.21), diastolic blood pressure (mean 81.1, SD 10.5 vs mean 82.7, SD 9.6 mm Hg; *P*=.49), and SpO_2 (mean 98.4%, SD 1.3% vs mean 98.5%, SD 1.2%; *P*=.71; Figure 3).

There were no significant differences in parecoxib and dexmedetomidine use between the intervention and control groups during ablation. The intervention group had significantly decreased fentanyl use compared to the control group (*P*=.003; Table 3). Additionally, patients in the intervention group reported significantly fewer times of pain than those in the control group during ablation (*P*<.001). The incidence of adverse events in the intervention group was lower than that in the control group, but the difference did not reach statistical significance (*P*=.15).

Figure 3. Hemodynamic parameters of the 2 groups. bpm: beats per minute; DBP: diastolic blood pressure; HR: heart rate; RFCA: radiofrequency catheter ablation; SBP: systolic blood pressure; SpO_2: peripheral oxygen saturation.

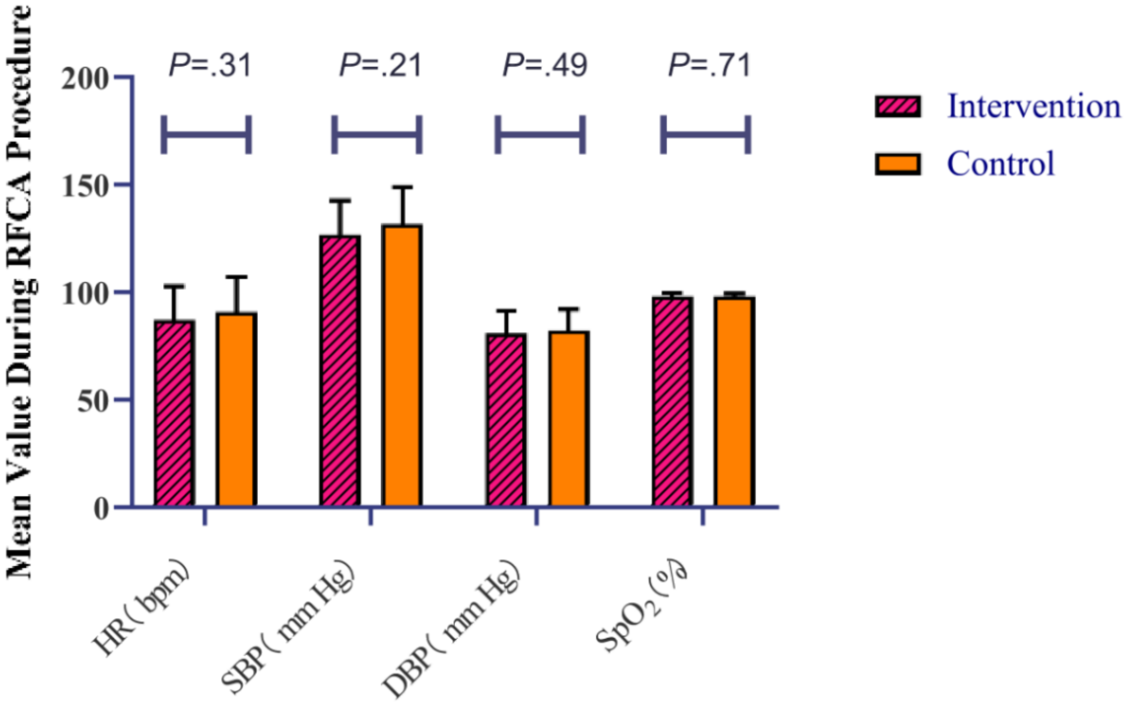


Table 3. Comparison of secondary outcomes between the intervention and control groups.

Medications and outcomes	Intervention group (n=40)	Control group (n=40)	*P* value
Parecoxib (mg), mean (SD)	29.0 (18.1)	27.0 (19.0)	.63
Fentanyl (mcg/kg), mean (SD)	3.96 (1.37)	4.85 (1.25)	.003
Dexmedetomidine (mcg/kg), mean (SD)	1.49 (0.82)	1.85 (1.00)	.57
Patient reports of pain (times), mean (SD)	1.5 (1.4)	2.8 (1.7)	<.001
Adverse events, n (%)	5 (13)	10 (25)	.15

Discussion

Principal Findings

The purpose of this study was to evaluate the effectiveness of a novel BCI-based mindfulness meditation app for patients with AF to improve their physical and psychological status in RFCA. The selected app provides EEG feedback for mindfulness meditation, which may help patients relax and reduce unpleasant experiences without interfering with the ablation process. The key findings showed significantly lower pain, anxiety, and fatigue scores in the intervention group than among those receiving conventional care. No significant differences were found in the mean heart rate, blood pressure, or SpO_2 between groups during RFCA. Additionally, the intervention group had a significant decrease in fentanyl use, while the differences in other sedative drugs were not significant. Although the incidence of adverse events was lower in the intervention group, the difference was not significant.

Mobile health, via wireless technologies such as smartphone apps and wearable devices, is currently used for patients with AF mostly for screening, management, and rehabilitation [36-39]. Numerous studies have investigated the impact of meditation on cardiac disease and suggest that it may offer potential benefits for cardiovascular health, including reducing blood pressure, improving psychological and physiological responses to stress, and possibly mitigating AF progression by modulating the autonomic nervous system [21,40]. Mobile app–based mindfulness meditation has the potential to serve as an adjunct to the RFCA procedure for patients with AF owing to its low risks, potential benefits, and relatively low cost. Nevertheless, it is essential to acknowledge the costs associated with using a commercialized app and headband device, along with the requirement for labor and its cost to monitor the app. Additionally, caution should be exercised in making claims about the low risk of this intervention, as it is based solely on the findings of this small pilot study. Further research is needed to explore the relationship between meditation and AF ablation. To our knowledge, this is the first study to assess the effectiveness of a mindfulness meditation app together with a BCI-based wearable device for patients with AF during RFCA. This study achieved promising results and indicated that this type of intervention could be easily integrated into the standard RFCA workflow.

It is well known that RFCA for AF can be accompanied by considerable pain and anxiety when conscious sedation is used, which, however, has potential side effects [15]. Anxiety is common in patients with AF [41]. Patients may experience pain and uncertainty for several hours during RFCA, which could amplify negative emotions such as anxiety and made them feel exhausted. Therefore, it is necessary to monitor and intervene in patients with AF's anxiety during ablation procedures [16]. Previous studies have indicated that apps based on nonpharmacological interventions could effectively reduce the fatigue, anxiety, pain, and the use of sedative drugs in invasive medical treatment [42-44]. Nørgaard et al [45] examined the effects of visualization together with usual pain medication in comparison with conventional care among patients with AF undergoing RFCA, and found significant reductions in the perception of pain and anxiety, as well as the doses of analgesics used in the intervention group [46]. Wearable devices are drastically changing medical practices nowadays. Roxburgh et al [47] implemented a virtual reality headset for patients with AF undergoing cryoballoon ablation under conscious sedation and found that the virtual reality group had a significantly lower perceived pain score and higher comfort score. In recent years, there has been a growing body of literature examining the effects of mindfulness training supported by EEG feedback. Crivelli et al [48] investigated the potential benefits of EEG-based brain-sensing device–supported mindfulness practices in individuals with mild stress levels and found a significant reduction in stress and anxiety. Similarly, Balconi et al [49] used mindfulness exercises in combination with wearable EEG information sensor devices to explore the effect of reducing overall stress levels in healthy individuals and found significant improvements in physiological (heart rate and variability) and subjective markers of stress (perceived stress, anxiety, and mood states). These findings are in line with those of this study, suggesting that EEG feedback may facilitate meditation by providing real-time information to aid users in achieving a mindfulness state. In our study, we assessed the effectiveness of an intervention based on a mindfulness meditation app combined with a BCI-based wearable device, which provides meditation guidance with synchronous EEG feedback to patients. This intervention has a minimal learning curve without requiring specialist training and provides personalized feedback, which may enhance patient engagement and adherence [50]. The potential mechanisms underlying pain, anxiety, and fatigue relief through mindfulness meditation are likely linked with the ability of meditation to change the activity of the insula, somatosensory cortex, anterior cingulate cortex, and prefrontal cortex. These changes may reduce patients' attention, memory, and perception of physical and psychological discomfort [51,52].

When patients experience physical and mental discomfort in RFCA, the levels of catecholamines, adrenocorticotropic hormone, prolactin, cortisol, and prostaglandins in their blood may increase [53], which could result in unstable hemodynamic parameters (heart rate, blood pressure, and SpO_2) [54] and, thereby, affect the performance of procedures and patient safety. Previous studies have indicated that nonpharmacological interventions could effectively stabilize hemodynamic parameters in invasive operative procedures [42,55]. In this study, however, the between-group differences in mean heart rate, blood pressure, or SpO_2 during RFCA did not reach statistical significance. The observed discrepancy between objective and subjective outcomes may be attributed to various factors. One possible explanation is that blood pressure changes induced by meditation through the autonomic nervous system may be a long-term process [21]. Furthermore, the timing of data collection may have influenced the findings, as objective measures were collected during the intervention, whereas subjective data were gathered through self-reports post intervention. It is worth noting that patients with AF frequently experience AF episodes during RFCA procedures, which may lead to variations in heart rate and blood pressure levels and contribute to the lack of significant differences between the groups. Additional research with larger sample sizes may help elucidate the potential impact of the intervention on these objective parameters. Although there was no significant difference, the incidence of adverse events was lower in the intervention group. This indicates a possibility for the potential protective effects of our intervention during RFCA, which merit further investigations.

Strengths and Limitations

A strength of this study is that it is the first randomized controlled trial, to our knowledge, to explore the effectiveness of a mindfulness meditation app together with a BCI-based wearable device among patients with AF during RFCA, which adds to the evidence base in the areas of meditation and mobile health. This study was designed rigorously. We provided the same care protocol and implemented the use of wearable devices for both groups, and recorded the time, energy, and temperature of ablation to ensure comparable conditions. In addition, this study was performed in a pragmatic setting and no maximum age for participation was stated, which adds to the generalizability of our findings.

This study also has some limitations. First, it was not a double-blind trial. Neither study staff nor patients were blinded to the intervention due to the nature of the study design. However, the data analysts were masked to group allocation. Additionally, meditation practice is a long-term process. Even with the help of apps and wearable devices, it takes time to master meditation techniques to reach a meditative state faster. At least one preoperative practice session could be added to the protocol for the intervention group; however, this might affect the comparability of the baseline measures. Lastly, this is a single-center study, thus limiting the generalizability of our findings.

Conclusions

In conclusion, this study shows that BCI-based app–delivered mindfulness meditation significantly relieved pain, anxiety, fatigue, and may reduce the doses of sedative medication used during RFCA for AF. Although no significant differences in hemodynamic parameters and the incidence of adverse events were observed, there was a decrease in the incidence of adverse events in the intervention group. Smartphone apps and wearable devices could serve as feasible and promising adjuncts to improve patients with AF's experience in RFCA.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (72074124), Young Scholars Fostering Fund of the First Affiliated Hospital of Nanjing Medical University (PY2022002), and the Clinical Ability Improvement Project of the First Affiliated Hospital of Nanjing Medical University (JSPH-NC-2021-16).

Authors' Contributions

YH and ZB contributed to the development and administration of the project. YH, ZB, ZT, CC, and GY assisted with implementing the protocol, collecting data, performing statistical analyses, and writing the manuscript. ZB and GS contributed to the conceptualization, methodology, supervision, and funding acquisition. All authors provided edits to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1295 KB - mhealth_v11i1e44855_app1.pdf]

References

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Revista Española de Cardiología (English Edition) 2021 May;74(5):437. [doi: 10.1016/j.rec.2021.03.009]
2. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines and the Heart Rhythm Society. J Am Coll Cardiol 2019 Jul 09;74(1):104-132 [FREE Full text] [doi: 10.1016/j.jacc.2019.01.011] [Medline: 30703431]
3. Magnussen C, Niiranen TJ, Ojeda FM, Gianfagna F, Blankenberg S, Njølstad I, BiomarCaRE Consortium. Sex differences and aimilarities in atrial fibrillation epidemiology, risk factors, and mortality in community cohorts: results from the BiomarCaRE Consortium (Biomarker for Cardiovascular Risk Assessment in Europe). Circulation 2017 Oct 24;136(17):1588-1597 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.028981] [Medline: 29038167]
4. Wang L, Ze F, Li J, Mi L, Han B, Niu H, et al. Trends of global burden of atrial fibrillation/flutter from Global Burden of Disease Study 2017. Heart 2021 Jun 04;107(11):881-887. [doi: 10.1136/heartjnl-2020-317656] [Medline: 33148545]
5. Sheikh A, Patel NJ, Nalluri N, Agnihotri K, Spagnola J, Patel A, et al. Trends in hospitalization for atrial fibrillation: epidemiology, cost, and implications for the future. Prog Cardiovasc Dis 2015 Sep;58(2):105-116. [doi: 10.1016/j.pcad.2015.07.002] [Medline: 26162957]
6. Rahman F, Kwan GF, Benjamin EJ. Global epidemiology of atrial fibrillation. Nat Rev Cardiol 2014 Nov 12;11(11):639-654. [doi: 10.1038/nrcardio.2014.118] [Medline: 25113750]
7. Chew DS, Li Y, Cowper PA, Anstrom KJ, Piccini JP, Poole JE, et al. Cost-effectiveness of catheter ablation versus antiarrhythmic drug therapy in atrial fibrillation: the CABANA randomized clinical trial. Circulation 2022 Aug 16;146(7):535-547. [doi: 10.1161/circulationaha.122.058575]
8. Liu Y, Liu Q, Yang Y, Zhang C, Yin H, Wu J, et al. Effect of radiofrequency catheter ablation on left atrial structure and function in patients with different types of atrial fibrillation. Sci Rep 2022 Jun 09;12(1):9511 [FREE Full text] [doi: 10.1038/s41598-022-13725-w] [Medline: 35681013]
9. Morillo CA, Verma A, Connolly SJ, Kuck KH, Nair GM, Champagne J, RAAFT-2 Investigators. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial. JAMA 2014 Feb 19;311(7):692-700. [doi: 10.1001/jama.2014.467] [Medline: 24549549]

10. Shi S, Tang Y, Zhao Q, Yan H, Yu B, Zheng Q, China Atrial Fibrillation Center Project Team). Prevalence and risk of atrial fibrillation in China: a national cross-sectional epidemiological study. Lancet Reg Health West Pac 2022 Jun;23:100439 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100439] [Medline: 35800039]
11. Writing committee of the report on cardiovascular health and diseases in China. Report on cardiovascular health and diseases in China 2021: an updated summary. Biomed Environ Sci 2022 Jul 20;35(7):573-603 [FREE Full text] [doi: 10.3967/bes2022.079] [Medline: 35945174]
12. Wang Z, Jia L, Shi T, Liu C. General anesthesia is not superior to sedation in clinical outcome and cost-effectiveness for ablation of persistent atrial fibrillation. Clin Cardiol 2021 Feb 29;44(2):218-221 [FREE Full text] [doi: 10.1002/clc.23528] [Medline: 33373042]
13. Hummel J, Elsayed-Awad H. Walking the tightrope between deep sedation and general anesthesia: by whom can this safely be done? J Cardiovasc Electrophysiol 2011 Dec;22(12):1344-1345. [doi: 10.1111/j.1540-8167.2011.02151.x] [Medline: 21848634]
14. Di Biase L, Saenz LC, Burkhardt DJ, Vacca M, Elayi CS, Barrett CD, et al. Esophageal capsule endoscopy after radiofrequency catheter ablation for atrial fibrillation. Circ: Arrhythmia and Electrophysiology 2009 Apr;2(2):108-112. [doi: 10.1161/circep.108.815266]
15. Smith HS, Laufer A. Opioid induced nausea and vomiting. Eur J Pharmacol 2014 Jan 05;722:67-78. [doi: 10.1016/j.ejphar.2013.09.074] [Medline: 24157979]
16. Koleck TA, Mitha SA, Biviano A, Caceres BA, Corwin EJ, Goldenthal I, et al. Exploring depressive symptoms and anxiety among patients with atrial fibrillation and/or flutter at the time of cardioversion or ablation. J Cardiovasc Nurs 2021;36(5):470-481. [doi: 10.1097/jcn.0000000000000723]
17. Wasserlauf J, Kaplan RM, Walega DR, Arora R, Chicos AB, Kim SS, et al. Patient-reported outcomes after cryoballoon ablation are equivalent between moderate sedation and general anesthesia. J Cardiovasc Electrophysiol 2020 Jul 27;31(7):1579-1584 [FREE Full text] [doi: 10.1111/jce.14547] [Medline: 32400079]
18. Attanasio P, Huemer M, Shokor Parwani A, Boldt LH, Mügge A, Haverkamp W, et al. Pain reactions during pulmonary vein isolation under deep sedation: cryothermal versus radiofrequency ablation. Pacing Clin Electrophysiol 2016 May 31;39(5):452-457. [doi: 10.1111/pace.12840] [Medline: 26926665]
19. Nørgaard MW, Pedersen PU, Bjerrum M. Visualisation during ablation of atrial fibrillation - stimulating the patient's own resources: patients' experiences in relation to pain and anxiety during an intervention of visualisation. Eur J Cardiovasc Nurs 2015 Dec 04;14(6):552-559. [doi: 10.1177/1474515114548643] [Medline: 25189476]
20. Lang EV, Benotsch EG, Fick LJ, Lutgendorf S, Berbaum ML, Berbaum KS, et al. Adjunctive non-pharmacological analgesia for invasive medical procedures: a randomised trial. Lancet 2000 Apr;355(9214):1486-1490. [doi: 10.1016/s0140-6736(00)02162-0]
21. Levine GN, Lange RA, Bairey - Merz CN, Davidson RJ, Jamerson K, Mehta PK, et al. Meditation and cardiovascular risk reduction. JAHA 2017 Oct 11;6(10):e002218. [doi: 10.1161/jaha.117.002218]
22. Izgu N, Gok Metin Z, Karadas C, Ozdemir L, Metinarikan N, Corapcıoglu D. Progressive muscle relaxation and mindfulness meditation on neuropathic pain, fatigue, and quality of life in patients with type 2 diabetes: a randomized clinical trial. J Nurs Scholarsh 2020 Sep;52(5):476-487. [doi: 10.1111/jnu.12580] [Medline: 32536026]
23. Thomas Z, Novak M, Platas SGT, Gautier M, Holgin AP, Fox R, et al. Brief mindfulness meditation for depression and anxiety symptoms in patients undergoing hemodialysis. CJASN 2017 Oct 12;12(12):2008-2015. [doi: 10.2215/cjn.03900417] [Medline: 29025788]
24. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
25. Mridha MF, Das SC, Kabir MM, Lima AA, Islam MR, Watanobe Y. Brain-computer interface: advancement and challenges. Sensors (Basel) 2021 Aug 26;21(17):5746 [FREE Full text] [doi: 10.3390/s21175746] [Medline: 34502636]
26. Daly JJ, Huggins JE. Brain-computer interface: current and emerging rehabilitation applications. Arch Phys Med Rehabil 2015 Mar;96(3 Suppl):S1-S7 [FREE Full text] [doi: 10.1016/j.apmr.2015.01.007] [Medline: 25721542]
27. Torres EP, Torres EA, Hernández-Álvarez M, Yoo SG. EEG-based BCI emotion recognition: a survey. Sensors (Basel) 2020 Sep 07;20(18):5083 [FREE Full text] [doi: 10.3390/s20185083] [Medline: 32906731]
28. Devipriya A, Nagarajan N. A novel method of segmentation and classification for meditation in health care systems. J Med Syst 2018 Sep 11;42(10):193. [doi: 10.1007/s10916-018-1062-y] [Medline: 30206721]
29. Pronk Y, Peters MCWM, Sheombar A, Brinkman J. Effectiveness of a mobile eHealth app in guiding patients in pain control and opiate use after total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e16415 [FREE Full text] [doi: 10.2196/16415] [Medline: 32167483]
30. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: A multicenter pilot clinical study. Pediatr Blood Cancer 2017 Oct;64(10):1-9. [doi: 10.1002/pbc.26554] [Medline: 28423223]
31. Smith JL, Allen JW, Haack C, Wehrmeyer K, Alden K, Lund MB, et al. The impact of app-delivered mindfulness meditation on functional connectivity and self-reported mindfulness among health profession trainees. Mindfulness (N Y) 2021 Oct 07;12(1):92-106 [FREE Full text] [doi: 10.1007/s12671-020-01502-7] [Medline: 33052251]

32. Chiu LYL, Sun T, Ree R, Dunsmuir D, Dotto A, Ansermino JM, et al. The evaluation of smartphone versions of the visual analogue scale and numeric rating scale as postoperative pain assessment tools: a prospective randomized trial. Can J Anaesth 2019 Jun 22;66(6):706-715. [doi: 10.1007/s12630-019-01324-9] [Medline: 30796700]
33. Blödt S, Pach D, Eisenhart-Rothe SV, Lotz F, Roll S, Icke K, et al. Effectiveness of app-based self-acupressure for women with menstrual pain compared to usual care: a randomized pragmatic trial. Am J Obstet Gynecol 2018 Feb;218(2):227.e1-227.e9 [FREE Full text] [doi: 10.1016/j.ajog.2017.11.570] [Medline: 29155036]
34. Liu W, Mao Y, Wei D, Yang J, Du X, Xie P, et al. Structural asymmetry of dorsolateral prefrontal cortex correlates with depressive symptoms: evidence from healthy individuals and patients with major depressive disorder. Neurosci Bull 2016 Jun 25;32(3):217-226 [FREE Full text] [doi: 10.1007/s12264-016-0025-x] [Medline: 27015663]
35. Park S, Sato Y, Takita Y, Tamura N, Ninomiya A, Kosugi T, et al. Mindfulness-based cognitive therapy for psychological distress, fear of cancer recurrence, fatigue, spiritual well-being, and quality of life in patients with breast cancer-a randomized controlled trial. J Pain Symptom Manage 2020 Aug;60(2):381-389. [doi: 10.1016/j.jpainsymman.2020.02.017] [Medline: 32105790]
36. Rizas KD, Freyer L, Sappler N, von Stülpnagel L, Spielbichler P, Krasniqi A, et al. Smartphone-based screening for atrial fibrillation: a pragmatic randomized clinical trial. Nat Med 2022 Sep 28;28(9):1823-1830. [doi: 10.1038/s41591-022-01979-w] [Medline: 36031651]
37. Cai C, Bao Z, Wu N, Wu F, Sun G, Yang G, et al. A novel model of home-based, patient-tailored and mobile application-guided cardiac telerehabilitation in patients with atrial fibrillation: A randomised controlled trial. Clin Rehabil 2022 Jan 16;36(1):40-50. [doi: 10.1177/02692155211032372] [Medline: 34266323]
38. Guo Y, Guo J, Shi X, Yao Y, Sun Y, Xia Y, mAF-App II Trial investigators. Mobile health technology-supported atrial fibrillation screening and integrated care: a report from the mAFA-II trial Long-term Extension Cohort. Eur J Intern Med 2020 Dec;82:105-111 [FREE Full text] [doi: 10.1016/j.ejim.2020.09.024] [Medline: 33067121]
39. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]
40. Bashir MU, Bhagra A, Kapa S, McLeod CJ. Modulation of the autonomic nervous system through mind and body practices as a treatment for atrial fibrillation. Rev Cardiovasc Med 2019 Sep 30;20(3):129-137 [FREE Full text] [doi: 10.31083/j.rcm.2019.03.517] [Medline: 31601087]
41. Thrall G, Lip GY, Carroll D, Lane D. Depression, anxiety, and quality of life in patients with atrial fibrillation. Chest 2007 Oct;132(4):1259-1264. [doi: 10.1378/chest.07-0036] [Medline: 17646231]
42. Guerrier G, Abdoul H, Jilet L, Rothschild P, Baillard C. Efficacy of a web app-based music intervention during cataract surgery: a randomized clinical trial. JAMA Ophthalmol 2021 Sep 01;139(9):1007-1013 [FREE Full text] [doi: 10.1001/jamaophthalmol.2021.2767] [Medline: 34323929]
43. Huberty J, Puzia ME, Green J, Vlisides-Henry RD, Larkey L, Irwin MR, et al. A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: Analysis from a randomized controlled trial. Gen Hosp Psychiatry 2021 Nov;73:30-37 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.004] [Medline: 34537477]
44. Chai PR, Schwartz E, Hasdianda MA, Azizoddin DR, Kikut A, Jambaulikar GD, et al. A brief music app to address pain in the emergency department: prospective study. J Med Internet Res 2020 May 20;22(5):e18537 [FREE Full text] [doi: 10.2196/18537] [Medline: 32432550]
45. Nørgaard MW, Werner A, Abrahamsen R, Larsen B, Darmer MR, Pedersen PU. Visualization and attentive behavior for pain reduction during radiofrequency ablation of atrial fibrillation. Pacing Clin Electrophysiol 2013 Feb 05;36(2):203-213. [doi: 10.1111/pace.12032] [Medline: 23127191]
46. Nørgaard MW, Pedersen PU, Bjerrum M. Understanding how patients use visualization during ablation of atrial fibrillation in reducing their experience of pain, anxiety, consumption of pain medication and procedure length: Integrating quantitative and qualitative results. Appl Nurs Res 2018 Feb;39:229-240. [doi: 10.1016/j.apnr.2017.11.026] [Medline: 29422164]
47. Roxburgh T, Li A, Guenancia C, Pernollet P, Bouleti C, Alos B, et al. Virtual reality for sedation during atrial fibrillation ablation in clinical practice: observational study. J Med Internet Res 2021 May 27;23(5):e26349 [FREE Full text] [doi: 10.2196/26349] [Medline: 34042589]
48. Crivelli D, Fronda G, Venturella I, Balconi M. Supporting mindfulness practices with brain-sensing devices. Cognitive and electrophysiological evidences. Mindfulness 2018 Jun 19;10(2):301-311. [doi: 10.1007/s12671-018-0975-3]
49. Balconi M, Fronda G, Crivelli D. Effects of technology-mediated mindfulness practice on stress: psychophysiological and self-report measures. Stress 2019 Mar 24;22(2):200-209. [doi: 10.1080/10253890.2018.1531845] [Medline: 30472901]
50. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
51. Zeidan F, Vago D. Mindfulness meditation-based pain relief: a mechanistic account. Ann N Y Acad Sci 2016 Jun;1373(1):114-127 [FREE Full text] [doi: 10.1111/nyas.13153] [Medline: 27398643]
52. Tang Y, Hölzel BK, Posner MI. The neuroscience of mindfulness meditation. Nat Rev Neurosci 2015 Apr 18;16(4):213-225. [doi: 10.1038/nrn3916] [Medline: 25783612]

53. Foji S, Tadayonfar MA, Mohsenpour M, Rakhshani MH. The study of the effect of guided imagery on pain, anxiety and some other hemodynamic factors in patients undergoing coronary angiography. Complement Ther Clin Pract 2015 May;21(2):119-123. [doi: 10.1016/j.ctcp.2015.02.001] [Medline: 25733083]
54. Bayrak A, Sagiroglu G, Copuroglu E. Effects of preoperative anxiety on intraoperative hemodynamics and postoperative pain. J Coll Physicians Surg Pak 2019 Sep 01;29(9):868-873. [doi: 10.29271/jcpsp.2019.09.868] [Medline: 31455484]
55. Angioli R, De Cicco Nardone C, Plotti F, Cafà EV, Dugo N, Damiani P, et al. Use of music to reduce anxiety during office hysteroscopy: prospective randomized trial. J Minim Invasive Gynecol 2014 May;21(3):454-459. [doi: 10.1016/j.jmig.2013.07.020] [Medline: 23962572]

Abbreviations

AF: atrial fibrillation
A-State: State Anxiety Inventory
BCI: brain-computer interface
BFI: Brief Fatigue Inventory
CONSORT: Consolidated Standards of Reporting Trials
EEG: electroencephalography
RFCA: radiofrequency catheter ablation
SpO^2: peripheral oxygen saturation

Edited by L Buis, G Eysenbach; submitted 06.12.22; peer-reviewed by X Gu, P Koopman, R Marshall, L Guo; comments to author 01.02.23; revised version received 28.03.23; accepted 13.04.23; published 03.05.23.

Please cite as:
He Y, Tang Z, Sun G, Cai C, Wang Y, Yang G, Bao Z
Effectiveness of a Mindfulness Meditation App Based on an Electroencephalography-Based Brain-Computer Interface in Radiofrequency Catheter Ablation for Patients With Atrial Fibrillation: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e44855
URL: https://mhealth.jmir.org/2023/1/e44855
doi:10.2196/44855
PMID:37133926

Original Paper

Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study

Hamish Fraser[1,2], MBChB, MSc; Daven Crossland[1,3], BA; Ian Bacher[1], MA, PhD; Megan Ranney[4], MPH, MD; Tracy Madsen[3,5], MD, PhD; Ross Hilliard[6], MD

[1]Brown Center for Biomedical Informatics, The Warren Alpert Medical School of Brown University, Providence, RI, United States
[2]Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, RI, United States
[3]Department of Epidemiology, Brown University School of Public Health, Providence, RI, United States
[4]School of Public Health, Yale University, New Haven, CT, United States
[5]Department of Emergency Medicine, The Warren Alpert Medical School of Brown University, Providence, RI, United States
[6]Department of Internal Medicine, Maine Medical Center, Portland, ME, United States

Corresponding Author:
Hamish Fraser, MBChB, MSc
Brown Center for Biomedical Informatics
The Warren Alpert Medical School of Brown University
233 Richmond Street
Providence, RI, 02912
United States
Phone: 1 401863 1815
Email: hamish_fraser@brown.edu

Abstract

Background: Diagnosis is a core component of effective health care, but misdiagnosis is common and can put patients at risk. Diagnostic decision support systems can play a role in improving diagnosis by physicians and other health care workers. Symptom checkers (SCs) have been designed to improve diagnosis and triage (ie, which level of care to seek) by patients.

Objective: The aim of this study was to evaluate the performance of the new large language model ChatGPT (versions 3.5 and 4.0), the widely used WebMD SC, and an SC developed by Ada Health in the diagnosis and triage of patients with urgent or emergent clinical problems compared with the final emergency department (ED) diagnoses and physician reviews.

Methods: We used previously collected, deidentified, self-report data from 40 patients presenting to an ED for care who used the Ada SC to record their symptoms prior to seeing the ED physician. Deidentified data were entered into ChatGPT versions 3.5 and 4.0 and WebMD by a research assistant blinded to diagnoses and triage. Diagnoses from all 4 systems were compared with the previously abstracted final diagnoses in the ED as well as with diagnoses and triage recommendations from three independent board-certified ED physicians who had blindly reviewed the self-report clinical data from Ada. Diagnostic accuracy was calculated as the proportion of the diagnoses from ChatGPT, Ada SC, WebMD SC, and the independent physicians that matched at least one ED diagnosis (stratified as top 1 or top 3). Triage accuracy was calculated as the number of recommendations from ChatGPT, WebMD, or Ada that agreed with at least 2 of the independent physicians or were rated “unsafe” or “too cautious.”

Results: Overall, 30 and 37 cases had sufficient data for diagnostic and triage analysis, respectively. The rate of top-1 diagnosis matches for Ada, ChatGPT 3.5, ChatGPT 4.0, and WebMD was 9 (30%), 12 (40%), 10 (33%), and 12 (40%), respectively, with a mean rate of 47% for the physicians. The rate of top-3 diagnostic matches for Ada, ChatGPT 3.5, ChatGPT 4.0, and WebMD was 19 (63%), 19 (63%), 15 (50%), and 17 (57%), respectively, with a mean rate of 69% for physicians. The distribution of triage results for Ada was 62% (n=23) agree, 14% unsafe (n=5), and 24% (n=9) too cautious; that for ChatGPT 3.5 was 59% (n=22) agree, 41% (n=15) unsafe, and 0% (n=0) too cautious; that for ChatGPT 4.0 was 76% (n=28) agree, 22% (n=8) unsafe, and 3% (n=1) too cautious; and that for WebMD was 70% (n=26) agree, 19% (n=7) unsafe, and 11% (n=4) too cautious. The unsafe triage rate for ChatGPT 3.5 (41%) was significantly higher (P=.009) than that of Ada (14%).

Conclusions: ChatGPT 3.5 had high diagnostic accuracy but a high unsafe triage rate. ChatGPT 4.0 had the poorest diagnostic accuracy, but a lower unsafe triage rate and the highest triage agreement with the physicians. The Ada and WebMD SCs performed better overall than ChatGPT. Unsupervised patient use of ChatGPT for diagnosis and triage is not recommended without improvements to triage accuracy and extensive clinical evaluation.

(JMIR Mhealth Uhealth 2023;11:e49995) doi:10.2196/49995

KEYWORDS

diagnosis; triage; symptom checker; emergency patient; ChatGPT; LLM; diagnose; self-diagnose; self-diagnosis; app; application; language model; accuracy; ChatGPT-3.5; ChatGPT-4.0; emergency; machine learning

Introduction

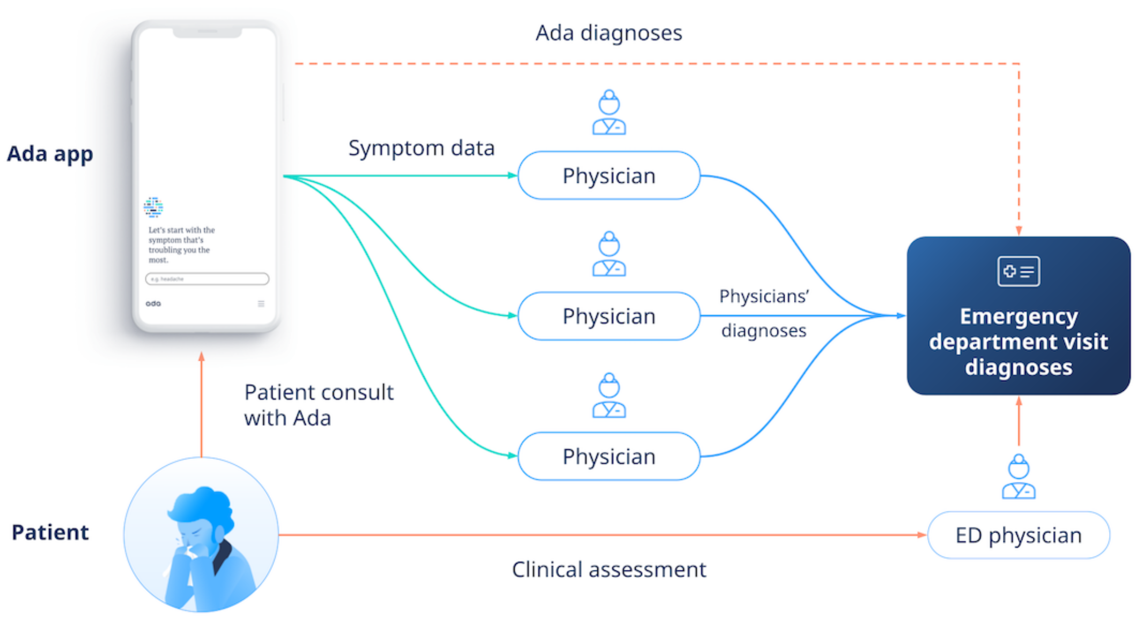
Accurate diagnosis is a key part of effective patient care, but misdiagnosis is common and can be harmful to patients [1]. Misdiagnosis can lead to delayed recognition of the true condition, which can misdirect the clinical history-taking, examination, and investigation. Diagnostic decision support systems have been developed for over 50 years, initially to support physicians [2-4]. However, over the last decade, their use has extended to direct and unsupervised use by patients with the development of diagnostic apps termed symptom checkers (SCs). Evidence from evaluation studies performed to date has shown that SCs have highly variable performance when tested with case vignettes (which typically include demographic data, some past medical history, and current symptoms). The best-performing systems have shown diagnostic and triage performance close to that of physicians using the same data, although poorer-performing systems have been shown to have lower accuracy than physicians or even patients tested on the same case vignettes [5-7].

Large language models (LLMs) are a new type of artificial intelligence technology designed primarily to predict the next words and phrases given some initial text. LLMs utilize neural networks that are inspired by neural structures seen in the human brain [8]. These networks are initially provided with certain tokens, which can consist of data points such as words or phrases, that allow the LLMs to "chunk" and analyze vast amounts of training data. As the LLM analyzes the data, patterns and associations between various tokens and other data contained in the training set build a map in which relationships between data points are quantified by differently weighted parameters. Once the model is appropriately trained and reliably produces the desired output, the outputs are reviewed and rated for accuracy by experts in the context the LLM is to be used, a process referred to as reinforcement learning with human feedback (RLHF) [9]. These ratings are then fed back to the LLM to improve the model's accuracy for a specific objective. LLMs have also demonstrated in-context learning, in which an existing model can complete a task after being prompted with a small number of examples, despite not originally being trained on these data [10].

GPT, or Generative Pre-trained Transformer (OpenAI Inc, San Francisco, CA), is a type of LLM that uses a transformer neural network with hundreds of billions of parameters to output human-like dialogue [11]. A recent version, GPT-3, was trained with a very large corpus of 570 Gigabytes of text collected from the internet. A newer version, GPT-3.5, has been shown to be capable of a wide range of tasks in different fields, including law and medicine. This model was linked to a chatbot to create ChatGPT 3.5, a version with which internet users can directly interact. Initial studies have shown that ChatGPT is capable of answering a wide range of questions about medicine and health. These include providing advice to patients on heart disease [12], performing tasks to support primary care physicians [13], and passing medical board exams based on performance in answering sample questions [14]. Studies have also shown that ChatGPT is capable of producing differential diagnoses when presented with medical data such as case summaries [15-17]. As with other uses of ChatGPT, there have been many examples reported of high accuracy or performance on medical tasks along with many examples of it outputting made-up data or "hallucinations" [18]. In many professional situations, an expert user can review the output of a system like ChatGPT and correct such errors; however, this is not generally the case if patients use the system directly. This situation is analogous to the use of SC apps, where poorly designed and evaluated SCs could put patients at risk [19]. ChatGPT is currently available (July 2023) in 2 versions: the original, publicly available 3.5 version and a subscription-based 4.0 version trained on an even larger set of data. It has not been reported whether either model has undergone specific RLHF for medicine or health care.

In a previous study [20], we evaluated a widely used SC from Ada Health, which in 2022 had been used by 11 million users carrying out 23 million health assessments [21]. The Ada app includes a chatbox functionality for eliciting a clinical history and a diagnostic and triage algorithm that has been extensively tested and updated based on preclinical and clinical testing. The system is based on a Bayesian network, which was built using extensive data from published clinical studies and input from expert clinicians. As with most available SCs, Ada is not based on a machine learning approach [7]. Ada has supported many clinical evaluation studies in-house and assisted independent researchers, providing extensive published data on its usability and its diagnostic and triage performance [6,7,20,22-24]. In the previous study, the Ada SC was used by patients waiting for clinical assessment in the emergency department (ED) of Rhode Island Hospital (RIH), Rhode Island, USA. The diagnoses of Ada were compared to the physicians' visit notes after they had seen the patient. The symptom data collected by Ada were then blindly reviewed by 3 independent ED physicians who provided their own diagnosis and triage results and then critiqued the Ada results. Triage is the process of assessing the risk of clinical deterioration of a patient and using this to prioritize care, which is typically carried out by a nurse or physician using a standard system soon after a patient's arrival at an ED. Figure 1 shows the study design and data collection and flow. Overall, Ada performed well with diagnostic accuracy close to the physician reviewers. The unsafe triage rate of 15% (with at least 2 of 3 physician reviewers agreeing) was similar to or better than that found in several studies of nurse triage [20].

Figure 1. Data collection and flow in the Rhode Island Hospital emergency department (ED) study [20].



While ChatGPT has been shown to have surprisingly good diagnostic accuracy on a range of clinical case descriptions, examples reported to date appear to use standardized case vignettes rather than real patient consultations. For example, Levine et al [15] tested ChatGPT on the 48 standard case vignettes that were previously used for a study of patients' performance in diagnosis and triage after reading these vignettes. The results showed high diagnostic accuracy, whereas triage accuracy was moderately good but no better than that of the patients overall (see the Discussion for more details).

Our objective in this study was to evaluate the diagnostic and triage accuracy of ChatGPT 3.5 and 4.0 using the real patient data collected by the Ada SC in the ED and the actual diagnoses from the ED physicians who saw the patients. We also evaluated the most commonly used SC app in the United States, WebMD [25], in an identical fashion for comparison.

Methods

Data Source

The data for this study came from a previous study performed at the RIH ED [20]. Patients in an assessment room waiting to see an ED physician were (1) recruited by a research assistant (RA), (2) completed a consent form, (3) used the Ada SC app on an iPad to self-report symptoms, and then (4) completed a short user survey in REDCap [26]. Patients with acute or serious medical conditions presenting to an ED were deemed eligible for inclusion in the study. In addition, the patients had to be English-speaking; aged ≥18 years; presenting for emergency evaluation of a medical (ie, nontrauma and nonmental health) problem; and deemed by the triage nurse to not be critically ill, defined as an Emergency Severity Index (ESI) score [27] of 2-5. Those with an ESI score of 1 were excluded. The diagnosis and triage results generated by Ada were sent by secure email to the study team and were not shown to the patient or the treating physician. Patients had to complete the study prior to being seen by the ED physician and were then given a US $20 gift card. Information on the final diagnoses from the ED physician's assessment was extracted by an RA from the EPIC electronic health record (EHR).

Study Design

In analysis of the prior study [20], after the completion of patient recruitment, the clinical data from Ada (age; sex; history of heart disease, diabetes, and hypertension; current pregnancy; and presenting symptoms and follow-up questions, in the form of a list of symptoms reported as present or absent) were reviewed by 3 independent ED physicians who had not seen the patient. The physicians were asked to provide 3 possible diagnoses and an appropriate triage level (ranging from ED visit to home care). They were also asked to critique the Ada diagnoses and triage suggestions. Overall, 30 cases had sufficient data for diagnostic analysis and 37 cases had sufficient data for triage analysis. A list of all 40 case presentations and diagnoses and an example of the symptom data collected by Ada are available as appendices from Fraser et al [20].

The diagnoses of Ada and of the independent physicians were compared to the final diagnoses from the ED physician. The process of data collection along with the diagnosis and triage comparison method were based on the previous study from Fraser et al [20], as shown in Figure 1. In summary, (1) the Ada app collects the *demographic and symptom data*, (2) these data are later reviewed by 3 independent (ED) physicians, and (3) the *Ada diagnoses* and the 3 *independent physicians' diagnoses* are compared to the *ED visit diagnoses* recorded in the EHR by the ED physician who carried out the *clinical assessment*. The primary metric for diagnostic accuracy was the percentage of cases where at least one of the Ada or independent physicians (the diagnostic agent) diagnoses matched at least one of the treating physicians' diagnoses (this could also be defined as the sensitivity of the diagnostic agent for at least one of the final diagnoses). The results were stratified into matches of the top diagnosis from the diagnostic agent and one of the top 3 diagnoses. These metrics were applied here to ChatGPT 3.5 and 4.0 and the WebMD SC.

Triage accuracy was calculated by comparing the triage level given by the diagnostic agent with the triage level of at least 2 of the 3 independent physicians. Triage was recorded as recommended levels of urgency of care: (1) home care, (2) routine primary care, and (3) urgent or emergency care. The independent physicians were asked for triage recommendations using a more detailed scale, which was then mapped to these 3 levels. ChatGPT has standard text for routine care: "It is recommended that the individual consults with a healthcare professional for a proper evaluation and diagnosis." The response from WebMD varies by diagnosis. An example of routine care triage for the diagnosis of chronic kidney disease is: "See your doctor if you have any symptoms of chronic kidney disease such as fatigue, decreased urination, swelling of legs, nausea, headache, and weakness." For urgent/emergency triage, ChatGPT outputs variations similar to the following text: "It is important for the patient to seek immediate medical attention," while WebMD displays an alert at the top of the diagnosis page that reads: "This is an URGENT CONDITION. Please contact your doctor or call 911."

This study design and the clinical data collected were used to evaluate the two versions of ChatGPT, versions 3.5 and 4.0. An RA (DC) entered the symptom data collected by the Ada SC for each case into ChatGPT and recorded the output. The RA was blinded to the actual diagnosis and triage results from the 2022 study [20]. For comparison, the Ada symptom data were also entered into the widely used SC from WebMD [25]. As with the previous study, diagnostic accuracy was evaluated by the number of cases where a ChatGPT or WebMD diagnosis matched at least one diagnosis from the ED physician who saw the patient. This was reported for matches for the top suggested diagnosis or any of the top 3 diagnoses. Matching was carried out by one author (HF) and then reviewed and verified by another author (RH), resulting in small changes to the overall scores.

The original study [20] used a version of Ada from 2018; however, the current version has undergone extensive updates, in part to qualify under EU regulations. As an approach to compare differences in the algorithms, we attempted to reenter data from the first 20 cases into the current version of Ada (released June/July 2023). The questions asked by Ada had been updated, resulting in modest differences in the questions regarding symptoms recorded as present, but large differences in questions regarding symptoms recorded as absent. It was therefore not possible to reliably compare the performance by this method; hence, only the original Ada performance is reported.

Ethical Approval

The original study was approved by the Institutional Review Board, Research Data Protection Office, Lifespan Healthcare, Providence, Rhode Island (1439681-3). Only secondary deidentified data were used for the current study.

Results

Diagnostic Accuracy

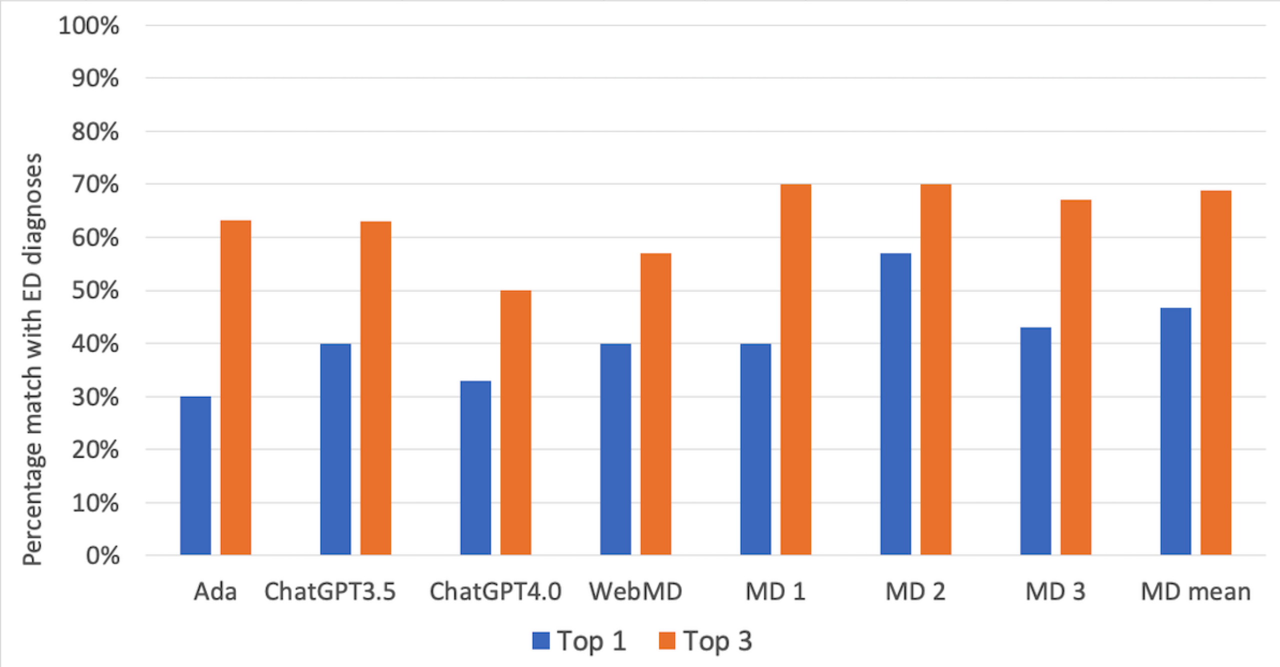
Results for diagnostic accuracy are shown in Table 1. From the original study [20], Ada had a match rate between its top-1 diagnosis and the final ED diagnosis of 9 (30%). The top match rate for ChatGPT 3.5, ChatGPT 4.0, WebMD, and the physicians mean rate was 40% (n=12), 33% (n=10), 40% (n=12), and 47%, respectively. The top-3 diagnostic match rate for Ada, ChatGPT 3.5, ChatGPT 4.0, WebMD, and physicians mean rate was 63% (n=19), 63% (n=19), 50% (n=15), 57% (n=17), and 69%, respectively.

Table 1. Diagnosis results for Ada, ChatGPT, and WebMD (N=30).

Matches	Ada, n (%)	ChatGPT 3.5, n (%)	ChatGPT 4.0, n (%)	WebMD, n (%)
Top 1	9 (30)	12 (40)	10 (33)	12 (40)
Top 3	19 (63)	19 (63)	15 (50)	17 (57)

Figure 2 shows the comparison of the 4 diagnostic tools with the 3 independent ED physicians using the same symptom data. Because the ED physician actually saw the patient and carried out a physical examination and likely investigations, their diagnostic accuracy will likely be higher than can be achieved with the Ada symptom data alone.

Figure 2. Diagnostic matches of Ada, ChatGPT 3.5 and 4.0, WebMD, and the independent emergency department (ED) physicians (MD 1, MD 2, MD 3). The gold standard is defined as diagnoses from ED discharge notes.



ChatGPT 3.5 had the strongest combined diagnostic performance of the 4 systems, whereas ChatGPT 4.0 had the weakest combined performance. The differences between the results for top-1 and top-3 diagnosis matches between the different systems and the mean physician performance were not significant. We also examined any additional diagnosis matches at the rank of 4th or 5th in cases that did not already have a match. ChatGPT 3.5 had no additional matching diagnoses below the rank of 3, ChatGPT 4.0 had 3 additional matches, WebMD had 1, and Ada had 2. The physicians were only asked for 3 possible diagnoses. Most matches were clear, with less than one-quarter requiring any discussion, primarily to determine whether the level of specificity of the diagnoses matched. For example "Lower Back Herniated Disk" from ChatGPT 3.5 and "Lumbar (Low-Back) Herniated Disc" from WebMD were not considered a match with "acute bilateral back pain" from the ED physician, whereas "lumbar strain" from ChatGPT 4.0 was rated a match. Another issue was that ChatGPT 4.0 appeared to behave differently when prompted with "what is the diagnosis" than if the case was entered without the "what is the diagnosis" prompt. A broader differential was often generated without the prompt. This issue was not seen with ChatGPT 3.5.

Triage Accuracy

As shown in Table 2, overall triage performance based on agreement between the systems and at least 2 of the 3 physicians was the highest for ChatGPT 4.0 (76%) and the lowest for ChatGPT 3.5 (59%). On the key statistic of unsafe triage, Ada performed the best with 14%, followed by WebMD with 19%, but ChatGPT 3.5 had much poorer performance at 41%. Comparing safe versus unsafe triage for Ada (5/37, 14%) and ChatGPT 3.5 (15/37, 41%), the difference was significant (P=.009; χ^2 test). The difference between WebMD and ChatGPT 3.5 was not significant (P=.08) (Multimedia Appendix 1 shows the detailed triage results including 95% CIs). ChatGPT 3.5 had no examples of triage that were too cautious and had a high rate of missing serious conditions requiring urgent or emergency care. ChatGPT 4.0 showed different behavior to ChatGPT 3.5. Agreement with the physician reviewers was high at 76% (28/37), the rate of *unsafe* triage fell to 22% (8/37), and the rate of *too cautious* triage increased slightly to 3% (1/37). Between ChatGPT 3.5 and 4.0, one case changed from *agree* to *unsafe* and one changed from *agree* to *too cautious*.

Table 2. Triage accuracy of symptom checkers and ChatGPT compared to the assessment of independent emergency department physicians (N=37).

Triage assessment	Ada, n (%)	ChatGPT 3.5, n (%)	ChatGPT 4.0, n (%)	WebMD, n (%)
Agree	23 (62)	22 (59)	28 (76)	26 (70)
Unsafe	5 (14)	15 (41)	8 (22)	7 (19)
Too cautious	9 (24)	0 (0)	1 (3)	4 (11)

Discussion

Principal Findings

This study builds on a previously validated set of cases entered by actual patients presenting for emergency or urgent care. We believe these results provide a unique assessment of the performance of WebMD, ChatGPT 3.5, and ChatGPT 4.0 on real patient data, as opposed to vignettes created by physicians (as used in most other studies that have assessed the performance of ChatGPT). Differences between patient-reported data and physician-created vignettes may include patients' medical knowledge or understanding of questions, the ability to distinguish symptoms not present versus unknown, and elicitation of unusual but potentially diagnostic symptoms. Additionally, the artificial limitation of most vignettes to point to a single diagnosis simplifies calculations of diagnostic accuracy but does not reflect most of the real clinical assessments seen in this study. A limitation here is that diagnostic accuracy is defined as sensitivity of the diagnostic agent's output to at least one of the reviewing ED physicians' diagnoses, rather than an assessment of the SC's concordance with the full differential given by the ED physician. Comprehensiveness and relevance scores (which are similar to sensitivity and positive predictive value, respectively) can be used to account for the proportion of correct diagnoses matched by a diagnostic tool, which correlated with the diagnosis matching scores in the earlier Ada study [20].

Fully assessing diagnostic tools such as SCs requires that all the stages of use are evaluated: (1) patients' ability to use the app, (2) ability of the app to collect a clinical history, (3) performance of the diagnostic and triage algorithms, and (4) patients' ability to interpret and act on the results presented. The results presented here primarily address point (3), the performance of the diagnostic algorithm. Usability of the Ada app (1) was demonstrated in the earlier study [20]. Ada has also been shown to collect a more comprehensive history than most SC apps (2) [23]. In the similar study of Ada performed in a primary care setting, the independent physicians reviewing the symptom data of the 201 cases were asked to indicate any questions they considered were missing from the Ada-collected history. In 134 (67%) of the cases, 0 or 1 of the clinicians suggested additional questions (these results are in preparation for publication). WebMD might perform less well in direct use by patients as it typically collects less clinical data. The performance of ChatGPT as a history-taking tool in the hands of patients has yet to be determined. We are planning additional studies to address (4), the effect of the outputs of SCs and LLMs such as ChatGPT on patient decision-making in seeking care.

In the earlier study of Ada that generated the data used here [20], the handling of cases with symptoms recorded as the final ED diagnosis was reviewed by the research team. In total, 33 cases had full data to carry out diagnostic analysis, but in 3 cases, the ED diagnosis remained unclear despite a second chart review (2 cases had a chief complaint of abdominal pain and the other had a chief complaint of dizziness). These 3 cases were excluded from the diagnostic analysis but included in the triage analysis. In 6 cases, the patient had chest pain as a symptom and underwent a screen for acute myocardial infarction (AMI). In those cases, a diagnosis of AMI or unstable angina was considered correct even if the patient ultimately had a negative screen. Two cases had a "diagnosis" listed as back pain; they were included in the analysis due to the high frequency of patients with back pain seen in the ED without specific diagnoses. Previous studies have shown that "symptom-based discharge" is common in EDs [28]. The process of defining one or more "correct diagnosis" from the ED physician's assessment and the EHR note they created was a significant challenge. It was previously noted that "In many cases, the ED physicians' role was to exclude serious causes for the presenting symptoms, with the patient potentially seeking investigation through their primary care physician or specialist at a later date" [20].

The result of triage performance presented a mixed picture. Compared to Ada, WebMD had a higher *agree* level with the independent physicians (70%) but also had a higher *unsafe* rate of 19%. ChatGPT 3.5 had a slightly lower *agree* score than Ada and a much higher *unsafe* triage rate. For the urgent and serious cases seen in this study, such an unsafe triage rate could put patients at genuine risk, discouraging them from seeking needed care quickly. ChatGPT 4.0 had a significantly better triage performance with a higher rate of *agree* compared to the two SCs and to previous vignette studies [5,6]. The unsafe triage rate of ChatGPT 4.0 was still higher than that for Ada or WebMD but much lower than that for ChatGPT 3.5.

There are a range of systems and scoring systems available for triage (such as the Manchester Triage System used by Cotte et al [22] and the ESI [27] used at RIH), which makes comparison with other studies more difficult. The earlier study [20] tested the effect of adding vital sign data on physicians' triage decisions, but the effect was small. Studies of how patients are currently using LLMs for medical advice would also strengthen the conclusions and recommendations obtained here for future development and use of these systems.

Comparison With Other Studies

In 2015, Semigran et al [5] compared the performance of 23 available SCs with 45 case vignettes. The average SC sensitivity for the correct diagnosis for their top diagnosis was 35.5% and that for the top 3 diagnoses was 51%. Independent physicians achieved a sensitivity of 72% for their top diagnosis and 84.3% for the top 3 diagnoses [29]. The 3 best-performing apps had a top-1 sensitivity of 43%-50% and top-3 sensitivity of 67%-70%, which are slightly higher than those obtained in this study. A repeat of that study 5 years later [6] showed an improvement in average diagnosis performance (of apps tested in both studies) with top-1 sensitivity of 45.5%. The top-1 sensitivity for Ada in that study was higher at 53%. These vignette studies differ from the real cases seen here in having only one "correct" diagnosis.

These studies also assessed triage accuracy for the SC apps that provided the output. In the 2015 study of Semigran et al [5], the mean correct triage (match between app and vignette) rate was 59.1%. For the Schmieding et al [6] study performed 5 years later, the rate was slightly lower at 55.8%. However, the authors noted that the apps were less risk-adverse, resulting in undertriaging more than 40% of the vignettes rated as an emergency, which was a similar level to the result with ChatGPT shown here. Schmieding et al [6] used the 45 case vignettes (with simplified language) to evaluate the triage accuracy of lay users. They showed a correct triage rate of 60.9% for lay users compared to only 58% for SCs. They noted that "most lay participants outperformed all but 5 SCs." Importantly, the SCs had higher scores for emergencies in this study but lower scores for low-risk cases. Three studies evaluated SC performance on actual emergency or urgent care patients. Cotte et al [22] (working for Ada) evaluated Ada's triage accuracy on 378 "walk-in" patients in urgent care. They compared its triage accuracy with the result from the Manchester Triage System and showed an undertriage rate of 8.9% of cases and an overtriage rate of 57.1%. A study of 2 SCs based on chart review of 100 records in an ED in Hong Kong showed triage accuracies of 50% and 74%, but noted poorer performance on more urgent cases [30]. A study in Canada of a locally developed SC showed significantly better triage accuracy than patients (73% vs 58%; *P*<.01), with better performance on emergency cases [31].

In 2023, Levine et al [15] used the 48 case vignettes developed for a previous study of patients directly assessing diagnosis and triage [32]. They evaluated GPT-3, the underlying model of ChatGPT. The results showed that GPT-3 had the correct diagnosis from the vignette in its top 3 88% of the time compared to the correct rate of lay individuals of 54%, (*P*<.001). this figure was 75%, and for lay individuals 43%. For triage, GPT-3 had an *agree* rate of 70%, which was similar to that of lay individuals of 74%. They did not report unsafe triage rates. Unlike the present study, Levine et al [15] were able to evaluate the performance of GPT-3 on low-risk "self-care" cases and showed that triage accuracy was much lower at 50%. SC evaluation studies have also shown lower accuracy on self-care cases [6]. This is a particular concern regarding the ability of diagnostic tools to help reduce overload on urgent and emergency care systems. The authors also note that their vignettes are "simulated cases" and that the diagnostic agents "may perform differently when presented with real-world symptoms." Another concern is that presenting data to ChatGPT in different ways may affect the output; therefore, if patients pose a question in a specific manner, they may receive less accurate results. More guidance is required regarding the standardization of medical queries and ideally specific entry modes for medical questions in LLMs.

A notable feature of the triage behavior of ChatGPT 3.5 shown here was that *in many cases it conflicted with the diagnoses it produced*. Cases rated as unsafe in this study included diagnoses that were correctly made for myocardial infarction, pyelonephritis, and head injury, yet urgent/emergency triage was not recommended. A simple triage rule that tied the minimum triage level to the most serious and urgent diagnoses listed would have the potential to correct these errors and reduce unsafe triage performance. Both Ada and WebMD appear to link diagnosis and triage levels in this way. Similar discordant results between diagnosis and triage have been seen with other SCs in a previous study [6].

The behavior of ChatGPT 4.0 was significantly different from that of version 3.5. There was a large improvement in triage performance, which at 76% was the highest agreement with the physicians' triage seen in our study. This is also very high compared to the performances from previous vignette studies [5,6,33]. Unsafe triage also improved with the updated version of ChatGPT, but was still almost twice that of Ada. In traditional diagnostic algorithms (such as Bayesian networks used by Ada), the output threshold can be adjusted to favor higher or lower levels of urgency of care overall. ChatGPT 3.5 and 4.0 would likely benefit from such an adjustment, reducing *unsafe* triage and increasing the *too cautious* category. Given the generally high diagnostic accuracy seen with ChatGPT 3.5, safety and consistency of triage and confidence in the system would seem to require such "guard rails." An important aspect of the performance of ChatGPT 4.0 was the substantial drop in diagnostic accuracy compared to that of version 3.5, which could make tying triage to diagnoses less effective. ChatGPT 4.0 also produced more examples of general diagnostic categories than found for version 3.5, such as "renal disease" rather than specific diseases that had strong matches to the ED diagnoses.

Chen et al [34] recently compared the performance of ChatGPT versions 3.5 and 4.0 on four standard problems both soon after the initial release of version 4.0 in March 2023 and again in June 2023. They reported large falls in performance on tasks such as math problems and programming with the newer version, but some improvements in the previous version. Both versions had more formatting mistakes in code generation in June. These differences were attributed to large changes in the LLM algorithms, which were made without notification to the users (who had paid a subscription for the version 4.0 service). We used ChatGPT 4.0 for the current ED study in June 2023 and may therefore have seen lower diagnostic and triage performance than studies carried out earlier. Gilbert et al [35] made an assessment of the potential regulation of LLMs as medical devices, particularly for tasks such as clinical decision support. For the categories of *Verification*, *Usability*, and *Surveillance*, they noted that the near-infinite range of possible inputs and outputs prevents standardized regulation of current approaches. For the categories of *Provenance* and *Changes*, they noted the lack of a stable controlled versioning (as also noted above) as a serious barrier. As there is no proven method currently available to prevent harmful outputs, they argue that current methods of *risk mitigation* are also ineffective. While LLMs used for decision support are required to be regulated as medical devices in the United States and European Union/United Kingdom [35], they do not seem to qualify in their current form. These concerns are all multiplied if the system is being used by patients without input or oversight from medical professionals.

Limitations

This study did not include examples of patients that would likely not have required any clinical care (ie, could have managed their symptoms at home). Future studies need to evaluate the question of overtriage and the risk of increasing unnecessary visits to health facilities, an area our primary care study should help to address. Other limitations of this study include the relatively small number of cases available for measurement of diagnosis. This is being addressed in the ongoing 201-patient primary care study, along with newly funded studies being developed in the ED and primary care settings with a plan to recruit an additional 700 patients. Larger studies will also help address questions regarding the performance and usability of systems by different patient groups. The usability of the Ada app by patients assessed through the questionnaire in the ED and primary care studies did not detect evidence of differences by race, ethnicity, or gender. However, lower usability for patients aged over 60 years was seen.

Another possible limitation is that changing the wording of prompts to ChatGPT can make a significant difference to the system's output. This has the potential to change measured diagnostic and triage performance. To address this, we intentionally used a simple approach to entering the data and simply prompted with "What is the diagnosis" to try and capture the likely behavior of patients. This prompt is similar to that used in the previous study of patient diagnosis and health information–seeking behavior by Levine and colleagues [32]. Further study is required to evaluate the effects of different prompts; as noted in the Results, behavior was different with ChatGPT 3.5 and 4.0. It is our view that the developers of ChatGPT (Open AI Inc and Microsoft Inc) and other LLMs have a responsibility to provide clear instructions, accept a range of prompts, and query the user if their intention is not clear, rather than place the responsibility and risk on the patient.

Conclusions

LLMs such as ChatGPT are a new technology with surprising performance in the analysis of medical data, including the provision of medical diagnosis and triage. The ease of use of these tools and the high quality and appealing textual output they produce are likely to make them attractive to people seeking answers in many areas, including medicine and health. As has been seen with the use of search engines such as Google [36] and existing SCs [7,20], a substantial proportion of the population are likely to use such tools. Now that ChatGPT is built into the search engine Bing (Microsoft Inc), it is likely to see much greater use. The results of this study show both the important potential of LLMs like ChatGPT and a number of serious concerns about their current performance and safety. In future studies, we plan to include other LLMs such as Bard and the medically focused example Med-PaLM, both from Google Inc. Rigorous clinical evaluation of this sort with a wide range of real patients and different types of potential illnesses is essential to prevent these exciting and attractive tools from achieving widespread and unregulated use without a clear understanding of the risks. Otherwise, it is likely they will provide unsafe advice to a subset of patients with potentially life-threatening conditions and increase unnecessary visits to urgent or emergency care for less serious conditions.

Acknowledgments

This study received funding from the Brown University Office of the Vice President of Research (OVPR) Seed award, (HF, MR, TM and RH), and from the Agency for Healthcare Research and Quality (R01HS029513; HF, MR, TM and RH). MR also received funding from COBRE NIGMS P20 GM139664. TM also received funding from NHLBI K23 HL140081.

None of the authors has any connections to OpenAI, Microsoft Inc, or WebMD. As described in the previous study [20], there was technical assistance to allow the use of the Ada symptom checker app but no funding was provided by Ada Health Inc and the company had no involvement in the current study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed results of the triage matches shown in Table 2, including 95% CIs.
[PDF File (Adobe PDF File), 53 KB - mhealth_v11i1e49995_app1.pdf]

References

1. Hamilton W. Diagnosis: Cancer diagnosis in UK primary care. Nat Rev Clin Oncol 2012 Apr 03;9(5):251-252. [doi: 10.1038/nrclinonc.2012.63] [Medline: 22473095]
2. Vardell E, Moore M. Isabel, a clinical decision support system. Med Ref Serv Q 2011 Apr 25;30(2):158-166. [doi: 10.1080/02763869.2011.562800] [Medline: 21534115]
3. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA 1987 Jul 03;258(1):67-74. [doi: 10.1001/jama.258.1.67] [Medline: 3295316]
4. Shortliffe EN, Shah NH. AI in medicine: some pertinent history. In: Cohen TA, Patel VL, Shortliffe EH, editors. Intelligent systems in medicine and health. The role of AI. Cham: SpringerLink; 2022:21-50.
5. Semigran H, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
6. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
7. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
8. Hardesty L. Explained: neural networks. MIT News. 2017. URL: https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414 [accessed 2023-06-13]

9. What is ChatGPT? Commonly asked questions about ChatGPT. OpenAI. 2023. URL: https://help.openai.com/en/articles/6783457-what-is-chatgpt [accessed 2023-06-13]
10. Akyurek E. What learning algorithm is in-context learning? Investigations with linear models. arXiv. 2023. URL: https://arxiv.org/abs/2211.15661 [accessed 2023-06-13]
11. Zewe A. Solving a machine-learning mystery. MIT News. 2023 Feb 07. URL: https://news.mit.edu/2023/large-language-models-in-context-learning-0207 [accessed 2023-06-13]
12. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). medRxiv. 2023. URL: https://www.medrxiv.org/content/10.1101/2023.03.25.23285475v1 [accessed 2023-06-13]
13. Thirunavukarasu AJ, Hassan R, Mahmood S, Sanghera R, Barzangi K, El Mukashfi M, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. JMIR Med Educ 2023 Apr 21;9:e46599 [FREE Full text] [doi: 10.2196/46599] [Medline: 37083633]
14. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
15. Levine D, Tuwani R, Kompa B, Varma A, Finlayson SG, Mehrotra A, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model. medRxiv. 2023 Feb 01. URL: https://www.medrxiv.org/content/10.1101/2023.01.30.23285067v1 [accessed 2023-06-13]
16. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
17. Benoit JR. ChatGPT for clinical vignette generation, revision, and evaluation. medRxiv. 2023 Feb 08. URL: https://www.medrxiv.org/content/10.1101/2023.02.04.23285478v1 [accessed 2023-06-13]
18. Alkaissi HS, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 2023 Feb;15(2):e35179 [FREE Full text] [doi: 10.7759/cureus.35179] [Medline: 36811129]
19. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Nov 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
20. Fraser HSF, Cohan G, Koehler C, Anderson J, Lawrence A, Pateña J, et al. Evaluation of diagnostic and triage accuracy and usability of a symptom checker in an emergency department: observational study. JMIR Mhealth Uhealth 2022 Sep 19;10(9):e38364 [FREE Full text] [doi: 10.2196/38364] [Medline: 36121688]
21. Ada Health. 2022. URL: https://ada.com/ [accessed 2023-06-13]
22. Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, et al. Safety of triage self-assessment using a symptom assessment App for Walk-in Patients in the emergency care setting: observational prospective cross-sectional study. JMIR Mhealth Uhealth 2022 Mar 28;10(3):e32340 [FREE Full text] [doi: 10.2196/32340] [Medline: 35343909]
23. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021 Jul 15;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
24. Montazeri M, Multmeier J, Novorol C, Upadhyay S, Wicks P, Gilbert S. Optimization of patient flow in urgent care centers using a digital tool for recording patient symptoms and history: simulation study. JMIR Form Res 2021 May 21;5(5):e26402 [FREE Full text] [doi: 10.2196/26402] [Medline: 34018963]
25. WebMD symptom checker. URL: https://symptoms.webmd.com/default.htm#/info [accessed 2022-06-13]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. Emergency Severity Index (ESI): a triage tool for emergency departments. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/patient-safety/settings/emergency-dept/esi.html [accessed 2021-02-01]
28. Slovis B, McCarthy D, Nord G, Doty A, Piserchia K, Rising K. Identifying Emergency Department Symptom-Based Diagnoses with the Unified Medical Language System. West J Emerg Med 2019 Oct 24;20(6):910-917 [FREE Full text] [doi: 10.5811/westjem.2019.8.44230] [Medline: 31738718]
29. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. JAMA Intern Med 2016 Dec 01;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]
30. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S, Leung L. Triage accuracy of online symptom checkers for accident and emergency department patients. Hong Kong J Emerg Med 2019 Apr 16;27(4):217-222. [doi: 10.1177/1024907919842486]
31. Chan F, Lai S, Pieterman M, Richardson L, Singh A, Peters J, et al. Performance of a new symptom checker in patient triage: Canadian cohort study. PLoS One 2021 Dec 1;16(12):e0260696 [FREE Full text] [doi: 10.1371/journal.pone.0260696] [Medline: 34852016]

32. Levine DM, Mehrotra A. Assessment of diagnosis and triage in validated case vignettes among nonphysicians before and after internet search. JAMA Netw Open 2021 Mar 01;4(3):e213287 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3287] [Medline: 33779741]
33. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun 11;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
34. Chen L, Zharia M, Zou J. How is ChatGPT's behavior changing over time? arXiv. 2023 Aug 01. URL: https://arxiv.org/abs/2307.09009 [accessed 2023-09-01]
35. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023 Jun 30:online ahead of print. [doi: 10.1038/s41591-023-02412-6] [Medline: 37391665]
36. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis 2018 Sep 25;5(3):95-105 [FREE Full text] [doi: 10.1515/dx-2018-0009] [Medline: 30032130]

Abbreviations

AMI: acute myocardial infarction
ED: emergency department
EHR: electronic health record
ESI: Emergency Severity Index
LLM: large language model
RA: research assistant
RIH: Rhode Island Hospital
RLHF: reinforcement learning with human feedback
SC: symptom checker

Edited by L Buis; submitted 16.06.23; peer-reviewed by M Millenson, M Sim; comments to author 06.07.23; revised version received 17.08.23; accepted 25.08.23; published 03.10.23.

Please cite as:
Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R
Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study
JMIR Mhealth Uhealth 2023;11:e49995
URL: https://mhealth.jmir.org/2023/1/e49995
doi:10.2196/49995
PMID:37788063

Evaluating the Impact of an mHealth Platform for Managing Acute Postoperative Dental Pain: Randomized Controlled Trial

Bunmi Tokede[1,*], DDS, DMSc; Alfa Yansane[2,*], PhD; Ana Ibarra-Noriega[1], DDS; Joanna Mullins[3], MHI, RDH; Kristen Simmons[3], PhD, RDH; Nicholas Skourtes[3], DMD; Urvi Mehta[1], MPH; Sayali Tungare[1], PhD; David Holmes[4], BDS; Joel White[2], DDS; Muhammad Walji[1], PhD; Elsbeth Kalenderian[2,5], DDS, PhD

1

2

3

4

5

*these authors contributed equally

Corresponding Author:
Bunmi Tokede, DDS, DMSc

Abstract

Background: Postoperative dental pain is pervasive and can affect a patient's quality of life. Adopting a patient-centric approach to pain management involves having contemporaneous information about the patient's experience of pain and using it to personalize care.

Objective: In this study, we evaluated the use of a mobile health (mHealth) platform to collect pain-related patient-reported outcomes over 7 days after the patients underwent pain-inducing dental procedures; we then relayed the information to the dentist and determined its impact on the patient's pain experience.

Methods: The study used a cluster-randomized experimental study design with an intervention arm where patients were prompted to complete a series of questions relating to their pain experience after receiving automated text notifications on their smartphone on days 1, 3, 5, and 7, with the resulting information fed back to dentists, and a control arm where patients received usual care. Providers were randomized, and patients subsequently assumed the enrollment status of their providers. Providers or their staff identified eligible patients and invited them to participate in the study. Provider interviews and surveys were conducted to evaluate acceptance of the mHealth platform.

Results: A total of 42 providers and 1525 patients participated. For the primary outcome (pain intensity on a 1 to 10 scale, with 10 being the most painful), intervention group patients reported an average pain intensity of 4.8 (SD 2.6), while those in the control group reported an average pain intensity of 4.7 (SD 2.8). These differences were not significant. There were also no significant differences in secondary outcomes, including pain interference with activity or sleep, patient satisfaction with pain management, or opioid prescribing. Patient surveys revealed reluctance to use the app was mostly due to technological challenges, data privacy concerns, and a preference for phone calls over texting. Providers had high satisfaction with the app and suggested integrating additional features, such as an in-system camera for patients to upload pictures and videos of the procedural site, and integration with the electronic health record system.

Conclusions: While the mHealth platform did not have a significant impact on acute postoperative pain experience, patients and providers indicated improvement in patient-provider communication, patient-provider relationship, postoperative complication management, and ability to manage pain medication prescribing. Expanded collaboration between mHealth developers and frontline health care providers can facilitate the applicability of these platforms, further help improve its integration with the normal clinic workflow, and assist in moving toward a more patient-centric approach to pain management.

Trial Registration: ClinicalTrials.gov NCT03881891; https://www.clinicaltrials.gov/study/NCT03881891

(JMIR Mhealth Uhealth 2023;11:e49677) doi:10.2196/49677

KEYWORDS

mobile health; patient-reported outcomes; acute pain; dentistry; dental; dentist; pain; mHealth; patient-reported outcome; PRO; patient-reported outcome measures; PROs; PROM; randomized controlled trial; RCT

Introduction

The experience of pain is a national and global public health problem with significant physical, cognitive, and emotional costs [1-3]. Postoperative dental pain, in particular, is pervasive and can significantly affect a patient's quality of life and ability to perform daily activities [4]. Untreated or poorly managed postoperative dental pain can lead to complications such as infection, delayed healing, and the need for additional dental treatment [5]. Effective pain management should prioritize the individual needs and preferences of each patient. This patient-centered approach may involve tailoring the treatment plan to the patient's specific needs, providing clear and concise information about pain management options, and actively involving the patient in the decision-making process. By taking a more holistic approach to acute pain management, health care professionals can help ensure that patients receive the most effective and personalized care possible [6].

While dentists are prescribing fewer postoperative opioids [7], current practice suggests that opioid prescriptions are often discordant with evidence-based prescription guidelines [8], especially after common oral surgery procedures such as dental extractions. Third molar extractions are the dental procedures most likely to be associated with an opioid prescription [9]. This is problematic because dentists are responsible for a disproportional share of opioids prescribed to adolescents, for whom even a single opioid prescription increases the lifetime risk of future opioid abuse [10]. One reason for the inappropriate prescribing of opioids is that oral health providers are unable to accurately predict or actively monitor postoperative pain. Their desire to prevent unwanted unscheduled visits leads them to pre-emptively prescribe opioids in an attempt to satisfy patients' short-term pain management expectations. Patient expectations of receiving the most effective pain relievers, coupled with diminished patient satisfaction and negative reviews if their expectations are not met, provide yet another perverse incentive for pre-emptive opioid prescriptions [11].

Adopting a patient-centric approach to pain management involves collecting valuable information about the patient's experience of pain and related factors, and providing the information back to the dentist to help manage the care [12]. Patient-reported outcomes (PROs) and PRO measures play a crucial role in this process. PROs refer to any report of the patient's health status that comes directly from the patient, while patient-reported outcome measures are validated questionnaires that patients complete to self-assess their health status [13]. Patient self-reporting is a critical part of comprehensive pain assessment [14], given pain's subjective and multidimensional nature. PROs allow clinicians to directly assess patients' symptoms, symptom burden, functional status, health behaviors, health-related quality of life, and care experiences [15], and deliver value-based care.

Against this backdrop, the use of mobile health (mHealth) systems for the collection of PROs is on the rise [16-21]. An mHealth system is a platform that incorporates mobile devices, wireless communication technologies, and software apps to deliver health care services and information to patients and health care providers. These platforms can be designed for various purposes such as remote monitoring of patients, disease management, or telemedicine and are potentially powerful platforms for the delivery of behavior change interventions because they can improve engagement with established strategies for prevention and treatment through personalized goal setting, individualized dosing reminders, and gamification [22]. By leveraging mHealth systems to collect, integrate, and analyze PROs, providers can efficiently gather valuable information about the patient's pain experience and improve the effectiveness of pain management strategies. In dentistry, the timely and efficient capture of PRO data, such as postoperative pain experience, through an mHealth system is lacking, and this represents a missed opportunity to improve patient outcomes, care experience, and provider performance.

Therefore, this study aimed to assess the impact of an mHealth platform on acute dental postoperative pain management in terms of pain experience and patient satisfaction. We also explored the providers' perspectives on the use of mobile technology in the management of acute postoperative dental pain.

Methods

Study Overview

A 24-month phase 2 cluster randomized controlled trial was conducted to evaluate the impact of using an mHealth platform on patient postoperative pain experiences, satisfaction with pain management, and dental provider satisfaction with the platform. The multicenter study was conducted at an academic dental institution and a large privately held dental group practice. Data collection spanned February 2020 through January 2022. Consented providers or staff identified eligible patients and invited them to participate in the study.

Study Sites and Participants

The study was conducted at two dental institutions. One is part of an academic dental site and the other is a large privately held dental group practice of around 50 offices across the Pacific Northwest region of the United States. The academic dental center comprises predoctoral, resident, and faculty clinics. The patient population provides a diverse sample in terms of demographics and socioeconomic status.

The provider inclusion criteria were being a general dentist or specialist in oral and maxillofacial surgery endodontics, or periodontics; performing any one or combination of the identified potentially pain-inducing procedures (see list below); practicing for a minimum of two clinic sessions per week (ie, one full clinic day); having a minimum of 6 months of practice experience; and having access to and willingness to use a smartphone.

The patient inclusion criteria were being 18 years or older and having access to and ability to use a smartphone.

Included "Pain-Inducing" Procedures

The core set of pain-associated dental procedure codes (Code to Dental Terminology; American Dental Association) included were endodontics: D3310, D3320, D3330, D3346, D3347, D3348, D3410, D3421, D3425, D3426, and D3450; periodontal surgery: D4210, D4211, D4212, D4240, D4241, D4249, D4260, D4261, and D4263; oral surgery: D7210, D7220, D7230, D7240, D7241, D7250, D7310, D7311, D7320, and D7321; and implant dentistry: D6010, D6011, D6012, D6013, D6040, D6050, D6100, D6101, D6102, D6103, D6104, and D6081.

Intervention

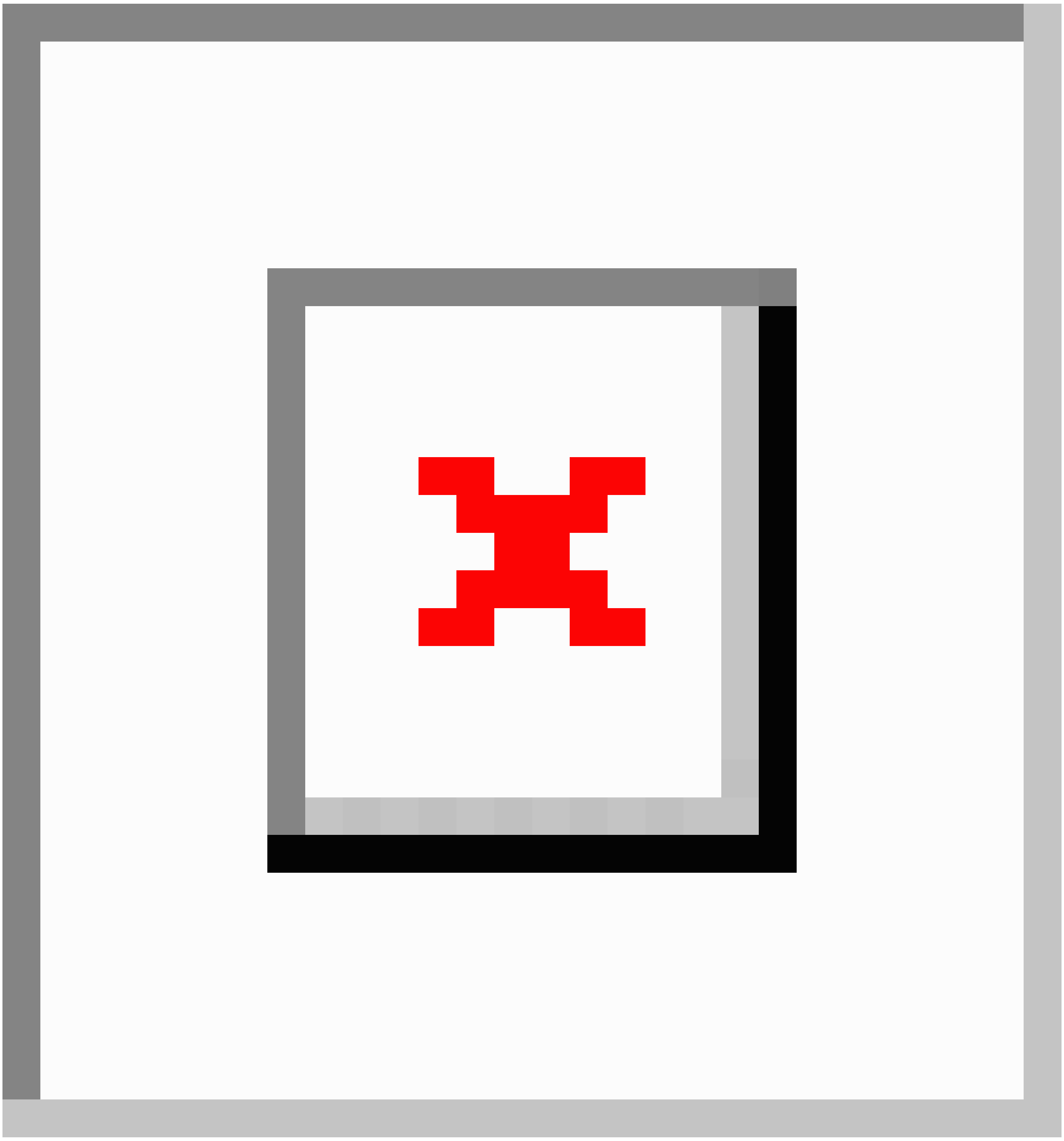
The mHealth platform deployed in this study was FollowApp.Care. A detailed description of the platform has been previously published [23]. Briefly, FollowApp.Care is a communications platform to collect patient-generated health data before or after a procedure. The platform is designed to inform treatment decisions, improve patient care, and generate performance reports. FollowApp.Care can be accessed through any SMS text message–enabled smartphone.

On completion of any of the eligible procedures, enrolled patients (including intervention and control groups) received postoperative care instructions and guidance according to each institution's standard practice (usual care). The intervention group received additional guidance about FollowApp.Care, including the timing and frequency of text notifications and when to expect a response from providers or office staff, if necessary. Patients in the intervention group received text notifications at predetermined time intervals (eg, 9 AM) on days 1, 3, 5, and 7 prompting them to complete a brief pain assessment survey covering the preceding 24-hour period. Additionally, a comment/chat feature enabled patients to securely communicate more information to their dental care team through FollowApp.Care when needed. The control group received usual care and was advised to contact their providers or dental offices through the usual channels if they experienced any unexpected symptoms or had any complaints or questions. Control participants filled out the PRO pain survey only on day 7. To ensure that FollowApp.Care was implemented as intended, fidelity was also measured.

Randomization

Each of the participating providers was randomized to one study arm (the mHealth intervention plus standard care vs standard care only), and each patient automatically assumed the randomization status of their provider. As such, each patient was nested within a specific provider (Figure 1). Randomization was conducted using pseudorandom number generation, and randomization codes were maintained by the statistician in a secure cloud-based storage system. Each week, consenting providers or their staff identified eligible patients and invited them to participate in the study. Using a standardized template provided by the research team, clinic staff members (who had undergone training in human subjects' protection) obtained informed consent from interested patients to confirm their willingness to participate in the study before their procedures. The intervention could not be masked.

Figure 1. Randomization scheme: each provider represents a single cluster that was randomized to either the intervention or control group. Each participating patient seen would then assume the randomization status of their provider.



Means of Data Collection

We used the mHealth platform (FollowApp.Care) to collect PRO data (pain experience) from patients after dental procedures. Electronic health record (EHR) data for postprocedure prescribing data was extracted using the patient enrollment data. EHR data was then merged with the mHealth survey response data for each patient.

Study Outcomes

The primary PRO of interest was pain intensity—an assessment of the worst severity of pain experienced during the 7 days after an eligible dental procedure—and the data was collected using an item from the validated Patient-Reported Outcomes Measurement Information System (PROMIS) Shortform 3A Version 1 questionnaire [24]. The response categories range from "No pain" to "Very severe" and are measured on a 0 to 10 rating scale. The outcome was treated as continuous.

Secondary Outcomes

Pain Interference

Pain interference, defined as interference with activity (walking, work, general activity, sleep) and interference with affect (mood, enjoyment of life), was captured using 3 items taken from the validated Revised American Pain Society Patient Outcome Questionnaire (APS-POQ-R) form [25]. Response categories ranged from "No interference" to "High interference" on a 0 to

10 rating scale. Each item queries how, in the last 7 days, pain interfered with doing activities such as walking, sitting in a chair, or standing at the sink; falling asleep; and staying asleep.

Patient Satisfaction

Satisfaction with how pain was managed was assessed with the following two statements from the validated APS-POQ-R form, which was measured on a 0 to 10 rating scale [25]:

1. Ability to participate in decisions about pain treatment
2. Satisfaction with the results of your pain treatment

Use of Opioid Medications

The proportion of participating patients who got a postoperative opioid prescription was assessed using data from the patient EHR. Through secondary analysis of the EHR, medication-prescribing patterns were collected by deploying query scripts to identify the patients who received the prescriptions postoperatively, including type, dosage, frequency, and duration.

Sample Size

Among the 2 included dental sites, a total of 42 providers were recruited to participate in the study over the 2-year study period. Each provider was expected to reasonably recruit 19 patients per year. The expected number of patients was 1596. Adjusting for a 60% response rate among recruited patients, we calculated a sample of 958. Given a total sample size of 958 patients, a standard significance level of .05 (α=.05), and a within-cluster correlation coefficient of 0.1 (ρ=0.1), we estimated that the power to detect a 2.0-unit effect difference in pain would be 80.7%. The minimum power achieved was derived using R (version 4.3.1 for Windows; R Foundation for Statistical Computing; longpower package). All statistical analyses were performed at the standard significance level (α=.05) using R.

Statistical Methods

Means and corresponding estimates of precision (eg, SDs and 95% CIs) and frequency distributions with percentage contributions were used to report the distribution of each variable included in the quantitative analyses. To test whether there was a difference in pain intensity, interference, or satisfaction with pain management between the study groups on day 7, a hierarchical model was performed that adjusted for within-clinic correlations and repeated measures over patient responses. Models included the procedure type, time, age, gender, and race/ethnicity.

Fidelity

Fidelity was measured using metrics as outlined in Textbox 1.

Textbox 1. Fidelity metrics for patients and providers.

Patient fidelity measures

- Provided verbal consent and received the information sheet.
- FollowApp.Care profile was created
- Received text notifications on day 0
- Patient response time
- Number of patients who have phone service provided by T-Mobile
- Response rate day 1
- Response rate day 3
- Response rate day 5
- Response rate day 7

Dentist fidelity measures

- Signed consent forms before training
- Completed 1-hr training
- Verified FollowApp.Care profile
- Unique identifiers provided
- Completed Unified Theory of Acceptance and Use of Technology survey
- Number of log-ins
- Number of successful log-ins
- Number of unsuccessful log-ins
- Number of alerts triggered
- Number of alerts resolved
- Number of alerts resolved by chat
- Number of alerts resolved by phone
- Number of alerts resolved by acknowledgment
- Number of alerts unresolved
- Average response time to alerts

Assessing Provider Acceptance

To assess whether practitioners were unduly burdened by the technology and whether it fit seamlessly into their workflow, the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire was administered to those in the intervention group. Four key constructs were measured: performance expectancy, effort expectancy, social influence, and facilitating conditions. A descriptive analysis was performed to describe the constructs of the UTAUT questionnaire.

Semistructured virtual interviews were also conducted with dental care providers from both study sites. Nine of these interviews were with dentists alone, and five were group interview sessions with dental care teams that consisted of dentists, dental assistants, dental hygienists, or dental clinic administrative staff. The main aim of these interviews was to evaluate the provider's experience with using the mHealth platform for managing their patients' postoperative pain, including its impact on their clinic workload, workflow patterns, and satisfaction with the effectiveness of pain management. Our analysis for this research focused on using this interview data to identify the barriers and facilitators for using an mHealth platform for postoperative acute pain management and communication. One trained interviewer conducted all the interviews. Each interview was audio or video recorded through Zoom video telephonic software (Zoom Video Communications) and then transcribed using Rev speech-to-text transcription services. For the qualitative analysis, a combination of deductive and inductive approaches was used. Independent coding of the transcripts was performed by 2 of the authors, the coding was assessed and discussed for variation and consensus, and codes were identified that fit into the predefined themes of barriers and facilitators for the use of the mHealth platform. The framework method of analysis [26] was used to organize and analyze the codes and themes.

Ethical Considerations

The study protocol was reviewed and approved by the University of Texas Institutional Review Board (IRB# 18-25477) and registered on ClinicalTrials.gov (NCT03881891). Using a standardized template provided by the research team, providers or clinic staff members obtained informed consent from interested patients before their respective surgical procedures.

Results

Patient and Provider Population

A total of 42 providers (intervention: n=24; control: n=18), consisting of 24 general dentists, 16 endodontists, and 2 oral surgeons, participated in the trial. The study included 1525 patients (intervention: n=851; control: n=674) with an average age of 44.5 (SD 14.3) years, of whom 675 (44.3%) were female and 865 (56.7%) were White (Table 1). The most common procedures were oral surgery procedures.

Table . Patient Characteristics.

Variable		Control (n=674)	Intervention (n=851)
Gender, n (%)			
	Male	255 (37.8)	292 (34.3)
	Female	313 (46.4)	362 (42.5)
	Other	106 (15.8)	197 (23.2)
Race n (%)			
	Asian	7 (1.0)	19 (2.2)
	American Indian/Alaska Native	3 (0.4)	0 (0.0)
	Black	5 (0.7)	9 (1.1)
	Hispanic/Latino	35 (5.2)	45 (5.3)
	Native Hawaiian/Pacific Islander	5 (0.7)	4 (0.5)
	White	432 (64.1)	433 (50.9)
	More than one race	18 (2.7)	29 (3.4)
	Other	5 (0.7)	4 (0.5)
	Unknown	164 (24.3)	308 (36.2)
Age of patients (years), mean (SD)		44.8 (13.8)	44.3 (14.6)

Fidelity

Response rates for the mHealth-administered surveys were 56.9% (484/851) on day 1 and 49.8% (424/851) on day 7 for intervention patients, and 42% (283/674) for control patients. All patients had a FollowApp.Care profile created, and 98.3% (1504/1525) received the day 1 SMS text message notification, with an average response time of about 7 hours. Over the study duration, 349 alerts were generated. Of these, 335 were resolved, 14 were unresolved, and the average response time to patient alerts was 8 hours and 48 minutes.

Postoperative Pain Experience

For the primary outcome "How intense was your pain at its worst following your procedure?" (pain intensity), intervention group patients reported an average pain intensity of 4.8 (SD 2.6), while those in the control group reported an average pain intensity of 4.7 (SD 2.8). These differences were not significant. Intervention group patients also responded to the following question: "What is your pain level right now?" The mean pain intensity ranged from 2.9 (SD 2.4) on day 1 to 1.2 (SD 1.8) on day 7 post procedure.

Table 2 shows that there were no significant differences in interference in falling asleep and staying asleep between the intervention and control groups.

Table . Pain interference scores for activities, falling asleep, and staying asleep.

	Pain interference score		*P* value
	Control (n=674)	Intervention (n=851)	
Activities	0.8	1	.19
Falling asleep	1.6	1.9	.08
Staying asleep	1.4	1.7	.36

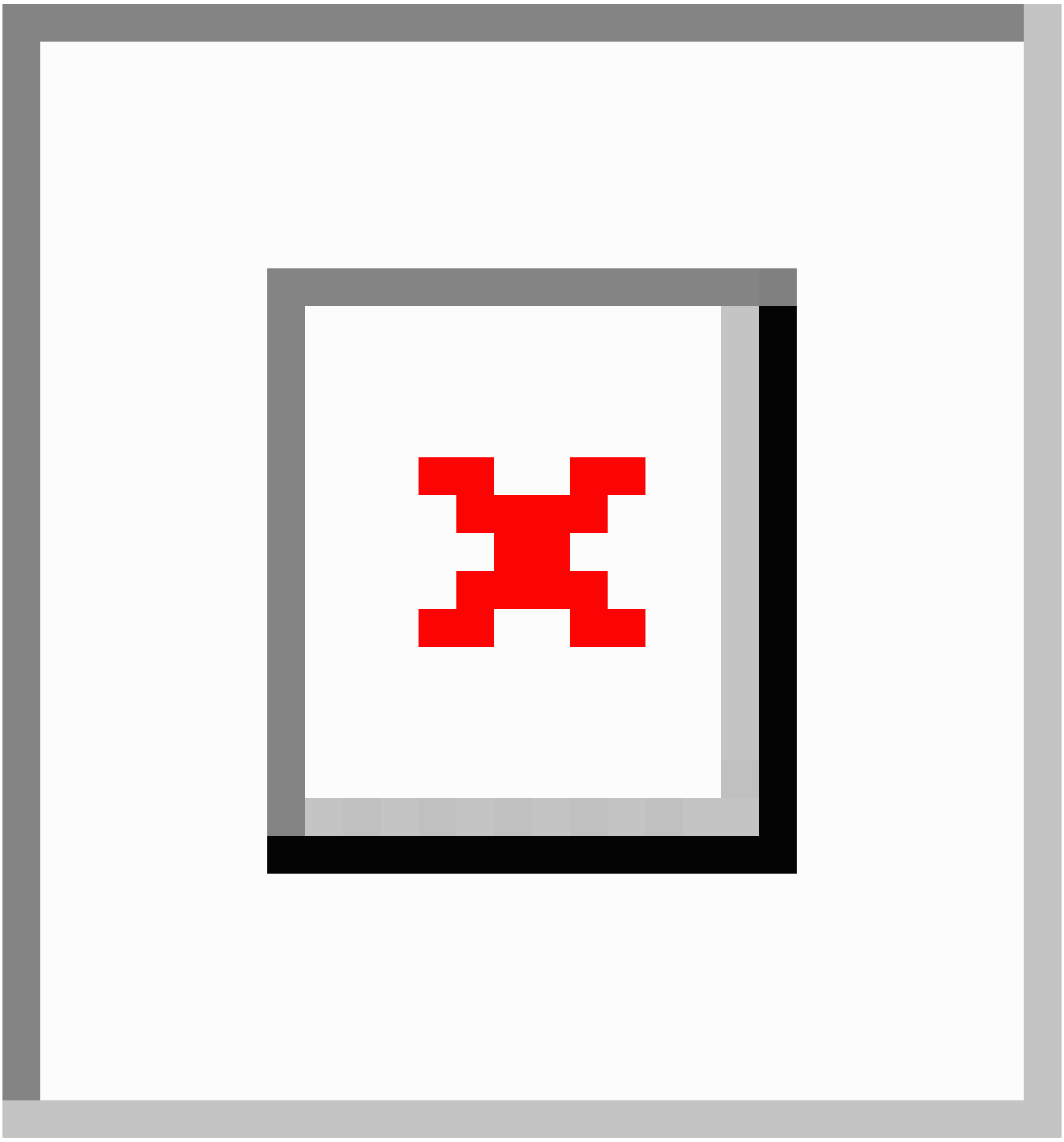
Patient Satisfaction

In response to the question "Were you allowed to participate in decisions about your pain treatment as much as you wanted to? (0, least to 10, most)," respondents in the intervention group reported an average of 7.7 (SD 3.5) out of 10 in participation in decision-making, while those in the control group reported an average of 8.3 (SD 3.0). When asked "Select the one number that best shows how satisfied you are with the results of your pain treatment," respondents in the intervention group reported an average of 8.6 (SD 2.2) out of 10, while those in the control group reported 8.9 (SD 2.0). There was no significant difference between the groups.

Use of Opioid Medications

Figure 2 displays the most frequently used patient-reported medications. Descriptive statistics were used to determine the distribution of opioids prescribed to the responding patients by the providers. In total, 26.4% (225/851) of patients in the intervention group were prescribed an opioid, while 16.8% (113/674) of those in the control group were prescribed an opioid. Nearly 50% of the opioid prescriptions were written by only 3 providers. Using the mHealth platform did not appear to have an effect on the odds of opioids prescribed after a dental procedure (odds ratio 1.17, 95% CI 0.61-1.64; *P*=.40) after adjusting for gender, procedure group, and provider.

Figure 2. Patient-reported medication use.



Adjusted Analysis

The regression analysis showed that there was no statistically significant difference between the intervention and control arms for all study outcomes, after adjusting for provider, gender, and procedure type.

Provider Experience With the mHealth App

Results from the UTAUT questionnaire indicated that most providers found the platform useful, clear, and understandable; that their organization in general thought they should use it; and that they have the necessary resources and knowledge to use the platform. The validated UTAUT questionnaire was administered to 18 intervention providers. The four UTAUT constructs are associated with a behavioral intention to use the new technology (FollowApp.Care app); high scores on each of the constructs are associated with a higher behavioral intention to use the FollowApp.Care platform. The responses to the four items that form the performance expectancy construct showed that most providers found FollowApp.Care useful, enabling them to perform tasks more quickly, increasing productivity, and increasing the chances of a positive performance review. Median scores for each item were ≥4 on a 7-point Likert scale. The responses to the four items that form the effort expectancy construct showed that most providers found that FollowApp.Care is clear and understandable, believing that they can become skillful and that the platform is easy to use and operate. Median scores for each item were ≥5 on a 7-point Likert scale. The responses to the four items that form the social influence construct showed that most providers found that those who influence their behavior, people who are important to them, and their clinical management as well as the organization in general thought that they should use the platform. Median scores for each item were ≥5 on a 7-point Likert scale. The responses to the four items that form the facilitating conditions construct show that most providers found that they have the necessary resources and knowledge to use FollowApp.Care, that it is generally compatible with other systems that they use, and that there is assistance for its operation. Median scores for each item were ≥4 on a 7-point Likert scale (see Multimedia Appendix 1 for descriptive tables).

Qualitative Analyses

Three main themes were identified from the perspective of providers regarding the use of the platform for postoperative acute pain management: (1) potential facilitators and barriers to adoption, (2) patient acceptance and hesitancy, and (3) future use of the platform (Multimedia Appendix 2). Providers seemed to appreciate the improved accessibility for patients.

> *It seemed like we were a lot more accessible. It would lower their anxiety or if they were scared that something was going on, they were able to get answers a lot quicker. I think most of the patients liked it from what I remember hearing.*

The chat feature was particularly helpful in facilitating direct communication between dentists and patients, resulting in improved patient care and stronger patient-dentist relationships. Dentists found the alert system useful for identifying patients with specific symptoms and reducing unnecessary postoperative appointments. However, they also reported feeling a lack of personal touch through mHealth SMS text messaging. Furthermore, they thought it represented an additional burden on their workload and an invasion of their time due to receiving messages and alerts after work hours.

> *Actually sometimes it (using the platform) was just an additional step, because I had to respond on the app and then call the patient because they still needed to talk to me. I still needed to talk to them. I felt like texting them wasn't enough. You know? So, yeah, I think in that way it was probably good for data collection and everything, but I think that that created an extra step for us.*

Patients' reluctance to use the app was mostly due to technological challenges, data privacy concerns, and a preference for phone calls over texting. Dentists suggested integrating additional features into the app, such as an in-system camera for patients to upload pictures and videos of the procedural site, integration with the EHR system, and including postoperative expectations and instructions in the platform for the patient to access after the procedure.

Discussion

Principal Findings

In this prospective, randomized, parallel-arm clinical trial evaluating the impact of an mHealth app on overall dental postoperative acute pain experience, we found no significant differences in pain experience or use of analgesic medication after painful dental procedures between the intervention (mHealth) and control (standard care) arms. Providers and patients, however, reported that the use of an mHealth platform had significant potential for improving patient-provider communication, patient-provider relationships, postoperative complication management, and the ability to customize pain medication prescribing. Almost all previously completed trials on the use of mobile technology apps for pain management in health care have focused on chronic pain. No trials have been reported in dentistry.

Several factors may have influenced the observed lack of difference between the study arms. First, comparisons between intervention and control groups were made on day 7. Previous studies suggest that postoperative dental pain is usually of short duration, reaching its maximum intensity in the early postoperative period (day 1) and petering out before day 7, regardless of the pain control technique [27]. Second, the included sites already have robust processes in place for postoperative patient care management (usual care). This includes 24-hour on-call dentists who are available by phone after standard hours for all questions and emergencies. This reduces the likelihood that patients in either arm would have experienced significant postoperative pain management issues. For the same reason, patients may have felt adequately involved in postoperative care decisions—hence, no difference was observed in patient satisfaction. Third, although there were four different types of "painful procedures" included in this study, the intervention arm seemed to have a lopsided proportion (681/851, 80% vs 359/674, 53.3% in the control arm) of patients undergoing the most painful procedures (oral surgery procedures). This uneven distribution resulted from the fact that patients assumed the randomization group of their providers. Fourth, the patient response rate decreased from day 1 to day 7 in the intervention group, with fewer patients responding to the surveys over time. This is a common issue with survey-based research, as patients may decline to respond to surveys, which could have introduced a selection bias. Finally, the importance of the design characteristics of mHealth apps should also be considered [28]. Characteristics such as reminders, notifications, incentives, follow-up, and the way these functions are provided can affect whether and how an app is used. It should also be noted that the timing and frequency of reminders must be well designed or they will be ignored by users [29]. A previous study found that users were most likely to use an app in 24 hours when the notification was sent at noon on weekends [30]. Gamification and incentive mechanisms such as virtual badges, unlocked levels, and behavior data comparison with other users are also considered driving forces for use [28].

Nevertheless, this study highlights the potential of PROs for providing valuable data for optimizing the delivery of care. Mobile phones have been shown to be an effective platform for assessing various aspects of patient health, including symptoms, symptom burden, health status, health behaviors, and health-related quality of life. In dentistry, mobile apps have been used to encourage evidence-based oral hygiene routines [31-34], triage emergencies [35], and prevention of dental caries [36]. Additionally, extensive evidence from systematic reviews and meta-analyses in medicine has demonstrated that mobile apps can effectively improve physical and mental health [37], medication adherence [16], and self-management of disease [17].

Our qualitative analysis revealed that the use of mHealth systems could be clinically useful in ways that have also been reported in other studies. For example, a study conducted in rural Ghana found that providers perceived the use of mHealth technology to be an approach to increasing health care access [38]. Similarly, another study reported that the use of mHealth technology improved patient communication [39]. In our study, providers perceived the mHealth app to be useful in guiding medication prescription, in contradiction to another study in which providers were concerned about overprescribing medication when using mHealth technology [40].

Postoperative pain measurement by recall is difficult to accurately determine. Research on autobiographical memory [41] indicates that recall is not just subject to random error but also is fraught with systematic bias, which can distort recall even after relatively short intervals. Many experiences are not retained in memory, so often the information we are asked to provide simply is not available for direct retrieval. A dramatic demonstration of the biases in recall—and an indication of how quickly these biases can set in—was reported by Redelmeier et al [42]. Summary ratings of pain by patients who had undergone a colonoscopy 20-30 minutes earlier were found to be unduly influenced by the peak level of pain (presumably because it was most salient) and the pain intensity at the end of the procedure (most recent). In other words, recall did not accurately represent the average pain over the interval because it was based on a few of the most memorable moments, essentially ignoring most of the experience. This shows the potential for bias even over short intervals. Besides being distorted by the operation of heuristic recall strategies, memory is also influenced by what we know and believe rather than actual recall. People unconsciously reorganize their "memories" to make them fit a coherent script or theory of events, or to reconcile events with what transpired subsequently [43]. Ecological momentary assessment (EMA) methods and technologies, designed to support the self-report of experience in the moment of daily life, are being considered poised to revolutionize human-centered research [44]. mHealth platforms could potentially be deployed more effectively if used in the context of EMA methods in which patients report their pain experience at the moment they are experiencing it and do not have to wait to receive survey prompts [45].

Limitations

This study was conducted at two sites where standard postoperative care is exemplary, with disciplined adherence to evidence-based guidelines. Future studies should focus on pragmatic trials including sites that are more similar to everyday dental clinics with less stringent protocols, processes, or guidelines in place. EMA approaches should also be incorporated. As the primary outcome was pain intensity, a more predictable pain model, such as one limited to impacted third molar surgeries, might have been, in hindsight, better suited for this study.

Conclusion

The study showed that using the mHealth platform did not have a significant impact on acute postoperative pain experience. However, patients and providers indicated increased improvements in patient-provider communication, patient-provider relationship, postoperative complication management, and the ability to manage pain medication prescribing.

Acknowledgments

This work was supported by the Agency for Healthcare Research and Quality grant 1U18HS026135.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive analysis for the Unified Theory of Acceptance and Use of Technology questionnaire.

[PDF File, 105 KB - mhealth_v11i1e49677_app1.pdf]

Multimedia Appendix 2
Thematic analysis: barriers and facilitators.
[PDF File, 120 KB - mhealth_v11i1e49677_app2.pdf]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1316 KB - mhealth_v11i1e49677_app3.pdf]

References

1. Simon LS. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. J Pain Palliat Care Pharmacother 2012 Jul 5;26(2):197-198. [doi: 10.3109/15360288.2012.678473]
2. Yuan H, Ahmed WL, Liu M, Tu S, Zhou F, Wang S. Contribution of pain to subsequent cognitive decline or dementia: a systematic review and meta-analysis of cohort studies. Int J Nurs Stud 2023 Feb;138:104409. [doi: 10.1016/j.ijnurstu.2022.104409] [Medline: 36527860]
3. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
4. Tuk JG, Lindeboom JA, van Wijk AJ. Effect of periapical surgery on oral health-related quality of life in the first postoperative week using the Dutch version of Oral Health Impact Profile. Oral Maxillofac Surg 2021 Dec;25(4):549-559. [doi: 10.1007/s10006-021-00954-y] [Medline: 33761022]
5. Gan TJ. Poorly controlled postoperative pain: prevalence, consequences, and prevention. J Pain Res 2017 Sep 25;10:2287-2298. [doi: 10.2147/JPR.S144066] [Medline: 29026331]
6. Edgman-Levitan S, Schoenbaum SC. Patient-centered care: achieving higher quality by designing care through the patient's eyes. Isr J Health Policy Res 2021 Mar 5;10(1):21. [doi: 10.1186/s13584-021-00459-9] [Medline: 33673875]
7. Yan CH, Lee TA, Sharp LK, Hubbard CC, Evans CT, Calip GS, et al. Trends in opioid prescribing by general dentists and dental specialists in the U.S., 2012-2019. Am J Prev Med 2022 Jul;63(1):3-12. [doi: 10.1016/j.amepre.2022.01.009] [Medline: 35232618]
8. Oral Analgesics for acute dental pain. American Dental Association. 2022. URL: www.ada.org/resources/research/science-and-research-institute/oral-health-topics/oral-analgesics-for-acute-dental-pain [accessed 2023-09-15]
9. Chua KP, Hu HM, Waljee JF, Brummett CM, Nalliah RP. Opioid prescribing patterns by dental procedure among US publicly and privately insured patients, 2013 through 2018. J Am Dent Assoc 2021 Apr;152(4):309-317. [doi: 10.1016/j.adaj.2021.01.001] [Medline: 33637299]
10. Lutfiyya MN, Gross AJ, Schvaneveldt N, Woo A, Lipsky MS. A scoping review exploring the opioid prescribing practices of US dental professionals. J Am Dent Assoc 2018 Dec;149(12):1011-1023. [doi: 10.1016/j.adaj.2018.07.017] [Medline: 30205900]
11. Moore PA, Dionne RA, Cooper SA, Hersh EV. Why do we prescribe Vicodin? J Am Dent Assoc 2016 Jul;147(7):530-533. [doi: 10.1016/j.adaj.2016.05.005] [Medline: 27350643]
12. NEJM Catalyst. What is patient-centered care? NEJM Catal 2017 Jan 1. [doi: 10.1056/CAT.17.0559]
13. Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. J Med Radiat Sci 2021 Jun;68(2):186-195. [doi: 10.1002/jmrs.421] [Medline: 32815314]
14. Carr DB, Jacox AK. Acute Pain Management: Operative or Medical Procedures and Trauma Clinical Practice Guideline: DIANE Publishing; 1997.
15. Cella D, Hahn E, Jensen S, Butt Z, Nowinski C, Rothrock N, et al. Patient-Reported Outcomes in Performance Measurement: RTI Press; 2015. [doi: 10.3768/rtipress.2015.bk.0014.1509]
16. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190. [doi: 10.2196/24190] [Medline: 34032583]
17. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940. [doi: 10.2196/10940] [Medline: 30026178]
18. Huang L, Li VW, Yang T, Liu J, Murphy J, Michalak EE, et al. Mobile health applications for depression in China: a systematic review. Cureus 2022 Jul 26;14(7):e27299. [doi: 10.7759/cureus.27299] [Medline: 35903483]
19. Lu DJ, Girgis M, David JM, Chung EM, Atkins KM, Kamrava M. Evaluation of mobile health applications to track patient-reported outcomes for oncology patients: a systematic review. Adv Radiat Oncol 2021 Jan-Feb;6(1):100576. [doi: 10.1016/j.adro.2020.09.016] [Medline: 33073061]

20. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231. [doi: 10.2196/11231] [Medline: 30348633]
21. Zheng C, Chen X, Weng L, Guo L, Xu H, Lin M, et al. Benefits of mobile apps for cancer pain management: systematic review. JMIR Mhealth Uhealth 2020 Jan 23;8(1):e17055. [doi: 10.2196/17055] [Medline: 32012088]
22. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
23. Kalenderian E, White J, Yansane AI, Urata J, Holmes D, Funkhouser K, et al. Study protocol: understanding pain after dental procedures, an observational study within the National Dental PBRN. BMC Oral Health 2022 Dec 9;22(1):581. [doi: 10.1186/s12903-022-02573-9] [Medline: 36494795]
24. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH roadmap cooperative group during its first two years. Med Care 2007 May;45(5 Suppl 1):S3-S11. [doi: 10.1097/01.mlr.0000258615.42478.55] [Medline: 17443116]
25. Gordon DB, Polomano RC, Pellino TA, Turk DC, McCracken LM, Sherwood G, et al. Revised American Pain Society Patient Outcome Questionnaire (APS-POQ-R) for quality improvement of pain management in hospitalized adults: preliminary psychometric evaluation. J Pain 2010 Nov;11(11):1172-1186. [doi: 10.1016/j.jpain.2010.02.012] [Medline: 20400379]
26. Smith J, Firth J. Qualitative data analysis: the framework approach. Nurse Res 2011;18(2):52-62. [doi: 10.7748/nr2011.01.18.2.52.c8284] [Medline: 21319484]
27. Seymour RA, Meechan JG, Blair GS. An investigation into post-operative pain after third molar surgery under local analgesia. Br J Oral Maxillofac Surg 1985 Dec;23(6):410-418. [doi: 10.1016/0266-4356(85)90025-7] [Medline: 2933061]
28. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021 Mar 22;9(3):357. [doi: 10.3390/healthcare9030357] [Medline: 33809828]
29. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1-11. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
30. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123. [doi: 10.2196/10123] [Medline: 30497999]
31. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]
32. Kanoute A, Carrouel F, Gare J, Dieng SN, Dieng A, Diop M, et al. Evaluation of oral hygiene-related mobile apps for children in sub-Saharan Africa. Int J Environ Res Public Health 2022 Oct 1;19(19):12565. [doi: 10.3390/ijerph191912565] [Medline: 36231862]
33. Carrouel F, Bourgeois D, Clément C, Tardivo D, Martinon P, Guiral S, et al. Oral-hygiene-related mobile apps in the French app stores: assessment of functionality and quality. Int J Environ Res Public Health 2022 Jun 14;19(12):7293. [doi: 10.3390/ijerph19127293] [Medline: 35742541]
34. Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, et al. The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth 2020 Jul 8;8(7):e16365. [doi: 10.2196/16365] [Medline: 32673235]
35. Stein CD, Xiao X, Levine S, Schleyer TKL, Hochheiser H, Thyvalikakath TP. A prototype mobile application for triaging dental emergencies. J Am Dent Assoc 2016 Oct;147(10):782-791. [doi: 10.1016/j.adaj.2016.03.021] [Medline: 27206728]
36. Chen R, Santo K, Wong G, Sohn W, Spallek H, Chow C, et al. Mobile apps for dental caries prevention: systematic search and quality evaluation. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e19958. [doi: 10.2196/19958] [Medline: 33439141]
37. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295. [doi: 10.2196/jmir.7740] [Medline: 28838887]
38. Peprah P, Abalo EM, Agyemang-Duah W, Budu HI, Appiah-Brempong E, Morgan AK, et al. Lessening barriers to healthcare in rural Ghana: providers and users' perspectives on the role of mHealth technology. A qualitative exploration. BMC Med Inform Decis Mak 2020 Feb 10;20(1):27. [doi: 10.1186/s12911-020-1040-4] [Medline: 32041608]
39. Kumar D, Arya M. mHealth is an innovative approach to address health literacy and improve patient-physician communication - an HIV testing exemplar. J Mob Technol Med 2015 Jan 1;4(1):25-30. [doi: 10.7309/jmtm.4.1.6] [Medline: 25729441]
40. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 4;9(3):e21795. [doi: 10.2196/21795] [Medline: 33661123]
41. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. Science 1987 Apr 10;236(4798):157-161. [doi: 10.1126/science.3563494] [Medline: 3563494]
42. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. Pain 2003 Jul;104(1-2):187-194. [doi: 10.1016/s0304-3959(03)00003-4] [Medline: 12855328]
43. Ross M. Relation of implicit theories to the construction of personal histories. Psychol Rev 1989 Apr;96(2):341. [doi: 10.1037/0033-295X.96.2.341]

44. Pentland A, Lazer D, Brewer D, Heibeck T. Improving public health and medicine by use of reality mining. Stud Health Technol Inform 2009;149:93-102. [Medline: 19745474]
45. Boukhechba M, Cai L, Chow PI, Fua K, Gerber MS, Teachman BA, et al. Contextual analysis to understand compliance with smartphone-based ecological momentary assessment. Presented at: The 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY p. 232-238. [doi: 10.1145/3240925.3240967]

Abbreviations

APS-POQ-R: Revised American Pain Society Patient Outcome Questionnaire
EHR: electronic health record
EMA: ecological momentary assessment
mHealth: mobile health
PRO: patient-reported outcome
PROMIS: Patient-Reported Outcomes Measurement Information System
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by L Buis; submitted 05.06.23; peer-reviewed by J Fredericks-Younger, S Li; revised version received 17.07.23; accepted 21.07.23; published 20.10.23.

Please cite as:
Tokede B, Yansane A, Ibarra-Noriega A, Mullins J, Simmons K, Skourtes N, Mehta U, Tungare S, Holmes D, White J, Walji M, Kalenderian E
Evaluating the Impact of an mHealth Platform for Managing Acute Postoperative Dental Pain: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e49677
URL: https://mhealth.jmir.org/2023/1/e49677
doi:10.2196/49677

The Role of Pain Duration and Pain Intensity on the Effectiveness of App-Delivered Self-Management for Low Back Pain (selfBACK): Secondary Analysis of a Randomized Controlled Trial

Anne Lovise Nordstoga[1,2,*], PhD; Lene Aasdahl[3,4,*], PhD; Louise Fleng Sandal[5], PhD; Tina Dalager[5], PhD; Atle Kongsvold[4], MSc; Paul Jarle Mork[4], PhD; Tom Ivar Lund Nilsen[4,6], PhD

1

2

3

4

5

6

*these authors contributed equally

Corresponding Author:
Anne Lovise Nordstoga, PhD

Abstract

Background: Clinical guidelines for nonspecific low back pain (LBP) recommend self-management tailored to individual needs and capabilities as a first-line treatment. Mobile health solutions are a promising method for delivering tailored self-management interventions to patients with nonspecific LBP. However, it is not clear if the effectiveness of such self-management interventions depends on patients' initial pain characteristics. High pain intensity and long-term symptoms of LBP have been associated with an unfavorable prognosis, and current best evidence indicates that long-term LBP (lasting more than 3 months) requires a more extensive treatment approach compared to more acute LBP. The artificial intelligence–based selfBACK app supports tailored and evidence-based self-management of nonspecific LBP. In a recent randomized controlled trial, we showed that individuals who received the selfBACK app in addition to usual care had lower LBP-related disability at the 3-month follow-up compared to those who received usual care only. This effect was sustained at 6 and 9 months.

Objective: This study aims to explore if the baseline duration and intensity of LBP influence the effectiveness of the selfBACK intervention in a secondary analysis of the selfBACK randomized controlled trial.

Methods: In the selfBACK trial, 461 adults (18 years or older) who sought care for nonspecific LBP in primary care or at an outpatient spine clinic were randomized to receive the selfBACK intervention adjunct to usual care (n=232) or usual care alone (n=229). In this secondary analysis, the participants were stratified according to the duration of the current LBP episode at baseline (≤12 weeks vs >12 weeks) or baseline LBP intensity (≤5 points vs >5 points) measured by a 0-10 numeric rating scale. The outcomes were LBP-related disability measured by the Roland-Morris Disability Questionnaire (0- to 24-point scale), average LBP intensity, pain self-efficacy, and global perceived effect. To assess whether the duration and intensity of LBP influenced the effect of selfBACK, we estimated the difference in treatment effect between the strata at the 3- and 9-month follow-ups with a 95% CI.

Results: Overall, there was no difference in effect for patients with different durations or intensities of LBP at either the 3- or 9-month follow-ups. However, there was suggestive evidence that the effect of the selfBACK intervention on LBP-related disability at the 3-month follow-up was largely confined to people with the highest versus the lowest LBP intensity (mean difference between the intervention and control group −1.8, 95% CI −3.0 to −0.7 vs 0.2, 95% CI −1.1 to 0.7), but this was not sustained at the 9-month follow-up.

Conclusions: The results suggest that the intensity and duration of LBP have negligible influence on the effectiveness of the selfBACK intervention on LBP-related disability, average LBP intensity, pain self-efficacy, and global perceived effect.

Trial Registration: ClinicalTrials.gov NTC03798288; https://clinicaltrials.gov/study/NCT03798288

International Registered Report Identifier (IRRID): RR2-10.2196/14720

(JMIR Mhealth Uhealth 2023;11:e40422) doi:10.2196/40422

KEYWORDS

low back pain; self-management; mHealth; mobile app; health app; mobile health; back; pain; randomized controlled trial; RCT; secondary analysis; secondary analyses; spine; effectiveness; digital intervention

Introduction

Low back pain (LBP) is a common reason for primary care visits globally [1,2], and more than 90% of cases among adults are defined as nonspecific LBP [3]. Recurrent episodes of LBP occur within a year in 30%-60% of patients [4], and 10%-30% of these develop persistent LBP [5]. Clinical guidelines recommend self-management tailored to individual needs and capabilities as a first-line treatment of nonspecific LBP [6,7]. Supporting self-management through digital interventions, such as mobile apps, has been suggested as a viable approach to improve and reinforce self-management interventions [8]. In a recent randomized controlled trial (RCT) among adults seeking care for LBP, we showed that those who received artificial intelligence (AI)–based individually tailored self-management support delivered via a mobile app (selfBACK) in addition to usual care had less LBP-related disability at the 3-, 6-, and 9-month follow-ups, compared with those who received usual care alone [9]. However, it is still unclear whether digital interventions to support self-management are more effective for specific subgroups [10].

Previous studies have shown that high LBP intensity and long-term symptoms are associated with less favorable prognosis and poorer outcomes among patients in a primary care setting [11-14]. Moreover, the current best evidence indicates that LBP lasting >3 months requires a broader and more extensive treatment approach than acute or subacute LBP [15]. It is therefore conceivable that the baseline duration and intensity of LBP influence the effectiveness of digital and individually tailored self-management interventions for nonspecific LBP, and such knowledge can assist clinicians in selecting patients best suited for this type of self-management support. In this secondary analysis of the selfBACK RCT [9], we explore if baseline LBP intensity and duration of the current LBP episode influence the effectiveness of the selfBACK intervention.

Methods

Study Design

This secondary analysis is based on data from the selfBACK multicenter RCT with two parallel arms (ClinicalTrials.gov NTC03798288). The trial investigated the effectiveness of evidence-based and individually tailored self-management support delivered via the selfBACK mobile app as an adjunct to usual care for adults with nonspecific LBP [9,16]. The methods and primary results of the RCT have been published in previous studies [9,16] and are only briefly described here.

Participants and Randomization

We recruited adults (18 years or older) with nonspecific LBP who had consulted a primary care clinician (ie, general practitioner, physiotherapist, or chiropractor) in the Trondheim Municipality in Norway or in the region of Southern Denmark, or who had undergone a clinical examination at an outpatient spine clinic (Spine Centre of Southern Denmark). The inclusion criteria were to have experienced LBP within the preceding 8 weeks, a score of mild to moderate LBP-related disability rated as 6 points or higher on the Roland-Morris Disability Questionnaire (RMDQ), a smartphone, and access to email. The 6-point cutoff on RMDQ defines mild to moderate disability due to LBP and is considered to have the potential for a clinically meaningful improvement. The exclusion criteria were the inability to carry out the intervention (ie, mental or physical conditions that limited participation; inability to perform physical exercise; or problems with speaking, reading, or understanding Danish or Norwegian), fibromyalgia, previous spinal surgery, current pregnancy, or current participation in other studies related to LBP. Participants were recruited from March 8 to December 14, 2019. After giving informed consent, the participants completed a web-based questionnaire and were then randomized to the selfBACK intervention or usual care (control group) in a 1:1 ratio using a computer-generated sequence stratified by country (Norway or Denmark) and type of care provider (general practitioner, physiotherapist, chiropractor, or outpatient clinic).

Intervention

The intervention was delivered as an adjunct to usual care and has been described in detail in a previous study [9]. In brief, participants randomized to the intervention group were instructed to install the selfBACK app on their smartphone and to wear a step-detecting wristband (Mi Band 3, Xiaomi) connected to the app. The selfBACK app contains three main components of self-management: recommendations of physical activity (ie, number of steps), video instructions for strength and flexibility exercises, and daily educational content. Weekly self-management recommendations are provided for each component, and the recommendations are tailored to individual characteristics, symptoms, and progression by using the case-based reasoning methodology, a branch of knowledge-driven AI [17,18]. The app also includes tools such as goal setting, mindfulness audios, pain-relieving exercises, and sleep reminders, as well as general educational content related to LBP. The participants received encouraging push notifications triggered by their self-management behavior to motivate and reinforce the desired behavior. The design, architecture, and functions of the selfBACK system, as well as the development of the evidence-based content, have been described in detail in previous studies [16,19-21]. Participants randomized to usual care were instructed to follow the advice and treatment provided by their clinician or health care provider.

Outcomes and Follow-Up

The outcomes were LBP-related disability measured by the RMDQ (range 0-24, higher scores indicating higher LBP-related disability) [22], average LBP intensity in the preceding week measured on a numeric rating scale (NRS) by the statement “Please indicate your average back pain level during the last week on a scale from 0 (no pain) to 10 (worst pain imaginable)”

[23], pain self-efficacy measured by the Pain Self-Efficacy Questionnaire (range 0-60, higher scores indicating greater confidence) [24], and overall improvement measured by the Global Perceived Effect scale (range −5, “very much worse,” to 5, “very much better”) [25]. The outcomes were measured at baseline, 6 weeks, 3 months, 6 months, and 9 months. The main outcome in the RCT was LBP-related disability measured by the RMDQ at the 3-month follow-up [9].

Stratification Variables

Subgroups were defined according to baseline reporting of the duration of the current LBP episode (≤12 weeks vs >12 weeks) and the average LBP intensity in the preceding week (≤5 vs >5 NRS points). The duration of LBP was assessed by the question “What is the length of time you have had LBP during this episode?” with four response options “less than 1 week,” “1-4 weeks,” “5-12 weeks,” and “more than 12 weeks,” whereas the average LBP intensity was measured by the NRS (range 0, “no pain,” to 10, “worst pain imaginable”).

Statistical Analyses

An intention-to-treat analysis was used to estimate mean group differences with 95% CIs from constrained longitudinal data analysis using a linear mixed model [26,27]. In this model, the intervention and control groups have a common baseline mean, and all follow-up time points are included in the analysis. Dependency between repeated measures was accounted for by including a random intercept for each participant. To assess whether the duration and intensity of LBP modified the effect of the intervention, we first estimated the stratified treatment effect (ie, within each duration and intensity group) at 3 and 9 months, and then calculated the between-group difference in treatment effect with 95% CI and associated *P* values. All analyses were adjusted for the two variables used to stratify the randomization (ie, country and care provider) as well as the level of education (<10 years, 10-12 years, ≥12 years of schooling), gender (male, female), and age (years). Additionally, estimates stratified by LBP duration were adjusted for baseline LBP intensity in the preceding week (continuous, 0-10 points), whereas estimates stratified by pain intensity were adjusted for the duration of the current LBP episode measured at baseline (<1 week, 1-4 weeks, 5-12 weeks, >12 weeks) as well as the average baseline LBP intensity (continuous, 0-10 points; the latter adjustment accounted for variation in LBP intensity within the strata).

All analyses were performed using Stata, version 16.1 (StataCorp).

Ethics Approval

All participants provided written informed consent before participation in the study; this also covered the secondary analyses performed in this study. There was no financial compensation for participants, but all participants got a ticket in a raffle for an iPad. The selfBACK RCT was approved by the Regional Committee for Medical and Health Research Ethics in Central Norway (No. 2017/923-6) and the Danish Data Protection Agency (201-57-0008) and regional ethics committee in Denmark (S-20182000-24). All study data was processed and analyzed without any personal information that could identify the participants.

Results

The flow of participants through the trial is shown in Figure 1.

Figure 1. Flow of participants through the trial. *The reasons for exclusion were being younger than 18 years (n=2); being unable to speak, read, or understand the national language (n=2); having mental or physical conditions that limited participation (n=12); being unable to take part in exercise or physical activity (n=5); having a fibromyalgia diagnosis (n=11); participating currently in other lower back research (n=2); and having previous back surgery (n=30).

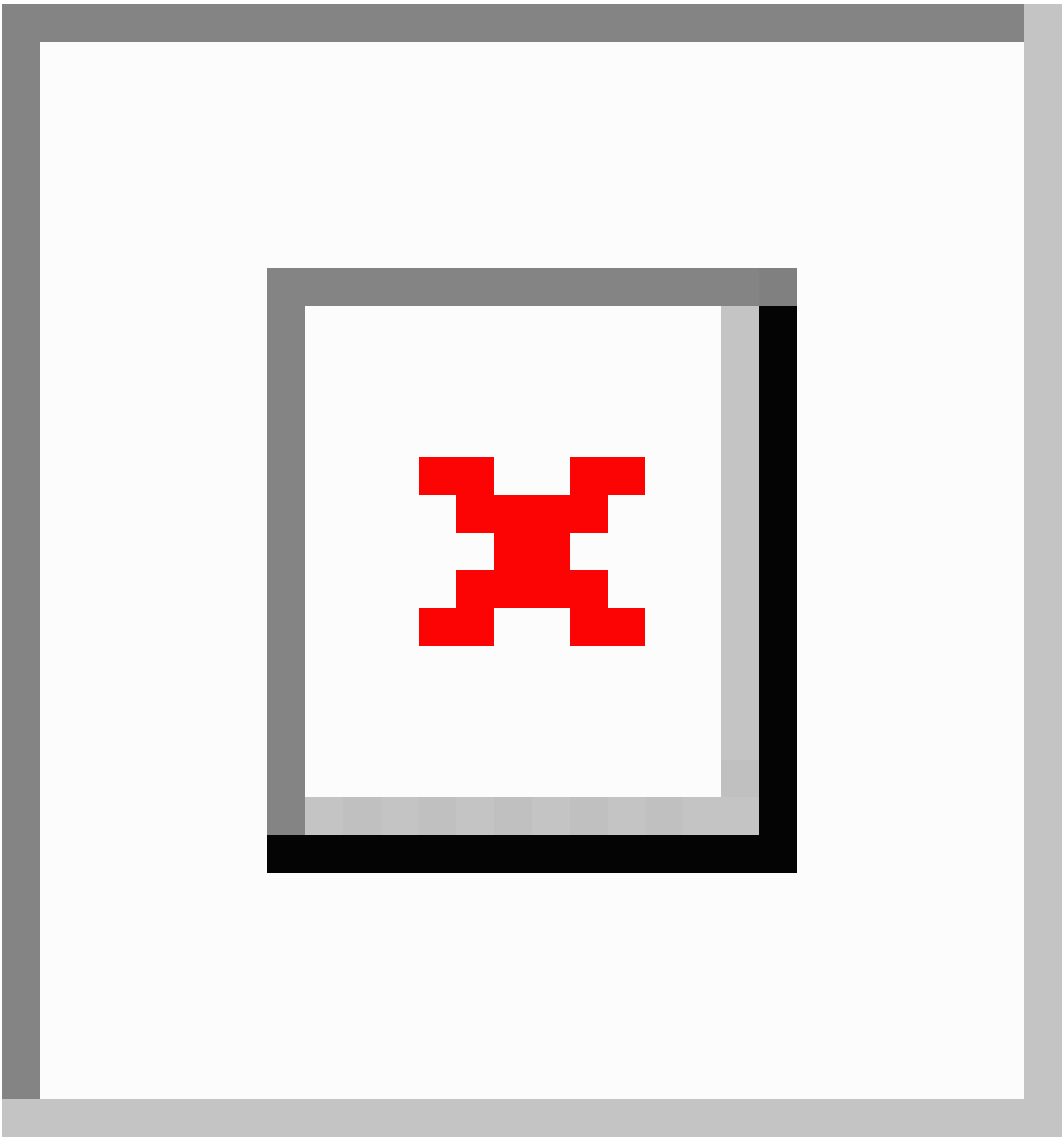


Table 1 shows the baseline characteristics of the study participants, stratified according to the duration of the current LBP episode and the average baseline LBP intensity in the preceding week. At baseline, 267 (57.9%) of the 461 participants reported the duration of their current LBP episode as >12 weeks, and 185 (40.1%) participants reported LBP intensity >5 on the NRS. Based on the means and proportions between the intervention and control groups, sociodemographic characteristics and the type of care provider for patient recruitment were largely similar, as well as within the LBP duration and intensity strata.

Table . Baseline characteristics of study participants stratified according to the duration of the current low back pain (LBP) episode and average LBP intensity in the preceding week (range 0-10).

Variable		Duration of current LBP episode				Average LBP intensity in the preceding week			
		≤12 weeks		>12 weeks		Low (≤5)		High (>5)	
		Control (n=93)	selfBACK (n=101)	Control (n=136)	selfBACK (n=131)	Control (n=131)	selfBACK (n=145)	Control (n=98)	selfBACK (n=87)
Age (years), mean (SD)		44.0 (13.2)	45.6 (15.0)	48.6 (15.0)	50.3 (14.8)	46.9 (13.9)	48.6 (14.4)	46.5 (15.2)	47.6 (16.0)
Female, n (%)		49 (52.7)	50 (49.5)	85 (62.5)	71 (54.2)	74 (56.5)	78 (53.8)	60 (61.2)	43 (49.4)
Male, n (%)		44 (47.3)	51 (50.5)	51 (37.5)	60 (45.8)	57 (43.5)	67 (46.2)	38 (38.8)	44 (50.6)
BMI (kg/m^2), mean (SD)		28.3 (4.8)	27.5 (5.4)	27.5 (5.7)	27.2 (4.2)	27.1 (4.4)	27.3 (4.3)	28.8 (6.3)	27.5 (5.4)
Education >12 years, n (%)		60 (64.5)	73 (72.3)	85 (62.5)	79 (60.3)	88 (67.2)	99 (68.3)	57 (58.2)	53 (60.9)
Full-time employment, n (%)		67 (72.0)	63 (62.4)	75 (55.1)	75 (57.3)	84 (64.1)	86 (59.3)	58 (59.2)	52 (59.8)
Clinical setting, n (%)									
	Physiotherapist	26 (28.0)	20 (19.8)	41 (30.1)	48 (36.6)	40 (30.5)	44 (30.3)	27 (27.6)	24 (27.6)
	Chiropractor	40 (43.0)	46 (45.5)	39 (28.7)	35 (26.7)	50 (38.2)	50 (34.5)	29 (29.6)	31 (35.6)
	General practitioner	21 (22.6)	22 (21.8)	13 (9.6)	12 (9.2)	16 (12.2)	21 (14.5)	18 (18.4)	13 (14.9)
	Outpatient back clinic	6 (6.5)	13 (12.9)	43 (31.6)	36 (27.5)	25 (19.1)	30 (20.7)	24 (24.5)	19 (21.8)
LBP characteristics									
	Average LBP intensity in the preceding week, mean (SD)	4.8 (2.0)	4.8 (2.2)	5.0 (1.8)	4.8 (1.8)	3.6 (1.2)	3.6 (1.2)	6.7 (0.9)	6.9 (1.1)
	Worst LBP intensity in the preceding week, mean (SD)	6.5 (2.1)	6.6 (2.0)	6.6 (1.9)	6.6 (1.8)	5.4 (1.8)	5.8 (1.7)	8.1 (1.0)	8.0 (1.2)
	Daily use of pain medication, n (%)	27 (29.0)	34 (33.7)	47 (34.6)	49 (37.4)	56 (42.7)	61 (42.1)	18 (18.4)	22 (25.3)
LBP duration current episode (weeks), n (%)									
	<1	9 (9.7)	9 (8.9)	—[a]	—	6 (4.6)	4 (2.8)	3 (3.1)	5 (5.7)
	1-4	46 (49.5)	49 (48.5)	—	—	27 (20.6)	34 (23.4)	19 (19.4)	15 (17.2)
	5-12	38 (40.9)	43 (42.6)	—	—	20 (15.3)	22 (15.2)	18 (18.4)	21 (24.1)
	>12	—	—	136 (100.0)	131 (100.0)	78 (59.5)	85 (58.6)	58 (59.2)	46 (52.9)

[a]Not available.

The results stratified by LBP duration (≤12 weeks vs >12 weeks) at baseline are shown in Figure 2 and Table S1 in Multimedia Appendix 1. At 3 months, the effectiveness of the intervention was largely similar in both strata of pain duration, indicating no or minor effect modification. For LBP-related disability, we observed a point difference of –0.9 and –0.6 between the selfBACK and control groups in those with ≤12 weeks and >12 weeks pain duration, respectively. The corresponding estimates were differences of −0.7 points versus −0.5 points in LBP intensity, 1.9 versus 2.9 points in pain self-efficacy, and no difference in global perceived effect (0.7 points vs 0.7 points).

Overall, the baseline LBP intensity was not found to influence the effectiveness of the selfBACK intervention (Figure 3 and Table S2 in Multimedia Appendix 1). However, at 3 months, those receiving selfBACK had LBP-related disability 1.8 (95% CI −3.0 to −0.7) points lower than the controls if LBP intensity was high (>5 NRS points) and 0.2 (95% CI −1.1 to 0.7) points lower if LBP intensity was low (≤5 NRS points). This corresponds to a mean difference in effect between the LBP intensity strata of −1.6 points (95% CI −3.1 to −0.2; $P_{interaction}$=.03). These differences were not sustained at 9 months.

Figure 2. Mean scores with 95% CIs for (A) LBP-related disability (RMDQ), (B) LBP intensity, (C) pain self-efficacy, and (D) global perceived effect at all time points for the selfBACK and control groups, stratified according to LBP duration at baseline (≤12 weeks vs >12 weeks). The right panel (E-H) shows the mean difference at the 3-month follow-up between the selfBACK and control groups within each LBP duration stratum, and *the corresponding difference due to interaction (ie, between strata difference). BL: baseline; LBP: low back pain; RMDQ: Roland-Morris Disability Questionnaire.

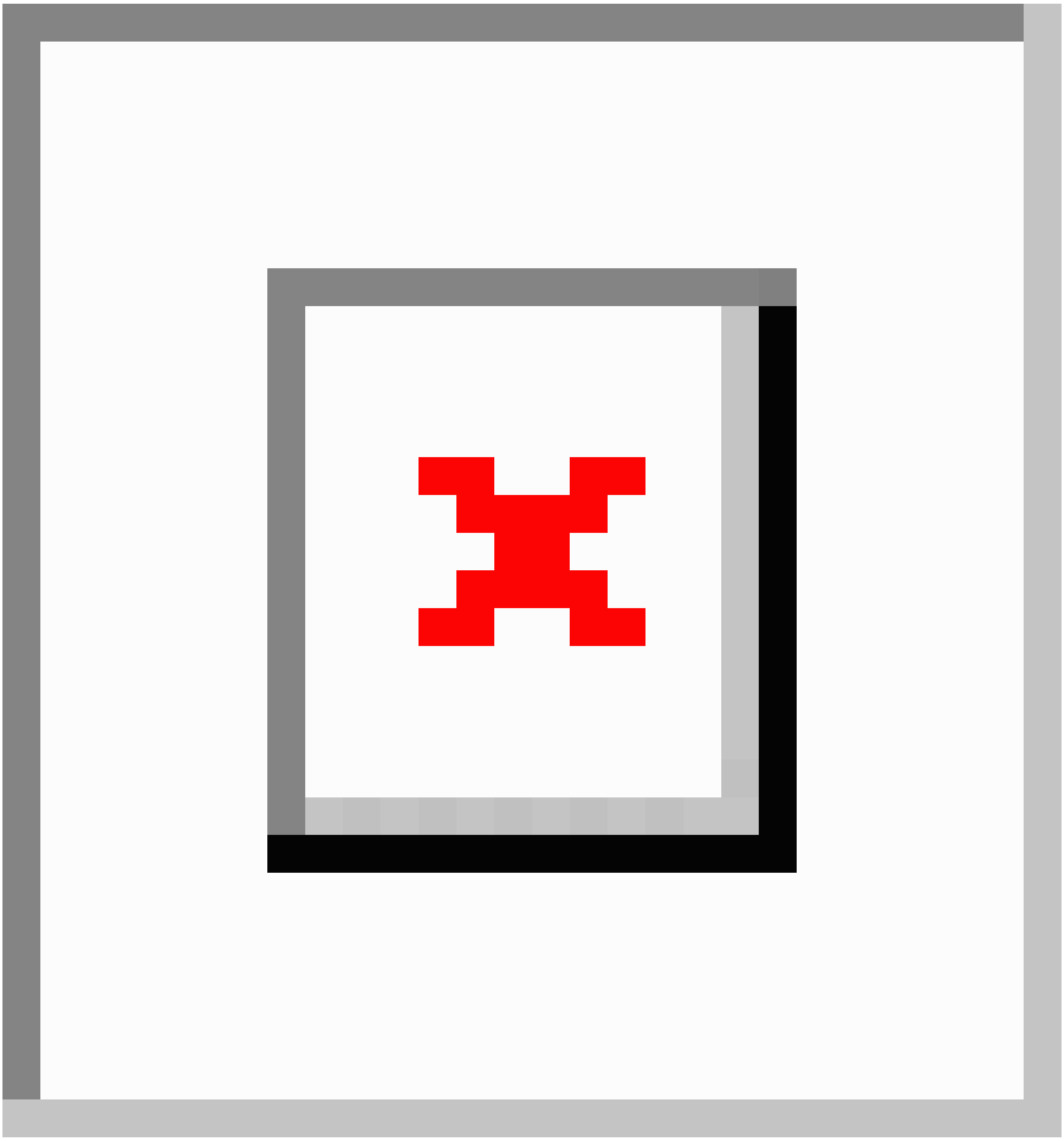
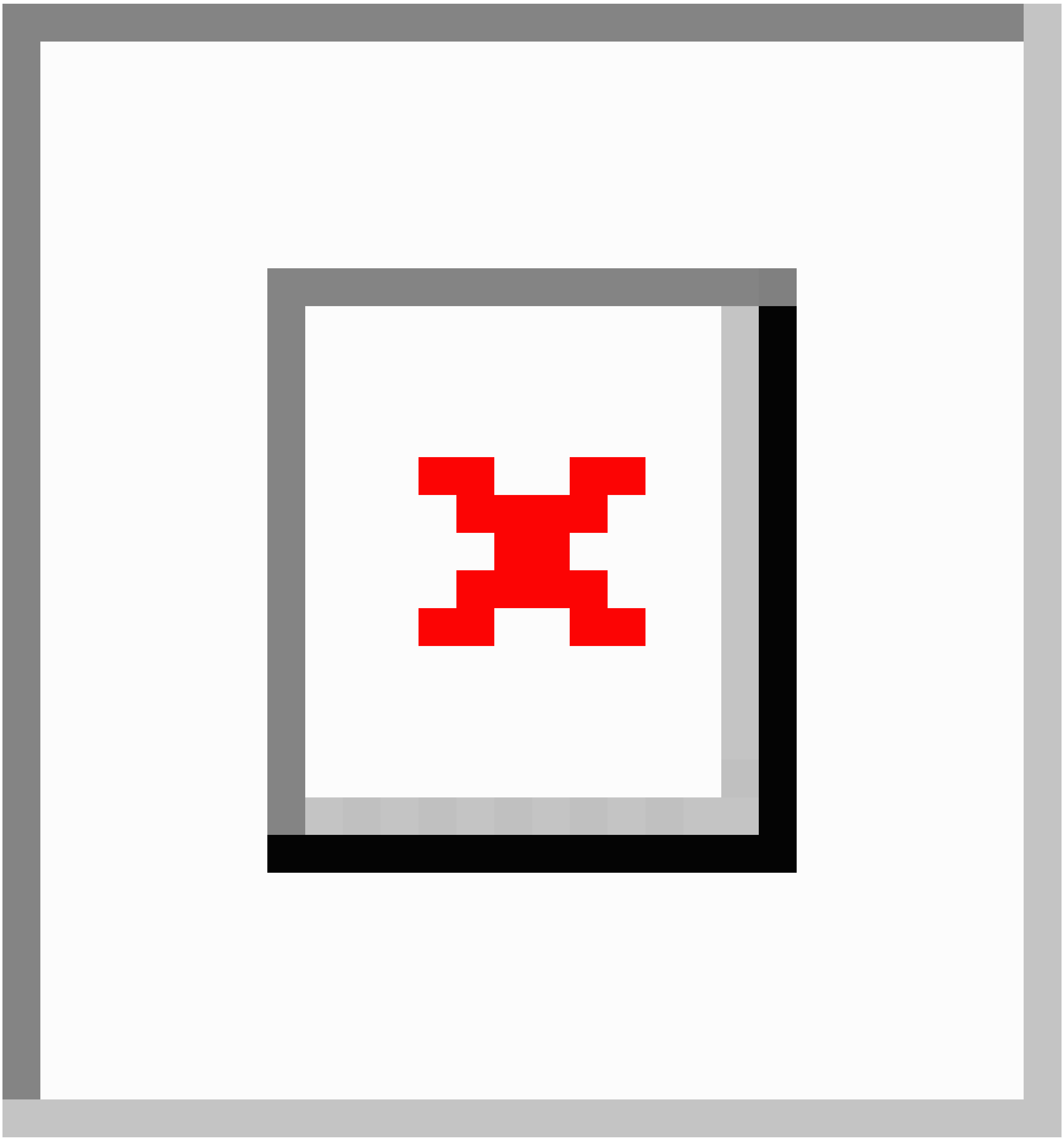


Figure 3. Mean scores with 95% CIs for (A) LBP-related disability (RMDQ), (B) LBP intensity, (C) pain self-efficacy, and (D) global perceived effect at all time points for the selfBACK and control groups, stratified according to LBP intensity at baseline (≤5 points vs >5 points). The right panel (E-H) shows the mean difference at the 3-month follow-up between the selfBACK and control groups within each LBP intensity stratum, and *the corresponding difference due to interaction (ie, between strata difference). BL: baseline; LBP: low back pain; RMDQ: Roland-Morris Disability Questionnaire.



Discussion

The duration and intensity of LBP did not have any major influence on the effectiveness of individually tailored self-management support as delivered via the selfBACK app on LBP-related disability, LBP intensity, pain self-efficacy, and global perceived effect. For participants with high LBP intensity at baseline, there was a larger difference between the control and the selfBACK group than for those with lower pain intensity. However, the numerical differences were small and with CIs that included the null value. These findings suggest that tailored and evidence-based self-management support along with usual care can be a useful supplement to usual care regardless of the duration and intensity of LBP.

Clinical practice guidelines recommend tailored self-management support as a first-line treatment for LBP regardless of symptom severity and duration [7,28]. To our knowledge, selfBACK represents the first attempt to use an AI-based app to deliver evidence-based and individually tailored self-management support. In line with previous observational studies, patients with higher pain intensity or long-lasting LBP at baseline were doing worse at follow-up for all outcomes

[11-13]. Accordingly, patients with long-lasting pain have been considered a more challenging subgroup, and current best evidence suggests that more extensive care is required to improve LBP and associated symptoms within this patient group [15]. Interestingly, the lack of difference between the selfBACK and control groups in the LBP duration strata in this study is, therefore, somewhat unexpected. However, the RCT was not specifically designed for subgroup analyses, and our explorative analyses are somewhat underpowered.

Furthermore, the lack of heterogenic effects between the LBP duration and intensity strata may in part be explained by the individually tailored intervention. The selfBACK system provides evidence-based and personalized recommendations for self-management by using information on personal characteristics, individual goals, and symptom progression assessed by weekly follow-up questions in the app [9]. Accordingly, the content of the intervention is adapted to patient characteristics, including their baseline LBP duration and intensity, and may thus reduce the difference between the groups throughout the follow-up period. Although further research is needed to assess the effect of individual tailoring, the lack of influence of the LBP duration and intensity on the effect may indicate that the AI-based tailoring implemented in the selfBACK intervention was successful.

We also observed a small increased effect on LBP-related disability for those with high LBP intensity at baseline; however, the clinical significance of this finding is questionable (1.6 points difference on a 0-24 scale). It is possible that the effect would have been larger among those with high LBP intensity and long-term LBP if the self-management intervention was combined with other components suggested by the guidelines, such as cognitive behavioral therapy [15]. Unfortunately, we do not have access to data on the usual care provided to the control group or intervention group. Thus, we cannot rule out the possibility that patients with long-lasting LBP or high LBP intensity received more extensive care compared to those with short-term LBP or low LBP intensity. However, the large amount of background information collected in this study was not available for the clinicians, and we do not expect the clinicians to have the time or resources to collect this information in a standard clinical setting.

Assessing if symptom severity such as the duration and intensity of LBP influences the effect of a given treatment is clinically relevant and may have important implications for implementing an intervention in clinical practice. Our results did not find an increased effect of the selfBACK intervention for a specific subgroup in this study, indicating that primary care clinicians should not restrict individually tailored and evidence-based self-management support to certain subgroups of patients based on the duration of the current LBP episode or level of LBP intensity.

The randomized design and the repeated measures of the outcomes are important strengths of this study. However, there are some limitations that should be considered when interpreting the results. First, the participants were stratified after randomization, which may potentially create an imbalance in baseline characteristics between the subgroups. However, by assessing the means and proportions of the sociodemographic variables and LBP characteristics, they remained largely equally distributed between the selfBACK and control groups for both the LBP duration and LBP intensity strata. Second, the number of comparisons performed in this study was large due to the two stratification variables and four outcome measures, which increases the risk of chance findings. Third, the RCT was not specifically powered for subgroup analyses, resulting in somewhat imprecise estimates with wide CIs. Fourth, the choice of the cutoff value for the stratification of LBP intensity was based on a pragmatic approach (ie, an approximately equal number of participants in the intervention and control group within each stratum). Using other cutoff values may yield different results. Finally, missing data at the 6- and 9-month follow-ups (~24% missing at both time points) is a possible source of bias. Although the analyses were conducted according to an intention-to-treat principle using a linear mixed model, the underlying and nonverifiable assumption is that data are missing at random.

The baseline duration and intensity of LBP had no or minor influence on the effectiveness of tailored and evidence-based self-management support as delivered via the selfBACK app. Our results, therefore, suggest that primary care clinicians should not restrict the use of tailored and evidence-based self-management support based on the duration and intensity of LBP.

Acknowledgments

This project has received funding from the European Union Horizon 2020 research and innovation program under grant agreement 689043. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability

The data are kept for 5 years (identifiable data), and the data will thereafter be anonymized and stored for up to 30 years. Researchers who want to use the data, and of whom the proposal to use the data has been approved by the data steering group, can access the data for relevant research purposes. Contact PJM (paul.mork@ntnu.no) or Karen Søgaard (ksogaard@health.sdu.dk) for more information.

Authors' Contributions

TILN, PJM, ALN, and LA contributed to the study conception and design. Data analyses were performed by ALN and LA. The first draft of the manuscript was written by ALN and LA. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables on the effect of the selfBACK intervention stratified by pain duration and intensity.
[DOCX File, 33 KB - mhealth_v11i1e40422_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 365 KB - mhealth_v11i1e40422_app2.pdf]

References

1. St Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree ME, Pankratz JJ, et al. Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. Mayo Clin Proc 2013 Jan;88(1):56-67. [doi: 10.1016/j.mayocp.2012.08.020] [Medline: 23274019]
2. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord 2010 Jul 2;11(1):1-10. [doi: 10.1186/1471-2474-11-144] [Medline: 20598124]
3. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
4. Itz CJ, Geurts JW, van Kleef M, Nelemans P. Clinical course of non-specific low back pain: a systematic review of prospective cohort studies set in primary care. Eur J Pain 2013 Jan;17(1):5-15. [doi: 10.1002/j.1532-2149.2012.00170.x] [Medline: 22641374]
5. Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? JAMA 2010 Apr 7;303(13):1295-1302. [doi: 10.1001/jama.2010.344] [Medline: 20371789]
6. Corp N, Mansell G, Stynes S, Wynne-Jones G, Morsø L, Hill JC, et al. Evidence-based treatment recommendations for neck and low back pain across Europe: a systematic review of guidelines. Eur J Pain 2021 Feb;25(2):275-295. [doi: 10.1002/ejp.1679] [Medline: 33064878]
7. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020 Jan;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
8. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287. [doi: 10.2196/jmir.5692] [Medline: 27806926]
9. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021 Oct 1;181(10):1288-1296. [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]
10. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179. [doi: 10.2196/jmir.7290] [Medline: 28550009]
11. Dunn KM, Jordan KP, Croft PR. Contributions of prognostic factors for poor outcome in primary care low back pain patients. Eur J Pain 2011 Mar;15(3):313-319. [doi: 10.1016/j.ejpain.2010.07.008] [Medline: 20728385]
12. Cruz EB, Canhão H, Fernandes R, Caeiro C, Branco JC, Rodrigues AM, et al. Prognostic indicators for poor outcomes in low back pain patients consulted in primary care. PLoS One 2020 Mar 27;15(3):e0229265. [doi: 10.1371/journal.pone.0229265] [Medline: 32218561]
13. Nieminen LK, Pyysalo LM, Kankaanpää MJ. Prognostic factors for pain chronicity in low back pain: a systematic review. Pain Rep 2021 Apr 1;6(1):e919. [doi: 10.1097/PR9.0000000000000919] [Medline: 33981936]
14. Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? Best Pract Res Clin Rheumatol 2010 Apr;24(2):167-179. [doi: 10.1016/j.berh.2009.12.005] [Medline: 20227639]
15. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 9;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]

16. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An app-delivered self-management program for people with low back pain: protocol for the selfBACK randomized controlled trial. JMIR Res Protoc 2019 Dec 3;8(12):e14720. [doi: 10.2196/14720] [Medline: 31793897]
17. Aamodt A, Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun 1994;7(1):39-59. [doi: 10.3233/AIC-1994-7104]
18. Bach K, Szczepanski T, Aamodt A, Gundersen OE, Mork PJ. Case representation and similarity assessment in the self BACK decision support system. In: Goel A, Díaz-Agudo M, Roth-Berghofer T, editors. Case-Based Reasoning Research and Development ICCBR 2016 Lecture Notes in Computer Science. Cham: Springer; 2016, Vol. Vol 9969. [doi: 10.1007/978-3-319-47096-2_3]
19. Mork PJ, Bach K, selfBACK Consortium. A decision support system to enhance self-management of low back pain: protocol for the selfBACK project. JMIR Res Protoc 2018 Jul 20;7(7):e167. [doi: 10.2196/resprot.9379] [Medline: 30030208]
20. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020 Sep 9;7(2):e18729. [doi: 10.2196/18729] [Medline: 32902393]
21. Svendsen MJ, Sandal LF, Kjær P, Nicholl BI, Cooper K, Mair F, et al. Using intervention mapping to develop a decision support system-based smartphone app (selfBACK) to support self-management of nonspecific low back pain: development and usability study. J Med Internet Res 2022 Jan 24;24(1):e26555. [doi: 10.2196/26555] [Medline: 35072645]
22. Roland M, Fairbank J. The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000 Dec 15;25(24):3115-3124. [doi: 10.1097/00007632-200012150-00006]
23. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov;63(S11):S240-S252. [doi: 10.1002/acr.20543] [Medline: 22588748]
24. Nicholas MK. The pain self-efficacy questionnaire: taking pain into account. Eur J Pain 2007 Feb;11(2):153-163. [doi: 10.1016/j.ejpain.2005.12.008] [Medline: 16446108]
25. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163-170. [doi: 10.1179/jmt.2009.17.3.163] [Medline: 20046623]
26. Liang KY, Zeger SL. Longitudinal data analysis of continuous and discrete responses for pre-post designs. Sankhyā Indian J Stat Ser B 2000 Apr;62(1):134-148.
27. Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? Stat Med 2009 Sep 10;28(20):2509-2530. [doi: 10.1002/sim.3639] [Medline: 19610129]
28. O'Sullivan K, O'Keeffe M, O'Sullivan P. NICE low back pain guidelines: opportunities and obstacles to change practice. Br J Sports Med 2017 Nov;51(22):1632-1633. [doi: 10.1136/bjsports-2017-097810] [Medline: 28663215]

Abbreviations

AI: artificial intelligence
LBP: low back pain
NRS: numeric rating scale
RCT: randomized controlled trial
RMDQ: Roland-Morris Disability Questionnaire

Edited by L Buis; submitted 21.06.22; peer-reviewed by C Sitges, D Dunsmuir; revised version received 20.12.22; accepted 09.06.23; published 31.08.23.

Please cite as:
Nordstoga AL, Aasdahl L, Sandal LF, Dalager T, Kongsvold A, Mork PJ, Nilsen TIL
The Role of Pain Duration and Pain Intensity on the Effectiveness of App-Delivered Self-Management for Low Back Pain (selfBACK): Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e40422
URL: https://mhealth.jmir.org/2023/1/e40422
doi:10.2196/40422

Blood Pressure Measurement Based on the Camera and Inertial Measurement Unit of a Smartphone: Instrument Validation Study

Yong-Hoon Yoon[1], MD, DPhil; Jongin Kim[2], DPhil; Kwang Jin Lee[2], DPhil; Dongrae Cho[2]; Jin Kyung Oh[1], MD, DPhil; Minsu Kim[1], MD, DPhil; Jae-Hyung Roh[1], MD, DPhil; Hyun Woong Park[1], MD, DPhil; Jae-Hwan Lee[1], MD, DPhil

1

2

Corresponding Author:
Jae-Hwan Lee, MD, DPhil

Abstract

Background: Even though several mobile apps that can measure blood pressure have been developed, the data about the accuracy of these apps are limited.

Objective: We assessed the accuracy of AlwaysBP (test) in blood pressure measurement compared with the standard, cuff-based, manual method of brachial blood pressure measurement (reference).

Methods: AlwaysBP is a smartphone software that estimates systolic blood pressure (SBP) and diastolic blood pressure (DBP) based on pulse transit time (PTT). PTT was calculated with a finger photoplethysmogram and seismocardiogram using, respectively, the camera and inertial measurement unit sensor of a commercially available smartphone. After calculating PTT, SBP and DBP were estimated via the Bramwell-Hill and Moens-Korteweg equations. A calibration process was carried out 3 times for each participant to determine the input parameters of the equations. This study was conducted from March to August 2021 at Chungnam National University Sejong Hospital with 87 participants aged between 19 and 70 years who met specific conditions. The primary analysis aimed to evaluate the accuracy of the test method compared with the reference method for the entire study population. The secondary analysis was performed to confirm the stability of the test method for up to 4 weeks in 15 participants. At enrollment, gender, arm circumference, and blood pressure distribution were considered according to current guidelines.

Results: Among the 87 study participants, 45 (52%) individuals were male, and the average age was 35.6 (SD 10.4) years. Hypertension was diagnosed in 14 (16%) participants before this study. The mean test and reference SBPs were 120.0 (SD 18.8) and 118.7 (SD 20.2) mm Hg, respectively (difference: mean 1.2, SD 7.1 mm Hg). The absolute differences between the test and reference SBPs were <5, <10, and <15 mm Hg in 57.5% (150/261), 84.3% (220/261), and 94.6% (247/261) of measurements. The mean test and reference DBPs were 80.1 (SD 12.6) and 81.1 (SD 14.4) mm Hg, respectively (difference: mean −1.0, SD 6.0 mm Hg). The absolute differences between the test and reference DBPs were <5, <10, and <15 mm Hg in 75.5% (197/261), 93.9% (245/261), and 97.3% (254/261) of measurements, respectively. The secondary analysis showed that after 4 weeks, the differences between SBP and DBP were 0.1 (SD 8.8) and −2.4 (SD 7.6) mm Hg, respectively.

Conclusions: AlwaysBP exhibited acceptable accuracy in SBP and DBP measurement compared with the standard measurement method, according to the Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization protocol criteria. However, further validation studies with a specific validation protocol designed for cuffless blood pressure measuring devices are required to assess clinical accuracy. This technology can be easily applied in everyday life and may improve the general population's awareness of hypertension, thus helping to control it.

(JMIR Mhealth Uhealth 2023;11:e44147) doi:10.2196/44147

KEYWORDS

blood pressure; camera; cuffless; inertial measurement unit; smartphone software; cardiovascular; hypertension; smartphone; measure; accuracy; accurate; software; mHealth; mobile health

Introduction

Hypertension is an established major risk factor for various cardiovascular diseases and mortality [1]. Significant efforts have been made to understand the epidemiology, pathophysiology, and associated risks of hypertension to reduce premature cardiovascular morbidity and mortality [2,3]. Thus, the proportion of patients with well-controlled hypertension has increased over the past few decades [4]. High blood pressure can be easily detected at home or in public health centers and is often effectively treated with medications. However, there is a high rate of unawareness of hypertension for a large portion of the total population with hypertension, and the blood pressure control rate of patients with hypertension, especially in low-income countries, remains unsatisfactory. For these people, the early detection and proper management of hypertension are important to prevent future cardiovascular disease, which depends on the accuracy and accessibility of blood pressure measurement methods.

To improve the accessibility of blood pressure measurement, many new devices are being developed for this purpose that are comparable to the conventional, cuff-based, manual method or oscillometric blood pressure measurement devices. Cuffless devices include wristbands, rings, patch systems, and smartphone software; these devices have exhibited acceptable accuracy when compared with conventional blood pressure measurements [5-9]. The principle of cuffless blood pressure measurement is mainly based on the analysis of the pulse transit time (PTT) or photoplethysmography waveform to estimate blood pressure via a linear regression model or machine learning techniques. These methods are easier to use in daily life, and there is no cuff-induced discomfort during their use. However, most new technologies require photoplethysmograms or other sensors to detect the signal waveforms, which may limit their widespread use because of the cost associated with the additional equipment that is required.

The number of smartphone users has increased worldwide, and many people in high-income and low-income countries use them [10]. Most smartphones are equipped with a camera, image sensor, and inertial measurement unit (IMU) sensor; thus, attempts have been made to measure biosignals by using these sensors [11]. If heart movements can be detected by using the IMU sensor of a smartphone and by placing it on an individual's chest, and if the pulse of the fingertip can be acquired by using the camera of a mobile phone, then the smartphone may be used to estimate the PTT. Furthermore, if blood pressure can be estimated through the PTT obtained by using a smartphone, then blood pressure can be monitored with a smartphone without the use of additional equipment. Therefore, this study assessed a smartphone-based software that estimates blood pressure by using the PTT obtained through the smartphone.

Methods

Study Design and Participants

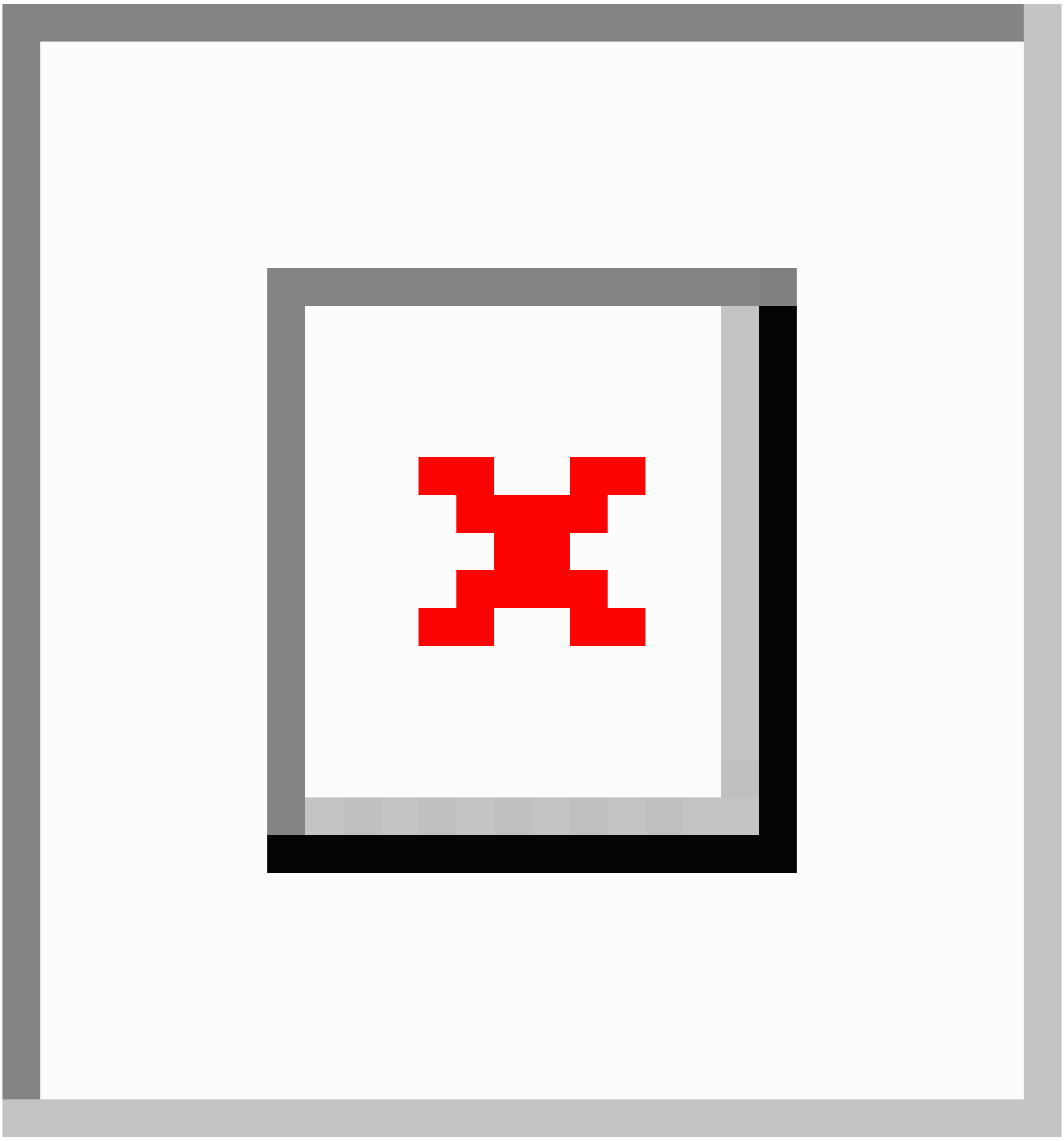
This study was conducted at Chungnam National University Sejong Hospital from March 2021 to August 2021. The enrollment targeted adults aged between 19 and 70 years. Those who met the following conditions were excluded: (1) individuals who were unable to use the test or reference blood pressure measurement device for up to 30 minutes in a sitting position (2) and individuals who had a history of cardiac arrhythmias or peripheral vascular disease. We determined the number of participants according to a validation guideline, which states that at least 85 patients are required for an Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) validation study [12]. At enrollment, gender, arm circumference, and blood pressure distribution were also considered according to the AAMI/ESH/ISO guideline [12,13].

The test device, AlwaysBP (Deepmedi Inc), is a smartphone software that estimates a person's systolic blood pressure (SBP) and diastolic blood pressure (DBP) based on the PTT. The blood pressure estimation model was designed to include the following two steps: (1) calculation of the PTT by using seismocardiography and photoplethysmography via the test device and (2) input of the acquired PTT into the blood pressure estimating equations (Figure 1). PTT was calculated as the time interval between the aortic valve opening on the seismocardiogram, as detected by the IMU sensor, and the onset of the photoplethysmography waveform on the index finger, as detected by the camera of the smartphone. These seismocardiography and photoplethysmography signals were acquired for 20 seconds in each measurement. An infinite impulse response band-pass filter was used for seismocardiography (5-45 Hz) and photoplethysmography (0.8-8 Hz) to denoise the raw signal. The filtered seismocardiography signal was then ensemble-averaged after segmenting the interval between the onset of each wave of the photoplethysmography signal and 450 milliseconds after the onset. This ensemble-averaged signal was input into a deep neural network model to determine the aortic valve opening time accurately. The manufacturer pretrained this deep neural network model, and the same model was used for all participants. After PTT calculation, SBP and DBP were estimated with the equations introduced by Bramwell-Hill and Moens-Korteweg, as follows [14]:

(1)

(2)

Figure 1. Mechanism and application of the test device. AO: aortic opening; DNN: deep neural network; IMU: inertial measurement unit; PPG: photoplethysmography; PTT: pulse transit time.



To determine the input parameters of the equations, a calibration process was carried out 3 times for each participant [14]. The PTT was measured by using the test device during calibration, and SBP and DBP were also measured simultaneously by using the cuff-based reference device. Pulse pressure (PP) and mean blood pressure (MBP) were calculated based on the SBP and DBP values. Based on the first and second calibration processes, the average PTT value obtained by the test device was entered as PTT_0, and the average PP and MBP measured by cuff-based blood pressure measurement device were entered as PP_0 and MBP_0, respectively. The PTT value obtained from the third calibration was input as the PTT of the formulas, and SBP and DBP values were input as the SBP and DBP, respectively; moreover, the γ value was updated according to the results of the calculated equations (before the correction, γ was determined to be 0.031 for individuals younger than 40 years and 0.09 for those older than 40 years) [15]. After the correction was completed, when measuring blood pressure for the primary analysis, the averages of the MBP, PP, and PTT values from the three calibration processes were input as MBP_0, PP_0, and PTT_0, respectively, and the γ obtained from the third calibration was used in the equations. This equation model was implemented in the AlwaysBP app and used to measure blood pressure at the initial visit and at follow-up. The test blood pressure measurement software was installed and used in a Galaxy S10 smartphone (Samsung Electronics) throughout this study. The test software has yet to be released on the app market for commercial use.

Data Acquisition and Blood Pressure Measurements

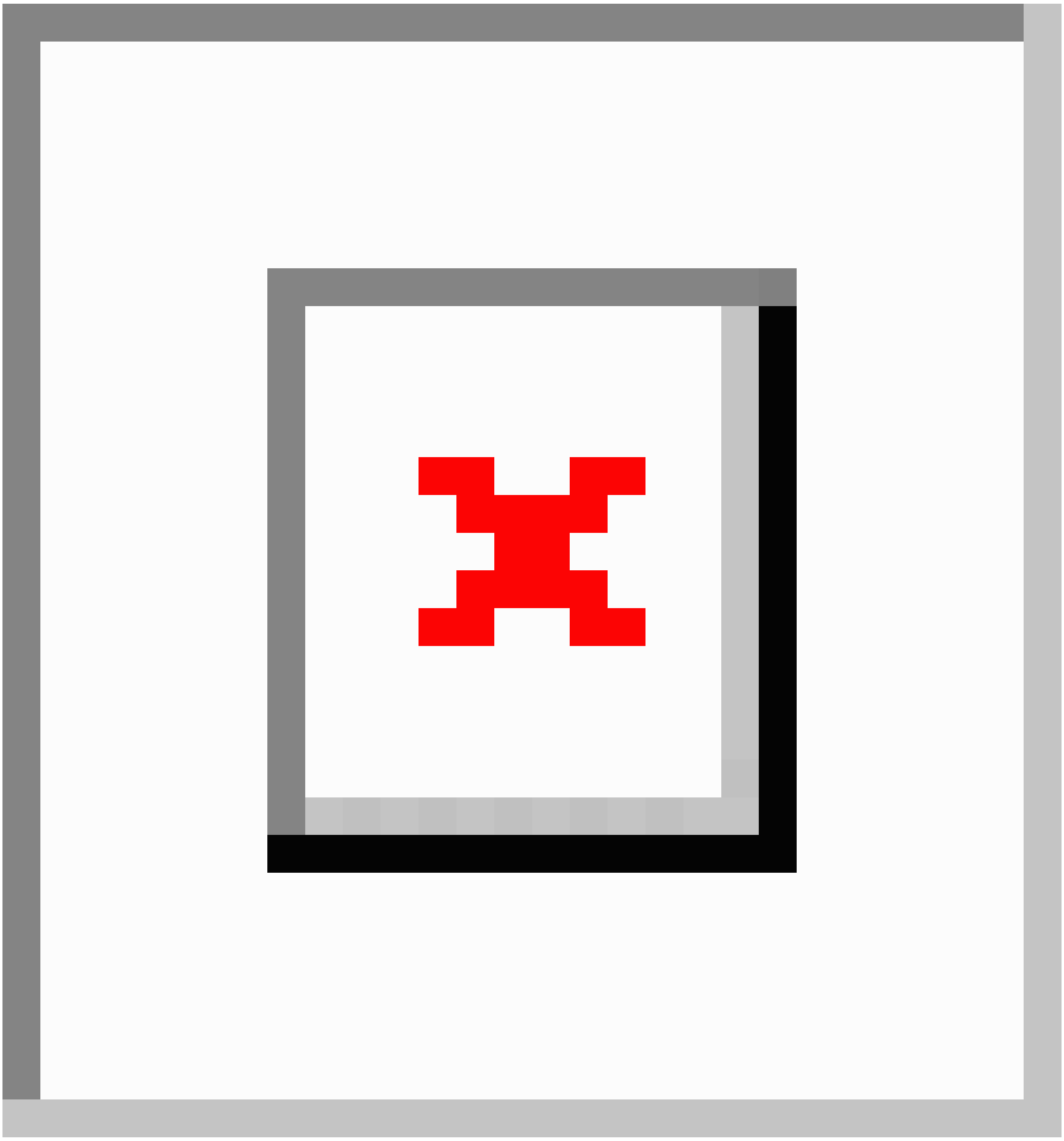
The basic demographics, including past medical history, comorbidities, and drug use, of the participants were assessed through an interview with the physician and recorded on a dedicated electronic case reporting form. The participants sat in a comfortable position for at least 5 minutes before measuring blood pressure. During the blood pressure measurement, the participant's back and arms were supported, their left arm was placed at heart level, their legs were not crossed, and their feet were placed flat on the floor. Dialogue and other interferences were prohibited throughout the measurement. Calibration of the test device was performed just before the measurement of the blood pressure.

The test blood pressure was defined as the blood pressure value measured by the test device (AlwaysBP), and the reference blood pressure was defined as the blood pressure value measured by the reference blood pressure device (InBody). The reference SBP and DBP were measured, using the participant's left arm, by 2 trained nurses who were blind to each other's readings. A double-headed stethoscope (Y-tube) was used for the simultaneous auscultation of the brachial pulse, and the blood pressures at the first and fifth Korotkoff sound were recorded as SBP and DBP, respectively. The average blood pressure values measured by the two observers were considered the reference blood pressures for the analysis. The test blood pressure was measured simultaneously with the reference blood pressure. The participant's right hand held the test device, and the index finger was positioned to cover the camera lens on the back of the smartphone to obtain the finger photoplethysmography signal. Concomitantly, the seismocardiography signal was obtained by placing the front of the smartphone on the participant's anterior chest wall. This position was maintained for 20 seconds, and if the participant's position was inadequate for proceeding with the study or blood pressure measurement, the research personnel advised the participant to return to the correct position. If an error occurred because of the poor signal quality of the test blood pressure device resulting from the insufficient color change of the finger to red (a result of the LED light being below a predetermined threshold), blood pressure measurement was attempted again until the error did not occur.

Study Protocol

This study consisted of 2 parts (Figure 2). The primary analysis aimed to evaluate the accuracy of the test device compared with the reference device for the entire study population (N=87). In turn, the secondary analysis was performed to confirm the stability of the test device for up to 4 weeks after the initial calibration and blood pressure measurement.

Figure 2. Study flowchart. BP: blood pressure.



For the primary analysis, blood pressure was measured 3 times for each participant after individualized calibration. Calibration was performed by entering the measured reference blood pressure value into the test device after applying the test and the reference devices at the same time. This process was performed 3 times consecutively, without a break. Afterward, 5 minutes of rest were allowed after completion of the calibration. Subsequently, blood pressure was measured 3 times per participant, with an interval of 5 minutes for resting. As a result, a total of 261 pairs of test-reference blood pressure sets from 87 participants were obtained for SBP and DBP, respectively.

Blood pressure was measured and recorded weekly for 4 weeks in 15 participants who were initially enrolled in this study to investigate the stability of the test device over time. Blood pressure measurements performed during the follow-up were carried out identically to the initial blood pressure measurements, with the exception that no calibration was performed.

Statistical Analysis

Baseline characteristics were expressed as percentages for categorical variables and as means with SDs for continuous variables. Differences between test and reference blood pressure values in the primary and secondary analyses were expressed as means and SDs. To evaluate the correlation between the test and reference blood pressures, the *P* value and Pearson correlation coefficient (*r*) were calculated. Bland-Altman plots were generated to compare the distribution of differences in SBP and DBP between the test and reference devices. All reported *P* values were 2-sided, and significance was set at *P*<.05. All statistical analyses were performed by using R software (version 4.1.1.; R Foundation for Statistical Computing).

Ethics Approval

All participants participated voluntarily in this study and provided written informed consent. The study protocol was approved by the Institutional Review Committee of Chungnam National University Sejong Hospital (approval number: 2020-10-023). This study was conducted in accordance with the relevant guidelines and regulations.

Results

Baseline Characteristics

Among the 87 study participants, 45 (52%) individuals were male, and the average age was 35.6 (SD 10.4) years. The mean height and weight were 168.2 (SD 8.1) cm and 75.2 (SD 17.9) kg, respectively. Further, 14 (16%) participants were diagnosed with hypertension and were taking antihypertensive drugs prior to enrollment. The most used drug to manage hypertension was an angiotensin II receptor blocker. The mean screening SBP and DBP were 134.2 (SD 20.3) mm Hg and 82.8 (SD 15.2) mm Hg, respectively (Table 1).

Table . Baseline characteristics of the participants (N=87).

Characteristics		Value
Age (years), mean (SD)		35.6 (10.4)
Men, n (%)		45 (52)
Height (cm), mean (SD)		168.2 (8.1)
Weight (kg), mean (SD)		75.2 (17.9)
Arm circumference (cm), mean (SD)		30.0 (3.8)
Hypertension with medications, n (%)		14 (16)
Angiotensin II receptor blockers, n (%)		11 (13)
Beta-blocker, n (%)		3 (3)
Calcium channel blocker, n (%)		9 (10)
Diuretics, n (%)		3 (3)
Diabetes mellitus, n (%)		3 (3)
Dyslipidemia, n (%)		5 (6)
Blood pressure at screening		
	SBP[a] (mm Hg), mean (SD)	134.2 (20.3)
	DBP[b] (mm Hg), mean (SD)	82.8 (15.2)
	HR[c] (beats per minute), mean (SD)	81.2 (10.4)
	SBP≥160 mm Hg, n (%)	9 (10)
	DBP≥100 mm Hg, n (%)	14 (16)
	SBP≤110 mm Hg, n (%)	9 (10)
	DBP ≤70 mm Hg, n (%)	17 (20)

[a]SBP: systolic blood pressure.
[b]DBP: diastolic blood pressure.
[c]HR: heart rate.

Primary Analysis

The mean test SBP was 120.0 (SD 18.8) mm Hg, and the mean reference SBP was 118.7 (SD 20.2) mm Hg (Table 2). The difference between the test and reference blood pressures was 1.2 (SD 7.1) mm Hg. The absolute differences between the test and baseline SBPs were <5, <10, and <15 mm Hg in 57.5% (150/261), 84.3% (220/261), and 94.6% (247/261) of the measurements, respectively (Figure 3). In the Bland-Altman plot, of the 261 measurements, 15 (5.7%) were outside 2 SDs of the differences between test and reference SBPs. The DBP values of the test and reference devices were 80.1 (SD 12.6) mm Hg and 81.1 (SD 14.4) mm Hg, respectively. The difference was −1.0 (SD 6.0) mm Hg, but the difference became progressively more pronounced at higher blood pressure levels. The absolute differences between the test and the reference DBPs were <5, <10, and <15 mm Hg in 75.5% (197/261), 93.9% (245/261), and 97.3% (254/261) of the measurements. In the Bland-Altman plot, of the 261 measurements, 10 (3.8%) were outside 2 SDs of the differences, mostly for high blood pressure values.

Table . Differences between the test and reference blood pressures among 87 participants (261 measurement pairs in the primary analysis).

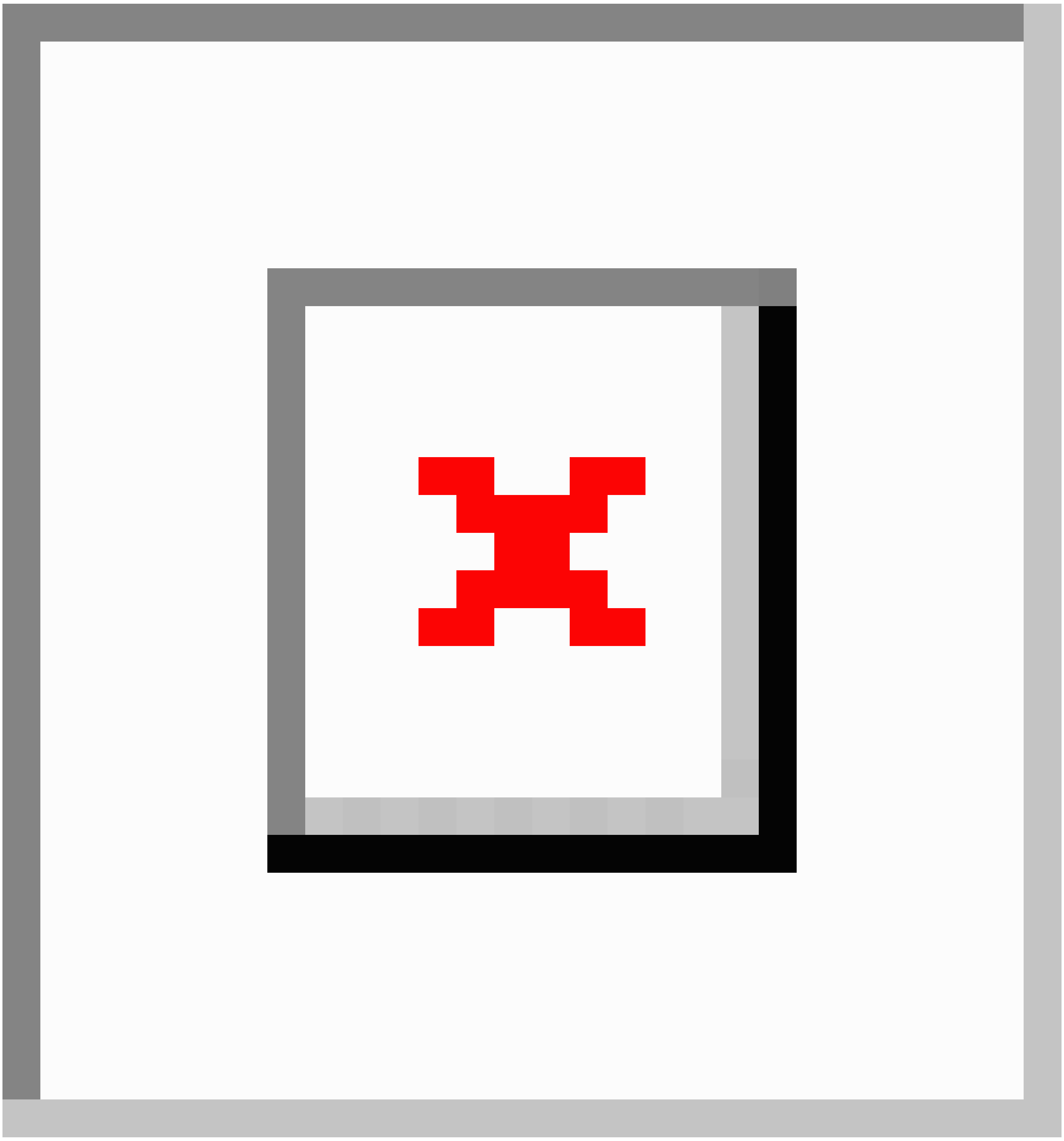
		SBP[a], mm Hg	DBP[b], mm Hg	HR[c], beats per minute
Test device, mean (SD)		120.0 (18.8)	80.1 (12.6)	75.2 (10.4)
Reference device, mean (SD)		118.7 (20.2)	81.1 (14.4)	75.4 (10.1)
Difference, mean (SD)		1.2 (7.1)	−1.0 (6.0)	−0.2 (6.9)
Coefficient of variation for each participant, %				
	Test device	1.3	0.4	4.9
	Reference device	3.1	3.2	4

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]HR: heart rate.

Figure 3. Bland-Altman plots of the SBPs and DBPs measured using the test and reference devices. (A) Bland-Altman plot of the test and reference SBPs. (B) Bland-Altman plot of the test and reference DBPs. DBP: diastolic blood pressure; SBP: systolic blood pressure.



Secondary Analysis

Table 3 lists the results of the follow-up measurements for the test and reference blood pressure values in 15 participants, which were used in the secondary analysis. The differences at the baseline were 1.7 (SD 6.9) mm Hg (*P*=.10) and −2.0 (SD 5.3) mm Hg (*P*=.01) for SBP and DBP, respectively. After 4 weeks, the differences between SBP and DBP were 0.1 (SD 8.8) mm Hg (*P*=.96) and −2.4 (SD 7.6) mm Hg (*P*=.04), respectively.

Table . Differences between the test and reference blood pressures among 15 participants during the 4-week follow-up (45 measurement pairs per week in the secondary analysis).

		Baseline	1 week	2 weeks	3 weeks	4 weeks
SBP[a]						
	Difference (mm Hg), mean (SD)	1.7 (6.9)	1.1 (6.8)	1.0 (7.6)	1.2 (8.7)	0.1 (8.8)
	P value	.10	.31	.36	.37	.96
DBP[b]						
	Difference (mm Hg), mean (SD)	−2.0 (5.3)	−1.6 (7.5)	−3.2 (6.8)	0 (8.1)	−2.4 (7.6)
	P value	.01	.15	.003	.97	.04

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

Discussion

Principal Findings

This study assessed the accuracy of a smartphone-based blood pressure measurement software (AlwaysBP) compared with the standard, cuff-based, manual method of brachial blood pressure measurement. Our results revealed that the test device showed acceptable accuracy for both SBP and DBP, and the difference between the test and reference SBP values was stable over the 1-month follow-up. The DBP values of the study device were not stable for 1 month.

The development of various cuffless blood pressure monitors has allowed for the measurement of blood pressure easily and comfortably. Such devices are expected to improve the penetration rate of blood pressure measurements, which is very important for lowering the rates of hypertension unawareness and uncontrolled hypertension after diagnosis. Several smart devices that can measure blood pressure without the use of a cuff have been developed. The most commercially available devices are smartwatch devices, which work mainly by acquiring photoplethysmography signals and estimating blood pressure via a machine learning algorithm based on these signals or PTT calculation [5,7,9,15,16]. The greatest advantage of smart devices is their ability to measure blood pressure continuously without interference; as such, they can be used to monitor blood pressure during rest, work, and exercise. In addition, they facilitate 24-hour ambulatory blood pressure monitoring by analyzing the trend and pattern of blood pressure change over 24 hours. Although controversies remain regarding their accuracy, they represent good alternatives to cuff-based, 24-hour ambulatory blood pressure monitoring if their accuracy is improved. Devices that measure blood pressure by using only a smartphone, without additional devices, have also been developed. With OptiBP (Biospectal), the photoplethysmography signal of the finger is measured by using a smartphone camera, and a pretrained deep learning model is used to estimate blood pressure based on the photoplethysmography signal [6,17]. Similar to the test device used in our study, the OptiBP software does not require an additional device; thus, it has the advantage of significantly improving the accessibility to blood pressure measurement.

In our study, the test device estimated SBP and DBP based on the PTT. The advantage of PTT-based blood pressure measurement over photoplethysmography analysis–based blood pressure estimation has not been elucidated fully. However, there is a huge body of evidence that PTT is associated with blood pressure [18-20]. Moreover, the PTT can be used to assess vascular stiffness, arterial atherosclerosis, and the risk of future cardiovascular events [21]. The software system used in our study does not require additional devices to measure both photoplethysmography signals and cardiac signals for PTT calculation. Moreover, because the AlwaysBP software collects photoplethysmography and seismocardiography signals simultaneously, it may detect an abnormal heart state more accurately than other devices that use photoplethysmography alone [22]. Therefore, this software may play a role in improving the early detection of hypertension and arterial stiffness and in the prediction of cardiovascular events among the general population.

The existing blood pressure measuring devices requiring calibration also have limitations and therefore need to be improved. Changes in blood pressure within an individual are not usually considered in calibration and testing. Thus, there have been debates regarding the problems in validating cuffless blood pressure measuring devices that require calibration, because the time between calibration and testing is insufficient to account for individual differences in blood pressure [23]. Our system also relies on a single individualized calibration, which may not be suitable for longtime use. Calibration was done to calculate γ in the equations, and this value can be affected by age, sex, height, and vascular stiffness. The single-calibration results of the initial measurements were used for a long time. Even though the differences in SBP values between the test and reference devices among 15 participants over 1 month showed satisfactory results, the SD tended to become more prominent at follow-up. Additionally, DBP values were not stable during follow-up in our study population. Our results showed that accuracy may decrease over time, suggesting that a regular calibration process for long-term users may be necessary. This decrease in accuracy may have been due to changes in physiological properties, such as vascular stiffness and atherosclerosis, that vary at different rates in individuals over time. There will also be issues about the necessity of regular calibration and how often those calibrations will be needed for accurate blood pressure measurements. Ultimately, a calibration-free system is needed, but the development of calibration-free devices is still ongoing [24]. To minimize the calibration issues, new devices that can be used without calibration should be developed. Moreover, the test DBPs showed relatively poor correlations at high blood pressures in this study, although the reasons underlying this finding are unclear. PTT calculation via smartphones is different from conventional PTT calculation, which involves electrocardiographic and photoplethysmographic sensors that are used in clinical practice and have higher resolutions. The inherent limitations of using a smartphone-only blood pressure measurement system result in potential accuracy issues, rendering such systems inferior in accuracy when compared to other smart devices, such as smartwatches with electrocardiographic and photoplethysmographic sensors. Further, we used a set of commonly used equations for calculating blood pressure with PTT. This equation set may not have been fit for our data, which were acquired via smartphones. Therefore, additional studies are needed, and an update of the algorithm may be necessary.

Due to the inherent features exhibited by cuffless blood pressure measuring devices, it is imperative to establish more specialized criteria for validating such products [25]. The existing criteria cannot account for the capability of these devices to accurately monitor blood pressure fluctuations in response to positional changes, physical activity, emotional stress, and prolonged periods, which can significantly influence individuals' blood pressure levels. Consequently, it is essential to subject future devices to rigorous testing with a dedicated validation protocol tailored for cuffless devices. These protocols should emphasize the ability to effectively track blood pressure variations for conditions commonly encountered in daily life.

Limitations

This study had several limitations. First, it was conducted at a single center in Korea. The device algorithm was based on the Korean population during development, and this study was also conducted with Koreans. Second, although study participants were enrolled according to the AAMI/ESH/ISO protocol, we understand that this protocol should not be used for a cuffless device; therefore, further study is required with a validation protocol specifically designed for cuffless blood pressure devices. Third, both arms were used to detect test and reference blood pressure values simultaneously. Although participants with a history of peripheral vascular disease were not enrolled, the difference in reference blood pressures between both arms was not assessed prior to this study. Fourth, despite meeting the validation criteria suggested by the AAMI/ESH/ISO guidelines, our device's accuracy may not be sufficient for populations with high DBPs, as shown by the study results. Fifth, our study primarily included young and healthy participants; thus, the accuracy and applicability of our device for older and comorbid populations may be limited. Sixth, our study did not include the average agreement between the two observers and the frequency of device errors; therefore, further study is required to assess the functional consistency and reliability of the AlwaysBP software. Lastly, our data did not include information regarding the skin pigmentation of participants, which can have effects on the signal to noise ratio. Although our study exclusively recruited individuals of East Asian ethnicity and used LED technology from smartphones to obtain high-quality photoplethysmography data from participants' fingers, the influence of an individual's skin pigmentation cannot be disregarded. Hence, an evaluation of the accuracy of the study device based on variations in skin pigmentation is necessary. Further investigations involving larger sample sizes with diverse comorbidities and skin colors are also required. Such additional studies should be conducted by an independent group for transparency and to mitigate potential bias [26].

Conclusion

This study assessed the accuracy of the AlwaysBP smartphone software, which was used to estimate blood pressure in 87 study participants. The cuffless smartphone software showed acceptable accuracy for SBP and DBP when compared with the reference cuff-based manual blood pressure measurement method, according to the AAMI/ESH/ISO protocol criteria. However, we understand that this protocol was only designed to validate cuff-based blood pressure measuring devices. Therefore, further validation studies with a specific validation protocol designed for cuffless blood pressure measuring devices are required to assess clinical accuracy.

The cuffless smartphone technology can be easily applied in everyday life, and it may improve the awareness and control of hypertension in the general population by enabling regular blood pressure monitoring, remote monitoring by health care providers, and programmed alerts and reminders for medication and measurement. Additionally, educational content can promote adherence to treatment plans and provide tips for managing hypertension.

Acknowledgments

This work was supported by the Technology Innovation Program (20014477; development of a noncontact artificial intelligence health monitoring system based on multimodal sensors), funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

YHY contributed to conceptualization, the methodology, validation, the formal analysis, the investigation, resource management, data curation, the writing of the original draft, and visualization. JK, KJL, and DC contributed to conceptualization, the methodology, the design of the software, and resource management. JKO, MK, and JHR contributed to resource management. HWP contributed to resource management and the review and editing of this paper. JHL contributed to conceptualization, the review and editing of this paper, supervision, project administration, and funding acquisition.

Conflicts of Interest

YHY and JHL serve as advisers for Deepmedi Inc, the owner of the AlwaysBP technology. KJL, JK, and DC are employees of Deepmedi Inc.

References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. J Am Coll Cardiol 2017 Jul 4;70(1):1-25. [doi: 10.1016/j.jacc.2017.04.052] [Medline: 28527533]
2. Bakris G, Ali W, Parati G. ACC/AHA versus ESC/ESH on hypertension guidelines: JACC guideline comparison. J Am Coll Cardiol 2019 Jun 18;73(23):3018-3026. [doi: 10.1016/j.jacc.2019.03.507] [Medline: 31196460]
3. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. ESC/ESH guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 1;39(33):3021-3104. [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]
4. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980. [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
5. Moon JH, Kang MK, Choi CE, Min J, Lee HY, Lim S. Validation of a wearable cuff-less wristwatch-type blood pressure monitoring device. Sci Rep 2020 Nov 4;10(1):19015. [doi: 10.1038/s41598-020-75892-y] [Medline: 33149118]
6. Degott J, Ghajarzadeh-Wurzner A, Hofmann G, Proença M, Bonnier G, Lemkaddem A, et al. Smartphone based blood pressure measurement: accuracy of the OptiBP mobile application according to the AAMI/ESH/ISO universal validation protocol. Blood Press Monit 2021 Dec 1;26(6):441-448. [doi: 10.1097/MBP.0000000000000556] [Medline: 34139747]
7. Vybornova A, Polychronopoulou E, Wurzner-Ghajarzadeh A, Fallet S, Sola J, Wuerzner G. Blood pressure from the optical Aktiia Bracelet: a 1-month validation study using an extended ISO81060-2 protocol adapted for a cuffless wrist device. Blood Press Monit 2021 Aug 1;26(4):305-311. [doi: 10.1097/MBP.0000000000000531] [Medline: 33675592]
8. Matsumura K, Rolfe P, Toda S, Yamakoshi T. Cuffless blood pressure estimation using only a smartphone. Sci Rep 2018 May 8;8(1):7298. [doi: 10.1038/s41598-018-25681-5] [Medline: 29740088]
9. Fortin J, Rogge DE, Fellner C, Flotzinger D, Grond J, Lerche K, et al. A novel art of continuous noninvasive blood pressure measurement. Nat Commun 2021 Mar 2;12(1):1387. [doi: 10.1038/s41467-021-21271-8] [Medline: 33654082]
10. Silver L, Pew Research Center. Smartphone ownership is growing rapidly around the world, but not always equally. 2019. URL: www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2023-08-17]
11. Scully CG, Lee J, Meyer J, Gorbach AM, Granquist-Fraser D, Mendelson Y, et al. Physiological parameter monitoring from optical recordings with a mobile phone. IEEE Trans Biomed Eng 2012 Feb;59(2):303-306. [doi: 10.1109/TBME.2011.2163157] [Medline: 21803676]
12. Stergiou GS, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, et al. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) collaboration statement. J Hypertens 2018 Mar;36(3):472-478. [doi: 10.1097/HJH.0000000000001634] [Medline: 29384983]
13. Stergiou GS, Palatini P, Asmar R, Ioannidis JP, Kollias A, Lacy P, et al. Recommendations and practical guidance for performing and reporting validation studies according to the Universal Standard for the validation of blood pressure measuring devices by the Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO). J Hypertens 2019 Mar;37(3):459-466. [doi: 10.1097/HJH.0000000000002039] [Medline: 30702492]
14. Shao J, Shi P, Hu S, Yu H. A revised point-to-point calibration approach with adaptive errors correction to weaken initial sensitivity of cuff-less blood pressure estimation. Sensors (Basel) 2020 Apr 13;20(8):2205. [doi: 10.3390/s20082205] [Medline: 32295090]
15. Zheng YL, Yan BP, Zhang YT, Poon CCY. An armband wearable device for overnight and cuff-less blood pressure measurement. IEEE Trans Biomed Eng 2014 Jul;61(7):2179-2186. [doi: 10.1109/TBME.2014.2318779] [Medline: 24760899]
16. Elgendi M, Fletcher R, Liang Y, Howard N, Lovell NH, Abbott D, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med 2019 Jun 26;2:60. [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]

17. Schoettker P, Degott J, Hofmann G, Proença M, Bonnier G, Lemkaddem A, et al. Blood pressure measurements with the OptiBP smartphone app validated against reference auscultatory measurements. Sci Rep 2020 Oct 20;10(1):17827. [doi: 10.1038/s41598-020-74955-4] [Medline: 33082436]
18. Mukkamala R, Hahn JO, Inan OT, Mestha LK, Kim CS, Töreyin H, et al. Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. IEEE Trans Biomed Eng 2015 Aug;62(8):1879-1901. [doi: 10.1109/TBME.2015.2441951] [Medline: 26057530]
19. Lee J, Yang S, Lee S, Kim HC. Analysis of pulse arrival time as an indicator of blood pressure in a large surgical biosignal database: recommendations for developing ubiquitous blood pressure monitoring methods. J Clin Med 2019 Oct 24;8(11):1773. [doi: 10.3390/jcm8111773] [Medline: 31653002]
20. Ding X, Zhang YT. Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm. Biomed Eng Lett 2019 Feb 18;9(1):37-52. [doi: 10.1007/s13534-019-00096-x] [Medline: 30956879]
21. Kim HL, Kim SH. Pulse wave velocity in atherosclerosis. Front Cardiovasc Med 2019 Apr 9;6:41. [doi: 10.3389/fcvm.2019.00041] [Medline: 31024934]
22. Rai D, Thakkar HK, Rajput SS, Santamaria J, Bhatt C, Roca F. A comprehensive review on seismocardiogram: current advancements on acquisition, annotation, and applications. Mathematics 2021;9(18):2243. [doi: 10.3390/math9182243]
23. Mukkamala R, Yavarimanesh M, Natarajan K, Hahn JO, Kyriakoulis KG, Avolio AP, et al. Evaluation of the accuracy of cuffless blood pressure measurement devices: challenges and proposals. Hypertension 2021 Nov;78(5):1161-1167. [doi: 10.1161/HYPERTENSIONAHA.121.17747] [Medline: 34510915]
24. Ibrahim B, Jafari R. Cuffless blood pressure monitoring from a wristband with calibration-free algorithms for sensing location based on bio-impedance sensor array and autoencoder. Sci Rep 2022 Jan 10;12(1):319. [doi: 10.1038/s41598-021-03612-1] [Medline: 35013376]
25. Stergiou GS, Mukkamala R, Avolio A, Kyriakoulis KG, Mieke S, Murray A, et al. Cuffless blood pressure measuring devices: review and statement by the European Society of Hypertension Working Group on Blood Pressure Monitoring and Cardiovascular Variability. J Hypertens 2022 Aug 1;40(8):1449-1460. [doi: 10.1097/HJH.0000000000003224] [Medline: 35708294]
26. Whelton PK, Picone DS, Padwal R, Campbell NRC, Drawz P, Rakotz MK, et al. Global proliferation and clinical consequences of non-validated automated BP devices. J Hum Hypertens 2023 Feb;37(2):115-119. [doi: 10.1038/s41371-022-00667-z] [Medline: 35279699]

Abbreviations

AAMI/ESH/ISO: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization
DBP: diastolic blood pressure
IMU: inertial measurement unit
MBP: mean blood pressure
PP: pulse pressure
PTT: pulse transit time
SBP: systolic blood pressure

Edited by L Buis; submitted 08.11.22; peer-reviewed by A Logan, B Merid, S Melville; revised version received 13.04.23; accepted 21.07.23; published 08.09.23.

Please cite as:
Yoon YH, Kim J, Lee KJ, Cho D, Oh JK, Kim M, Roh JH, Park HW, Lee JH
Blood Pressure Measurement Based on the Camera and Inertial Measurement Unit of a Smartphone: Instrument Validation Study
JMIR Mhealth Uhealth 2023;11:e44147
URL: https://mhealth.jmir.org/2023/1/e44147
doi:10.2196/44147

Original Paper

The Treatment Outcome of Smart Device–Based Tinnitus Retraining Therapy: Prospective Cohort Study

Myung-Whan Suh[1,2*], MD, PhD; Moo Kyun Park[1,2*], MD, PhD; Yoonjoong Kim[3,4], MD; Young Ho Kim[2,3], MD, PhD

[1]Department of Otorhinolaryngology, Seoul National University Hospital, Seoul, Republic of Korea

[2]Sensory Organ Research Institute, Seoul National University Medical Research Center, Seoul, Republic of Korea

[3]Department of Otorhinolaryngology-Head and Neck Surgery, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[4]Department of Otorhinolaryngology-Head and Neck Surgery, Chungbuk National University Hospital, Cheongju-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Young Ho Kim, MD, PhD
Department of Otorhinolaryngology-Head and Neck Surgery
Seoul Metropolitan Government-Seoul National University Boramae Medical Center
20 Boramae-ro 5-gil
Dongjak-gu
Seoul, 07061
Republic of Korea
Phone: 82 28702442
Fax: 82 28703863
Email: yhkiment@gmail.com

Abstract

Background: Tinnitus retraining therapy (TRT) is a standard treatment for tinnitus that consists of directive counseling and sound therapy. However, it is based on face-to-face education and a time-consuming protocol. Smart device–based TRT (smart-TRT) seems to have many advantages, but the efficacy of this new treatment has been questioned.

Objective: The aim of this study was to compare the efficacy between smart-TRT and conventional TRT (conv-TRT).

Methods: We recruited 84 patients with tinnitus. Results were compared between 42 patients who received smart-TRT and 42 control participants who received conv-TRT. An interactive smart pad application was used for directive counseling in the smart-TRT group. The smart pad application included detailed education on ear anatomy, the neurophysiological model of tinnitus, concept of habituation, and sound therapy. The smart-TRT was bidirectional: There were 17 multiple choice questions between each lesson as an interim check. The conv-TRT group underwent traditional person-to-person counseling. The primary outcome measure was the Tinnitus Handicap Inventory (THI), and the secondary outcome measure was assessed using a visual analogue scale (VAS).

Results: Both treatments had a significant treatment effect, which comparably improved during the first 2 months. The best improvements in THI were –23.3 (95% CI –33.1 to –13.4) points at 3 months and –16.8 (95% CI –30.8 to –2.8) points at 2 months in the smart-TRT group and conv-TRT group, respectively. The improvements on the VAS were also comparable: smart-TRT group: –1.2 to –3.3; conv-TRT: –0.7 to –1.7.

Conclusions: TRT based on smart devices can be an effective alternative for tinnitus patients. Considering the amount of time needed for person-to-person counseling, smart-TRT can be a cost-effective solution with similar treatment outcomes as conv-TRT.

(JMIR Mhealth Uhealth 2023;11:e38986) doi:10.2196/38986

KEYWORDS

tinnitus; tinnitus retraining therapy; smart device; sound therapy; rehabilitation; therapy; tablet application; app-based; therapy; digital therapy; device-based therapy

Introduction

Tinnitus retraining therapy (TRT) is a habituation therapy that can alleviate tinnitus-induced distress by means of directive counseling and sound therapy [1,2]. Despite the various attempts to cure tinnitus, currently there is no treatment that can completely eliminate tinnitus. Psychological and behavioral interventions such as cognitive behavioral therapy and TRT have been applied as alternatives. According to the 2014 guideline from the American Academy of Otolaryngology–Head and Neck Surgery, clinicians should “educate” patients with persistent, bothersome tinnitus about management strategies [3]. Also, clinicians may recommend “sound therapy” to patients with persistent, bothersome tinnitus [3]. These 2 strategies constitute the basis of TRT. Although high-quality randomized clinical trials are still lacking, it has been proposed in a Cochrane Database systematic review that TRT is more effective than sound masking [4].

One problem with TRT is that it is very time-consuming. With classic TRT, patients undergo at least 3 or 4 sessions of extensive, one-on-one, 60-minute, directive counseling that includes education about the auditory system, brain function, and Jastreboff’s neurophysiological model [2,5]. Due to the long and extensive counseling, it takes a lot of manpower and time. This problem is also associated with the high cost of TRT. In order to overcome this problem, group counseling [6] and simplified TRT [7] have been proposed by several researchers. Simplified TRT is efficient in that the counseling is short (30 minutes) and it omits the lengthy explanation about the anatomy and physiology of hearing. According to previous results, it seems that the treatment effect of simplified TRT is similar to that of classic TRT [7].

To further save manpower and time, we have come up with a TRT using interactive applications and smart devices (smart pad and smartphone). Instead of the classic one-on-one, 60-minute counseling, patients may engage with an application that delivers all the directive counseling content. A tablet computer or smart pad such as an iPad can be used for the bidirectional and intuitive operation. Traditional audiovisual devices such as a television are unidirectional: They can provide information to the user but cannot give any feedback. With the help of recent technology, we can now build applications that are bidirectional: The system can ask questions and assess progress, mimicking one-on-one counseling [8]. This system may also be beneficial for patients, since the cost of TRT can be reduced and there are fewer constraints around time and space [9,10]. Sound therapy can also be provided using an application or mp3 file installed on the patient’s smartphone.

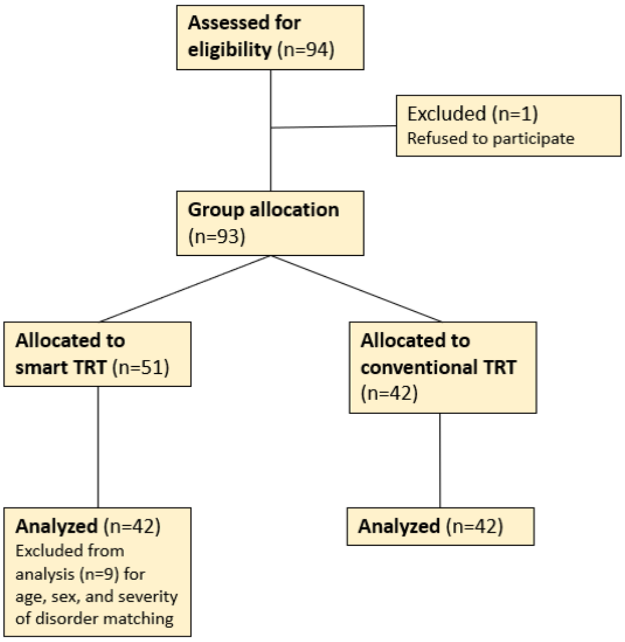
Smart device–based TRT (smart-TRT) seems to have many advantages, but the efficacy of this new treatment has been questioned. Especially, the low level of human contact during TRT may lead to insufficient engagement with the educational intervention [11]. If this new technology is to be recommended to patients, the treatment outcome should be similar or better than conventional TRT (conv-TRT). The aim of this study was to compare the efficacy between smart-TRT and conv-TRT.

Methods

Participants

This study was a prospective trial with 2 arms. Both groups were recruited from the same institute during the same period, but group allocation was not randomized. The participants freely selected between the newly developed intervention or the conventional treatment. Of the 94 participants assessed for eligibility, 1 participant was excluded because the person refused to participate. Of the remaining 93 participants, 51 participants were allocated to the smart-TRT group, and 42 participants were allocated to the conv-TRT group (Figure 1). All participants underwent the intervention, and no one dropped out. We excluded 9 participants from the final analysis in order to match the 2 groups in terms of age, sex, and severity of tinnitus. The inclusion criteria were patients who were older than 20 years and had experienced chronic essential tinnitus for more than 6 months. Patients with vascular tinnitus, posttraumatic tinnitus, psychological disorders, severe hearing loss, sleep disorders, dementia, organic brain disorders, substance abuse, chronic renal failure, or uncontrolled malignancy were excluded from the study. Normal middle ear status was confirmed via audiogram and otoscopy, and we screened for abnormal psychological conditions such as depression, anxiety, and insomnia using the validated version of the Beck Depression Inventory (BDI) for depression [12,13], the State-Trait Anxiety Inventory (STAI) for anxiety [14], and the Pittsburgh Sleep Quality Index (PSQI) for sleep quality [15].

Figure 1. Flow diagram of the study design. TRT: tinnitus retraining therapy.



Ethics Approval

This study was approved by the Seoul National University Hospital Institutional Review Board (IRB no. 1207-112-419) and was conducted according to the tenets of the Declaration of Helsinki. All participants provided written informed consent.

Directive Education

For the smart-TRT group, 3 different interactive smart pad applications were prepared for the 3 directive education sessions. Each application was a composite of numerous video clips. Two screens were displayed on the smart pad: a big screen that showed illustrations or cartoons and a small screen that showed the face of the speaker (1st session, MWS; 2nd session, MKP; 3rd session, YHK). The 3 smart pad applications were presented to the patients with a 1-month interval between each at the clinic. We did not allow the participants to use these applications at home by themselves, to allow a fair comparison with the control group.

The smart pad applications included detailed education on the anatomy and physiology of the ear and auditory pathways, the perception of sound in the auditory cortex, "selective" listening, why tinnitus becomes a problem, the misconception that tinnitus causing hearing difficulties, an explanation of habituation as a goal, subconscious processing of auditory stimuli, "filtering" and "blocking" auditory stimuli from reaching consciousness, how to apply "sound therapy," the neurophysiological model, and homework for the patient. Although the first session explained every aspect of these points, the second and third sessions reviewed the first session and added some new points with further examples. The smart-TRT was bidirectional: There were 17 multiple choice questions between each lesson for an interim check. The questions were mandatory, and the education session did not proceed if the patient did not respond. After the patient's response, the correct answer was provided with further explanation why the answer was correct or incorrect. Since the patient's response to each question was quite variable, the duration of the directive education differed between participants. It took at least 45 minutes for a patient to complete the first education session if the patient answered all the questions correctly and quickly. It took at least 25 minutes for a patient to complete the second and third education sessions. For some patients who had difficulty understanding the directive education, it could take more than 1 hour to complete 1 session.

For the conv-TRT group, simplified group (1-4 patients/session) counseling was provided by a single clinician (MWS). The counselor in the conv-TRT group was identical to the first speaker in the smart-TRT group (MWS). The contents and teaching materials for the directive counseling were also identical between the 2 groups. It took about 45 minutes to 60 minutes for a patient to complete the first session. It took about 10 minutes to 20 minutes for a patient to complete the second and third education sessions. The second and third sessions were rather short in the conv-TRT group, because most of the essential information and strategies were already well-covered. It took less time to review the last session and add new knowledge and encourage higher levels of motivation. Other than these points, all the other treatment and follow-up conditions were the same in the 2 groups.

Sound Therapy

The same sound source file (white noise) was provided to the patients in both groups. The patients used their own smartphone or a portable mp3 player to play the sound. If the patient was familiar with using smartphone applications, a sound therapy application that had been built by our group was installed on their smartphone. We instructed the patients in both groups to use the sound therapy device at the level of the mixing point for at least 6 hours a day.

Outcomes and Statistical Analysis

To calculate the sample size, the study was powered at 80% with a type I error of 5%. We assumed that a 5.9 difference in the Tinnitus Handicap Inventory (THI) score with an SD of 8 between the treatment groups was significant based on the study by Kaldo et al [16]. Assuming a loss of 30%, the number of patients needed for each group was 42 (84 patients total).

The primary outcome measure was the THI score. The change in the THI score was defined as ΔTHI = postTHI score – preTHI score. The secondary outcome measure was a visual analogue scale (VAS) score to quantify awareness of tinnitus, loudness of tinnitus, annoyance caused by tinnitus, and the effects of tinnitus on daily life [17-20]. The change in the VAS score was defined as ΔVAS = postVAS score – preVAS score. The effects of TRT were assessed based on changes in the THI and VAS scores at 0, 1, 2, and 3 months after the TRT. Continuous variables are expressed as mean (SD), and all statistical analyses were performed using SPSS version 16.0 (SPSS Inc, Chicago, IL). A repeated measure analysis of variance was used to evaluate the effect of time, group, and interaction between time and group. We used *t* tests to compare continuous variables and chi-square tests to compare categorical variables. *P* values <.05 were considered to indicate statistical significance.

Results

Demographics

Among the 84 participants (mean age 57.9, SD 11.1 years; 40 men and 44 women), the mean baseline THI was 48.7 (SD 20.9). The baseline clinical characteristics of the smart-TRT group and conv-TRT group are summarized in Table 1. There were no differences in age, gender, affected ear, baseline THI, baseline VAS, baseline STAI, baseline BDI, baseline PSQI, loss to follow-up rate, and pure tone audiometry threshold.

Table 1. Baseline characteristics.

Characteristics	Smart TRT[a] (n=42)	Conventional TRT (n=42)	*P* value
Age (years), mean (SD)	55.8 (12.0)	59.9 (9.9)	.09
Gender, n (%)			.38
Male	22 (52)	18 (43)	
Female	20 (48)	24 (57)	
Side, n (%)			.72
Right	9 (21)	10 (24)	
Left	12 (29)	16 (38)	
Both	15 (36)	12 (29)	
Head or unclear	6 (14)	4 (10)	
Baseline Tinnitus Handicap Inventory, mean (SD)	46.9 (20.7)	50.5 (21.1)	.43
Baseline VAS[b] (awareness), mean (SD)			
Awareness	7.1 (3.4)	7.2 (2.9)	.92
Annoyance	5.9 (2.7)	6.6 (2.8)	.22
Loudness	5.8 (2.4)	6.8 (2.3)	.06
Effect on daily life	4.4 (2.3)	4.9 (2.9)	.41
State-Trait Anxiety Inventory, mean (SD)			
X1	44.1 (9.1)	46.0 (10.9)	.39
X2	43.0 (7.2)	45.7 (12.0)	.24
Beck Depression Inventory, mean (SD)	13.4 (9.6)	15.1 (10.0)	.43
Pittsburgh Sleep Quality Index, mean (SD)	7.8 (3.9)	8.9 (5.8)	.35
Loss to follow-up at 3 months, n (%)	19 (55)	19 (55)	.99
PTA[c] threshold (dB HL[d], mean (SD)			
Right	23.2 (14.6)	25.6 (16.8)	.49
Left	24.0 (15.4)	28.9 (20.6)	.22

[a]TRT: tinnitus retraining therapy.

[b]VAS: visual analogue scale.

[c]PTA: pure-tone audiometry; mean 6-tone average = (500 Hz + 2*1000 Hz + 2*2000 Hz + 4000 Hz)/6.

[d]HL: hearing loss.

Primary Outcome Measure: THI

Figure 2 shows the mean ΔTHI as a function of time. The best ΔTHI score was –23.3 points (95% CI –33.1 to –13.4) at 3 months and –16.8 points (95% CI –30.8 to –2.8) at 2 months in the smart-TRT group and conv-TRT group, respectively. In both groups, the THI score significantly improved over time (within-participant effect: $F_{1.8,42.1}$=10.741, *P*<.001), but there was no difference in the treatment outcome between the 2 groups (between-participant effect: $F_{1,24}$=0.094, *P*=.76). Also, there was no interaction between time and group (time x group effect: $F_{1.8,42.1}$=0.773, *P*=.45). That is, the pattern of gradual improvement as well as the outcome were similar between smart-TRT and conv-TRT.

When each time point was evaluated, a significant reduction in the THI score was found in the smart-TRT group at 1 month (t_{27}=–3.312, *P*=.003), 2 months (t_{23}=–5.040, *P*<.001), and 3 months (t_{18}=–4.947, *P*<.001). A significant reduction was also found in the conv-TRT group at 1 month (t_{30}=2.183, *P*=.04) and 2 months (t_{16}=–2.549, *P*=.02). The treatment effect was marginal (t_{18}=–2.037, *P*=.057) after 3 months of conv-TRT (Table 2).

Figure 2. Mean (SD: error bars) change in the Tinnitus Handicap Inventory (THI) score (ΔTHI) as a function of time. Conv-TRT: conventional tinnitus retraining therapy; Smart-TRT: smart device tinnitus retraining therapy. *P<.05.

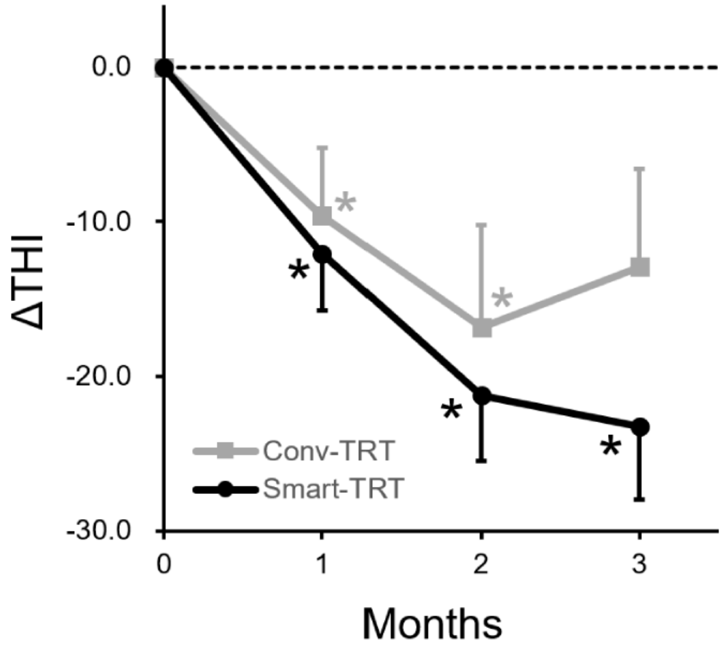


Table 2. Treatment outcome at 3 months.

Outcomes	Smart TRT[a] group		Conventional TRT group	
	Results, mean (95% CI)	*P* value[b]	Results, mean (95% CI)	*P* value[b]
Primary outcome measure: change in the Tinnitus Handicap Inventory	–23.3 (–33.1 to –13.4)	<.001	–12.9 (–26.3 to 0.41)	.057
Secondary outcome measure: change in the visual analogue scale				
Awareness of tinnitus	–3.28 (–4.99 to –1.57)	.001	–1.37 (–2.86 to 0.12)	.07
Annoyance due to tinnitus	–2.89 (–4.24 to –1.54)	<.001	–1.74 (–3.19 to –0.28)	.02
Loudness of tinnitus	–1.22 (–2.41 to –0.03)	.045	–0.74 (–1.80 to 0.33)	.16
Effect on daily life by tinnitus	–2.67 (–3.93 to –1.40)	<.001	–1.26 (–2.47 to –0.05)	.04

[a]TRT: tinnitus retraining therapy.

[b]1-sample *t* test compared with 0.

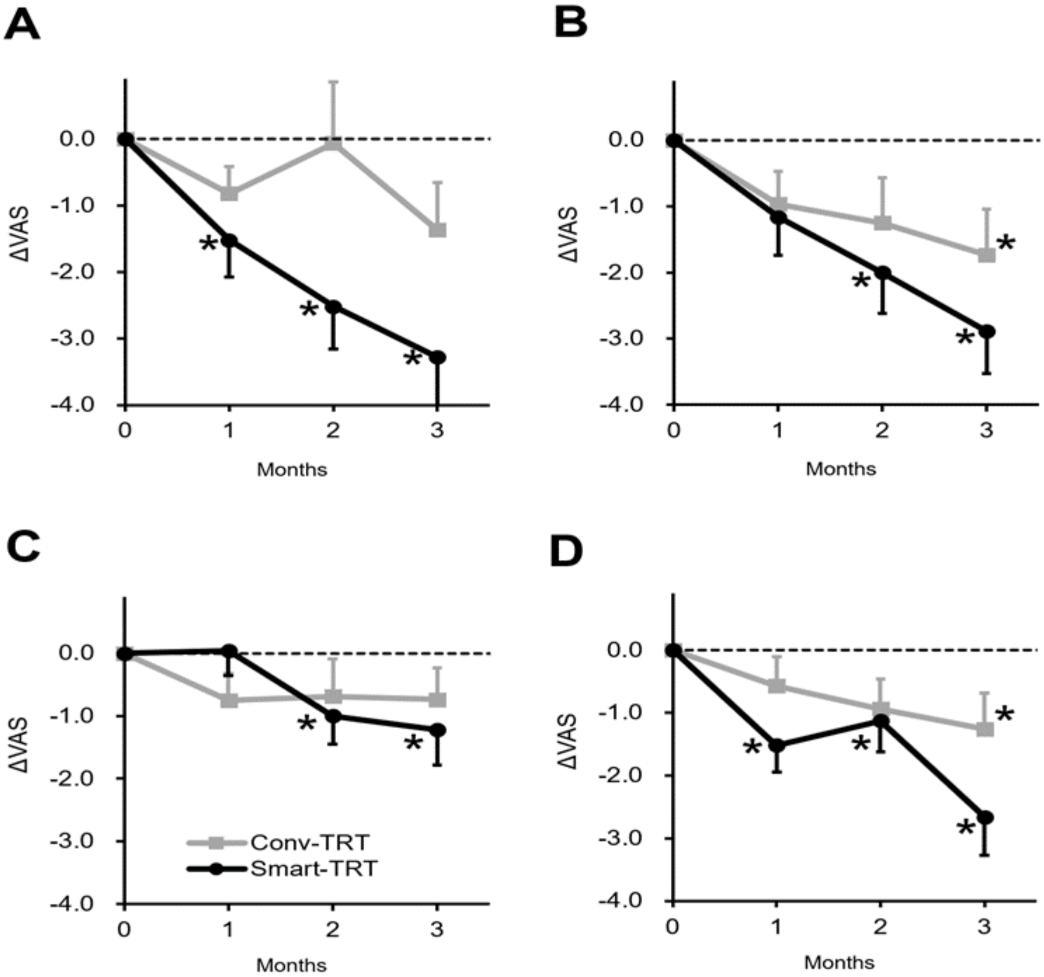
Secondary Outcome Measure: VAS

Figure 3 shows the mean ΔVAS as a function of time. In both groups, the VAS significantly improved over time (within-participant effect) in 3 VAS categories: awareness of tinnitus ($F_{2.2,4.4}$=6.667, *P*=.002), annoyance caused by tinnitus ($F_{3,66}$=4.358, *P*=.007), and effect of tinnitus on daily life ($F_{3,66}$=4.288, *P*=.008). There was no significant change in the loudness of tinnitus over time ($F_{2.4,51.9}$=0.795, *P*=.48).

There were no differences in the treatment outcome between the 2 groups (between-participant effect) in all 4 VAS categories: awareness of tinnitus ($F_{1,22}$=1.196, *P*=.29), annoyance caused by tinnitus ($F_{1,22}$=2.507, *P*=.13), loudness of tinnitus ($F_{1,22}$=0.163, *P*=.96), and effect of tinnitus on daily life ($F_{1,22}$=2.518, *P*=.13). Also, there were no significant interactions between time and group (time x group effect): awareness of tinnitus ($F_{2.2,47.7}$=0.790, *P*=.47), annoyance caused by tinnitus ($F_{3,66}$=1.371, *P*=.26), loudness of tinnitus ($F_{2.4,51.9}$=0.568, *P*=.60), and effect of tinnitus on daily life ($F_{3,66}$=1.606, *P*=.20).

When the ΔVAS was evaluated within each time point, a significant treatment effect was found in the smart-TRT group within 1 month to 2 months. The maximum treatment effect was found at the last follow-up time point (3 months; Table 2). That is, ΔVAS for awareness of tinnitus (t_{17}=–4.038, *P*=.001), annoyance caused by tinnitus (t_{17}=–4.506, *P*=.045), loudness of tinnitus (t_{17}=–2.170, *P*<.001), and effect of tinnitus on daily life (t_{17}=–4.038, *P*=.001) were significantly different from 0 (1-sample *t* test) at 3 months. A similar pattern was found in the conv-TRT group, but the significant improvements were only found in ΔVAS for annoyance caused by tinnitus (t_{10}=–2.511, *P*=.02) and effect of tinnitus on daily life (t_{18}=–2.191, *P*=.04) after 3 months (Figure 3).

Figure 3. Mean (SD: error bars) change in the visual analogue scale (VAS) scores (ΔVAS) as a function of time: (A) awareness of tinnitus, (B) annoyance caused by tinnitus, (C) loudness of tinnitus, (D) effect of tinnitus on daily life. Conv-TRT: conventional tinnitus retraining therapy; Smart-TRT: smart device tinnitus retraining therapy. **P*<.05.



Discussion

Principal Findings

From this study, we were able to show that the treatment outcome of smart-TRT is similar to that of conv-TRT. That is, both treatments had a significant treatment effect that comparably improved over time. As for the primary outcome measure (ΔTHI), the improvement was –23.3 in the smart-TRT group and –16.8 in the conv-TRT group (significant improvement over time within both groups, but no difference between groups). These improvements are not only statistically significant but also clinically significant, given that a change in THI score of 7 or greater is clinically meaningful [21]. The secondary outcome (ΔVAS) was also similar between the 2 groups, with a slight advantage in the smart-TRT group (between –1.2 and –3.3) compared with the conv-TRT group (between –0.7 and –1.7). At 3 months, the THI score slightly deteriorated in the conv-TRT group, while the effect lasted in the smart-TRT group. Although the *P* value was marginal, the ΔTHI score at 3 months was not different from baseline. This finding may imply a deterioration of the treatment result after 3 months. Meanwhile, the secondary outcome measure (VAS scores for annoyance and effect on daily life) showed a steady decrease at 3 months in both groups. It seems that TRT delivered via smart devices can be an alternative treatment for tinnitus patients with similar treatment effects.

The attempt to use smart devices [22] and internet-based audiovisual media [8] for educational interventions for patients with chronic health conditions is a common trend. For example, video-based education and interactive games have been used for patients with type 2 diabetes mellites [23,24]. Multimedia-based animations and quizzes have helped patients with obesity [25]. Osteoarthritis has also been managed with educational modules consisting of text and video [26]. The use of information and communication technology (ICT) for health-related purposes was able to help disease management by facilitating access to health information and helping to increase understanding of the disease [8,27]. Directive counseling in TRT is more complicated than these examples because it is bidirectional. The counselor must understand the individual condition of each patient and tailor the instructions and counseling content depending on the patient's response. However, recent studies postulated that ICT-based interaction can provide not only 1-way educational interventions but also 2-way, interactive counseling via electronic counseling or e-counseling [8]. We agree that smart-TRT can be limited in this bidirectional interactive communication. Also, the low level of human contact may decrease the efficacy and motivation [11]. However, there seems to be other advantages that can balance these limitations.

The biggest advantage of smart-TRT is that it is cost-effective. Since the directive counseling is performed by an application installed on a smart pad, the health provider does not have to spend much time with each patient. Delivering the main idea of TRT and checking whether the patient is making good progress can be done via the smart pad. During this study, the health provider only had to answer questions after each smart-TRT course. The high cost-effectiveness of ICT-based management has been proven for other interventions such as lifestyle modification [9], weight gain prevention [10], and smoking cessation [28]. There is no cost-effectiveness analysis for smart-TRT yet. However, we think it will at least reduce medical personnel expenses for each counseling session. Moreover, by using multiple smart pad devices, many patients can simultaneously undergo directive counseling under the supervision of 1 health provider. Another advantage of smart-TRT is that it can be used for contactless health care and telemedicine. W Beukes et al [29] reported that Internet-based cognitive behavioral therapy for tinnitus could overcome accessibility barriers. The COVID-19 outbreak has greatly changed our way of living as well as how we obtain medial information [30]. We think smart-TRT can be a contactless solution for tinnitus during such difficult times.

The smart pad interface and interactive nature of the multimedia content seem to be critical to the outcome of smart-TRT. The high efficacy of ICT-based education is now generally accepted [31]. As a result, ICT-based courses in college and university have increased by 440% during the past decade [32]. However, there are 2 differences between patients with tinnitus and higher education students. First, the input method and device interface can be a barrier to some patients with tinnitus. Most patients with tinnitus are old and not completely comfortable with a computer, mouse, trackpad, and keyboard. Smart devices using a touch screen interface may play an important role in such situations. That is, given the simplicity and self-explanatory nature, most patients can easily learn how to operate and interact with a smart pad. During our study, it took less than 2 minutes to explain how to use the device, despite some patients having no experience with smart pads. Second, the desire to adopt new information, attitudes, and ideas is much lower in patients with tinnitus. Simply delivering information via a single medium (especially text) does not sustain attention nor positively influence patients with tinnitus. Thanks to the flexibility and expandability of smart pad applications, we can now mix graphic, audio, video, and text content to maximize treatment effect even in less motivated or elderly patients.

Interestingly, the smart-TRT group had a slightly better treatment outcome (3-month ΔTHI of –23.3 points) than the conv-TRT group and in former publications on conv-TRT. Other studies reported a ΔTHI outcome of conv-TRT of –8.3 or –14.5 points at 3 months [33,34]. This is very similar to our results in the conv-TRT group (ΔTHI of –16.8 points). The interpretation of why smart-TRT is slightly better can be controversial. The most probable explanation is that this difference is not statistically nor clinically significant. Another explanation can be that smart-TRT is better because the educational interventions were delivered by 3 different specialists. For conv-TRT, a single health provider was in charge and provided a continued series of directive counseling. This can ensure a good patient-doctor relationship, but the counseling technique and content can become monotonous after several visits. In contrast, 3 different specialists can provide different insights and opportunities for motivation, despite delivering the content though a smart pad. The basic idea and general approach in treating tinnitus were the same, but details on how to deliver the idea were different between specialists. Also, the patients might have felt more confident about their treatment program because 3 different specialists spoke with one voice and repeated the main idea every time.

Limitations

There are several limitations in this study. First, this was not a randomized trial. Although this was a prospective study and the smart-TRT group was enrolled according to the predetermined plan, the conv-TRT group had pre-existing data that were closely matched to the smart-TRT group. However, we believe this point did not undermine the main idea of this study. This is because (1) all the baseline demographics were very similar between the 2 groups and (2) the patients were recruited from the same institute during a similar period, managed by the same medical personnel, and evaluated using identical questionnaires at identical time points. Second, the follow-up duration was not long enough. Since tinnitus is a chronic disorder, the treatment effect should be followed for several months to years. Unfortunately, we were only able to follow the patients up to 3 months. There is a possibility that the results could be different in the long term. However, according to our previous studies, the short-term effect of TRT can also provide clinically important information [1,35]. Third, both smart-TRT and conv-TRT were not able to decrease the perceived sound itself. That is, there was no treatment effect in the VAS category of tinnitus loudness. TRT may only be effective in controlling the distress caused by tinnitus. This is different from recently introduced treatments that can also control the loudness of tinnitus [17,36].

Conclusions

TRT could be effectively delivered to patients with tinnitus using smart devices. TRT-based smart devices could save the time and cost associated with conventional in-person therapy. These methods have gained attention during the COVID-19 era for the potential to decrease the chance of viral spread.

Acknowledgments

This study was supported by a clinical research grant provided from the Seoul Metropolitan Government-Seoul National University (SMG-SNU) Boramae Medical Center, Seoul, Republic of Korea. This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT; number NRF-2021R1A2C1094147). This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number HC19C0128).

Conflicts of Interest

None declared.

References

1. Kim BJ, Chung S, Jung JY, Suh M. Effect of different sounds on the treatment outcome of tinnitus retraining therapy. Clin Exp Otorhinolaryngol 2014 Jun;7(2):87-93 [FREE Full text] [doi: 10.3342/ceo.2014.7.2.87] [Medline: 24917903]
2. Jastreboff PJ. 25 years of tinnitus retraining therapy. HNO 2015 Apr;63(4):307-311. [doi: 10.1007/s00106-014-2979-1] [Medline: 25862626]
3. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical practice guideline: tinnitus. Otolaryngol Head Neck Surg 2014 Oct;151(2 Suppl):S1-S40. [doi: 10.1177/0194599814545325] [Medline: 25273878]
4. Phillips JS, McFerran D. Tinnitus retraining therapy (TRT) for tinnitus. Cochrane Database Syst Rev 2010 Mar 17;2010(3):CD007330 [FREE Full text] [doi: 10.1002/14651858.CD007330.pub2] [Medline: 20238353]
5. Bauer CA, Berry JL, Brozoski TJ. The effect of tinnitus retraining therapy on chronic tinnitus: A controlled trial. Laryngoscope Investig Otolaryngol 2017 Aug;2(4):166-177 [FREE Full text] [doi: 10.1002/lio2.76] [Medline: 28894836]
6. Henry JA, Loovis C, Montero M, Kaelin C, Anselmi K, Coombs R, et al. Randomized clinical trial: group counseling based on tinnitus retraining therapy. J Rehabil Res Dev 2007;44(1):21-32 [FREE Full text] [doi: 10.1682/jrrd.2006.02.0018] [Medline: 17551855]
7. Aazh H, Moore BCJ, Glasberg BR. Simplified form of tinnitus retraining therapy in adults: a retrospective study. BMC Ear Nose Throat Disord 2008 Nov 03;8:7 [FREE Full text] [doi: 10.1186/1472-6815-8-7] [Medline: 18980672]
8. Xie LF, Itzkovitz A, Roy-Fleming A, Da Costa D, Brazeau A. Understanding self-guided web-based educational interventions for patients with chronic health conditions: systematic review of intervention features and adherence. J Med Internet Res 2020 Aug 13;22(8):e18355 [FREE Full text] [doi: 10.2196/18355] [Medline: 32788152]
9. Schulz DN, Smit ES, Stanczyk NE, Kremers SPJ, de Vries H, Evers SMAA. Economic evaluation of a web-based tailored lifestyle intervention for adults: findings regarding cost-effectiveness and cost-utility from a randomized controlled trial. J Med Internet Res 2014 Mar 20;16(3):e91 [FREE Full text] [doi: 10.2196/jmir.3159] [Medline: 24650860]
10. Krishnan A, Finkelstein EA, Levine E, Foley P, Askew S, Steinberg D, et al. A digital behavioral weight gain prevention intervention in primary care practice: cost and cost-effectiveness analysis. J Med Internet Res 2019 May 17;21(5):e12201 [FREE Full text] [doi: 10.2196/12201] [Medline: 31102373]
11. Hutchesson MJ, Morgan PJ, Callister R, Pranata I, Skinner G, Collins CE. Be Positive Be Healthe: development and implementation of a targeted e-Health weight loss program for young women. Telemed J E Health 2016 Jun;22(6):519-528 [FREE Full text] [doi: 10.1089/tmj.2015.0085] [Medline: 26701611]
12. Chung YC, Rhee MK, Lee YH, Park SH, Sohn CH, Hong SK, et al. A standardization study of Beck Depression Inventory 1-Korean version (K-BDI): reliability and factor analysis. Korean Journal of Psychopathology 1995;4(1):77-95 [FREE Full text]
13. Rhee MK, Lee YH, Jung HY, Choi JH, Kim SH, Kim YK, et al. A standardization study of Beck Depression Inventory (II): Korean version (K-DBI): validity. Korean Journal of Psychopathology 1995;4(1):96-104 [FREE Full text]
14. Hahn DW, Lee CH, Chon KK. Korean adaptation of Spielberger's STAI (K-STAI). Korean Journal of Health Psychology 1996;1:1-14 [FREE Full text]
15. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. Sleep Breath 2012 Sep;16(3):803-812. [doi: 10.1007/s11325-011-0579-9] [Medline: 21901299]

16. Kaldo V, Haak T, Buhrman M, Alfonsson S, Larsen HC, Andersson G. Internet-based cognitive behaviour therapy for tinnitus patients delivered in a regular clinical setting: outcome and analysis of treatment dropout. Cogn Behav Ther 2013;42(2):146-158. [doi: 10.1080/16506073.2013.769622] [Medline: 23432207]
17. Noh T, Rah Y, Kyong JS, Kim JS, Park MK, Lee JH, et al. Comparison of treatment outcomes between 10 and 20 EEG electrode location system-guided and neuronavigation-guided repetitive transcranial magnetic stimulation in chronic tinnitus patients and target localization in the Asian brain. Acta Otolaryngol 2017 Sep;137(9):945-951. [doi: 10.1080/00016489.2017.1316870] [Medline: 28471721]
18. Noh T, Kyong JS, Chang MY, Park M, Lee J, Oh S, et al. Comparison of treatment outcomes following either prefrontal cortical-only or dual-site repetitive transcranial magnetic stimulation in chronic tinnitus patients: A double-blind randomized study. Otol Neurotol 2017 Feb;38(2):296-303. [doi: 10.1097/MAO.0000000000001266] [Medline: 28068305]
19. Park JH, Noh T, Lee JH, Oh S, Kim JS, Chung CK, et al. Difference in tinnitus treatment outcome according to the pulse number of repetitive transcranial magnetic stimulation. Otol Neurotol 2015 Sep;36(8):1450-1456. [doi: 10.1097/MAO.0000000000000802] [Medline: 26154843]
20. Kyong J, Noh T, Park MK, Oh S, Lee JH, Suh M. Phantom perception of sound and the abnormal cortical inhibition system: an electroencephalography (EEG) study. Ann Otol Rhinol Laryngol 2019 Jun;128(6_suppl):84S-95S. [doi: 10.1177/0003489419837990] [Medline: 31092043]
21. Zeman F, Koller M, Figueiredo R, Aazevedo A, Rates M, Coelho C, et al. Tinnitus handicap inventory for evaluating treatment effects: which changes are clinically relevant? Otolaryngol Head Neck Surg 2011 Aug;145(2):282-287. [doi: 10.1177/0194599811403882] [Medline: 21493265]
22. Sleath B, Carpenter DM, Davis SA, Sayner R, Lee C, Loughlin CE, et al. Provider-adolescent discussion and provider education about asthma triggers during pediatric visits: results of a randomized trial. J Asthma 2021 Dec;58(12):1565-1573 [FREE Full text] [doi: 10.1080/02770903.2020.1817936] [Medline: 32867560]
23. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and HbA1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (The ANODE Study). J Med Internet Res 2017 Nov 08;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
24. Kerfoot BP, Gagnon DR, McMahon GT, Orlander JD, Kurgansky KE, Conlin PR. A team-based online game improves blood glucose control in veterans with type 2 diabetes: a randomized controlled trial. Diabetes Care 2017 Sep 08;40(9):1218-1225. [doi: 10.2337/dc17-0310] [Medline: 28790131]
25. Thomas JG, Leahey TM, Wing RR. An automated internet behavioral weight-loss program by physician referral: a randomized controlled trial. Diabetes Care 2015 Jan;38(1):9-15 [FREE Full text] [doi: 10.2337/dc14-1474] [Medline: 25404659]
26. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, et al. The web-based osteoarthritis management resource My Joint Pain improves quality of care: a quasi-experimental study. J Med Internet Res 2015 Jul 07;17(7):e167 [FREE Full text] [doi: 10.2196/jmir.4376] [Medline: 26154022]
27. Win KT, Hassan NM, Bonney A, Iverson D. Benefits of online health education: perception from consumers and health professionals. J Med Syst 2015 Mar;39(3):27. [doi: 10.1007/s10916-015-0224-4] [Medline: 25666928]
28. Smit ES, Evers SMAA, de Vries H, Hoving C. Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. J Med Internet Res 2013 Mar 14;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
29. W Beukes E, Andersson G, Fagelson M, Manchaiah V. Internet-based audiologist-guided cognitive behavioral therapy for tinnitus: randomized controlled trial. J Med Internet Res 2022 Feb 14;24(2):e27584 [FREE Full text] [doi: 10.2196/27584] [Medline: 35156936]
30. Gravas S, Ahmad M, Hernández-Porras A, Furriel F, Alvarez-Maestro M, Kumar A, Office of Education and SIU Board of Directors. Impact of COVID-19 on medical education: introducing homo digitalis. World J Urol 2021 Jun 29;39(6):1997-2003 [FREE Full text] [doi: 10.1007/s00345-020-03417-3] [Medline: 32860535]
31. Gernsbacher MA. Why internet-based education? Front Psychol 2014;5:1530 [FREE Full text] [doi: 10.3389/fpsyg.2014.01530] [Medline: 25653625]
32. Allen IE, Seaman J. Grade Change: Tracking Online Education in the United States. Babson Survey Research Group. 2014. URL: https://www.bayviewanalytics.com/reports/gradechange.pdf [accessed 2022-12-23]
33. Lee K, Makino K, Yamahara K. Evaluation of tinnitus retraining therapy for patients with normal audiograms versus patients with hearing loss. Auris Nasus Larynx 2018 Apr;45(2):215-221. [doi: 10.1016/j.anl.2017.03.009] [Medline: 28434756]
34. Henry JA, Schechter MA, Zaugg TL, Griest S, Jastreboff PJ, Vernon JA, et al. Clinical trial to compare tinnitus masking and tinnitus retraining therapy. Acta Otolaryngol Suppl 2006 Dec(556):64-69. [doi: 10.1080/03655230600895556] [Medline: 17114146]
35. Han JJ, Lee JH, Oh SH, Chang SO, Suh M. Assessing the effects of tinnitus retraining therapy in patients lost to follow-up: a telephone survey. Otol Neurotol 2015 Apr;36(4):581-587. [doi: 10.1097/MAO.0000000000000716] [Medline: 25756459]
36. Zeng F, Richardson M, Tran P, Lin H, Djalilian H. Tinnitus treatment using noninvasive and minimally invasive electric stimulation: experimental design and feasibility. Trends Hear 2019;23:2331216518821449 [FREE Full text] [doi: 10.1177/2331216518821449] [Medline: 30803416]

Abbreviations

BDI: Beck Depression Inventory
conv-TNT: conventional TRT
ICT: information and communication technology
KHIDI: Korea Health Industry Development Institute
NRF: National Research Foundation
PSQI: Pittsburgh Sleep Quality Index
smart-TNT: smart device–based TRT
SMG-SNU: Seoul Metropolitan Government-Seoul National University
STAI: State-Trait Anxiety Inventory
THI: Tinnitus Handicap Inventory
TRT: tinnitus retraining therapy
VAS: visual analogue scale

Edited by L Buis; submitted 05.05.22; peer-reviewed by JJ Song, M Masalski; comments to author 29.06.22; revised version received 28.08.22; accepted 20.12.22; published 12.01.23.

Please cite as:
Suh MW, Park MK, Kim Y, Kim YH
The Treatment Outcome of Smart Device–Based Tinnitus Retraining Therapy: Prospective Cohort Study
JMIR Mhealth Uhealth 2023;11:e38986
URL: https://mhealth.jmir.org/2023/1/e38986
doi:10.2196/38986
PMID:36633890

Original Paper

Impact of an eHealth Smartphone App on Quality of Life and Clinical Outcome of Patients With Hand and Foot Eczema: Prospective Randomized Controlled Intervention Study

Wanja Alexander Weigandt[1], MD; Yannic Schardt[1], MD; Aimee Bruch[1]; Raphael Herr[2], PhD; Matthias Goebeler[3], MD; Johannes Benecke[1], MD; Astrid Schmieder[3], MD

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

[2]Center for Preventive Medicine and Digital Health Baden-Württemberg (CPD-BW), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[3]Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Würzburg, Germany

Corresponding Author:
Astrid Schmieder, MD
Department of Dermatology
Venereology and Allergology
University Hospital Würzburg
Josef-Schneider Strasse 2
Würzburg, 97080
Germany
Phone: 49 931 201 26234
Fax: 49 931 201 26700
Email: schmieder_a@ukw.de

Abstract

Background: Chronic hand and foot eczema is a polyetiological dermatological condition. Patients experience pain, itching, and sleep disturbances and have a reduced quality of life. Skin care programs and patient education can improve the clinical outcome. eHealth devices offer a new opportunity to better inform and monitor patients.

Objective: This study aimed to systematically analyze the effect of a monitoring smartphone app combined with patient education on the quality of life and clinical outcome of patients with hand and foot eczema.

Methods: Patients in the intervention group received an educational program; attended study visits on weeks 0, 12, and 24; and had access to the study app. Patients in the control group attended the study visits only. The primary end point was a statistically significant reduction in Dermatology Life Quality Index, pruritus, and pain at weeks 12 and 24. The secondary end point was a statistically significant reduction in the modified Hand Eczema Severity Index (HECSI) score at weeks 12 and 24. This is an interim analysis at week 24 of the 60-week randomized controlled study.

Results: In total, 87 patients were included in the study and randomized to the intervention group (n=43, 49%) or control group (n=44, 51%). Of the 87 patients, 59 (68%) completed the study visit at week 24. There were no significant differences between the intervention and control groups regarding quality of life, pain, itch, activity, and clinical outcome at weeks 12 and 24. Subgroup analysis revealed that, compared with the control group, the intervention group with an app use frequency of fewer than once every 5 weeks had a significant improvement in the Dermatology Life Quality Index at weeks 12 (P=.001) and 24 (P=.05), in pain measured on a numeric rating scale at weeks 12 (P=.02) and 24 (P=.02), and in the HECSI score at week 12 (P=.02). In addition, the HECSI scores assessed on the basis of pictures taken by the patients of their hands and feet correlated strongly with the HECSI scores recorded by physicians during regular personal visits (r=0.898; P=.002) even when the quality of the images was not that good.

Conclusions: An educational program combined with a monitoring app that connects patients with their treating dermatologists can improve quality of life if the app is not used too frequently. In addition, telemedical care can at least partially replace personal care in patients with hand and foot eczema because the analysis of the pictures taken by the patients correlates strongly with that of the in vivo images. A monitoring app such as the one presented in this study has the potential to improve patient care and should be implemented in daily practice.

Trial Registration: Deutsches Register Klinischer Studien DRKS00020963; https://drks.de/search/de/trial/DRKS00020963

(JMIR Mhealth Uhealth 2023;11:e38506) doi:10.2196/38506

KEYWORDS
hand and foot eczema; eHealth; mobile health; mHealth; telemedicine; disease management; smartphone app; mobile phone

Introduction

Background

The prevalence of combined chronic hand and foot eczema in industrialized cities is 5.4% [1]. Women are more frequently affected than men, with an incidence of 9.6 per 1000 compared with 4.0 per 1000 [2].

Hand and foot eczema is considered to be chronic if it persists for >3 months despite adequate therapy or recurs with a frequency of more than twice a year [3]. It does not represent a homogeneous disease entity. The clinical picture, morphology, localization, and etiology can be very different. In general, 4 different etiologies of hand and foot eczema exist: allergic contact, acute-toxic, cumulative-toxic, and atopic hand and foot eczema [4]. Allergic contact hand and foot eczema is typically a type IV sensitization to diverse allergens such as nickel, cobalt, chromates, and fragrancies [5]. Cumulative-toxic hand and foot eczema occurs after repeated exposure to substances that only mildly irritate the skin. Over time, the regenerative capacity of the skin is exceeded, and the eczematous reaction becomes visible. Atopic hand and foot eczema develops on the basis of a genetic predisposition called atopic diathesis. It is therefore a localized variant of atopic eczema with a corresponding etiology [3,6].

The severity of eczema ranges from very mild to very severe, with therapy-refractory courses associated with intense pain and itching [7]. In addition, patients with eczema often have to face social stigmatization and struggle with feelings of shame [8]. These physical and psychological circumstances often lead to a radical reduction in quality of life and may even result in depression [9].

More often than not, patients with eczema have limited knowledge of the pathogenesis of their skin condition and the correct disease management [10]. In many other diseases such as type 2 diabetes mellitus, patient education has proven to be an effective method to increase knowledge of the disease, thereby improving the clinical outcome. Coppola et al [11] have shown that patient education is usually associated with an improvement in clinical knowledge, lifestyle, and psychosocial outcomes in comparison with usual care. In Germany, there are skin protection seminars run by employers' liability insurance associations, but these are reserved for people whose eczema is caused or exacerbated by their professional activity.

In our department of dermatology, patient education alone for patients with psoriasis had no significant effect on the clinical outcome [12]. We therefore assume that one-time education of patients with chronic inflammatory skin conditions may not suffice to ameliorate the disease in the long term.

eHealth-based supporting systems for patients are becoming popular and are incorporated more frequently into patient care. Germany recently set up the German acronym for Digital Health Applications (DiGA) directory, which lists Conformité Européenne–marked medical devices that aim to detect, monitor, treat, or alleviate diseases or to detect, treat, alleviate, or compensate for injuries or disabilities [13]. Physicians (MDs) in Germany can prescribe eHealth devices listed in the DiGA directory. There are currently no DiGA directory–listed eHealth devices for patients experiencing hand and foot eczema in Germany, and scientific data on the beneficial effect of eHealth applications for these patients are missing.

Objectives

The aim of this prospective randomized controlled intervention study was to analyze whether a monitoring smartphone app combined with patient education would improve the quality of life and clinical outcome of patients with hand and foot eczema. The study app was developed specifically for this study. With the app, our patients were able to periodically measure Dermatology Life Quality Index (DLQI) and Hand Eczema Severity Index (HECSI; modified version for foot eczema) scores, as well as the impact on activity and pain (both measured on a numeric rating scale [NRS]), and document the progression of their disease through photographs [14-16]. In addition, the app allowed patients to directly contact their own treating physicians through a chat function.

Furthermore, the DLQI, HECSI, and NRS (for activity and pain) scores were assessed by the treating physicians during personal visits at weeks 0 (before the intervention), 12, and 24.

The final aim behind the development of the app was to reduce waiting time for a physician's appointment in case of an emergency by expanding teledermatological services for patients with hand and foot eczema and to allow precise self-monitoring by the patients.

Methods

Study Design

The aim of this 60-week randomized controlled intervention study was to investigate the effect of patient education in combination with a monitoring smartphone app on patients experiencing chronic hand and foot eczema. This is an interim analysis of the data from study weeks 0, 12, and 24.

The study was carried out at the department of dermatology, venereology, and allergology at the University Medical Center Mannheim in Mannheim, Germany, from August 13, 2018, to August 30, 2021. The inclusion criteria included a physician-confirmed diagnosis of chronic hand and foot eczema, ability to give informed consent, access to a smartphone, and patient age between 18 and 75 years. During the first study visit (week 0 [V1]), the study participants were randomly assigned to the control or intervention group in a ratio of 1:1.

To assign patients to a group, we created 50 lots for the intervention group and 50 lots for the control group. These were sealed in an urn, and the patients were asked to draw lots.

In total, 90 participants were included in the study, but 3 (3%) dropped out of the study before they were assigned to a group. Of the 87 remaining participants, 43 (49%) were assigned to the intervention group and 44 (51%) to the control group.

The control group started the first study visit at week 0. Information on sociodemographic data, preexisting conditions, and previous and current therapies were collected, and standardized questionnaires such as the DLQI administered. In addition, patients' current level of knowledge about their disease, severity of the disease measured using the HECSI or a modified form of the HECSI for foot eczema, and the intensity of the pain and itch measured using an NRS ranging from 0 to 10 were recorded. Furthermore, the negative impact on the activity measured using the NRS of patients was assessed. In-person follow-up visits were carried out at V2 and V3. The same parameters were recorded for the intervention group. In addition, these patients received a 2-hour detailed training session on pathogenesis, classification, therapeutic options, and behavioral recommendations from 2 dermatological specialists at our clinic. Each patient also received a personal access code to our app, DermaScope Mobile. Using this app, patients were able to take pictures of their hands and feet, use a chat function to ask questions that were answered by their treating dermatologists, and complete questionnaires on quality of life (DLQI) and current symptoms (NRS for itch and pain). Screenshots of the app can be found in the paper by Domogalla et al [17]. The highest possible app use frequency was once a week.

The quality of each image uploaded in the app by the patients was categorized by the rater (YS) as good or bad based on the following criteria: well-illuminated picture, sharp and focused image, and complete presentation of the hands and feet. All 3 criteria had to be met for the image to be rated as good. Each image was assigned to the rater (YS), who checked its quality based on these 3 criteria. If all criteria were met, the image was rated as of good quality. We then calculated an electronic HECSI (eHECSI) score based on these images and statistically examined the extent to which this score correlated with the HECSI score collected in person.

The primary end point of the study was to determine the effect of extensive patient training, physician-patient contact on demand, and our app on quality of life as well as itching and pain at weeks 12 and 24. The secondary end points were the effect on the disease outcome assessed with the HECSI at weeks 12 and 24. Modulating effects of sex, age, and disease duration were evaluated for each end point.

Ethics Approval

The medical ethics committee of the Medical Faculty Mannheim, Heidelberg University, approved the study (2017-655N-MA), and the implementation complied with the Declaration of Helsinki. All participants were instructed in detail regarding the study design and gave their informed consent before participating in the study.

Statistical Analysis

Linear panel data regression analyses estimated the trajectories in the outcomes. Random effect regressions determined the main and interaction effects of group membership (intervention vs control group) and visit time point (V1, V2, and V3) on DLQI, pain, daily activity, and HECSI scores. Two models of adjustment were calculated. The first model was unadjusted, whereas the second model was adjusted for sex, age, and disease duration. In additional analyses, the effects of app use frequency over 24 weeks were included (group membership: control vs <20% app use frequency vs ≥20% app use frequency). Therefore, the intervention group was divided into 2 groups: one comprised patients with app use frequency <20%, and the other was made up of patients with app use frequency ≥20% during the observation period of 24 weeks. The chosen cutoff of 20% equals app use frequency of once every 5 weeks. Variables were tested for normal distribution, and where relevant, they were transformed to approach normal distribution (power transform square root of DLQI and $\log_{10}$ of HECSI). All statistical analyses were performed using STATA Special Edition (version 14.0; StataCorp LLC).

To determine the extent to which the eHECSI score correlated with the HECSI score assessed at the face-to-face visit, we calculated Spearman correlation coefficients.

We also examined within the intervention group the socioeconomic factors that influenced the course of HECSI and DLQI.

Table 1 shows mean values of the scales, Figure 1 shows the flowchart of the study, and Figures 2-4 show predictive margins (delta method).

Table 1. Patient characteristics.

	Week 0 (V1[a])			Week 24 (V3)		
	Overall (n=87)	Control group (n=44)	Intervention group (n=43)	Overall (n=59)	Control group (n=36)	Intervention group (n=23)
Sex, n (%)						
Female	51 (59)	25 (57)	26 (61)	33 (56)	21 (58)	12 (52)
Male	36 (41)	19 (43)	17 (40)	26 (44)	15 (42)	11 (48)
Age (years)						
Mean (SD)	47.07 (15.42)	48.05 (14.09)	46.07 (16.78)	51.05 (13.98)	50.28 (12.84)	52.26 (15.82)
Median	50	51	49	54	53	54
BMI (kg/m^2)[b]						
Mean (SD)	27.62 (7.53)	26.45 (5.7)	28.82 (9.2)	27.60 (8.17)	26.63 (5.70)	29.11 (10.97)
Median	25.78	25.16	26.81	25.82	25.3	26.81
Smoker, n (%)[b]	29 (33)	16 (36)	13 (30)	20 (34)	13 (37)	7 (30)
Duration of eczema (years)[b]						
Mean (SD)	6.9 (8.23)	6.0 (8.47)	7.81 (7.98)	8.02 (9.27)	6.67 (9.09)	10.13 (9.34)
Median	4	3	6	4	4	8
Antieczema therapy, n (%)						
Topical urea	78 (90)	41 (93)	37 (86)	56 (95)	35 (97)	21 (91)
Topical glucocorticoids	57 (66)	30 (68)	27 (63)	41 (70)	26 (72)	15 (62)
Topical calcineurin inhibitor	14 (16)	9 (21)	5 (12)	12 (20)	8 (22)	4 (17)
Systemic therapy	10 (12)	3 (7)	7 (16)	7 (12)	3 (8)	4 (17)
DLQI[c] (scores range from 0 to 30)						
Mean (SD)	7.97 (6.38)	7.73 (7.16)	8.21 (5.55)	4.71 (5.38)	5 (5.39)	4.26 (5.5)
Median	6	6	8	3	3.5	2
Pain (scores range from range 0 to 10)						
Mean (SD)	1.94 (2.67)	2.14 (2.77)	1.74 (2.59)	1.78 (2.61)	2.33 (2.97)	0.91 (1.65)
Median	0	1	0	0	1	0
Activity (scores range from range 0 to 10)						
Mean (SD)	4.02 (3.23)	3.95 (3.37)	4.09 (3.12)	2.08 (2.60)	2.50 (2.90)	1.43 (1.93)
Median	4	4	4	1	1.5	1
HECSI[d] (scores range from0 to 360)						
Mean (SD)	22.53 (21.29)	20.93 (20.72)	24.16 (21.99)	14.58 (18.44)	17.00 (21.67)	10.80 (11.16)
Median	18	15	19	8	9	6
App use frequency[e], n (%)						
<20%	N/A[f]	N/A	N/A	8 (14)	0 (0)	8 (35)
≥20%	N/A	N/A	N/A	15 (25)	0 (0)	15 (65)

[a]V: visit time point.

[b]Data for BMI, smoking, and eczema duration were collected at the first visit only.

[c]DLQI: Dermatology Life Quality Index.

[d]HECSI: Hand Eczema Severity Index.

[e]Data for app use frequency were calculated over the whole 24 weeks.

[f]N/A: not applicable.

Figure 1. Flow chart of the study cohort and subcohorts.

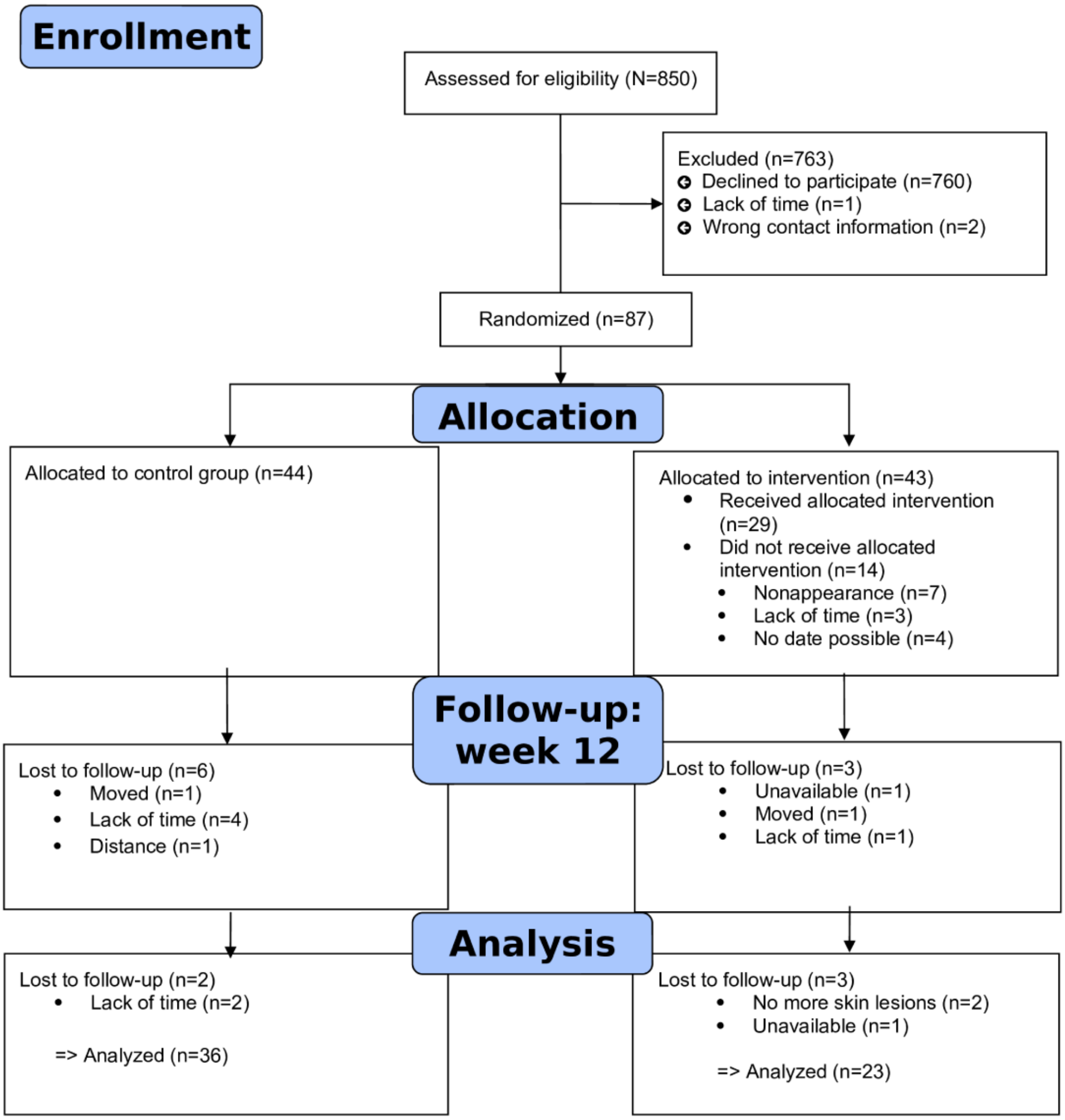


Figure 2. Progression of Dermatology Life Quality Index (DLQI), pain, activity, and Hand Eczema Severity Index (HECSI) in the control group (n=36) versus that in the intervention group (n=23). (A) Progression of DLQI over 24 weeks in the intervention group compared with that in the control group. Changes in both groups from baseline were significant (week 12: *P*=.006; week 24: *P*<.001). There were no significant differences between the groups (week 12: *P*=.09; week 24: *P*=.11). (B) Progression of pain scores over 24 weeks in the intervention group compared with that in the control group. Changes in both groups from baseline were not significant (week 12: *P*=.48; week 24: *P*=.28). There were no differences between the groups (week 12: *P*=.90; week 24: *P*=.27). (C) Progression of activity scores over 24 weeks in the intervention group compared with that in the control group. Changes in both groups from baseline were significant (week 12: *P*=.04; week 24: *P*=.001). There were no significant differences between the groups (week 12: *P*=.21; week 24: *P*=.26). (D) Progression of HECSI over 24 weeks in the intervention group compared with that in the control group. Changes in both groups from baseline were significant (week 12: *P*=.03; week 24: *P*=.002). There were no significant differences between the groups (week 12: *P*=.26; week 24: *P*=.14). Significance at *P*<.05. NRS: numeric rating scale.

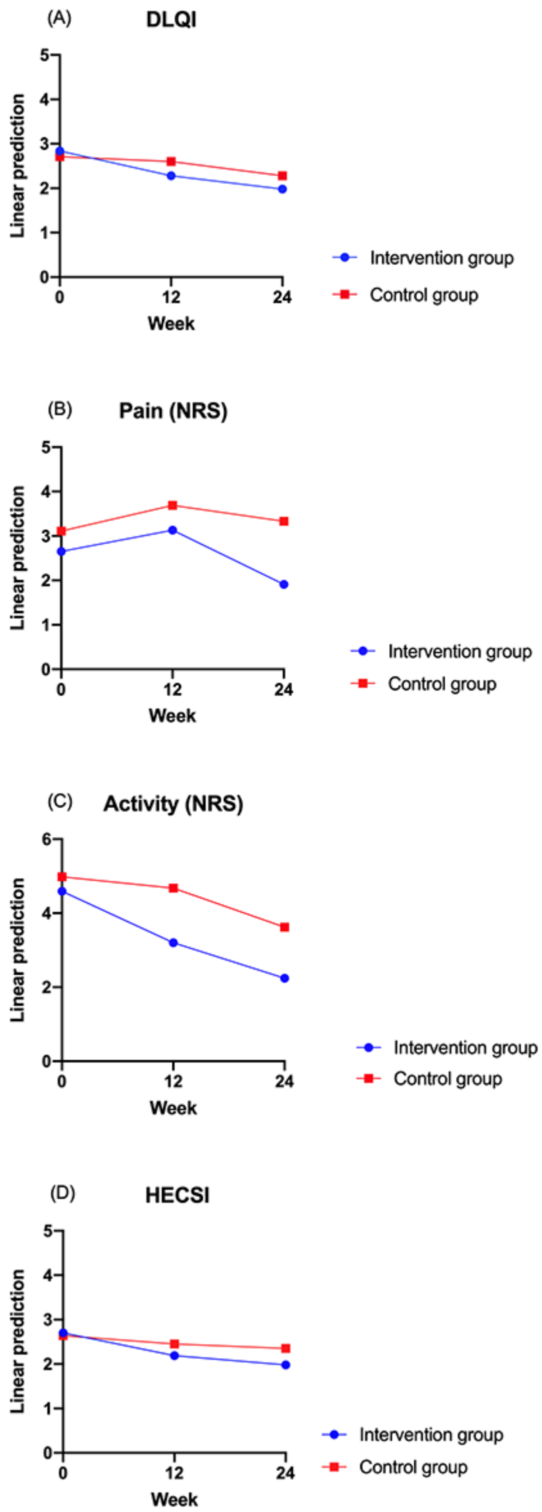


Figure 3. Progression of Dermatology Life Quality Index (DLQI), pain, activity, and Hand Eczema Severity Index (HECSI) in the control group (n=36) versus that in the intervention group with <20% app use frequency (n=8) versus that in the intervention group with ≥20% app use frequency (n=15). (A) Progression of DLQI over 24 weeks in the intervention group with <20% app use frequency compared with that in the intervention group with ≥20% app use frequency compared with that in the control group. Changes were significant in the <20% app use frequency group (week 12: *P*=.001; week 24: *P*=.049) but not in the ≥20% app use frequency group (week 12: *P*=.91; week 24: *P*=.39) compared with controls. (B) Development of pain scores over 24 weeks in the intervention group with <20% app use frequency compared with that in the intervention group with ≥20% app use frequency compared with that in the control group. Changes were significant in the <20% app use frequency group (week 12: *P*=.02; week 24: *P*=.02) but not in the ≥20% app use frequency group (week 12: *P*=.14; week 24: *P*=.91). (C) Development of activity scores over 24 weeks in the intervention group with <20% app use frequency compared with that in the intervention group with ≥20% app use frequency compared with that in the control group. Changes in the <20% app use frequency group were significant at week 12 but not at week 24 (week 12: *P*=.01; week 24: *P*=.17), whereas in the ≥20% app use frequency group (week 12: *P*=.98; week 24: *P*=.56), there were no significant differences. (D) Progression of HECSI over 24 weeks in the intervention group with <20% app use frequency compared with that in the intervention group with ≥20% app use frequency compared with that in the control group. Changes in the <20% app use frequency group were significant at week 12 but not at week 24 (week 12: *P*=.02; week 24: *P*=.12). There were no significant differences in the ≥20% app use frequency group (week 12: *P*=.94; week 24: *P*=.35). Significance at *P*<.05. NRS: numeric rating scale.

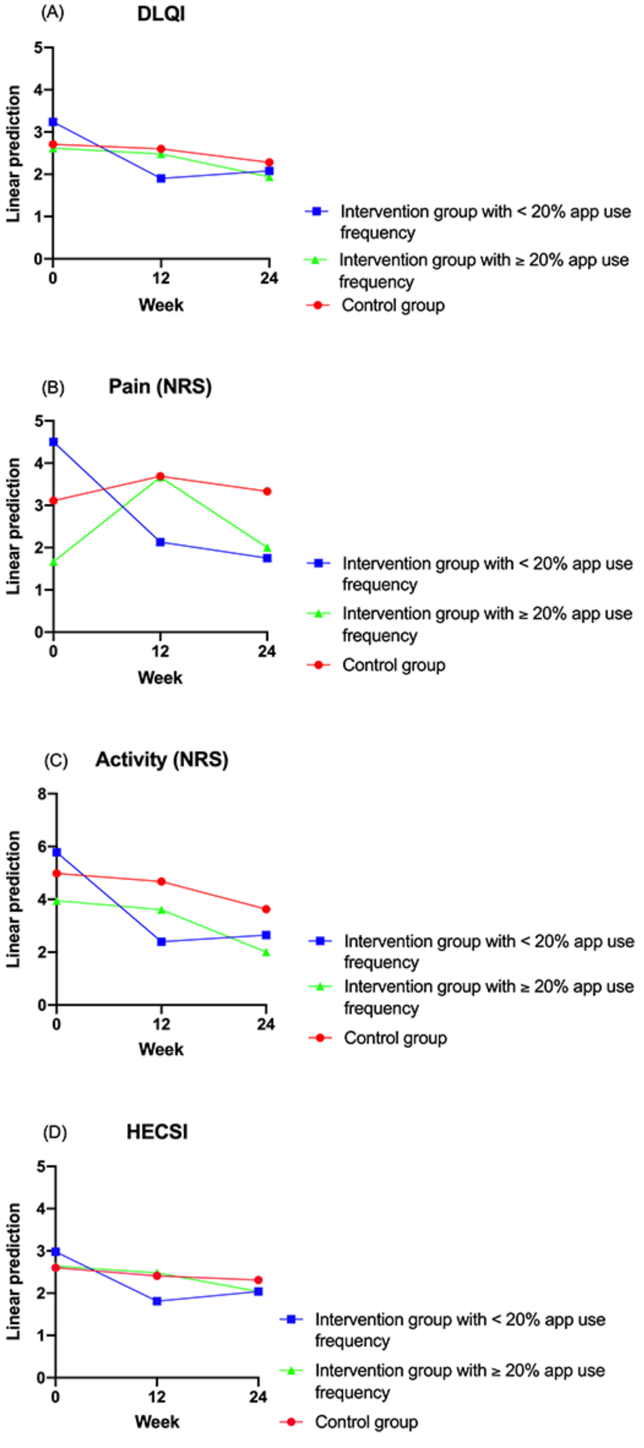
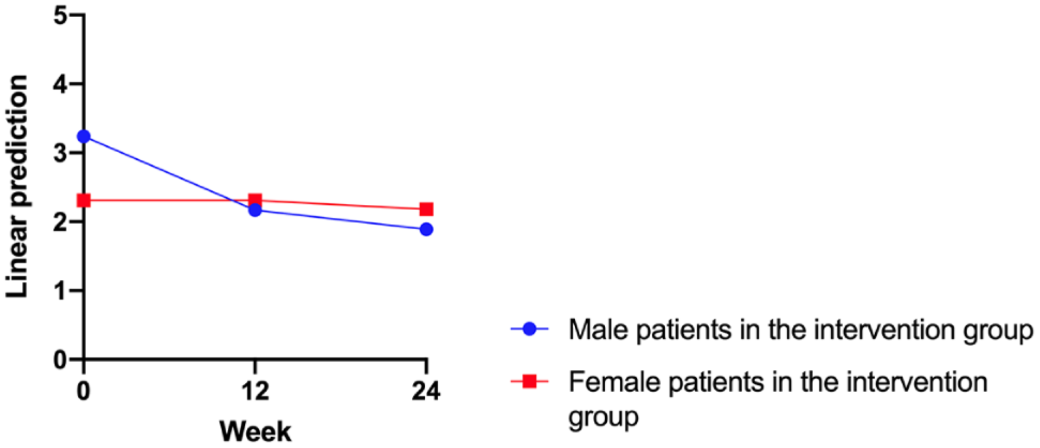


Figure 4. Sex-specific progression of the Hand Eczema Severity Index in the intervention group over 24 weeks. Female participants in the intervention group were compared with male participants. Changes were significant only for the male participants (week 12: *P*=.008; week 24: *P*=.003). Significance at *P*<.05.



Results

Patient Demographics

In total, 90 patients were included in the study. The main reasons for declining participation were lack of time, amelioration of hand and foot eczema, or distance to our outpatient clinic.

Of the 90 patients who signed the informed consent form, 87 (97%) took part in the baseline visit and were randomized 1:1 to the intervention (n=43, 49%) or control (n=44, 51%) groups. Of the 90 patients initially included in the study, 3 (3%) dropped out of the study before the baseline visit. Of the 87 remaining patients, 23 (26%) discontinued the study after the baseline visit or the educational program (intervention group: 17/43, 40%, and control group: 6/44, 14%). Leading up to week 24, of the 64 remaining patients, 5 (8%) discontinued the study, resulting in 59 (92%) patients completing the week 24 visit (Figure 1).

Effects of the Intervention on Quality of Life, Pain, Activity, and Clinical Outcome

Patients in both the intervention and control groups showed an improvement in quality of life (DLQI) at weeks 12 (V2) and 24 (V3; week 12 [V2]: *r*=–0.56; *P*=.006; week 24 [V3]: *r*=–0.86; *P*<.001; Figure 2; Table 2) compared with the baseline visits. No significant differences were observed between the control and intervention groups (*r*=–0.23; *P*=.42) and their progress (week 12 [V2]: *r*=0.45; *P*=.09; week 24 [V3]: *r*=0.42; *P*=.11; Table 2), although the intervention group showed a greater improvement than the control group.

Regarding pain, patients in both groups showed no significant amelioration over time compared with the baseline visits (V2: *r*=0.48; *P*=.48; V3: *r*=–0.74; *P*=.28; Figure 2; Table 2). There were no significant differences between the intervention and control groups (*r*=0.46; *P*=.53) and their trajectories (V2: *r*=0.11; *P*=.90; V3: *r*=0.96; *P*=.27; Table 2).

A significant improvement was observed in the activity score from V1 until V3 (V2: *r*=–1.39; *P*=.04; V3: *r*=–2.35; *P*=.001; Figure 2; Table 2). There was no difference between the 2 groups (*r*=0.08; *P*=.92) and their progress (V2: *r*=1.09; *P*=.21; V3: *r*=0.99; *P*=.26; Table 2).

There was also a significant improvement in the severity of eczema as assessed by the HECSI in both groups compared with the baseline visits (V2: *r*=–0.51; *P*=.02; V3: *r*=–0.72; *P*=.002; Figure 2; Table 2). There was no difference between the groups (*r*=–0.16; *P*=.56) or their trajectories (V2: *r*=0.33; *P*=.26; V3: *r*=0.43; *P*=.14; Table 2). All results were independent of sex, age, or disease duration (model 1; Table 2). Table 1 shows mean values of the scales, whereas Figure 2 shows predictive margins (delta method).

Table 2. Random effect regression models over 24 weeks. Model 0 unadjusted, and model 1 adjusted for age, sex, and disease duration (n=59; observations=177).

Assessment	Model 0		Model 1	
	r (SE)	*P* value	*r* (SE)	*P* value
DLQI[a]				
Week 0	Ref[b]	Ref	Ref	Ref
Week 12	–0.561 (0.205)	.006	–0.561 (0.205)	.006
Week 24	–0.855 (0.205)	<.001	–0.855 (0.205)	<.001
Intervention group	Ref	Ref	Ref	Ref
Control group	–0.234 (0.288)	.42	–0.128 (0.288)	.66
Week 0 × control	Ref	Ref	Ref	Ref
Week 12 × control	0.450 (0.262)	.09	0.450 (0.262)	.09
Week 24 × control	0.420 (0.262)	.11	0.420 (0.262)	.11
R^2: within	0.184 (N/A[c])	N/A	0.184 (N/A)	N/A
R^2: between	0.01 (N/A)	N/A	0.097 (N/A)	N/A
R^2: overall	0.059 (N/A)	N/A	0.125 (N/A)	N/A
Pain				
Week 0	Ref	Ref	Ref	Ref
Week 12	0.478 (0.676)	.48	0.478 (0.676)	.48
Week 24	–0.739 (0.676)	.28	–0.739 (0.676)	.28
Intervention group	Ref	Ref	Ref	Ref
Control group	0.459 (0.723)	.53	0.468 (0.737)	.53
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × control group	0.105 (0.866)	.90	0.105 (0.866)	.90
Week 24 × control group	0.961 (0.866)	.27	0.961 (0.866)	.27
R^2: within	0.038 (N/A)	N/A	0.038 (N/A)	N/A
R^2: between	0.040 (N/A)	N/A	0.086 (N/A)	N/A
R^2: overall	0.039 (N/A)	N/A	0.064 (N/A)	N/A
Activity				
Week 0	Ref	Ref	Ref	Ref
Week 12	–1.390 (0.677)	.04	–1.390 (0.677)	.04
Week 24	–2.350 (0.677)	.001	–2.350 (0.677)	<.001
Intervention group	Ref	Ref	Ref	Ref
Control group	0.079 (0.765)	.92	0.393 (0.755)	.60
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × control group	1.090 (0.867)	.21	1.090 (0.867)	.21
Week 24 × control group	0.987 (0.867)	.26	0.987 (0.867)	.26
R^2: within	0.144 (N/A)	N/A	0.144 (N/A)	N/A
R^2: between	0.030 (N/A)	N/A	0.157 (N/A)	N/A
R^2: overall	0.082 (N/A)	N/A	0.151 (N/A)	N/A
HECSI[d]				
Week 0	Ref	Ref	Ref	Ref

Assessment	Model 0		Model 1	
	r (SE)	*P* value	*r* (SE)	*P* value
Week 12	–0.513 (0.229)	.03	–0.513 (0.229)	.03
Week 24	–0.715 (0.229)	.002	–0.715 (0.229)	.002
Intervention group	Ref	Ref	Ref	Ref
Control group	–0.158 (0.273)	.56	–0.062 (0.254)	.81
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × control group	0.327 (0.293)	.26	0.327 (0.293)	.26
Week 24 × control group	0.429 (0.293)	.14	0.429 (0.293)	.14
R^2: within	0.102 (N/A)	N/A	0.102 (N/A)	N/A
R^2: between	0.01 (N/A)	N/A	0.286 (N/A)	N/A
R^2: overall	0.044 (N/A)	N/A	0.211 (N/A)	N/A

[a]DLQI: Dermatology Life Quality Index.
[b]Ref: reference value.
[c]N/A: not applicable.
[d]HECSI: Hand Eczema Severity Index.

An App Use Frequency of Fewer Than Once Every 5 Weeks Leads to a Significant Amelioration of Quality of Life, Pain, Activity, and Extent of Eczema

When analyzing the outcomes in regard to app use frequency, the subgroup with an app use frequency of <20% showed a highly significant improvement in quality of life (DLQI) compared with the control group (V2: *r*=–1.23; *P*=.001; V3: *r*=–0.73; *P*=.05; Figure 3; Table 3). Overall, <20% app use means an app use frequency of <5 times over the study period. For the subgroup with ≥20% app use, there was no significant difference in the DLQI score compared with the control group (V2: *r*=–0.03; *P*=.91; V3: *r*=0.25; *P*=.39; Figure 3; Table 3).

The pain also improved significantly in the subgroup with <20% app use frequency compared with the control group (V2: *r*=–2.96; *P*=.02; V3: *r*=–2.97; *P*=.02; Figure 3; Table 3). In the subgroup with ≥20% app use frequency, there was again no significant effect (V2: *r*=1.41; *P*=.14; V3: *r*=–0.11; *P*=.91; Figure 3; Table 3).

In regard to the activity score of the patients, a significant improvement in the subgroup with <20% app use frequency in comparison with the control group was noted for V2, but not for V3 (V2: *r*=–3.07; *P*=.01; V3: *r*=–1.76; *P*=.17; Figure 3; Table 3). There were no significant differences in the subgroup with ≥20% app use frequency (V2: *r*=–0.03; *P*=.98; V3: *r*=–0.57; *P*=.56; Figure 3; Table 3).

The HECSI showed a significant improvement in the subgroup with <20% app use frequency in comparison with the control group at V2 but again not at V3 (V2: *r*=–0.99; *P*=.02; V3: *r*=–0.65; *P*=.12; Figure 3; Table 3). There were again no significant differences in the subgroup with ≥20% app use frequency in comparison with the control group (V2: *r*=0.03; *P*=.94; V3: *r*=–0.31; *P*=.35; Figure 3; Table 3). Again, all results were independent of sex, age, or disease duration (model 1; Table 3).

Table 3. Random effect regression models of the app use frequency subgroups <20% and ≥20% over 24 weeks. Model 0 unadjusted, and model 1 adjusted for age, sex, and disease duration (n=59; observations=177).

Assessment	Model 0		Model 1	
	r (SE)	*P* value	*r* (SE)	*P* value
DLQI[a]				
Week 0	Ref[b]	Ref	Ref	Ref
Week 12	–0.110 (0.159)	.49	–0.110 (0.159)	.49
Week 24	–0.435 (0.159)	.006	–0.440 (0.159)	.006
Control group	Ref	Ref	Ref	Ref
Intervention group with <20% app use frequency	0.524 (0.421)	.21	0.531 (0.419)	.21
Intervention group with ≥20% app use frequency	0.079 (0.331)	.81	–0.089 (0.337)	.79
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × intervention group with <20% app use frequency	–1.230 (0.373)	.001	–1.230 (0.373)	.001
Week 12 × intervention group with ≥20% app use frequency	–0.032 (0.293)	.91	–0.032 (0.292)	.91
Week 24 × intervention group with <20% app use frequency	–0.733 (0.373)	.049	–0.733 (0.373)	.049
Week 24 × intervention group with ≥20% app use frequency	–0.253 (0.293)	.39	–0.253 (0.292)	.39
Pain				
Week 0	Ref	Ref	Ref	Ref
Week 12	0.583 (0.521)	.26	0.583 (0.521)	.26
Week 24	0.222 (0.521)	.67	0.222 (0.521)	.67
Control group	Ref	Ref	Ref	Ref
Intervention group with <20% app use frequency	1.390 (1.050)	.19	1.600 (1.060)	.13
Intervention group with ≥20% app use frequency	–1.440 (0.827)	.08	–1.590 (0.846)	.06
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × intervention group with <20% app use frequency	–2.960 (1.220)	.02	–2.960 (1.220)	.02
Week 12 × intervention group with ≥20% app use frequency	1.420 (0.961)	.14	1.420 (0.961)	.14
Week 24 × intervention group with <20% app use frequency	–2.970 (1.220)	.02	–2.970 (1.220)	.02
Week 24 × intervention group with ≥20% app use frequency	0.111 (0.961)	.91	0.111 (0.961)	.91
Activity				
Week 0	Ref	Ref	Ref	Ref
Week 12	–0.306 (0.535)	.57	–0.306 (0.535)	.57
Week 24	–1.360 (0.534)	.01	–1.360 (0.535)	.01
Control group	Ref	Ref	Ref	Ref
Intervention group with <20% app use frequency	0.764 (1.120)	.50	0.792 (1.100)	.47
Intervention group with ≥20% app use frequency	–0.572 (0.880)	.55	–1.04 (0.880)	.24
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × intervention group with <20% app use frequency	–3.070 (1.250)	.01	–3.070 (1.250)	.01
Week 12 × intervention group with ≥20% app use frequency	–0.028 (0.986)	.98	–0.028 (0.986)	.98
Week 24 × intervention group with <20% app use frequency	–1.760 (1.250)	.17	–1.760 (1.250)	.16
Week 24 × intervention group with ≥20% app use frequency	–0.572 (0.986)	.56	–0.572 (0.986)	.56
HECSI[c]				
Week 0	Ref	Ref	Ref	Ref
Week 12	–0.185 (0.181)	.31	–0.185 (0.181)	.31
Week 24	–0.286 (0.181)	.11	–0.286 (0.181)	.11

Assessment	Model 0		Model 1	
	r (SE)	*P* value	*r* (SE)	*P* value
Control group	Ref	Ref	Ref	Ref
Intervention group with <20% app use frequency	0.383 (0.399)	.34	0.466 (0.371)	.21
Intervention group with ≥20% app use frequency	0.037 (0.314)	.91	–0.160 (0.296)	.59
Week 0 × control group	Ref	Ref	Ref	Ref
Week 12 × intervention group with <20% app use frequency	–0.990 (0.423)	.02	–0.990 (0.423)	.02
Week 12 × intervention group with ≥20% app use frequency	0.026 (0.333)	.94	0.026 (0.333)	.94
Week 24 × intervention group with <20% app use frequency	–0.652 (0.423)	.12	–0.652 (0.423)	.12
Week 24 × intervention group with ≥20% app use frequency	–0.310 (0.333)	.35	–0.310 (0.333)	.35

[a]DLQI: Dermatology Life Quality Index.

[b]Ref: reference value.

[c]HECSI: Hand Eczema Severity Index.

Male Patients Profit More From the Intervention Regarding the Clinical Outcome

In a further subgroup analysis of the intervention group in regard to the sex-specific development of the HECSI, we found a significant improvement in the HECSI compared with baseline only for male participants (V2: *r*=–1.06; *P*=.008; V3: *r*=–1.21; *P*=.003).

Correlation of the eHECSI With the HECSI

Correlating the eHECSI assessed on the basis of pictures taken by the patients of their hands and feet with the HECSI recorded by physicians during regular personal visits, the eHECSI correlated strongly with the in-person–assessed HECSI (*r*=0.898; *P*=.002) even when the quality of the images was not that good. If the pictures were of good quality, the correlation of the eHECSI with the HECSI was also highly significant (*r*=0.875; *P*<.001).

Discussion

Principal Findings

In our intervention study, we showed that the use of our monitoring app in combination with a patient education session has a significant effect on quality of life, pain, activity, and clinical outcome if the app is not used more than once every 5 weeks. In addition, men seem to profit more from app use frequency than women regarding the clinical outcome.

We first analyzed differences between the intervention and control groups in regard to amelioration of quality of life, pain, activity, and eczema. All our study patients, independent of group membership, had less pain, showed an enhanced quality of life, and participated more actively in life; in addition, their skin condition improved over time. Although the intervention group showed a stronger improvement at all times, the difference between the 2 groups never reached significance. As our patients received a physician's appointment every 3 months regardless of their skin condition, we conclude that the regular physician-patient contact was crucial for the amelioration of the disease in both groups. This aligns with the observations of Riedl et al [18] who showed that regular physician-patient contact leads to improvement in subjective and objective symptoms. Direct physician-patient contact seems to be more effective than an educational program combined with a monitoring app in the short term regarding our whole study population. In this case, the final evaluation of the study data at week 60 will provide better knowledge about the long-term effects achieved by our intervention.

In our previous intervention study involving a 60-week monitoring app for patients with psoriasis, we were able to show that patient education in combination with a monitoring app resulted in a significant amelioration of depressive and anxiety symptoms in patients who used the app fewer than once a month [17]. In that study, we concluded that patients who were chronically ill do not wish to be reminded of their disease too often. Moreover, it seemed that patients do not want to invest too much time in documenting their disease because they already need to spend considerable time in taking care of their eczematous skin. Furthermore, in this study, an app use frequency of fewer than once every 5 weeks led to a significant amelioration of quality of life, pain, activity, and extent of eczema in the subgroup using the app fewer than once a month (<20% app use frequency) compared with the control group. The mainstay of hand and foot eczema management is still topical therapy, which needs to be applied several times a day. For patients with psoriasis, process aspects such as application time have been associated with nonadherence and a negative impact on quality of life [19,20]. In line with this observation, Retzler et al [21] showed that topical treatment regimens in patients with atopic dermatitis have a detrimental effect on quality of life that increases with treatment duration and frequency of application. Therefore, an additional time-consuming burden imposed on patients with hand and foot eczema such as a too-frequent app-based documentation of their skin disease might generate no additional benefits regarding quality of life and disease outcome. It should be noted that the collected data do not allow differentiating whether patients who used the app less frequently simply experienced an improvement in their skin condition. This group could have profited solely from the patient education, which enhanced knowledge, provided in the study. This observation is in concordance with the study by Ahn et al [22], who were able to show that patient education and web-based resources in dermatology increase compliance and adherence to therapy. We cannot rule out that the education provided by the 2 dermatological specialists led to the assessed significant improvement in the subgroup using the app fewer than once every 5 weeks, but in our previous study [12] for patients with psoriasis, the education alone had no effect on the outcome. Therefore, we assume that the same is true for patients with chronic hand and foot eczema.

We additionally assessed whether patients reduce the app use frequency as their outcomes improve, but the subgroup with <20% app use frequency showed lower app use frequency from the start, with no decrease in the use in the course of time.

By contrast, the app provided in the study allowed direct contact between patients and their treating physicians, which probably reassured patients and improved quality of life in the intervention group when using the app fewer than once a week. We believe that the mere possibility of being able to contact the supervising physician if needed rather than the frequency of physician-patient contact is decisive to improved quality of life. In our clinical perception, frequent physician's appointments to obtain a follow-up prescription may become a burden, in particular for younger patients who have less time because of their jobs. Such patients might benefit significantly from additional teledermatological care.

Another finding of our study was that the HECSI of male participants decreased faster than those of female participants, independent of app use frequency, although women show higher adherence to topical therapy [23]. We assume that men may benefit more from a constant reminder to apply their topical therapy provided by an eHealth device even when they avoid frequent documentation of their eczema in the app. A positive benefit for reminder apps has already been demonstrated for therapy adherence in patients with cardiovascular disease [24]. Further studies addressing this point are needed in patients with hand and foot eczema.

One of the study's great strengths was that we were able to show that telemedical care can at least partially replace personal care in patients with hand and foot eczema because the analysis of pictures taken by the patients correlates strongly with that of the in vivo images. Therefore, the HECSI assessed in the face-to-face visit correlated significantly with the eHECSI. This is surely not the case for all dermatological diseases in which the disease can affect the whole body, especially the genital area and the capillitium. A study by Zabludovska et al [25] concluded that only significant changes were detected by photographs; however, in the study, the number of participants was very small (N=33). Whether photographs can be used to monitor the progression of chronic hand eczema and reliably determine HECSI should be further investigated.

Our study includes some limitations. A major limitation is the monocentric design and the small study cohort, which limits generalizability of the results. In particular, the group with <20% app use frequency is very small, which could have led to missed or overinterpreted differences between the groups, especially as we compared this subgroup of the intervention group with the control group. Further studies are necessary to verify our findings on a broader scale.

Conclusions

Overall, our intervention had a positive effect on quality of life, pain, activity, and possibly the clinical outcome in a subgroup of patients with hand and foot eczema.

We were able to show that a monitoring app for patients with hand and foot eczema that allows direct contact with their treating physicians combined with patient education may have the potential to improve the eczema outcome of these patients, especially if the app is not used too frequently. We believe that a monitoring app such as the one presented in this study has the potential to improve patient care and should be implemented in daily practice. However, because of the small number of participants, especially in the subgroups of the intervention group, as well as missing data on treatment adherence of the control group, these data need to be re-examined in a larger sample with consideration of individual factors.

Acknowledgments

The authors express special thanks to the team of the dermatology department of the University Hospital Mannheim. This work was sponsored by Novartis Pharma GmbH.

Conflicts of Interest

AS received honoraria for presentations and as a member of the advisory boards of LEO Pharma, Novartis GmbH, and Almirall Hermal GmbH. AS is the CEO and owner of Derma Intelligence GmbH, which developed DermaScope Mobile.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2536 KB - mhealth_v11i1e38506_app1.pdf]

References

1. Brans R, Hübner A, Gediga G, John SM. Prevalence of foot eczema and associated occupational and non-occupational factors in patients with hand eczema. Contact Dermatitis 2015 Aug 25;73(2):100-107. [doi: 10.1111/cod.12370] [Medline: 25716740]

2. Thyssen J, Johansen J, Linneberg A, Menné T. The epidemiology of hand eczema in the general population--prevalence and main findings. Contact Dermatitis 2010 Feb;62(2):75-87. [doi: 10.1111/j.1600-0536.2009.01669.x] [Medline: 20136890]
3. Gorris A, Kinaciyan T. Das Handekzem. hautnah 2020 Nov 02;19(4):162-170. [doi: 10.1007/s12326-020-00407-5]
4. Braun-Falco's Dermatologie, Venerologie und Allergologie. Berlin, Heidelberg: Springer; 2018.
5. Agner T, Aalto-Korte K, Andersen K, Foti C, Gimenéz-Arnau A, Goncalo M, et al. Factors associated with combined hand and foot eczema. J Eur Acad Dermatol Venereol 2017 May 15;31(5):828-832. [doi: 10.1111/jdv.14016] [Medline: 27768241]
6. Palmer C, Irvine A, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee S, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet 2006 Apr;38(4):441-446. [doi: 10.1038/ng1767] [Medline: 16550169]
7. Diepgen T. [Chronic hand eczema: epidemiology and therapeutic evidence]. Hautarzt 2008 Sep 17;59(9):683-4, 686. [doi: 10.1007/s00105-008-1560-9] [Medline: 18709342]
8. Agrawal P, Kumar A, Sharma Y, Deora M, Ranpariya R. Comparative analysis of epidemiological data as well as quality of life in patients having hand eczema vis-à-vis foot eczema. Indian Dermatol Online J 2019;10(5):519. [doi: 10.4103/idoj.IDOJ_487_18]
9. Boehm D, Schmid-Ott G, Finkeldey F, John S, Dwinger C, Werfel T, et al. Anxiety, depression and impaired health-related quality of life in patients with occupational hand eczema. Contact Dermatitis 2012 Oct;67(4):184-192. [doi: 10.1111/j.1600-0536.2012.02062.x] [Medline: 22564098]
10. Fisker MH, Ebbehøj NE, Jungersted JM, Agner T. What do patients with occupational hand eczema know about skin care? Contact Dermatitis 2013 Aug 22;69(2):93-98. [doi: 10.1111/cod.12060] [Medline: 23869728]
11. Coppola A, Sasso L, Bagnasco A, Giustina A, Gazzaruso C. The role of patient education in the prevention and management of type 2 diabetes: an overview. Endocrine 2016 Jul;53(1):18-27. [doi: 10.1007/s12020-015-0775-7] [Medline: 26494579]
12. Bubak C, Schaarschmidt M, Schöben L, Peitsch WK, Schmieder A. Analyzing the value of an educational program for psoriasis patients: a prospective controlled pilot study. BMC Public Health 2019 Nov 19;19(1):1535 [FREE Full text] [doi: 10.1186/s12889-019-7778-x] [Medline: 31744488]
13. Hersteller EL, Anwender L. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V. Bundesinstitut für Arzneimittel und Medizinprodukte. 2022 Mar 18. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf?__blob=publicationFile [accessed 2023-01-10]
14. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol 1994 May;19(3):210-216. [doi: 10.1111/j.1365-2230.1994.tb01167.x] [Medline: 8033378]
15. Held E, Skoet R, Johansen J, Agner T. The hand eczema severity index (HECSI): a scoring system for clinical assessment of hand eczema. A study of inter- and intraobserver reliability. Br J Dermatol 2005 Feb;152(2):302-307. [doi: 10.1111/j.1365-2133.2004.06305.x] [Medline: 15727643]
16. Hartrick CT, Kovan JP, Shapiro S. The numeric rating scale for clinical pain measurement: a ratio measure? Pain Pract 2003 Dec;3(4):310-316. [doi: 10.1111/j.1530-7085.2003.03034.x] [Medline: 17166126]
17. Domogalla L, Beck A, Schulze-Hagen T, Herr R, Benecke J, Schmieder A. Impact of an eHealth smartphone app on the mental health of patients with psoriasis: prospective randomized controlled intervention study. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e28149 [FREE Full text] [doi: 10.2196/28149] [Medline: 34431478]
18. Riedl D, Schüßler G. The influence of doctor-patient communication on health outcomes: a systematic review. Z Psychosom Med Psychother 2017 Jun 01;63(2):131-150. [doi: 10.13109/zptm.2017.63.2.131] [Medline: 28585507]
19. Fouéré S, Adjadj L, Pawin H. How patients experience psoriasis: results from a European survey. J Eur Acad Dermatol Venereol 2005 Nov;19 Suppl 3(s3):2-6. [doi: 10.1111/j.1468-3083.2005.01329.x] [Medline: 16274404]
20. Schmitt J, Wozel G, Garzarolli M, Viehweg A, Bauer M, Leopold K. Effectiveness of interdisciplinary vs. dermatological care of moderate-to-severe psoriasis: a pragmatic randomised controlled trial. Acta Derm Venereol 2014 Mar;94(2):192-197 [FREE Full text] [doi: 10.2340/00015555-1697] [Medline: 23995849]
21. Retzler J, Smith A, Reaney M, Rout R, Hudson R. Process utilities for topical treatment in atopic dermatitis. Qual Life Res 2019 Sep 3;28(9):2373-2381 [FREE Full text] [doi: 10.1007/s11136-019-02174-0] [Medline: 30945132]
22. Ahn CS, Culp L, Huang WW, Davis SA, Feldman SR. Adherence in dermatology. J Dermatolog Treat 2017 Mar 15;28(2):94-103. [doi: 10.1080/09546634.2016.1181256] [Medline: 27180785]
23. Zaghloul S, Cunliffe W, Goodfield M. Objective assessment of compliance with treatments in acne. Br J Dermatol 2005 May;152(5):1015-1021. [doi: 10.1111/j.1365-2133.2005.06357.x] [Medline: 15888162]
24. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
25. Zabludovska K, Ibler KS, Jemec GB, Agner T. Photographic documentation and hand eczema severity index for severity assessment of hand eczema. Dermatitis 2017;28(4):280-283. [doi: 10.1097/der.0000000000000306]

Abbreviations

DiGA: German acronym for Digital Health Applications

DLQI: Dermatology Life Quality Index
eHECSI: electronic Hand Eczema Severity Index
HECSI: Hand Eczema Severity Index
NRS: numeric rating scale
V: visit time point

Edited by L Buis; submitted 05.04.22; peer-reviewed by KL Mauco, I Tency, A Cobo, E Nabovati; comments to author 25.07.22; revised version received 26.08.22; accepted 21.09.22; published 07.03.23.

Please cite as:
Weigandt WA, Schardt Y, Bruch A, Herr R, Goebeler M, Benecke J, Schmieder A
Impact of an eHealth Smartphone App on Quality of Life and Clinical Outcome of Patients With Hand and Foot Eczema: Prospective Randomized Controlled Intervention Study
JMIR Mhealth Uhealth 2023;11:e38506
URL: https://mhealth.jmir.org/2023/1/e38506
doi:10.2196/38506
PMID:36881465

Original Paper

Effectiveness of a Sodium-Reduction Smartphone App and Reduced-Sodium Salt to Lower Sodium Intake in Adults With Hypertension: Findings From the Salt Alternatives Randomized Controlled Trial

Helen Eyles[1], PhD; Jacqueline Grey[1], BPED; Yannan Jiang[1], PhD; Elaine Umali[1], MPH; Rachael McLean[2], PhD; Lisa Te Morenga[3], PhD; Bruce Neal[4], PhD; Anthony Rodgers[4], PhD; Robert N Doughty[5], MD; Cliona Ni Mhurchu[1], PhD

[1]National Institute for Health Innovation, The University of Auckland, Auckland, New Zealand
[2]Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand
[3]Research Centre for Hauora and Health, Massey University Wellington, Wellington, New Zealand
[4]The George Institute for Global Health, University of New South Wales, Sydney, Australia
[5]Department of Medicine, The University of Auckland, Auckland, New Zealand

Corresponding Author:
Helen Eyles, PhD
National Institute for Health Innovation
The University of Auckland
School of Population Health, Grafton Campus, Faculty of Medical and Health Sciences
22-30 Park Avenue, Private Bag 92019, Auckland Mail Centre 1142
Auckland, 1023
New Zealand
Phone: 64 9 923 4658
Email: h.eyles@auckland.ac.nz

Abstract

Background: Even modest reductions in blood pressure (BP) can have an important impact on population-level morbidity and mortality from cardiovascular disease. There are 2 promising approaches: the SaltSwitch smartphone app, which enables users to scan the bar code of a packaged food using their smartphone camera and receive an immediate, interpretive traffic light nutrition label on-screen alongside a list of healthier, lower-salt options in the same food category; and reduced-sodium salts (RSSs), which are an alternative to regular table salt that are lower in sodium and higher in potassium but have a similar mouthfeel, taste, and flavor.

Objective: Our aim was to determine whether a 12-week intervention with a sodium-reduction package comprising the SaltSwitch smartphone app and an RSS could reduce urinary sodium excretion in adults with high BP.

Methods: A 2-arm parallel randomized controlled trial was conducted in New Zealand (target n=326). Following a 2-week baseline period, adults who owned a smartphone and had high BP (≥140/85 mm Hg) were randomized in a 1:1 ratio to the intervention (SaltSwitch smartphone app + RSS) or control (generic heart-healthy eating information from The Heart Foundation of New Zealand). The primary outcome was 24-hour urinary sodium excretion at 12 weeks estimated via spot urine. Secondary outcomes were urinary potassium excretion, BP, sodium content of food purchases, and intervention use and acceptability. Intervention effects were assessed blinded using intention-to-treat analyses with generalized linear regression adjusting for baseline outcome measures, age, and ethnicity.

Results: A total of 168 adults were randomized (n=84, 50% per group) between June 2019 and February 2020. Challenges associated with the COVID-19 pandemic and smartphone technology detrimentally affected recruitment. The adjusted mean difference between groups was 547 (95% CI −331 to 1424) mg for estimated 24-hour urinary sodium excretion, 132 (95% CI −1083 to 1347) mg for urinary potassium excretion, −0.66 (95% CI −3.48 to 2.16) mm Hg for systolic BP, and 73 (95% CI −21 to 168) mg per 100 g for the sodium content of food purchases. Most intervention participants reported using the SaltSwitch app (48/64, 75%) and RSS (60/64, 94%). SaltSwitch was used on 6 shopping occasions, and approximately 1/2 tsp per week of RSS was consumed per household during the intervention.

Conclusions: In this randomized controlled trial of a salt-reduction package, we found no evidence that dietary sodium intake was reduced in adults with high BP. These negative findings may be owing to lower-than-anticipated engagement with the trial intervention package. However, implementation and COVID-19–related challenges meant that the trial was underpowered, and it is possible that a real effect may have been missed.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000352101; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377044 and Universal Trial U1111-1225-4471

(JMIR Mhealth Uhealth 2023;11:e43675) doi:10.2196/43675

KEYWORDS

mobile health; mHealth; smartphone; smartphone app; cardiovascular disease; sodium; salt; blood pressure; technology; reduced-sodium salt; mobile phone

Introduction

Background

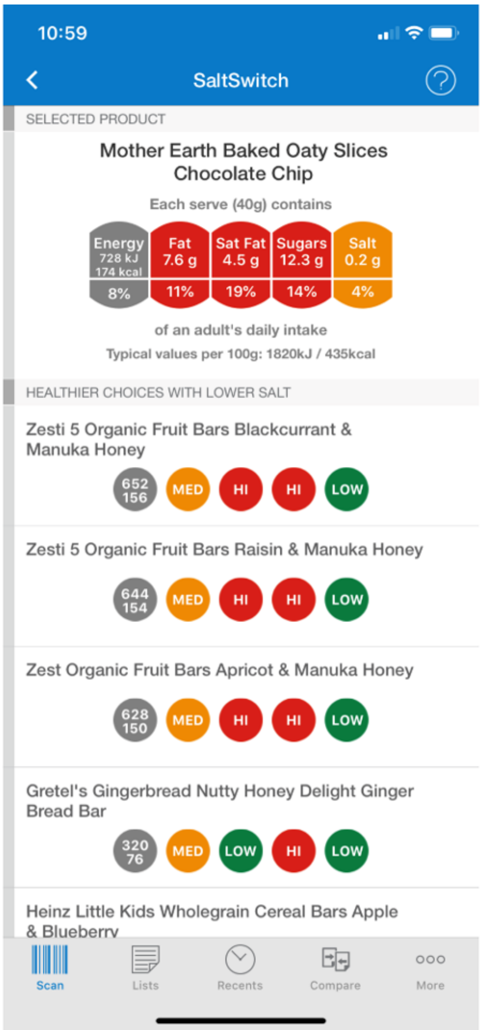
High blood pressure (BP) is the leading cause of premature and preventable death worldwide [1], mostly owing to its effect on cardiovascular disease (CVD). The relationship between high BP and sodium intake is widely recognized, with long-term reduction of dietary sodium resulting in a decrease in BP regardless of hypertension status, sex, ethnic group, or use of BP-lowering medication [2].

Even modest reductions in BP can have important impacts on population-level morbidity and mortality from CVD [2]. Therefore, in 2013, the World Health Organization (WHO) proposed a target for member states to achieve a 30% relative reduction in population sodium intake toward 2000 mg per day by 2025 [3], and at least 96 countries worldwide are working toward this target through a formal national sodium-reduction strategy [4]. In Aotearoa New Zealand (NZ), an ethnically diverse country of approximately 5 million, adults consume 40% more sodium than WHO recommendations (3373 mg per day) [5,6], and 1 in 5 adults has high BP [7]. Furthermore, high BP and cardiovascular conditions are unequally distributed, with populations traditionally underserved by the health care system, including those from lower-income groups, Māori (indigenous New Zealanders) whānau (families), and Pacific communities, having a higher burden [7]. Although NZ does not have a national sodium-reduction strategy, The Heart Foundation has been working with the food industry for more than a decade to remove sodium from low-cost, high-volume packaged foods [8]. In addition, the Ministry for Primary Industries launched the Health Star Rating front-of-pack nutrition label in 2014 to help consumers make healthier food choices and encourage reformulation [9]. Although some progress has been made from these voluntary initiatives, their impact on total population sodium intake is limited [10,11]. Therefore, effective, scalable, and equitable interventions are urgently needed.

There are 2 promising approaches: mobile health (mHealth) interventions and reduced-sodium salts (RSSs). There is a growing body of evidence suggesting that electronic health and mHealth interventions can support individual changes toward healthier diets [12]. However, there have been few robust randomized controlled trials (RCTs) of app-based interventions, particularly those related to dietary sodium reduction [13,14]. In 2014 and 2015, we conducted a 6-week pilot trial of the effects of the SaltSwitch smartphone app to support adults with diagnosed CVD to make lower-sodium food choices [15]. SaltSwitch (Figure 1) helps consumers choose packaged foods that are lower in sodium; these foods make up approximately 75% of the sodium consumed in NZ, with discretionary salt added during cooking and at the table contributing approximately 15% and the remainder being naturally present in fresh foods [16]. Intervention households in the pilot trial (n=33) purchased significantly less salt from packaged foods (mean 0.3, 95% CI 0.58-0.03 g/MJ) than control households (n=33), supporting a larger trial of the SaltSwitch app with longer-term follow-up (ACTRN12614000206628) [15].

However, SaltSwitch does not address discretionary salt. In contrast, RSSs, or salt substitutes, provide an alternative to regular table salt as some of the sodium chloride has been replaced by potassium salts or other minerals; they are lower in sodium and higher in potassium but have a similar mouthfeel, taste, and flavor. There is evidence from a 2022 Cochrane meta-analysis including 26 RCTs and 34,961 participants showing that the use of an RSS can reduce sodium chloride in the diet by 3% to 77% [16]. A subset of 12 RCTs in the review measured the effects of RSSs on 24-hour urinary sodium and potassium excretion, which ranged from −1730 to +460 mg per day and −170 to +720 mg per day, respectively. A total of 25 RCTs in the review measured BP, with 20 reporting data appropriate for meta-analysis; these studies found that RSSs can reduce systolic BP (SBP) by a mean difference of −4.76 (95% CI −6.01 to −3.5) mm Hg [17]. The wide range of effects was investigated in subgroup analyses, but there was low statistical power, and it was not possible to determine whether some types of RSS interventions were likely to be more effective than others or whether particular populations were most likely to benefit. Furthermore, none of the included studies were from countries where discretionary salt use contributed <25% to dietary sodium intake, such as in NZ.

Figure 1. The SaltSwitch smartphone app.



Objectives

The primary aim of the Salt Alternatives Study (SALTS) was to determine whether 12 weeks of intervention with a sodium-reduction package (SaltSwitch app + RSS) could reduce estimated 24-hour urinary sodium excretion in adults with high BP (ACTRN12619000352101; Universal Trial U1111-1225-4471).

Methods

Study Design

SALTS was a 2-arm parallel RCT conducted in NZ between May 2019 and February 2021. A 2-week baseline period was followed by a 12-week intervention period.

Ethics Approval

The trial protocol [18] was approved by the NZ Health and Disability Ethics Committees in February 2019 for a period of 3 years (18/NTB/239), and the trial was prospectively registered in the Australian New Zealand Clinical Trials Registry (ACTRN12619000352101).

Participants and Recruitment

Participants

Eligible participants were adults aged ≥18 years who owned a smartphone, had a seated SBP of ≥140 mm Hg or diastolic BP (DBP) of ≥85 mm Hg, planned to undertake household grocery shopping during the trial period, and could read and understand English. The exclusion criteria were SBP of >200 mm Hg; DBP of >120 mm Hg; using an RSS; using the SaltSwitch app; contraindication to altering sodium or potassium intake in the diet; taking furosemide, regular prednisone, or nonsteroidal anti-inflammatory drugs; having had a stroke or cardiovascular event in the previous 6 months; diagnosis of heart failure; planning on being away from home for ≥2 of the subsequent 14 weeks; or inability to provide informed consent. Participants were also excluded at the end of the baseline period if they did not return a spot (casual) urine sample and provide at least 6 home-based BP measures during the baseline period.

Recruitment

Participants were recruited from 2 large NZ cities: Auckland and Wellington. Recruitment settings were (1) face to face at community events such as night markets, outside pharmacies, in shopping malls, and via a mobile BP clinic run by the Stroke Foundation of NZ; (2) referrals from general practitioners (GPs) and pharmacists; (3) email invitations sent to staff at the University of Auckland; (4) six Facebook advertising campaigns; (5) two market research panels, Dynata and Horizon; and (6) HealthMatch, a clinical trial participant recruitment company. Specific engagement strategies were adopted to attract participants from Māori whānau and Pacific communities, including attendance at events with Ngāti Whātua Ōrākei (tangata whenua [indigenous people] of Tāmaki Makaurau or Auckland), hauora (well-being) health checks, local markets, and working directly with Pacific health organizations. All participants provided informed consent via the study smartphone app.

Randomization and Blinding

Eligible participants were randomly assigned in a 1:1 ratio to receive either the sodium-reduction intervention package (SaltSwitch smartphone app + RSS) or the control (generic heart-healthy eating information). Randomization was stratified by ethnicity (Māori and non-Māori) and age (<55 and ≥55 years) using permuted block randomization with variable block sizes of 2 or 4. Participants from Māori whānau and Pacific communities were not grouped for randomization as Pacific communities comprise a smaller proportion of the population and have a lower response rate [19] and Māori are the tangata whenua (original inhabitants) of Aotearoa NZ. The allocation sequence was generated by the study statistician (YJ) using computer-generated randomization lists and concealed in a secure database hosted on REDCap (Research Electronic Data Capture; Vanderbilt University) [20] until the point of randomization. Participants were assigned to trial groups by study research assistants using a REDCap software survey form. As the intervention required dietary change from participants and technology support from the study staff, it was not possible to blind participants or all study staff members to the allocation group. However, the lead study researchers (HE, RM, LTM, BN, AR, RND, and CNM) and trial statistician (YJ) were blinded until trial completion.

Intervention and Control

Intervention

Participants randomized to the intervention received a dietary sodium-reduction package including (1) access to the SaltSwitch smartphone app and (2) a supply of an RSS (as a salt substitute). To encourage the use of SaltSwitch and the RSS, intervention participants were sent weekly reminder notifications to their smartphones. Participants were advised to use the SaltSwitch app whenever they shopped for packaged food brought into the home and to use the RSS in all instances where they would usually use traditional table salt. However, no further dietary advice was provided.

The SaltSwitch app (Figure 1) enables users to scan the bar code of a packaged food using their smartphone camera and receive an immediate, interpretive traffic light nutrition label on-screen alongside a list of healthier, lower-salt options in the same food category. Users can also directly compare the salt content and healthiness of 2 or more foods and create a list of frequently scanned products. SaltSwitch was developed by the George Institute for Global Health [21] and adapted for NZ using the brand-specific Nutritrack (National Institute for Health Innovation) food composition database [22]. Nutritrack is updated annually via cross-sectional surveys of all packaged foods displaying a nutrition information panel sold at the 4 main supermarket chains in NZ (Countdown, New World, PAK'nSAVE, and Four Square) [22]. The Nutritrack database covers approximately 75% of all supermarket food purchases each year. The SaltSwitch food composition data were updated once during the trial. Once downloaded, the SaltSwitch app guided participants through a brief tutorial on how to use the app but did not provide any information on which products to scan. An older, out-of-date version of the SaltSwitch app was available in the NZ Apple and Android app stores during the trial as a component of the NZ FoodSwitch app [21].

The RSS (salt substitute) was manufactured by NuTek Food Science and was a blend of potassium and sea salt, which provided a 75% reduction in sodium compared with regular table salt (74.5% potassium chloride, 24.5% sodium chloride, and 1% silicon dioxide). Intervention participants were sent two 79-g containers of the RSS in plain packaging. The RSS provided to trial participants was not available for commercial sale in NZ during the trial. However, Mrs Rodgers Low Sodium Salt, comprising 49% sodium chloride and 46% potassium and magnesium chloride, was available for sale in some supermarkets.

Control

Participants randomized to the control group received a link to generic heart-healthy eating advice developed by the Heart Foundation of NZ sent to control participants' smartphones during week 1 of the 12-week intervention period. The generic advice was centered on a heart-healthy visual food guide showing the proportion of each type of food to eat each day. The web pages and links also included examples of food types such as grain foods and starchy vegetables, tips on how to achieve a heart-healthy eating pattern, how to read food labels, and how to cut back on salt.

Study Procedures

The Study Smartphone App

A customized study smartphone app was created to assist with the self-return of urine and BP measures and self-completion of questionnaires and support participants with their trial journey (Figures 2 and 3). The following features were included: consent, questionnaires, video tutorials, notification reminders for urine and BP collection, a barcode scanner for packaged foods, study contact information, and (posttrial) information about the intervention package.

Figure 2. The Salt Alternatives Study (SALTS) smartphone app part one.

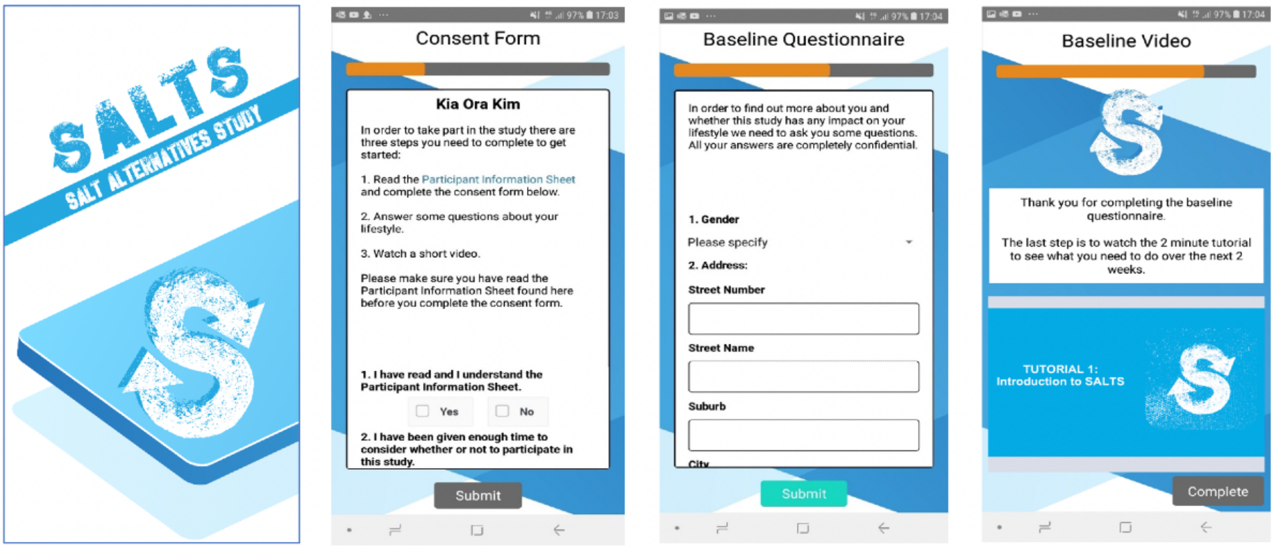
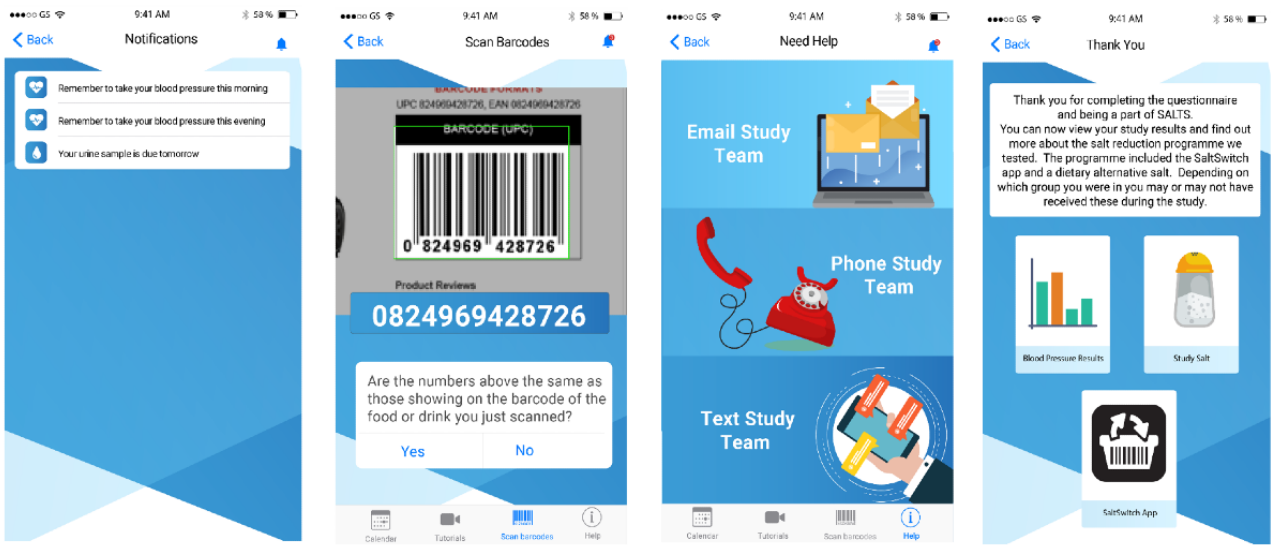


Figure 3. The Salt Alternatives Study (SALTS) smartphone app part two.



Referral, Screening, and Consent

Referrals were completed by study research assistants and health care providers using a web form [23] with fields for name, mobile number, email address, smartphone ownership, height, weight, and BP. Height was recorded to the nearest 1 cm, and weight was recorded to the nearest 100 g. Following 5 minutes of rest, the referrers took 3 BP measures on the left arm, and the average of the last 2 was automatically calculated. Individuals were advised to follow up their BP measurements with their GP if their measured SBP was ≥200 mm Hg or DBP was ≥85 mm Hg. Researchers used a standard stadiometer to measure height, a Salter electronic scale to measure weight, and an automated BP monitor [24] to measure BP. The equipment used by health care providers varied. Verbal consent was requested to enable the completed referral forms to be sent to the study researchers.

Early referrals did not attend any trial visits in person. However, from August 2019, referrals were offered a screening visit to assist with the use of the study technology. Screening and enrollment were completed by study research assistants via phone or in person using a web form [23]. Participants who met all screening criteria were sent an SMS text message with a link to download the study smartphone app (Figures 2, 3, and the following sections) and complete consent, after which they were provided with a Blipcare Wi-Fi–enabled BP monitor manufactured by Carematix Inc [24], equipment to collect and return 2 spot urine samples, and instructions for data collection. Phone support was also provided.

Baseline

The 2-week baseline period was designed to familiarize participants with trial technologies and collect baseline outcome data. The baseline questionnaire was hosted on REDCap [20] and included date of birth, address, ethnicity, qualifications, employment, household income, behavior regarding dietary salt (excluding total discretionary salt use), existing health conditions, number of household members sharing groceries, concurrent medications, and preferred times for BP measurement reminders. At baseline, participants were asked to scan the bar codes of all packaged foods purchased during the 2-week period, take BP measures in the morning and evening during the second week, and return a spot urine sample from any day during the second week. Potential participants who failed to return all baseline data by 2 weeks after enrollment received a follow-up support phone call; nonresponders 2 weeks after this call were considered lost to follow-up.

Follow-up

During the 12-week intervention period, participants were asked to scan the bar codes of all packaged food purchases (weeks 11 and 12), take BP measures in the morning and evening, collect and return a spot urine sample, and complete the follow-up questionnaire (all during week 12). The follow-up questionnaire was hosted on REDCap [20] and included all baseline questions in addition to questions related to the use of meal kits, recent cardiovascular or adverse events, and self-measured body weight. Intervention participants also answered questions about the use and acceptability of the intervention package and the amount of leftover RSS. All participants were provided with a summary of their BP measures, information on where to purchase an RSS, and access to the SaltSwitch smartphone app (removed 3 months after the last participant completed the trial) on trial completion.

Outcomes

Primary Outcome: Estimated 24-Hour Urinary Sodium Excretion

Urinary sodium excretion was measured as a proxy for sodium intake. To reduce participant burden, under- and overcollection, and a low response rate, urinary sodium excretion was estimated via a spot (casual) urine sample rather than measured using a gold-standard 24-hour urine collection [25]. Spot urine samples were collected at any time of day except the first void, chilled by participants, and frozen at −18 °C on receipt. Urine samples were thawed at room temperature, vortexed, and analyzed in batches. Urinary sodium and potassium levels were determined on a Roche Hitachi Cobas C311 unit biochemical analyzer using an ion-selective electrode. Urinary creatinine level was determined through Jaffe reaction using alkaline picrate (Roche Hitachi Cobas C311 analyzer). The concentration of sodium was converted to an estimated 24-hour sodium excretion using a standard urine volume of 1.99 L based on previously reported data for approximately 100 NZ adults [26].

Secondary Outcomes

Estimated 24-Hour Potassium Excretion

The 24-hour potassium excretion was estimated using the same methods as for the estimated 24-hour sodium excretion.

BP: SBP, DBP, and BP Control

BP was measured using a Blipcare Wi-Fi–enabled BP monitor programmed to automatically send readings back to study servers via an application programming interface [24]. Participants collected BP measurements in triplicate 1 minute apart on the left arm after 5 minutes of rest [27] in the morning and evening. Reminder notifications were sent to participants' smartphones, and if no measures were received, researchers followed up with a phone call and additional notifications. Participants who returned <6 BP measures at baseline were excluded. The definition for BP control (≤135/85 mm Hg) was lower than that used for referral purposes as the latter was taken at home and the former was taken in the community [28].

Sodium Content of Packaged Food Purchases

Bar codes for packaged foods purchased for home consumption were collected using a scanning feature in the study smartphone app (Figures 2 and 3). The sodium content of household food purchases was calculated by linking bar codes with Nutritrack [22], the brand-specific NZ food composition database used in the SaltSwitch app (see the *Intervention* section). Weekly reminder notifications were sent to the participants' smartphones, and if no measures were received, researchers followed up with a phone call and additional notifications.

Use and Acceptability of the Intervention Package

Data on the use and acceptability of the SaltSwitch smartphone app and the RSS were collected via the follow-up questionnaire. Bar codes scanned when using the SaltSwitch app were monitored using Google Analytics (for participants who had mobile data available). All intervention participants were asked to record how many teaspoons of RSS they had left at the end of the intervention period.

Safety and Adverse Events

Participants who reported abnormal BP measures after randomization were telephoned or sent an SMS text message advising them to visit their GP. Abnormal BP measures were defined as (1) consistently elevated SBP (>180 mm Hg for 3 consecutive days, including any missing days), (2) consistently low SBP (<90 mm Hg for 3 consecutive days, including any missing days), or (3) major changes in SBP from baseline (>20 mm Hg). The salt-reduction package was considered low risk. Therefore, only serious adverse events were collected via the follow-up questionnaire and reported to the Ethics Committee annually. A qualified medical representative was authorized to determine whether adverse events were considered serious.

Statistical Analysis

Sample Size

A total of 326 participants (163 per group) were estimated to provide 80% power at a 5% level of significance (2-sided) to detect a minimum effect size of 462 mg of sodium in the primary outcome between the 2 groups, allowing for a 10% loss to follow-up. The expected effect size was estimated from the SaltSwitch pilot study data, where estimations of 24-hour urinary sodium excretion were calculated using spot urine samples and a standard urine volume of 1.99 L with an SD of 1400 mg per day (ACTRN12614000206628) [15].

Main Comparative Analyses

All participant data collected at baseline and week 12 were summarized using descriptive statistics for the intervention and control groups separately. Continuous variables were presented as mean and SD, whereas categorical variables were reported as frequencies and percentages.

The trial evaluation was performed on an intention-to-treat basis, including all eligible participants in the group to which they were randomized. Multiple imputation methods were used for missing primary outcome data in the primary intention-to-treat analysis using the Markov chain Monte Carlo method and assuming that the data were missing at random. No imputation was considered on secondary outcomes. Sensitivity analysis was conducted on the primary outcome (1) without imputation and (2) using the International Cooperative Study on Salt and Blood Pressure formula [29] rather than a standard volume to estimate 24-hour sodium excretion. Linear regression was used for continuous outcomes adjusting for baseline outcome value, age, and ethnicity (stratification factors). The model-adjusted mean difference between the 2 groups was estimated with a 95% CI and *P* value. Logistic regression was used for categorical outcomes, and the estimated group difference was reported as the odds ratio. Owing to the small sample size, no subgroup analysis was considered.

The definition of valid data for spot (casual) urine samples was a collection at baseline during week −1 (−2 weeks to +2 weeks) and at follow-up during week 12 (−1 week to +1 week). Valid BP measurements at baseline were those taken during weeks −2 and −1 (−2 weeks to +2 weeks) and at follow-up during week 12 (−1 week to +1 week). The average SBP and DBP were calculated using a minimum of 6 readings at each time point. Valid bar code data to estimate the sodium content of household food purchases at baseline were scanned during week −1 (−2 weeks to +2 weeks) and at follow-up during week 12 (−1 week to +1 week). The average sodium content of food purchases was calculated for all bar codes received.

Statistical analyses were performed using SAS (version 9.4; SAS Institute). All statistical tests were 2-sided at a 5% significance level.

Changes in Response to the Challenges of the COVID-19 Pandemic

As recommended by Perlis et al [30], we outline the challenges associated with the COVID-19 pandemic and how these affected the SALTS trial. In NZ, there were strict lockdown periods from March 2020 to June 2020, from August 2020 to October 2020, in February 2021, and from August 2021 to November 2021. During these times, most postreferral and data collection procedures could be completed using remote technology. However, lockdown periods prevented enrolled participants from returning spot urine samples as couriers were only available for essential activities, and the university campus was closed, meaning that samples could not be received. Lockdown periods also substantially compromised recruitment as they prevented the collection of face-to-face BP measures necessary for new referrals. Consequently, recruitment was put on hold during these times. Furthermore, potential participants who had been referred and identified as eligible were unable to start the trial during lockdowns as it was not possible for researchers to courier the Wi-Fi–enabled BP monitor and equipment to collect urine samples; as a result, a considerable number of eligible participants lost interest and declined to take part, and recruitment was further compromised (see the *Recruitment* section).

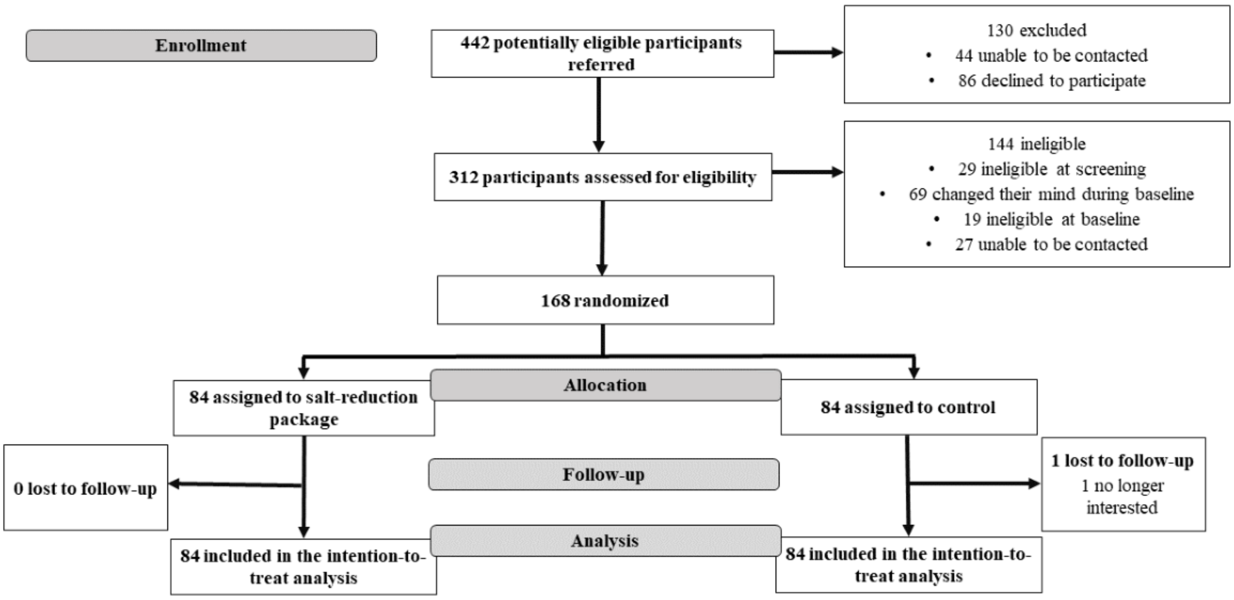
Results

Trial results are reported according to the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines for parallel-group randomized trials [31].

Recruitment

Recruitment took place over 16 months, starting on May 30, 2019, and finishing on October 2, 2019. The last participant was randomized in February 2020. A total of 442 potentially eligible participants were referred, of whom 312 (70.6%) were screened for initial eligibility, 86 (19.5%) declined to participate, and 44 (10%) were unable to be contacted (Figure 4). Of the 312 screened participants, 144 (46.2%) were ineligible as they changed their mind during baseline (n=69, 22.1%), did not meet the screening criteria (n=29, 9.3%), were unable to be contacted (n=27, 8.7%), or did not provide the required baseline data (n=19, 6.1%). The remaining 53.8% (168/312) of the initially screened eligible participants were randomized and took part in the trial; of those, the largest number was from market research panels (55/168, 32.7%) followed by face-to-face events (24/168, 14.3%). The final participant completed the trial on March 20, 2021 [31,32].

Figure 4. Trial profile.



Baseline Characteristics of Trial Participants

A total of 84 (50%) of the 168 participants were randomized to the intervention group, and the remaining 84 (50%) were randomized to the control group. A participant in the control group withdrew, stating that they were no longer interested in taking part. A total of 14.3% (24/168) of the participants identified as Māori, and 7.1% (12/168) identified as Pacific (7/84, 8% in the control group and 5/84, 6% in the intervention group). All participant characteristics were similar between groups (Table 1).

Table 1. Baseline characteristics of the trial participants (N=168).

	Control group (n=84)	Intervention group (n=84)
Baseline characteristics		
Age (years), mean (SD)	55 (13)	54 (13)
18 to 54, n (%)	38 (45)	37 (44)
≥55, n (%)	46 (55)	47 (56)
Gender, n (%)		
Men	41 (49)	36 (43)
Women	36 (43)	38 (45)
Nonbinary or not specified	7 (8)	10 (12)
Region, n (%)		
Auckland	71 (85)	76 (90)
Other New Zealand	13 (15)	8 (10)
Smartphone ownership, n (%)		
iPhone	32 (38)	41 (49)
Android	52 (62)	43 (51)
Prioritized ethnicity[a], n (%)		
Māori	12 (14)	12 (14)
Pacific	7 (8)	5 (6)
Asian	14 (17)	15 (18)
European or other	51 (61)	52 (62)
Highest qualification, n (%)		
None	5 (6)	1 (1)
Secondary	11 (13)	10 (12)
University degree, polytechnic, trade, or diploma	44 (52)	43 (51)
Postgraduate degree	21 (25)	27 (32)
Other	3 (4)	3 (4)
Employment status, n (%)		
Full- or part-time employment	57 (68)	58 (69)
Retired or full-time homemaker	17 (20)	14 (17)
Unemployed or student	9 (11)	12 (14)
Decline to answer	1 (1)	0 (0)
Annual household income (NZ $ [US $]), n (%)		
≤60,000 (US $37,826.7)	20 (24)	18 (21)
60,001 to 100,000 (US $37,827.33 to US $63,036.80)	21 (25)	21 (25)
≥100,001 (≥US $63,037.43)	30 (36)	34 (40)
Declined to answer	13 (15)	11 (13)
Behaviors regarding dietary salt, n (%)		
Add salt to food		
Always or often	35 (42)	35 (42)
Sometimes	23 (27)	19 (23)
Rarely or never	26 (31)	30 (36)
Salt added during cooking		
Always or often	52 (62)	54 (64)

	Control group (n=84)	Intervention group (n=84)
Sometimes	23 (27)	16 (19)
Rarely or never	9 (11)	14 (17)
Saltshaker placed on table		
Always or often	27 (32)	28 (33)
Sometimes	16 (19)	14 (17)
Rarely or never	40 (48)	42 (50)
Do not know	1 (1)	0 (0)
Trying to cut down the amount of salt consumed		
No	39 (46)	34 (40)
Yes	40 (48)	36 (43)
Do not know	5 (6)	14 (17)
Look at nutrition information on food packages		
Always or more often than not	29 (35)	25 (30)
Occasionally	38 (45)	50 (60)
Never	17 (20)	9 (11)
Baseline clinical measures		
Height (cm), mean (SD)	169 (10)	168 (9)
Weight (kg), mean (SD)	88 (19)	89 (18)
BMI (kg/m^2), mean (SD)	31 (6)	31 (5)
Blood pressure (mm Hg)[b], mean (SD)		
Systolic	142 (13)	143 (16)
Diastolic	86 (9)	86 (9)
Estimated 24-hour urine excretion (mg per day)[c], mean (SD)		
Sodium	3107 (1917)	3616 (2280)
Potassium	3601 (2202)	4232 (2659)
Current health condition[d], n (%)		
Diabetes	12 (14)	10 (12)
High cholesterol	26 (31)	29 (35)
High blood pressure	54 (64)	60 (71)

[a]Participants were allocated to a single ethnic group in the following order of priority even if they identified with more than one ethnicity: Māori, Pacific, Asian, and European or other.

[b]Valid blood pressure data only (ie, received within 1 week before and 2 weeks after randomization date and including a minimum of 6 readings).

[c]Valid urine data only (ie, received within 1 week before and 2 weeks after randomization date); n=73 for the control group and n=77 for the intervention group. Estimated from the concentration of 1 spot urine sample and a standard volume of 1.99 L, with no adjustment for electrolytes not excreted via urine.

[d]As advised by a health professional.

Return of Trial Outcome Data

Valid urine samples for the estimation of sodium and potassium excretion were returned by 89.3% (150/168) of the participants at baseline and 45.2% (76/168) of the participants at follow-up. More participants in the control group compared with the intervention group returned a valid urine sample at baseline (77/84, 92% vs 73/84, 87%, respectively; Table 2). Valid BP data were returned by 93.5% (157/168) of the participants at baseline and 83.3% (140/168) of the participants at follow-up. Valid bar code data for the estimation of the sodium content of household food purchases were returned by 76.2% (128/168) of the participants at baseline and 22% (37/168) of the participants at follow-up. The baseline questionnaire was completed by 100% (168/168) of the participants, and the follow-up questionnaire was completed by 76.2% (128/168) of the participants. The rate of return of follow-up data was consistent across ethnic groups except for the follow-up questionnaire, which was returned by 67% (24/36) of the participants identifying as Māori or Pacific and 78.8% (104/132) of the participants identifying as all other ethnicities.

Table 2. Estimates of the effect of the salt-reduction intervention package on urinary sodium and potassium excretion, blood pressure, and the sodium content of household packaged food purchases at 12 weeks (N=168).

	Control group (n=84)				Intervention group (n=84)				Adjusted difference at 12 weeks[a] (95% CI)
	Baseline		12 weeks		Baseline		12 weeks		
	Participants with valid data, n (%)[b]	Mean (SD)	Participants with valid data, n (%)[b]	Mean (SD)	Participants with valid data, n (%)[b]	Mean (SD)	Participants with valid data, n (%)[b]	Mean (SD)	
Primary outcome: estimated 24-hour sodium excretion (mg per day)									
Standard volume	73 (87)	3107 (1917)	38 (45)	3193 (2284)	77 (92)	3616 (2280)	38 (45)	3935 (2268)	
Multiple imputations (primary)[c]									547 (−331 to 1424)
No imputation[d]									670 (−304 to 1645)
INTERSALT[e] formula	65 (77)	3324 (813)	35 (42)	3222 (857)	66 (79)	3320 (782)	32 (38)	3289 (792)	
No imputation[f]									24 (−254 to 302)
Secondary outcomes									
Estimated 24-hour potassium excretion (mg per day)	73 (87)	3601 (2202)	38 (45)	4078 (2945)	77 (92)	4232 (2659)	38 (45)	4210 (2334)	132 (−1083 to 1347)
Systolic blood pressure (mm Hg)[g]	78 (93)	142 (13)	60 (71)	138 (15)	79 (94)	143 (15)	60 (71)	139 (12)	−0.66 (−3.48 to 2.16)
Diastolic blood pressure (mm Hg)[g]	78 (93)	86 (9)	60 (71)	84 (10)	79 (94)	86 (9)	60 (71)	84 (8)	−0.35 (−2.20 to 1.50)
Sodium content of household food purchases (mg per 100 g)[h]	67 (80)	346 (427)	20 (24)	253 (326)	61 (73)	350 (536)	17 (20)	262 (151)	73 (−21 to 168)

[a]Linear regression models adjusted for baseline outcome, age in years, and Māori or Pacific ethnicity.

[b]Valid urine, blood pressure (BP), and food purchasing data were collected within 1 week before or 2 weeks after randomization (for baseline) and 1 week before or 2 weeks after week 12 (for follow-up). For BP, a minimum of 6 readings during these time frames was considered valid. For food purchases, a minimum of 10 products scanned during these time frames was considered valid.

[c]Estimated 24-hour sodium excretion from spot urine using a standard volume of 1.99 L. Multiple imputations used on missing primary outcome data through an intention-to-treat analysis using the Markov chain Monte Carlo method and assuming data were missing at random.

[d]24-hour sodium excretion from spot urine using a standard volume of 1.99 L. No imputation for missing data.

[e]INTERSALT: International Cooperative Study on Salt and Blood Pressure.

[f]24-hour sodium excretion estimated using the INTERSALT formula. No imputation for missing data.

[g]BP control was defined as <135/85 mm Hg. The mean number of valid days for all BP measures at baseline was 12 (SD 6) in the control group and 13 (SD 7) in the intervention group. The corresponding values at week 12 were 8 (SD 3) and 9 (SD 5), respectively. The mean number of valid readings for all BP measures at baseline was 31 (SD 21) in the control group and 37 (SD 24) in the intervention group. The corresponding values at week 12 were 19 (SD 12) and 24 (SD 16), respectively.

[h]The mean number of food products scanned at baseline was 27 (SD 18) in the control group and 24 (SD 18) in the intervention group. The corresponding values at week 12 were 18 (SD 16) and 13 (SD 12), respectively.

Primary Outcome: Estimated 24-Hour Urinary Sodium Excretion

The mean estimated 24-hour urinary sodium excretion at 12 weeks was 3935 (SD 2268) mg per day in the intervention group and 3193 (SD 2284) mg per day in the control group (Table 2). There was no significant difference between the groups in estimated 24-hour sodium excretion at 12 weeks (adjusted mean difference=547 mg per day, 95% CI −331 to 1424; Table 2).

Sensitivity analyses were consistent with the primary analysis, with no significant differences observed in the mean difference between groups where no imputation was used or where 24-hour urinary sodium excretion was estimated using the International Cooperative Study on Salt and Blood Pressure formula [29] rather than a standard volume of 1.99 L (Table 2).

Secondary Outcomes

Estimated 24-Hour Urinary Potassium Excretion

The mean estimated 24-hour urinary potassium excretion at 12 weeks was 4210 (SD 2334) mg per day in the intervention group and 4078 (SD 2945) mg per day in the control group (Table 2). There was no significant difference between the groups in estimated 24-hour potassium excretion at 12 weeks (adjusted mean difference=132 mg per day, 95% CI −1083 to 1347; Table 2).

BP: SBP, DBP, and BP Control

The mean SBP for the intervention and control groups at 12 weeks was 139 (SD 12) mm Hg and 138 (SD 15) mm Hg, respectively (Table 2). The corresponding figures for DBP were 84 (SD 8) mm Hg and 84 (SD 10) mm Hg, respectively. No significant difference was observed between the groups for SBP or DBP at 12 weeks. The adjusted mean difference between groups was −0.7 (95% CI −3.5 to 2.2) mm Hg for SBP and −0.4 (95% CI −2.2 to 1.5) mm Hg for DBP (Table 2). The mean number of participants achieving BP control at 12 weeks in the intervention and control groups was 23 (SD 27) and 17 (SD 20), respectively. There was no significant difference between the groups in the odds of achieving BP control (adjusted odds ratio 1.0, 95% CI 0.45-2.1).

Sodium Content of Packaged Food Purchases

The mean number of all bar codes scanned at baseline was 24 (SD 18) for the intervention group and 27 (SD 18) for the control group. The corresponding mean values for the follow-up period were 13 (SD 12) and 18 (SD 16), respectively. There was no significant difference between the groups in the sodium content of packaged foods purchased (adjusted mean difference=73, 95% CI −21 to 168 mg per 100 g; Table 2).

Use and Acceptability of the Intervention Package

A total of 76% (64/84) of the intervention participants provided use and acceptability data; of these 64 participants, 48 (75%) reported using SaltSwitch when shopping, with 25 (52%) of them reporting that they used the app “at least half to every time” they shopped (Table 3). Google Analytics data were available for 96% (46/48) of the SaltSwitch users, who scanned a mean of 29 (SD 40) products during the 12-week intervention period over a mean of 6 (SD 6) shopping occasions. The most common responses from the 56% (27/48) of participants who reported what they “liked most” or “least” about SaltSwitch were that it helped with making lower-salt food choices and thinking about salt in food in general (5/27, 19%) but needed more products to be available in the app to scan (10/27, 37%).

A total of 94% (60/64) of the intervention participants who provided data used the RSS, with 69% (44/64) stating that between half and all the discretionary salt they consumed during the 12-week study period was the RSS. Of those who reported using less than half or none of the RSS (20/64, 31%), 20% (4/20) stated that this was because the taste was unacceptable (Table 3). Participants used a mean of approximately 37.2 g (6.5 tsp) of RSS over the 12-week intervention period, and 44% (28/64) stated that their study salt was consumed by other household members.

Table 3. Use and acceptability of the salt-reduction intervention package (n=64[a]).

	Values
SaltSwitch smartphone app, n (%)	
Used the SaltSwitch app when grocery shopping over the past 12 weeks (n=64)	48 (75)
How often used? (n=48)	
More than half to every time	15 (31)
Half of the time	10 (21)
A handful of times to less than half of the time	22 (46)
Did not answer	1 (2)
How easy to use? (n=48)	
Very easy to somewhat easy	33 (69)
Neither easy nor difficult	9 (19)
Somewhat difficult to very difficult	5 (10)
Did not answer	1 (2)
Think SaltSwitch is a good way to help shoppers make lower-salt food choices (n=64)	52 (81)
Reduced-sodium salt (study salt)	
Amount of salt consumed over the past 12 weeks that was study salt (n=64), n (%)	
All or nearly all	35 (55)
Half	9 (14)
Less than half	16 (25)
None	4 (6)
If less than half or none, what was the main reason for this? (n=20), n (%)	
Taste unacceptable	4 (20)
Unwilling or other reason	16 (80)
Teaspoons of salt left at end of study (n=56), mean (SD)[b]	19.8 (19.9)
Other household members used the study salt (n=64), n (%)	28 (44)
How many household members used the study salt? (n=28), n (%)	
1 to 2	17 (61)
3	4 (14)
≥4	7 (25)

[a]A total of 64 intervention participants returned the follow-up questionnaire.

[b]Intervention participants were provided with 158 g or approximately 26.3 tsp of salt.

Effects for Māori and Pacific Participants

Owing to the low engagement and recruitment of participants from Māori whānau (28/168, 16.7%) and Pacific communities (13/168, 7.7%), it was not possible to estimate differences in effects separately for these groups.

Adverse Events

No serious adverse events were reported during the trial period.

Challenges Associated With the Use of Trial Technology

Technology was used in the SALTS trial to streamline study processes, deliver the intervention, collect outcome data, and communicate with participants. Information on the challenges encountered owing to the use of technology, which affected all 4 stages of the SALTS trial and CONSORT flow diagram, [31] is summarized in Table 4 and has been reported elsewhere [32]. However, briefly, during enrollment and allocation of participants (stages 1 and 2), many participants lacked confidence in their ability to download the study smartphone app and connect the Wi-Fi–enabled BP monitor. During this stage, inefficiencies were also experienced by the researchers as the study data management system could not directly exchange information with the referral form and participant tracking systems. During stage 3 (follow-up and collection of outcome data), the study app performed inconsistently across different smartphone models and operating systems, and some participants did not switch on their phone notifications, meaning that they missed important study reminders. Finally, during stage 4 (data analysis), some participants did not complete the follow-up questionnaire or return BP measures within the time frames for valid data as prespecified in the study protocol [31].

Table 4. Challenges associated with the use of technology and future recommendations.

Trial stage	Technology challenge	Future recommendations
Enrollment and allocation of participants	• Not all smartphone owners use smartphone apps, and use may be lower in older populations. • Face-to-face support may be required for confident connection and use of technologies such as smartphone apps and other Wi-Fi–enabled devices. • Interoperability, or the exchange of information between technologies, is critical to harness the efficiencies they offer.	• Complete background research on the population of interest to understand their use of smartphone technology before using it widely in a research study. • Plan for flexibility in the study design to enable face-to-face support for familiarization with study technology, particularly during the early phases. • Incorporate funds and time in the study setup phase to ensure that technologies that need to exchange information with one another can do so correctly and efficiently. For example, ensure that web-based forms exchange data with data management systems and data management systems exchange data with participant booking systems. If funds and time cannot be included, consider the use of simple existing tools such as survey software and an ad hoc SMS text messaging service.
Following up participants and collecting outcome data	• Technology can behave in unanticipated ways in response to the variety of smartphone models and operating systems on the market, and it can be difficult to replicate “live” trial conditions for all individual circumstances. • Not all smartphone users like or read notifications.	• Create technology test plans and implement them during all phases of the trial, from early development to the completion of the last participant. When testing technology, use Apple and Android phones and include different operating systems. Have a “soft” launch to enable rigorous early testing with a small group of real participants. To avoid the impacts of software fixes on unrelated functionality, build technology in separate blocks of code that only connect where necessary. • Where possible, use SMS text messages rather than notifications to convey key study information to participants, particularly for those with limited Wi-Fi or data.
Data analysis	• The flexibility that technology provides to return outcome data at the participants’ convenience can increase the time frame for data return and the variability in measures.	• Set realistic time frames or windows for participants to return remote data to researchers. For example, for participants returning a casual urine sample by courier, a realistic number of days will be needed to provide the participant with options, they may need a reminder messages, or there could be courier delays. Set time frames for each outcome that is collected remotely and specify these before study start in the statistical analysis plan. • In addition to using standardized methods for the collection of clinical outcome data, consider whether other aspects of outcome data collection should also be standardized. For example, blood pressure measures vary considerably between and within individuals and from day to day and even hour to hour; in this case, standardizing the time for data collection (eg, 8 AM), rather than allowing participants to choose a time in the morning that suits them, will result in reduced variation in blood pressure measures across the sample.

Discussion

Principal Findings

In this RCT, we found no evidence that 12 weeks of intervention with a salt-reduction package reduced estimated 24-hour urinary sodium excretion in adults with high BP. The estimated mean sodium excretion was higher in the intervention group than in the control group at 12 weeks; however, the CI for the mean difference was wide, suggesting no real difference. In addition, we found no effect of the intervention package on any secondary outcome, including estimated 24-hour urinary potassium excretion and BP. Although most intervention participants reported using the SaltSwitch app (48/64, 75%) and the RSS (60/64, 94%) during the 12-week intervention period and acceptability of the intervention was high, the intervention dose was low; participants reported using SaltSwitch on less than half of shopping occasions and consuming only approximately 1/2 tsp of RSS per household per week during the intervention period. The low recruitment of participants from Māori whānau and Pacific communities meant that it was not possible to estimate differences in effects separately for these groups.

Limitations

In addition to the low intervention dose, important limitations of the SALTS trial were the reduced study power, low number of participants from Māori whānau and Pacific communities, lower-than-anticipated engagement with trial technologies, and use of a spot rather than 24-hour urine sample for estimation of the primary outcome. Implementation and COVID-19 challenges meant that the trial was substantially underpowered, and thus, it is possible that a real effect may have been missed. Although a larger trial would have enabled the study hypothesis to be tested as intended, it is possible that the null trial findings would have been similar because of the limited use of intervention components. Given the high acceptability, the reason for the low consumption of the RSS by trial participants is not clear, but the low use of discretionary salt at baseline by trial participants may have been a factor. The ease of use and acceptability of SaltSwitch were also high, but reported use was low, with the reasons most reported by participants for not using the intervention app more often being the use of web-based grocery shopping instead, others completing the grocery shopping, difficulty in downloading or using SaltSwitch, and lack of time.

Despite adopting specific recruitment strategies to attract and engage Māori whānau and Pacific communities, we randomized only 14.3% (24/168) and 7.1% (12/168) of the total participants from these groups, respectively. Although we were able to attend numerous face-to-face events with Māori communities and work directly with Pacific health organizations, a capacity-building approach led by Māori for Māori and by Pacific for Pacific where these groups take an active part in the research would likely have been more effective [33].

Challenges associated with the considerable use of technology in the trial also affected recruitment, engagement, and return of outcome data. However, it is difficult to know the magnitude of the effect related to these challenges as technology provides certain inherent efficiencies and enabled researchers to continue some aspects of the trial during COVID-19 lockdowns.

The use of spot urine rather than gold-standard 24-hour urine samples may have affected the ability to identify a difference in sodium intake between the intervention and control groups [34]. Although 24-hour urine samples are considered too burdensome for nonclinical study populations, future similar research would benefit from a subsample of participants providing a 24-hour urine sample, which would provide some information on the consistency of effects.

Caution should also be exercised when generalizing the findings of the SALTS trial to other population groups. In addition to the low number of participants from Māori whānau and Pacific communities, the study sample was highly educated; many were already trying to cut down on the amount of salt they consumed and using nutrition information on food packages to make healthier choices. Furthermore, 1 in 3 referrals declined to take part or were unable to be contacted, and the inclusion-to-randomization rate (calculated from the 59 referrals who were initially eligible and completed the baseline questionnaire but were not randomized) was lowest for those from Māori whānau (24/52, 46%) and Pacific communities (12/17, 71%; conversion rates for Asian and European or other were 29/36, 81% and 103/132, 78%, respectively). Possible reasons for why almost half of Māori referrals (24/52, 46%) and approximately 40% of Pacific referrals (12/17, 71%) did not continue in the study include the collective cultures of these groups misaligning with the individual framing of the intervention and the lack of face-to-face contact [32]. Although smartphone apps offer benefits, including lower scale-up costs, personalized health information, real-time delivery of advice, and remote assessment of outcomes, there is evidence indicating that face-to-face relationships are a key component for achieving social connection and digital inclusion, both of which are vital to reducing inequities [35,36].

Strengths

Nonetheless, the SALTS trial is one of the few RCTs of a smartphone app to promote dietary sodium reduction in adults [13]. To date, most mHealth interventions aimed at reducing sodium consumption have focused on improving knowledge and awareness of dietary salt intake using SMS text messaging. The use of more innovative technologies and rigorously designed trials were identified as key recommendations for future research in a 2019 systematic review [14]. Furthermore, SALTS is one of a limited number of trials testing an RSS in a country with a predominantly Western diet where packaged foods contribute most (>50%) of sodium to dietary intake [37].

Comparison With Prior Work

The null findings of the SALTS trial are inconsistent with those of 2 recent systematic reviews of mHealth interventions that specifically target dietary salt reduction. The first was a 2019 systematic review of the effectiveness of mHealth technologies for salt reduction and included 6 RCTs and 5 quasi-experimental studies; 8 of 11 studies produced positive results [14], and 2 of the RCTs stated salt consumption as the primary outcome (one of which was a pilot of the SaltSwitch app [15]), finding significant reductions in intake as estimated by a spot (casual) urine sample. However, both trials were small (ie, <100 participants). The results of a more recent (2020) RCT of the "LowSalt4Life" just-in-time adaptive mobile app for adults with hypertension were also positive, with a significant reduction in estimated 24-hour urinary sodium excretion compared with usual dietary advice [38]. A 2019 systematic review examining the effectiveness of electronic health interventions for BP control also found a significant overall reduction in sodium intake (n=15 trials) [13].

The reasons why the SaltSwitch app was ineffective compared with previous studies are difficult to determine as the intervention components associated with effectiveness were not investigated. However, in contrast to the SALTS intervention, most previous studies were co-designed or included at least one of the following behavior change techniques: system-generated feedback based on current behaviors, goal setting, regular motivational SMS text messages, or face-to-face support from a health care provider in addition to interaction through an electronic device [13]. A recent (2020) systematic review of mHealth RCTs supports the inclusion of behavior change techniques, with prompts or cues, general personalization, goal setting, and action planning found to be significantly associated with positive change [39]. The SaltSwitch app included only one of these techniques (prompts or cues to use the app), which may help explain its lack of efficacy. The feasibility of an app called SaltSwap combined with a brief behavioral intervention was recently explored in adults with high BP in the United Kingdom. Although researchers found no evidence that the intervention reduced dietary salt intake, findings of a future adequately powered trial will provide further information on the effectiveness of mHealth interventions based on behavioral theory [40]. The contrasting positive findings of the SALTS pilot trial also suggest that individual-level behavior change interventions may be more beneficial for highly motivated clinical populations [15,41]. Approaches likely to be more successful at the population level include those outlined by the WHO in their Surveillance, Harness Industry, Adopt Standards for Labelling and Marketing, Knowledge, Environment technical package for salt reduction (ie, regular measuring and monitoring of population salt use, reformulation of foods and meals to contain less salt, effective food labeling and marketing, education campaigns, and supporting settings such as hospitals and universities to promote healthy eating [42]).

The null findings of the SALTS trial are also inconsistent with evidence from a 2022 Cochrane meta-analysis, which showed that the use of an RSS can reduce urinary sodium excretion by up to 1730 mg per day [17]. However, none of the included studies were from countries such as NZ, where discretionary salt use contributes <25% to dietary sodium intake, and there were insufficient data in the review to determine whether the type of RSS or study population affected effectiveness. The limited use of the RSS by participants in the SALTS trial suggests that, in Aotearoa NZ, it may be more efficacious to focus on the use of RSSs in packaged foods rather than or as an adjunct to a replacement for traditional table salt. That said, RSSs may still be helpful for specific communities or in settings where most food is cooked or prepared in the home or on-site. In settings where RSSs are found to be effective, political actions to support implementation include understanding the path to market and removing cost and accessibility barriers for consumers and food companies [43].

Conclusions

In summary, our trial found no evidence of the effectiveness of a salt-reduction intervention package comprising a smartphone app and RSS on estimated 24-hour sodium excretion (or any secondary outcome assessed) in adults with high BP. The trial was underpowered because of challenges associated with the implementation of trial technologies and the COVID-19 pandemic, meaning that it is possible that a real effect may have been missed. Furthermore, because of low engagement and recruitment, it was not possible to determine potential effects for Māori whānau or Pacific communities. However, it is also possible that a larger trial in the same study population would produce similar results given the low intervention dose. Nonetheless, further research may be warranted to explore the efficacy of SaltSwitch for secondary prevention in highly motivated clinical populations such as those who have had a recent cardiac event. Further research should also be undertaken to explore the use of RSSs in packaged foods, especially for countries such as NZ, where these foods contribute >50% to population sodium intake, and in specific communities or in settings where most food is cooked in the home or prepared on-site. Finally, the challenges associated with the design and delivery of effective individual-level behavioral interventions highlight the need for comprehensive policies and programs, including improvements to food environments and systems, in addition to supportive tools for behavior change—this is critical if we are to achieve meaningful reductions in population sodium intake.

Acknowledgments

The authors are indebted to their trial participants. They would also like to thank their study project coordinator (Diane Wood) and research assistants (Neela Bhana, Bruce Kidd, and Shistata Shrestha); Nhung Nghiem, who helped with the development of an early plan for economic analysis; Tahu-Pōtiki Te Maro-Doran, who assisted with recruitment in Wellington; Green Cross Health and The Stroke Foundation of New Zealand for support with recruitment; student and staff volunteers who helped with recruitment during the Stroke Foundation "Big Blood Pressure Check" and at Ngāti Whaāua events (Shai Alsaie, Daria Faulconbridge, Bruce Kidd, Xinye Li, Jenny Lin, Magda Rosin, Huang Silver, Summer Wright, and Lina Yousif); Koda Web Design who developed the Salt Alternatives Study smartphone app and database; staff at the National Institute for Health Innovation involved in data management, software, budgets, and contracts (Manoj Alwis, Karen Carter, Sarah Douglas, John Faatui, Michelle Jenkins, Pragya Nandan, Deepak Pandey, Mahfuz Rahman, and Peter Tsai); and Gina Gastrillon and Fraser Taylor, who facilitated the development of the New Zealand version of the SaltSwitch smartphone app. This research was also supported (in part) by the provision of Salt for Life 100 Dietary Alternative Salt free of charge by NuTek. Funding was provided by the Health Research Council of New Zealand program grant (#18/672). HE is funded by a Heart Foundation of New Zealand Senior Fellowship grant (#1843). RND is supported by the Heart Foundation Chair of Heart Health. The funder had no role in the study design; collection, analysis, or interpretation of the data; writing of the report; or decision to submit the manuscript for publication.

Authors' Contributions

HE, BN, and CNM contributed to conceptualization. YJ, JG, EU, and RM contributed to data curation. YJ contributed to formal analysis. HE and CNM acquired funding. HE, JG, EU, and LTM contributed to investigation. HE, JG, EU, RM, LTM, BN, AR, RND, and CNM contributed to methodology. JG and EU were responsible for project administration. JG and EU contributed to resources. HE, BN, AR, RND, and CNM supervised the study. YJ and RM validated the data. HE and YJ contributed to visualization. HE wrote the original draft. HE, JG, UM, RM, LTM, BN, AR, RND, and CNM reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) questionnaire V1.6.1.

[PDF File (Adobe PDF File), 60136 KB - mhealth_v11i1e43675_app1.pdf]

References

1. GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. . [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
2. He FJ, Tan M, Ma Y, MacGregor GA. Salt reduction to prevent hypertension and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 2020 Feb 18;75(6):632-647 [FREE Full text] [doi: 10.1016/j.jacc.2019.11.055] [Medline: 32057379]
3. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization. Geneva, Switzerland: World Health Organization; 2013 Nov 14. URL: https://www.who.int/publications/i/item/9789241506236 [accessed 2023-02-09]
4. Santos JA, Tekle D, Rosewarne E, Flexner N, Cobb L, Al-Jawaldeh A, et al. A systematic review of salt reduction initiatives around the world: a midterm evaluation of progress towards the 2025 global non-communicable diseases salt reduction target. Adv Nutr 2021 Oct 01;12(5):1768-1780 [FREE Full text] [doi: 10.1093/advances/nmab008] [Medline: 33693460]
5. Skeaff SA, McLean R, Mann J, Williams S. The impact of mandatory fortification of bread with iodine. Ministry of Primary Industries. Wellington, New Zealand: University of Otago; 2013 Aug. URL: https://www.mpi.govt.nz/dmsdocument/4097-the-impact-of-mandatory-fortification-of-bread-with-iodine [accessed 2023-02-09]
6. Guideline: sodium intake for adults and children. World Health Organization. 2012 Dec 25. URL: https://www.who.int/publications/i/item/9789241504836 [accessed 2013-02-18]
7. Annual Update of Key Results 2019/20: New Zealand Health Survey. Ministry of Health. Wellington, New Zealand: Ministry of Health; 2020. URL: https://www.health.govt.nz/publication/annual-update-key-results-2019-20-new-zealand-health-survey [accessed 2021-05-31]
8. Food reformulation programme. Heart Foundation of New Zealand. 2022. URL: https://assets.heartfoundation.org.nz/documents/food-industry/food-reformulation/food-reformation-programme.pdf?mtime=1667526614?1670205225 [accessed 2022-12-05]
9. How Health Star Ratings work. Ministry for Primary Industries. 2021. URL: https://www.mpi.govt.nz/food-business/labelling-composition-food-drinks/health-star-ratings-food-labelling/how-health-star-ratings-work/ [accessed 2022-08-04]
10. Wang NX, Skeaff S, Cameron C, Fleming E, McLean RM. Sodium in the New Zealand diet: proposed voluntary food reformulation targets will not meet the WHO goal of a 30% reduction in total sodium intake. Eur J Nutr 2022 Sep;61(6):3067-3076 [FREE Full text] [doi: 10.1007/s00394-022-02864-5] [Medline: 35353200]
11. Bablani L, Ni Mhurchu C, Neal B, Skeels CL, Staub KE, Blakely T. Effect of voluntary Health Star Rating labels on healthier food purchasing in New Zealand: longitudinal evidence using representative household purchase data. BMJ Nutr Prev Health 2022 Dec;5(2):227-234 [FREE Full text] [doi: 10.1136/bmjnph-2022-000459] [Medline: 36619324]
12. de David CN, Ziegelmann PK, Goveia P, Silvani J, Silveira LR, Fuchs SC. The effect of mobile health focused on diet and lifestyle on blood pressure: a systematic review and meta-analysis. Eur J Prev Cardiol 2022 May 25;29(7):1142-1155. [doi: 10.1093/eurjpc/zwac017] [Medline: 35078238]
13. Ma Y, Cheng HY, Cheng L, Sit JW. The effectiveness of electronic health interventions on blood pressure control, self-care behavioural outcomes and psychosocial well-being in patients with hypertension: a systematic review and meta-analysis. Int J Nurs Stud 2019 Apr;92:27-46. [doi: 10.1016/j.ijnurstu.2018.11.007] [Medline: 30690164]
14. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of mobile health technologies aimed at salt reduction: systematic review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250 [FREE Full text] [doi: 10.2196/13250] [Medline: 30994467]
15. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep;24(13):1435-1444 [FREE Full text] [doi: 10.1177/2047487317715713] [Medline: 28631933]
16. Bhat S, Marklund M, Henry ME, Appel LJ, Croft KD, Neal B, et al. A systematic review of the sources of dietary salt around the world. Adv Nutr 2020 May 01;11(3):677-686 [FREE Full text] [doi: 10.1093/advances/nmz134] [Medline: 31904809]
17. Brand A, Visser ME, Schoonees A, Naude CE. Replacing salt with low-sodium salt substitutes (LSSS) for cardiovascular health in adults, children and pregnant women. Cochrane Database Syst Rev 2022 Aug 10;8(8):CD015207 [FREE Full text] [doi: 10.1002/14651858.CD015207] [Medline: 35944931]
18. Eyles H. SALTS: Salt ALTernatives Study: A randomised controlled trial to determine the effects of a 12-week dietary salt reduction intervention (SaltSwitch app + dietary salt substitute) on urinary sodium excretion in adults with high blood pressure. National Institute for Health Innovation, The University of Auckland. Auckland, New Zealand: The University of Auckland; 2020. URL: https://diet.auckland.ac.nz/sites/default/files/2022-03/SALTS%20Protocol_v4.0.pdf [accessed 2022-03-29]

19. Ethnic group summaries reveal New Zealand's multicultural make-up. Statistics New Zealand. Wellington, New Zealand: Statistics New Zealand; 2020. URL: https://www.stats.govt.nz/news/ethnic-group-summaries-reveal-new-zealands-multicultural-make-up/ [accessed 2022-12-07]
20. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
21. FoodSwitch. The George Institute for Global Health. 2022. URL: https://www.georgeinstitute.org/projects/foodswitch [accessed 2022-08-10]
22. The Nutritrack Database. The National Institute for Health Innovation, Auckland UniServices Limited. 2022. URL: https://diet.auckland.ac.nz/sites/default/files/2021-10/Nutritrack_print_v3_Sept%2021.pdf [accessed 2022-03-10]
23. Drupal 8 User Guide. Drupal. 2000. URL: https://www.drupal.org/node/3058590 [accessed 2022-12-13]
24. WiFi Blood Pressure Monitor. Blipcare. 2017. URL: http://www.blipcare.com/wifi-bp.html [accessed 2022-03-15]
25. Cogswell ME, Maalouf J, Elliott P, Loria CM, Patel S, Bowman BA. Use of urine biomarkers to assess sodium intake: challenges and opportunities. Annu Rev Nutr 2015;35:349-387 [FREE Full text] [doi: 10.1146/annurev-nutr-071714-034322] [Medline: 25974702]
26. McLean R, Williams S, Mann J. Monitoring population sodium intake using spot urine samples: validation in a New Zealand population. J Hum Hypertens 2014 Nov;28(11):657-662. [doi: 10.1038/jhh.2014.10] [Medline: 24573130]
27. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, Authors/Task Force Members:. 2018 ESC/ESH guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. J Hypertens 2018 Oct;36(10):1953-2041 [FREE Full text] [doi: 10.1097/HJH.0000000000001940] [Medline: 30234752]
28. Sharman JE, Howes FS, Head GA, McGrath BP, Stowasser M, Schlaich M, et al. Home blood pressure monitoring: Australian expert consensus statement. J Hypertens 2015 Sep;33(9):1721-1728 [FREE Full text] [doi: 10.1097/HJH.0000000000000673] [Medline: 26136205]
29. Intersalt Cooperative Research Group. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. BMJ 1988 Jul 30;297(6644):319-328 [FREE Full text] [doi: 10.1136/bmj.297.6644.319] [Medline: 3416162]
30. Perlis RH, Haneuse SJ, Rubenfeld GD, Fihn SD, Rivara FP. Reporting clinical studies affected by the COVID-19 pandemic: guidelines for authors. JAMA Netw Open 2021 Jan 04;4(1):e2036155 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.36155] [Medline: 33443576]
31. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
32. Eyles H, Grey J, Umali E, Ni Mhurchu C. Lessons learned from a remote blood pressure lowering trial in Aotearoa. Public Health Expert. Wellington, New Zealand: University of Otago; 2021 Sep 16. URL: https://blogs.otago.ac.nz/pubhealthexpert/lessons-learned-from-a-remote-blood-pressure-lowering-trial-in-aotearoa/ [accessed 2023-02-09]
33. Te Morenga L, Pekepo C, Corrigan C, Matoe L, Mules R, Goodwin D, et al. Co-designing an mHealth tool in the New Zealand Māori community with a “Kaupapa Māori” approach. AlterNative 2018 Mar;14(1):90-99 [FREE Full text] [doi: 10.1177/1177180117753169]
34. Huang L, Woodward M, Stepien S, Tian M, Yin X, Hao Z, et al. Spot urine samples compared with 24-h urine samples for estimating changes in urinary sodium and potassium excretion in the China Salt Substitute and Stroke Study. Int J Epidemiol 2018 Dec 01;47(6):1811-1820 [FREE Full text] [doi: 10.1093/ije/dyy206] [Medline: 30307504]
35. Grimes A, White D. Digital inclusion and wellbeing in New Zealand. Motu Working Paper 19-17. Motu Economic and Public Policy Research. Wellington, New Zealand: Motu Economic and Public Policy Research; 2019 Oct. URL: http://motu-www.motu.org.nz/wpapers/19_17.pdf [accessed 2021-07-21]
36. Ni Mhurchu C, Te Morenga L, Tupai-Firestone R, Grey J, Jiang Y, Jull A, et al. A co-designed mHealth programme to support healthy lifestyles in Māori and Pasifika peoples in New Zealand (OL@-OR@): a cluster-randomised controlled trial. Lancet Digit Health 2019 Oct;1(6):e298-e307 [FREE Full text] [doi: 10.1016/S2589-7500(19)30130-X] [Medline: 33323252]
37. Yin X, Neal B, Tian M, Li Z, Petersen K, Komatsu Y, et al. Sodium and potassium content of 24 h urinary collections: a comparison between field- and laboratory-based analysers. Public Health Nutr 2018 Apr;21(6):1036-1042 [FREE Full text] [doi: 10.1017/S1368980017003731] [Medline: 29306343]
38. Dorsch MP, Cornellier ML, Poggi AD, Bilgen F, Chen P, Wu C, et al. Effects of a novel contextual just-in-time mobile app intervention (LowSalt4Life) on sodium intake in adults with hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e16696 [FREE Full text] [doi: 10.2196/16696] [Medline: 32663139]
39. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020 Feb 20;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]

40. Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Albury C, et al. A mobile health salt reduction intervention for people with hypertension: results of a feasibility randomized controlled trial. JMIR Mhealth Uhealth 2021 Oct 21;9(10):e26233 [FREE Full text] [doi: 10.2196/26233] [Medline: 34673535]
41. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 07;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
42. SHAKE the Salt Habit: The SHAKE technical package for salt reduction. World Health Organization. 2016. URL: https://apps.who.int/iris/rest/bitstreams/1061141/retrieve [accessed 2022-12-12]
43. Bernabe-Ortiz A, Trieu K, Markland M, Hy Wu J, Saavedra-Garcia L, Hart AC, et al. The potential scalability of salt substitutes as an intervention to lower population sodium levels. Salud Publica Mex 2022 Jun 13;64:S6-13. [doi: 10.21149/12788] [Medline: 36130398]

Abbreviations

BP: blood pressure
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
DBP: diastolic blood pressure
GP: general practitioner
mHealth: mobile health
NZ: New Zealand
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RSS: reduced-sodium salt
SALTS: Salt Alternatives Study
SBP: systolic blood pressure
WHO: World Health Organization

Edited by G Eysenbach, L Buis; submitted 19.10.22; peer-reviewed by M Dorsch, S Payne; comments to author 09.11.22; revised version received 13.12.22; accepted 19.01.23; published 09.03.23.

Please cite as:
Eyles H, Grey J, Jiang Y, Umali E, McLean R, Te Morenga L, Neal B, Rodgers A, Doughty RN, Ni Mhurchu C
Effectiveness of a Sodium-Reduction Smartphone App and Reduced-Sodium Salt to Lower Sodium Intake in Adults With Hypertension: Findings From the Salt Alternatives Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e43675
URL: https://mhealth.jmir.org/2023/1/e43675
doi:10.2196/43675
PMID:36892914

Original Paper

The Effect of a mHealth App (KENPO-app) for Specific Health Guidance on Weight Changes in Adults With Obesity and Hypertension: Pilot Randomized Controlled Trial

Naoki Sakane[1], MD, PhD; Akiko Suganuma[1], BB; Masayuki Domichi[1], BBIS; Shin Sukino[1], BNS; Keiko Abe[1], BHNS; Akiyoshi Fujisaki[2], MS; Ai Kanazawa[3], BS; Mamiko Sugimoto[3], BS

[1]Division of Preventive Medicine, Clinical Research Institute, National Hospital Organization Kyoto Medical Center, Kyoto, Japan

[2]Technology Development HQ, OMRON HEALTHCARE Co., Ltd., Kyoto, Japan

[3]Cardiovascular Disease Business Planning Strategy HQ, OMRON HEALTHCARE Co., Ltd., Kyoto, Japan

Corresponding Author:
Naoki Sakane, MD, PhD
Division of Preventive Medicine
Clinical Research Institute
National Hospital Organization Kyoto Medical Center
1-1 Mukaihata-cho, Fukakusa
Fushimi-ku
Kyoto, 612-8555
Japan
Phone: 81 75 641 9161
Fax: 81 75 645 2781
Email: nsakane@gf6.so-net.ne.jp

Abstract

Background: Commercial smartphone apps that promote self-monitoring of weight loss are widely available. The development of disease-specific apps has begun, but there is no app for specific health guidance (SHG) to prevent metabolic syndrome, type 2 diabetes, and cardiovascular diseases in middle-aged adults in Japan.

Objective: This study aimed to determine the efficacy of an SHG mobile health app in facilitating weight loss in Japanese adults with obesity and hypertension.

Methods: In a 12-week, statistician-blinded, randomized parallel controlled trial, 78 overweight and obese men aged 40-69 years were assigned in a 1:1 ratio to either the usual support plus KENPO-app group (intervention group) or the active control group. KENPO-app (release April 10, 2019; OMRON Healthcare Co., Ltd.) was developed by the study team and focus groups and uses behavior change techniques (ie, self-monitoring and goal-setting theory). This app was developed for SHG based on the four specific health checkups and guidance system in Japan: (1) focusing primarily on achieving the target (weight loss of ≥2 kg); (2) assessing healthy eating, exercise habits, smoking habits, relaxation, and self-weighing; (3) providing information on the results of specific health checkups; and (4) starting an intervention period of 6 months with the interim assessment at 3 months. The initial assessment explored the following: personality traits (4 types), health checkup data concerns (10 items), symptom concerns (10 items), and the aim of the intervention (weight loss, improving fitness, symptoms, laboratory data). Chatbot-supported health information on health and health behavior was selected from 392 quizzes based on app data and was provided to participants. The KENPO-app had chatbot-supported feedback and information provision combined with a self-monitoring tool (weight, steps, and blood pressure). Data on active exercise, healthy eating, and healthy lifestyle habits were obtained using a web-based self-administered questionnaire at baseline and 12 weeks.

Results: The trial's retention rate was 95% (74/78). The adherence to daily self-weighing, wearing the pedometer, and blood pressure monitoring in the KENPO-app group was significantly higher than those in the active control group. Compared with the active control group, the median body weight and BMI of the intervention group significantly decreased at 3 months (–0.4, IQR –2.0 to 0.6 kg vs –1.1, IQR –2.7 to –0.5 kg; *P*=.03; –0.1, IQR –0.6 to 0.3 kg vs –0.4, IQR –0.8 to –0.2 kg; *P*=.02, respectively). The intervention increased the percentage of participants who self-reported taking ≥8000 steps, eating vegetables before rice, eating slowly, and relaxing. Personality traits were associated with the degree of weight loss in the intervention group.

Conclusions: The SHG-specific KENPO-app was feasible and induced modest but significant weight loss in adults with obesity.

Trial Registration: University Hospital Medical Information Network Center UMIN000046263; https://tinyurl.com/bderys3b

(JMIR Mhealth Uhealth 2023;11:e43236) doi:10.2196/43236

KEYWORDS

obesity; hypertension; mobile health care app; specific health guidance; obese; weight; mHealth; mobile health; mobile app; health app

Introduction

In 2008, all health insurers in Japan were mandated to provide specific health guidance (SHG) to prevent metabolic syndrome, type 2 diabetes, and cardiovascular diseases in middle-aged adults in Japan [1-6]. During the first implementation stage, between 2008 and 2012, the nationwide implementation goal was 45%. However, after reassessment in 2019, the actual implementation rate was far lower at 23.2%. During the second stage, between 2013 and 2017, the flow of the SHG process was reviewed. During the COVID-19 pandemic, self-quarantine was associated with unhealthy eating habits, sedentary behavior, and weight gain [7,8]. In addition, the efficacy of SHG was small, and repeater eligibility for SHG was a problematic issue [9]. During the third stage, between 2019 to 2023, SHG using information and communications technology (ICT) was initially introduced in 2021. For the fourth stage (2024-), the evaluation of the outcome (ie, ≥2 kg of weight loss) will be performed, and a mobile health (mHealth) care app will be introduced for SHG. Overall, research evidence suggests that mobile apps and wearables are effective self-regulating tools for weight loss in the Western population, but a discrepancy exists [10,11]. Commercial smartphone apps that promote the self-monitoring of weight loss are widely available. The development of disease-specific apps has begun. Several apps are used for real-world SHG, but there is no app specified for SHG. Therefore, we developed the SHG-specific mHealth app (KENPO-app), and this study aimed to determine its efficacy in facilitating weight loss in Japanese adults with obesity and hypertension.

Methods

Study Design

This study was a 12-week, statistician-blinded, parallel-group, randomized controlled trial (RCT) of adults with obesity and hypertension. The data were obtained between October 2021 and May 2022.

Ethics Approval

The study protocol was approved by the institutional review board of Kyoto Medical Center (NO.21-057), and the protocol of the study was registered at the University Hospital Medical Information Network Center (UMIN000046263).

Participants

Participants were recruited from the screening panel of the Omron monitor recruitment site for product development and research in Japan. Therefore, participants may have relatively higher computer/internet literacy. We held an online information session for this trial. Inclusion criteria were as follows: age 40-64 years, BMI≥25 kg/m^2, systolic blood pressure (SBP) ≥130 mm Hg or diastolic blood pressure (DBP) ≥85 mm Hg, smartphone users capable of installing apps, and individuals capable of communicating online. Exclusion criteria were as follows: receiving SHG at present, taking antihypertensive medicine, contraindication for healthy eating and active exercise by a doctor, pregnant or breastfeeding women, and inappropriate cases (ie, severe psychiatric disorders) as determined by a research doctor.

Randomization and Masking

The independent statistician randomly allocated participants into one of two intervention arms according to sex-, age-, SBP-, and BMI-stratified block randomization (seed=1221 and block size=2). We adopted a single-blind approach; thus, the effectiveness was assessed by blinded researchers who were unaware of the randomization results.

Self-monitoring Tool

Participants in both groups received a Bluetooth weighing scale (HBF-227T, OMRON Healthcare Co, Ltd), pedometer (HJA-405T, OMRON Healthcare Co, Ltd), and upper arm blood pressure (BP) monitor (HCR-7501T, OMRON Healthcare Co, Ltd).

The Active Control Group

The participants in the active control group received the usual support. The usual support was based on intensive SHG, and the participants in both groups received initial online face-to-face counseling by a health care professional (ie, a registered dietician) who had completed the established Ministry of Health, Labor and Welfare (MHLW) training course. Participants were briefed about their health condition and lifestyle through a review of their SHG results. They were instructed to set achievable personalized behavioral goals. After the initial counseling, a health care professional provided email support three times at 2, 6, and 12 weeks. Implementation points according to the MHLW in the active control group were comparable to the required points of ≥180 in the SHG. Daily recording of body weight and steps were recommended. Measurements of BP in the morning and evening were also recommended.

mHealth KENPO-app

KENPO-app (release April 10, 2019; OMRON Healthcare Co, Ltd) was developed using behavior change techniques (ie, self-monitoring and goal-setting theory) by the study team and focus groups. This app was developed for SHG based on the four specific health checkups and guidance system in Japan: (1) focusing on achieving the primary target (weight loss of ≥2 kg); (2) assessing healthy eating, exercise habits, smoking habits, relaxation, and self-weighing; (3) providing information on the results of specific health checkups; and (4) starting an intervention period of 6 months with the interim assessment at 3 months. The initial assessment explored the following: personality traits (4 types), concerns about health checkup data (10 items), concerns about symptoms (10 items), and the aim of the intervention (weight loss, improving fitness, symptoms, laboratory data; Figures 1 and 2). Saeki et al [12] reported a cluster analysis that showed 4 clusters in a total of 1500 people aged 15-75 years. They classified personality traits into four types: “Challenger” (self-realization and a sense of growth; fact-based extrovert), “Entertainer” (connection and gratitude; relationship introvert), “Communicator” (optimism; relationship introvert), and “Walker” (do things at my own pace; fact-based introvert). The targeted behavioral goals were as follows: exercise habits (10 items), healthy eating habits (10 items), lifestyle habits (10 items), and daily steps (5000, 7000, 8000, and 10,000 steps). The self-administrated questionnaire on exercise, healthy eating, and lifestyle habits had three choices: “not confident,” “ready to change,” and “already have been.” Chatbot-supported health information on health and health behavior was selected from 392 quizzes based on app data that was provided to participants (Textbox 1). Participants accessed the KENPO-app from the App Store or Google Play. Self-weighing twice was recommended. We did not perform revisions or updates during the study period. Safety and security procedures included privacy considerations and the availability of a hotline.

Figure 1. Screenshot and initial assessment of KENPO-app system. (A) KENPO-app at APP Store; (B) KENPO-app at Google Play; (C) initial assessment items.

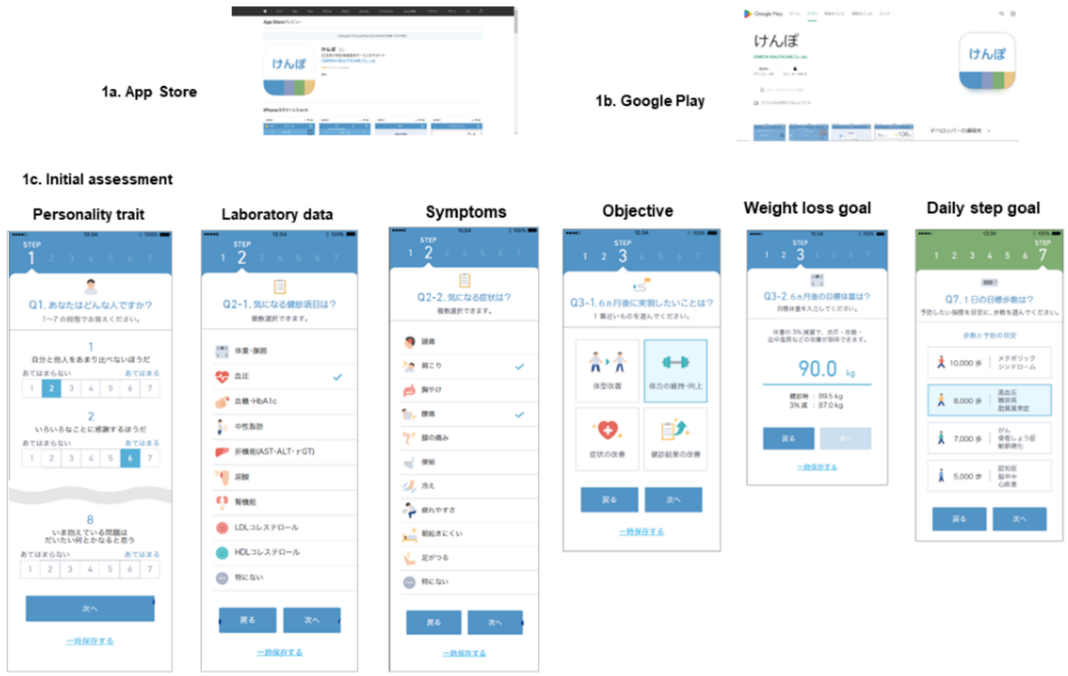
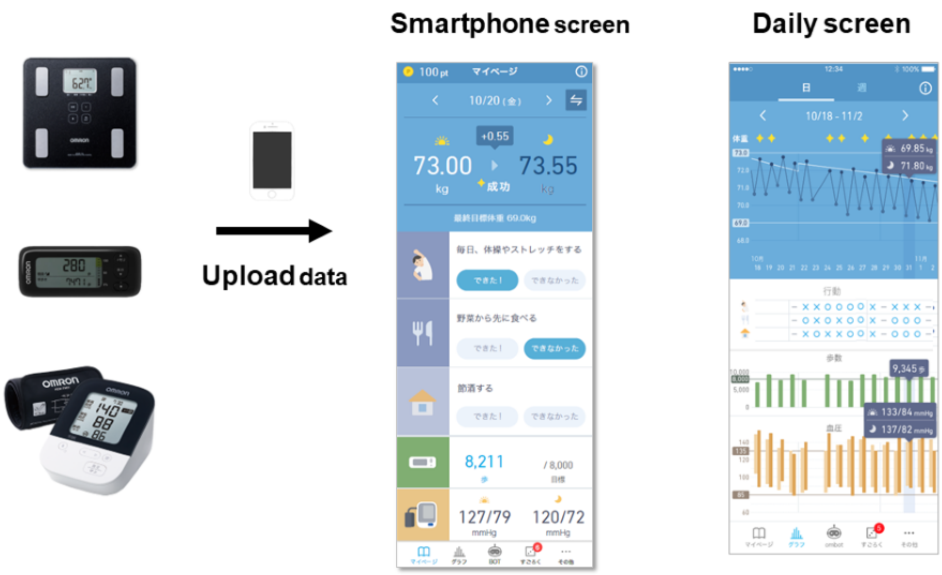


Figure 2. Outline of KENPO-app system.



Textbox 1. The contents of the KENPO-app study.

Input to the app

- Initial assessment (laboratory data, symptoms, and clinical goals)
- Personality trait
- Setting weight loss goal
- Targeting behavioral agenda (exercise, dietary, lifestyle habits, daily steps)

Self-monitoring (before initial specific health guidance [SHG])

- Pedometer, weight, and blood pressure, and behavioral agenda

Online initial SHG

- SHG based on app data and personality trait

Self-monitoring (after initial SHG)

- Pedometer, weight, and blood pressure, and behavioral agenda
- Chatbot-supported feedback of app data and sign of weight regain

Quiz on health

- Chatbot-supported quiz on health and health behavior

Input to the app (12 weeks)

- Final assessment (laboratory data, symptoms, objective)
- Behavioral agenda (exercise, dietary, and lifestyle habits)

Outcome

The outcome included changes in body weight and BMI. Weight measurements were uploaded to the cloud where the data were obtained, and the 7-day average weight was calculated. Other outcomes included changes in SBP, DBP, and adherence to the device. The frequency of weight, BP, and step uploads was recorded as a measure of adherence. Data on active exercise habits (10 items), healthy eating habits (10 items), and healthy lifestyle habits (10 items) were obtained using a web-based self-administered questionnaire at baseline and 12 weeks. Quality assurance was performed through standard operating procedures and benchmarking. Adherence to the apps was defined based on the sending rate of body weight measurements, and an attrition diagram was made.

Statistical Analysis

Data are expressed as the mean (SD), median (IQR), range, or number. Blinded data were analyzed using the R software version 4.0.0. (R Foundation for Statistical Computing) on an intention-to-treat basis by the statistician. Statistical analysis of quantitative data was performed using the Shapiro-Wilk test, Mann-Whitney *U* test, Student *t* test, and Spearman rank test. Categorical data were analyzed using Fisher exact test or exact binomial test. Those cases with missing data were omitted in the relevant analysis. Sensitivity analysis was performed using multiple imputation. There was no previous study related to our hypothesis. Therefore, the sample size was estimated as 52 people with an effect size of 0.8 (large effect size) for a pilot or feasibility study. With a dropout rate of approximately 30%, 80 people were required. The level of statistical significance was set at $P<.05$.

Results

Participants and Adherence

After 80 participants were screened, we enrolled 78 participants and excluded 2. Of the 78 participants, the mean age was 52.0 (SD 6.5) years, and 55% (n=43) were male. Those who had full- and part-time jobs accounted for 65% (n=51) and 19% (n=15) of the sample, respectively. Participants were assigned to either the KENPO-app group (intervention group; n=39) or the active control group (n=39). There were no differences in age (mean 52.5, SD 6.6 years vs mean 51.4, SD 6.4 years; $P=.47$), male sex ratio (56.4% vs 53.8%; $P>.99$), BMI (mean 27.5, SD 1.9 km/m2 vs mean 27.9, SD 2.3 kg/m2; $P=.41$), or SBP (mean 138.3, SD 12.7 mm Hg vs mean 136.2, SD 17.5 mm Hg; $P=.55$) between the groups. Two participants in the intervention group did not receive the intervention before the initial web-based consultation. We could not contact 1 participant after the intervention for unknown reasons in the active control group. As indicated in the CONSORT flow diagram, one participant was deleted from the analysis due to device failures in the active control group (Figure 3). The trial retention was 95% (74/78). Figure 4A shows the self-weighing attrition diagram. The adherence to daily self-weighing, wearing the pedometer, and BP monitoring in the KENPO-app group was significantly higher than those in the active control group. Participants in the KENPO-app group had the following health checkup data concerns: hyperglycemia 27% (n=10), hypertriglyceridemia 57% (n=21), and low high-density lipoprotein cholesterol 24% (n=9).

Figure 3. CONSORT flow diagram of KENPO-app study.

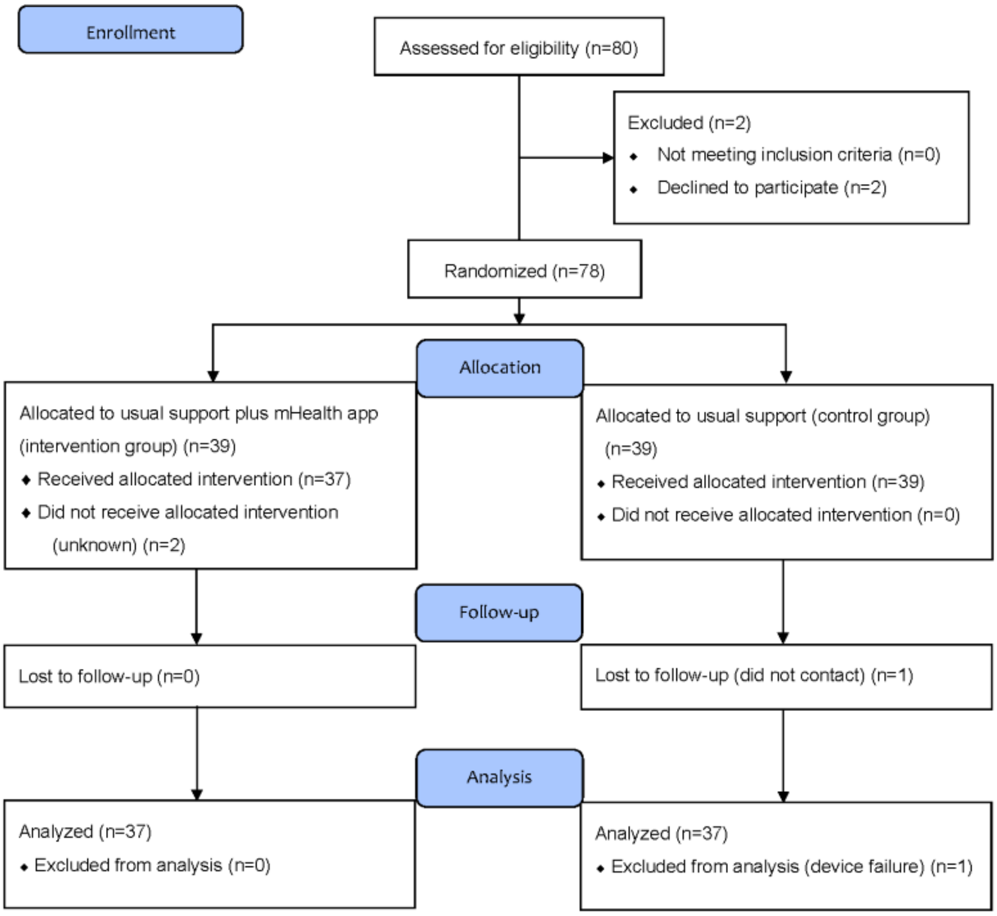
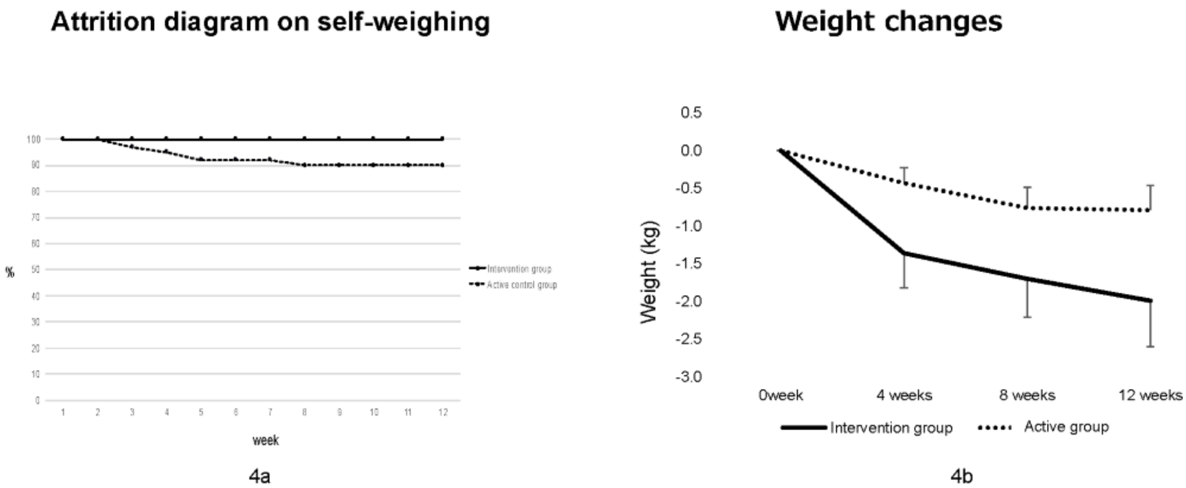


Figure 4. Attrition diagram on self-weighing and mean changes in body weight during the study period.



Outcome

Figure 4 shows the weight changes during the study (mean –2.0, SD 0.6 kg in the intervention group vs mean –0.8, SD 0.3 kg in the active control group at the 12-week follow-up). The distribution of the weight changes was not normal as assessed by the Shapiro-Wilk test. Compared with the active control group, the median body weight and BMI of the intervention group significantly decreased at 3 months (–0.4, IQR –2.0 to 0.6 kg vs –1.1, IQR –2.7 to –0.5 kg; *P*=.03; –0.1, IQR –0.6 to 0.3 kg vs mean –0.4, IQR –0.8 to –0.2 kg; *P*=.02, respectively). The sensitivity analysis confirmed the results. The odds ratio for achieving ≥3% and 2% weight loss was 1.58 (95% CI 0.47-5.63) and 2.27 (95% CI 0.79-6.85), respectively. Personality traits were associated with the degree of weight loss in the KENPO-app group. Compared with "challenger" (n=7), “walker” (n=6) had significantly greater weight loss (median –0.50, IQR –0.65 to 0.40 kg vs median –3.10, IQR –4.42 to –2.00; *P*=.02), but there was no difference in weight change among “challenger,” “communicator,” and “entertainer.” The adherence to daily self-weighing and BP monitoring in the intervention group was significantly higher than in the active control group (daily self-weighing: mean 79.3%, SD 10.5% vs mean 68.6%, SD 22.1%; *P*=.01; BP monitoring: mean 78.0%, SD 11.4% vs mean 58.8%, SD 31.6%; *P*=.001, respectively). Similarly, adherence to daily steps was higher in the intervention group than in the active control group (mean 75.7%, SD 13.4% vs mean 68.2%; *P*=.09). Only adherence to self-weighing in the morning and evening was negatively correlated with changes in body weight in the intervention group, although adherence to daily self-weighing and daily steps in both groups was not. SBP decreased in the intervention group (from mean 138, SD 13 mm Hg to mean 135, SD 10 mm Hg; *P*=.02), but this was not significantly different from the active control group. There were no changes in DBP observed in the intervention group.

Healthy Behavior

The percentage of participants who reported ≥8000 steps per day, slow eating speed, vegetable intake before rice, selecting a healthy menu, and relaxation in the intervention group increased after the 12-week study period, while the percentage of eating breakfast and reducing snacks in the active control group increased (Table 1). The rate of achieving ≥8000 steps based on the pedometer after the intervention tended to be higher than that in the active control group (58.8% vs 32.4%; *P*=.05). There were no severe adverse events or technical problems.

Table 1. Healthy lifestyle and symptoms during the study period according to the group.

Item	Intervention group (n=39), n (%)			Control group (n=39), n (%)		
	Pre	Post	*P* value	Pre	Post	*P* value
Active exercise habits						
Use of stairs instead of the escalators	5 (14)	8 (22)	.45	9 (24)	14 (38)	.13
Walking ≥8000 steps	6 (16)	13 (35)	.046	9 (24)	8 (22)	>.99
At least 30 min of brisk daily walks	4 (11)	9 (24)	.18	6 (16)	6 (16)	>.99
Do gymnastics/stretching everyday	3 (8)	6 (16)	.37	2 (5)	8 (22)	.08
Do not stay home and do nothing on holiday	3 (8)	7 (19)	.22	7 (19)	12 (32)	.18
Stand up and exercise once an hour	8 (22)	13 (35)	.18	9 (24)	9 (24)	>.99
Do housework (cooking, cleaning, etc)	15 (41)	18 (49)	.37	24 (65)	25 (68)	>.99
Resistance training ≥3 times per week	1 (3)	2 (5)	>.99	1 (3)	2 (5)	>.99
Use the gym or pool at least once per week	1 (3)	1 (3)	>.99	0 (0)	1 (3)	>.99
Play sports at least once a week	0 (0)	4 (11)	.13	3 (8)	4 (11)	>.99
Healthy eating habits						
Eat moderately	1 (3)	7 (19)	.08	6 (16)	10 (27)	.22
Eat breakfast	22 (60)	22 (60)	>.99	25 (68)	31 (84)	.04
Eat vegetable first	12 (32)	23 (62)	.01	22 (60)	26 (70)	.22
Eat slowly and well	1 (3)	9 (24)	.01	13 (35)	16 (43)	.37
Eat with nutritional balance in mind	4 (11)	9 (24)	.13	12 (32)	15 (41)	.45
Do not overeat carbohydrates	1 (3)	5 (14)	.13	9 (24)	14 (38)	.13
Eat fried food up to 3 times per week	1 (3)	6 (16)	.07	12 (32)	16 (43)	.29
Reduce salt	1 (3)	6 (16)	.07	11 (30)	12 (32)	>.99
Reduce sweet buns and delicatessen bread	7 (19)	9 (24)	.72	11 (30)	14 (38)	.37
Reduce eating in dinner	1 (3)	4 (11)	.25	8 (22)	13 (35)	.13
Healthy lifestyle						
Reduce sugar-sweetened beverages	8 (22)	14 (38)	.15	20 (54)	23 (62)	.45
Reduce sweets	4 (11)	5 (14)	>.99	5 (14)	15 (41)	.02
Check food labels	7 (19)	12 (32)	.13	11 (30)	14 (38)	.45
Choose healthy menu	0 (0)	8 (19)	.02	9 (24)	9 (24)	>.99
Do not eat anything 2 h before bedtime	9 (24)	15 (41)	.08	13 (35)	14 (38)	>.99
Go to bed early	8 (22)	12 (32)	.39	10 (27)	10 (27)	>.99
Try to have alcohol-free days	26 (70)	30 (81)	.29	27 (73)	29 (78)	.62
Reduce alcohol drinks	27 (73)	28 (76)	>.99	24 (65)	25 (68)	>.99
Stop smoking	33 (89)	32 (87)	>.99	33 (89)	33 (89)	>.99
Relaxation	6 (16)	16 (43)	.02	15 (41)	13 (35)	.79
Subjective symptoms						
Headache	10 (27)	3 (8)	.07	8 (22)	7 (19)	>.99
Shoulder stiffness	23 (62)	12 (32)	.006	21 (57)	13 (35)	.04
Heartburn	6 (16)	1 (3)	.07	7 (19)	1 (3)	.04
Lumbago	17 (46)	9 (24)	.10	17 (46)	15 (41)	.75
Knee pain	13 (35)	8 (22)	.12	6 (16)	5 (14)	>.99
Constipation	7 (19)	0 (0)	.02	8 (22)	8 (22)	>.99
Chillness	9 (24)	2 (5)	.02	13 (35)	4 (11)	.02

Item	Intervention group (n=39), n (%)			Control group (n=39), n (%)		
	Pre	Post	*P* value	Pre	Post	*P* value
Fatigue	11 (30)	8 (22)	.37	14 (38)	10 (27)	.29
Difficulty in waking up	3 (8)	3 (8)	>.99	9 (24)	7 (19)	.68
Leg cramps	6 (16)	2 (5)	.13	6 (16)	5 (14)	>.99

Discussion

Principal Findings

This is the first study to confirm the effectiveness of the SHG-specific KENPO-app in obese adults with hypertension. The mHealth KENPO-app is feasible and can produce modest but significant weight loss. Compared with standard care, the mHealth app produced modest weight loss (–1.0 kg to –2.4 kg of body weight) in obese adults with diabetes [13]. The meta-analysis by Ang et al [14], including 17 articles for Asian populations, indicated that the effect size of the RCTs for weight change was small to moderate. The effect size of our results was also small to moderate. In this study, we observed a significant change in SBP. Although mHealth apps are effective in reducing weight, they were ineffective in lowering BP in 160 adults with ≥2 cardiovascular risk factors [15]. Further examination, including a large sample size, is required to confirm these issues in the future.

Adherence and Weight Change

Self-monitoring of weight was a significant predictor of weight loss [16-18]. In this study, self-weighing twice a day (in the morning and evening) was negatively correlated with weight change, although daily self-weighing was not. We previously reported that self-weighing twice per day plus daily target setting and feedback were more effective in promoting weight loss than one daily self-measurement [19]. Self-weighing twice a day is a recommendation in clinical practice.

Healthy Behavior and Weight Loss

The current consensus states that obtaining less than 5000 steps per day is sedentary behavior, whereas obtaining >8000 steps indicates an active exercise habit [20]. A meta-analysis by Flores Mateo et al [21] indicated that the mHealth app was associated with significant changes in body weight and BMI of –1.04 kg and –0.43 kg/m^2, respectively, compared with the control group. However, there was no significant difference in physical activity between the groups. On the other hand, Richardson et al [22] performed a meta-analysis on pedometer-based interventions without caloric restriction, with a pooled estimated change in body weight of −1.3 kg. In this study, the KENPO-app increased the proportion of ≥8000 steps (self-reported).

Moreover, slow eating speed, vegetable intake before rice, selection of a healthy menu, and relaxation in the intervention group increased after the intervention. Fast eating speed is positively associated with obesity [23,24]. For weight loss strategies, the recommendation of increased vegetable consumption is often used [25]. Meal sequence, such as eating vegetables before rice, reduces postprandial glycemic and weight loss effects [26]. In a meta-analysis by Tapsell et al [27], 5 participants reported greater weight loss, 9 reported no difference, 1 showed weight gain, and 1 reported a positive association between weight loss and high vegetable consumption. Comprehensive healthy eating may have resulted in significant weight loss in the study.

Personality Traits and Weight Changes

Personality traits are an important factor in health behaviors. Interestingly, personality traits were associated with the degree of weight change in this study. Specific aspects of personality (ie, agreeableness) are relevant to weight loss maintenance [28,29]. People with greater openness and conscientiousness were associated with greater compliance with self-care [30]. Personality traits such as neuroticism, agreeableness, and conscientiousness are associated with self-weighing frequency, dietary habits, support, and difficulties during the weight loss process [31]. Further examinations including the big five personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) and large sample sizes are required to confirm these issues in the future.

Strength and Limitations

The strengths of this study include the SHG-specific mHealth app, objective measurement of data, and a high retention rate. Although mHealth apps for weight management are popular and widely available, many apps lack professional content expertise. Encouraging app development based on evidence-based online approaches would ensure content quality, allowing health care professionals to recommend their use [32]. However, there are several limitations, including the short-term (12 weeks) and lack of laboratory data. Careful attention should be paid to interpretations regarding the results because of the lack of blinding. In this study, health and ICT literacy may be higher compared to participants in the real-world needing SHG. We did not analyze the cost-effectiveness. Further examinations including cost-effective analysis are required to confirm these issues in real-world SHG. The generalizability of the findings is limited to other populations due to it being in the Japanese language and being an SHG-specific app.

In conclusion, the SHG-specific KENPO-app was feasible and induced significant weight loss in Japanese adults with obesity and hypertension.

Acknowledgments

This study was partly supported by the Japan Society for the Promotion of Science KAKENHI (grant 22K11253) and the Japan Agency for Medical Research and Development DOUKI-APP Study (grant 21ek0210124). This study was funded by OMRON Healthcare Co, Ltd.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1203 KB - mhealth_v11i1e43236_app1.pdf]

References

1. Tsushita K, S Hosler A, Miura K, Ito Y, Fukuda T, Kitamura A, et al. Rationale and descriptive analysis of specific health guidance: the Nationwide Lifestyle Intervention Program targeting metabolic syndrome in Japan. J Atheroscler Thromb 2018 Apr 01;25(4):308-322. [doi: 10.5551/jat.42010] [Medline: 29238010]
2. Ichikawa D, Saito T, Oyama H. Impact of predicting health-guidance candidates using massive health check-up data: a data-driven analysis. Int J Med Inform 2017 Oct;106:32-36. [doi: 10.1016/j.ijmedinf.2017.08.002] [Medline: 28870381]
3. Nakao YM, Miyamoto Y, Ueshima K, Nakao K, Nakai M, Nishimura K, et al. Effectiveness of nationwide screening and lifestyle intervention for abdominal obesity and cardiometabolic risks in Japan: The metabolic syndrome and comprehensive lifestyle intervention study on nationwide database in Japan (MetS ACTION-J study). PLoS One 2018;13(1):e0190862 [FREE Full text] [doi: 10.1371/journal.pone.0190862] [Medline: 29315322]
4. Muramoto A, Matsushita M, Kato A, Yamamoto N, Koike G, Nakamura M, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
5. Enomoto N, Nakamura S, Kanda S, Endo H, Yamada E, Kobayashi S, et al. Efficacy of additional intervention to the specific health guidance in Japan: The Takahata GENKI Project. Risk Manag Healthc Policy 2021;14:3935-3943 [FREE Full text] [doi: 10.2147/RMHP.S323444] [Medline: 34584471]
6. Iseki K, Konta T, Asahi K, Yamagata K, Fujimoto S, Tsuruya K, et al. Impact of metabolic syndrome on the mortality rate among participants in a specific health check and guidance program in Japan. Intern Med 2020 Nov 01;59(21):2671-2678. [doi: 10.2169/internalmedicine.4975-20] [Medline: 32669499]
7. Zachary Z, Brianna F, Brianna L, Garrett P, Jade W, Alyssa D, et al. Self-quarantine and weight gain related risk factors during the COVID-19 pandemic. Obes Res Clin Pract 2020;14(3):210-216 [FREE Full text] [doi: 10.1016/j.orcp.2020.05.004] [Medline: 32460966]
8. Zeigler Z. COVID-19 self-quarantine and weight gain risk factors in adults. Curr Obes Rep 2021 Sep;10(3):423-433 [FREE Full text] [doi: 10.1007/s13679-021-00449-7] [Medline: 34251647]
9. Fukuma S, Iizuka T, Ikenoue T, Tsugawa Y. Association of the National Health Guidance Intervention for Obesity and Cardiovascular Risks With Health Outcomes Among Japanese Men. JAMA Intern Med 2020 Dec 01;180(12):1630-1637 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4334] [Medline: 33031512]
10. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563. [doi: 10.1038/srep34563] [Medline: 27819345]
11. Wang E, Abrahamson K, Liu PJ, Ahmed A. Can mobile technology improve weight loss in overweight adults? A systematic review. West J Nurs Res 2020 Sep;42(9):747-759. [doi: 10.1177/0193945919888224] [Medline: 31762402]
12. Saeki M, Hasunuma R, Maeno T. The relationships among psychological factors on subjective well-being. Proc Annu Convention JPA 2012;76:1PMB06-1PMB06. [doi: 10.4992/pacjpa.76.0_1pmb06]
13. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
14. Ang SM, Chen J, Liew JH, Johal J, Dan YY, Allman-Farinelli M, et al. Efficacy of interventions that incorporate mobile apps in facilitating weight loss and health behavior change in the Asian population: systematic review and meta-analysis. J Med Internet Res 2021 Nov 16;23(11):e28185 [FREE Full text] [doi: 10.2196/28185] [Medline: 34783674]
15. Cho SMJ, Lee JH, Shim J, Yeom H, Lee SJ, Jeon YW, et al. Effect of smartphone-based lifestyle coaching app on community-dwelling population with moderate metabolic abnormalities: randomized controlled trial. J Med Internet Res 2020 Oct 09;22(10):e17435 [FREE Full text] [doi: 10.2196/17435] [Medline: 33034564]

16. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
17. Han M, Rhee SY. Effectiveness of mobile health applications for 5% body weight reduction in obese and overweight adults. J Obes Metab Syndr 2021 Dec 30;30(4):354-364 [FREE Full text] [doi: 10.7570/jomes21014] [Medline: 34853189]
18. Sakane N, Oshima Y, Kotani K, Suganuma A, Nirengi S, Takahashi K, et al. Self-weighing frequency and the incidence of type 2 diabetes: post hoc analysis of a cluster-randomized controlled trial. BMC Res Notes 2020 Aug 08;13(1):375 [FREE Full text] [doi: 10.1186/s13104-020-05215-x] [Medline: 32771041]
19. Oshima Y, Matsuoka Y, Sakane N. Effect of weight-loss program using self-weighing twice a day and feedback in overweight and obese subject: a randomized controlled trial. Obes Res Clin Pract 2013;7(5):e361-e366. [doi: 10.1016/j.orcp.2012.01.003] [Medline: 24304478]
20. Swift DL, Johannsen NM, Lavie CJ, Earnest CP, Church TS. The role of exercise and physical activity in weight loss and maintenance. Prog Cardiovasc Dis 2014;56(4):441-447 [FREE Full text] [doi: 10.1016/j.pcad.2013.09.012] [Medline: 24438736]
21. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
22. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. Ann Fam Med 2008;6(1):69-77 [FREE Full text] [doi: 10.1370/afm.761] [Medline: 18195317]
23. Zhang T, Cai L, Ma L, Jing J, Chen Y, Ma J. The prevalence of obesity and influence of early life and behavioral factors on obesity in Chinese children in Guangzhou. BMC Public Health 2016 Sep 09;16(1):954 [FREE Full text] [doi: 10.1186/s12889-016-3599-3] [Medline: 27613102]
24. Kolay E, Bykowska-Derda A, Abdulsamad S, Kaluzna M, Samarzewska K, Ruchala M, et al. Self-reported eating speed is associated with indicators of obesity in adults: a systematic review and meta-analysis. Healthcare (Basel) 2021 Nov 16;9(11):1559 [FREE Full text] [doi: 10.3390/healthcare9111559] [Medline: 34828605]
25. Paixão C, Dias CM, Jorge R, Carraça EV, Yannakoulia M, de Zwaan M, et al. Successful weight loss maintenance: a systematic review of weight control registries. Obes Rev 2020 May;21(5):e13003 [FREE Full text] [doi: 10.1111/obr.13003] [Medline: 32048787]
26. Kubota S, Liu Y, Iizuka K, Kuwata H, Seino Y, Yabe D. A review of recent findings on meal sequence: an attractive dietary approach to prevention and management of type 2 diabetes. Nutrients 2020 Aug 19;12(9):2502 [FREE Full text] [doi: 10.3390/nu12092502] [Medline: 32825124]
27. Tapsell LC, Dunning A, Warensjo E, Lyons-Wall P, Dehlsen K. Effects of vegetable consumption on weight loss: a review of the evidence with implications for design of randomized controlled trials. Crit Rev Food Sci Nutr 2014;54(12):1529-1538. [doi: 10.1080/10408398.2011.642029] [Medline: 24580555]
28. Koutras Y, Chrysostomou S, Giannakou K, Kosmidis MH, Yannakoulia M. Personality traits and weight loss maintenance: a cross-sectional study. Front Nutr 2021;8:702382 [FREE Full text] [doi: 10.3389/fnut.2021.702382] [Medline: 34277692]
29. Sullivan S, Cloninger CR, Przybeck TR, Klein S. Personality characteristics in obesity and relationship with successful weight loss. Int J Obes (Lond) 2007 Apr;31(4):669-674 [FREE Full text] [doi: 10.1038/sj.ijo.0803464] [Medline: 16953251]
30. Mendoza-Catalán G, Rodríguez-Santamaría Y, Domínguez-Chávez CJ, Juárez-Medina LL, Villa-Rueda AA, González-Ramírez J, et al. Personality traits and self-care behaviors in adults with type 2 diabetes mellitus. Diabetes Metab Syndr Obes 2022;15:1-6 [FREE Full text] [doi: 10.2147/DMSO.S340277] [Medline: 35018104]
31. Soini S, Mustajoki P, Eriksson JG, Lahti J. Personality traits associated with weight maintenance among successful weight losers. Am J Health Behav 2018 Nov 01;42(6):78-84. [doi: 10.5993/AJHB.42.6.8] [Medline: 30158003]
32. Nikolaou CK, Lean MEJ. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2017 Jan;41(1):200-202. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]

Abbreviations

BP: blood pressure
DBP: diastolic blood pressure
ICT: information and communications technology
mHealth: mobile health
MHLW: Ministry of Health, Labor and Welfare
RCT: randomized controlled trial
SBP: systolic blood pressure
SHG: specific health guidance

Edited by L Buis; submitted 07.10.22; peer-reviewed by K Tsushita, R Abd Jalil; comments to author 09.12.22; revised version received 13.01.23; accepted 25.01.23; published 12.04.23.

Please cite as:
Sakane N, Suganuma A, Domichi M, Sukino S, Abe K, Fujisaki A, Kanazawa A, Sugimoto M
The Effect of a mHealth App (KENPO-app) for Specific Health Guidance on Weight Changes in Adults With Obesity and Hypertension: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e43236
URL: https://mhealth.jmir.org/2023/1/e43236
doi:10.2196/43236
PMID:37043287

Original Paper

Improving Kidney Outcomes in Patients With Nondiabetic Chronic Kidney Disease Through an Artificial Intelligence–Based Health Coaching Mobile App: Retrospective Cohort Study

Wei Liu[1,2*], MD; Xiaojuan Yu[3,4,5,6*], MD; Jiangyuan Wang[7*], MD; Tianmeng Zhou[7], MS; Ting Yu[2,8], MM; Xuyong Chen[7], MD; Shasha Xie[7], MD; Fuman Han[7], MD; Zi Wang[3,4,5,6], MD

[1]Department of Nephropathy, Anqing Municipal Hospital, Anqing, China
[2]Anqing Medical Center of Anhui Medical University, Anqing, China
[3]Renal Division, Department of Medicine, Peking University First Hospital, Beijing, China
[4]Institute of Nephrology, Peking University, Beijing, China
[5]Key Laboratory of Renal Disease, Ministry of Health, Beijing, China
[6]Key Laboratory of Chronic Kidney Disease Prevention and Treatment, Ministry of Education, Beijing, China
[7]Beijing Kidney Health Technology Co., Ltd, Beijing, China
[8]Fifth Clinical Medical College, Anhui Medical University, Hefei, China
[*]these authors contributed equally

Corresponding Author:
Zi Wang, MD
Renal Division
Department of Medicine
Peking University First Hospital
No. 8. Xishiku Street, Xicheng District
Beijing, 100034
China
Phone: 86 01083575685
Email: tonypedia@126.com

Abstract

Background: Chronic kidney disease (CKD) is a global health burden. However, the efficacy of different modes of eHealth care in facilitating self-management for patients with CKD is unclear.

Objective: The aim of this study was to evaluate the effectiveness of a mobile app–based intelligent care system in improving the kidney outcomes of patients with CKD.

Methods: Our study was a retrospective analysis based on the KidneyOnline intelligent system developed in China. Patients with CKD but not dependent on dialysis who registered on the KidneyOnline app between January 2017 and January 2021 were screened. Patients in the the KidneyOnline intelligent system group and those in the conventional care group were 1:1 matched according to their baseline characteristics. The intervention group received center-based follow-up combined with the KidneyOnline intelligent patient care system, which was a nurse-led, patient-oriented collaborative management system. Health-related data uploaded by the patients were integrated using deep learning optical character recognition (OCR). Artificial intelligence (AI)–generated personalized recipes, lifestyle intervention suggestions, early warnings, real-time questions and answers, and personalized follow-up plans were also provided. Patients in the conventional group could get professional suggestions from the nephrologists through regular clinical visits, but they did not have access to the service provided by AI and the health coach team. Patients were followed for at least 3 months after recruitment or until death or start of renal replacement therapy.

Results: A total of 2060 eligible patients who registered on the KidneyOnline app from 2017 to 2021 were enrolled for the analysis. Of those, 902 (43.8%) patients were assessed for survival analysis after propensity score matching, with 451(50%) patients in the KidneyOnline intelligent patient care system group and 451(50%) patients in the conventional care group. After a mean follow-up period of 15.8 (SD 9.5) months, the primary composite kidney outcome occurred in 28 (6%) participants in the KidneyOnline intelligent patient care system group and 32 (7%) in the conventional care group, with a hazard ratio of 0.391 (95% CI 0.231-0.660; *P*<.001). Subgroup survival analysis demonstrated that the KidneyOnline care system significantly reduced the risk of composite kidney outcome, irrespective of age, sex, baseline estimated glomerular filtration rate (eGFR), and proteinuria. In addition, the mean arterial pressure (MAP) significantly decreased from 88.9 (SD 10.5) mmHg at baseline to 85.6 (SD 7.9) mmHg at 6 months (*P*<.001) in the KidneyOnline intelligent patient care system group and from 89.3 (SD 11.1) mmHg to 87.5 (SD 8.2) mmHg (*P*=.002) in the conventional CKD care group.

Conclusions: The utilization of the KidneyOnline intelligent care system was associated with reduced risk of unfavorable kidney outcomes in nondiabetic patients with CKD.

(JMIR Mhealth Uhealth 2023;11:e45531) doi:10.2196/45531

KEYWORDS

chronic kidney disease; self-management; mobile apps; end-stage kidney disease; eHealth intervention; kidney; efficacy; eHealth care; dialysis; deep-learning; artificial intelligence; patient care

Introduction

Chronic kidney disease (CKD) is a worldwide public health concern and is increasingly becoming a global economic burden [1]. Both end-stage kidney disease (ESKD) and a reduction of estimated glomerular filtration rate (eGFR) are associated with hospitalization, cardiovascular events, and risk of death [2]. The prevalence of CKD in China was reported to be 10.8% in 2012, while only 12.5% of Chinese adults were aware of CKD as a medical problem [3]. Patients with CKD are often accompanied by comorbidities such as hypertension, diabetes, and heart disease. Providing care for patients with CKD involves a multidisciplinary team of physicians, nurses, dieticians, and social workers. Interventions to improve the outcomes of patients with CKD include lifestyle modification, antihypertensive medication, lipid modification, and glycemic control in patients with diabetes mellitus [4]. Achieving better health outcomes for patients with CKD requires patients to be aware of CKD and to engage in treatment and management plans [5]. Telehealth apps provide new opportunities to enhance self-management, behavior change, and medication adherence in a convenient way, thus reducing health complications and improving the overall kidney survival time.

There has been growing interest from physicians and other health care providers to use novel health-related mobile apps for patients in their fields of interest. Telehealth educational apps are more flexible and adaptable to patients' preferences than paper-based or in-person health education. However, according to a systematic review, CKD-related apps accounted for only 1% of the total number of available apps in 2017 [6]. In addition, as opposed to in-person health consultations, telehealth programs usually do not offer patients the chance to ask their providers or educators questions immediately, which hampers the popularity and uptake of newly developed mobile apps.

Up to now, most studies describing the effectiveness of eHealth interventions on patients with CKD were from North America [7]. Moreover, the studies had relatively small sample sizes and only investigated changes in participants' clinical parameters in a short period of time. Importantly, there has been a lack of studies assessing eHealth interventions using hard kidney end points. The objective of this study is to evaluate the effectiveness of a novel nurse-led, real-time communicating, app-based intelligent patient care system in China between 2017 and 2021 from real-world data.

Methods

Study Design and Participants

This study was a retrospective cohort analysis based on a mobile app called KidneyOnline in China. The KidneyOnline app offered a platform with free materials and resources about CKD. Patients received recruitment information about the KidneyOnline app within WeChat; subsequently, they downloaded the app and became active app users. The KidneyOnline intelligent patient care system was embedded in the app, and users made their own decision about whether to join the intelligent care system. Informed patient consent forms were signed by the patients within the KidneyOnline app. All app users between January 2017 and January 2021 were screened according to the recruitment criteria, and eligible patients were grouped according to their original choice.

Patients were enrolled if they (1) were over 18 years old; (2) fulfilled the diagnosis of CKD (ie, eGFR <60 mL/min/1.73 m^2 or eGFR < 90 mL/min/1.73 m^2 but with albuminuria or hematuria for at least 3 months or as defined using other clinically indicated criteria); (3) uploaded 2 pieces of data upon registration with an interval of at least 3 months; (4) were free of dialysis, with a baseline eGFR >15mL/min/1.73 m^2; and (5) were able and willing to provide informed consent. Patients were excluded if (1) they or their relatives were unable to use a smartphone, (2) they intended to start dialysis or have a kidney transplant within the next 3 months, (3) there was a lack of baseline or follow-up data, or (4) they refused to communicate with the health coach team. Study participants were followed for at least 3 months after recruitment until death or the start of renal replacement therapy.

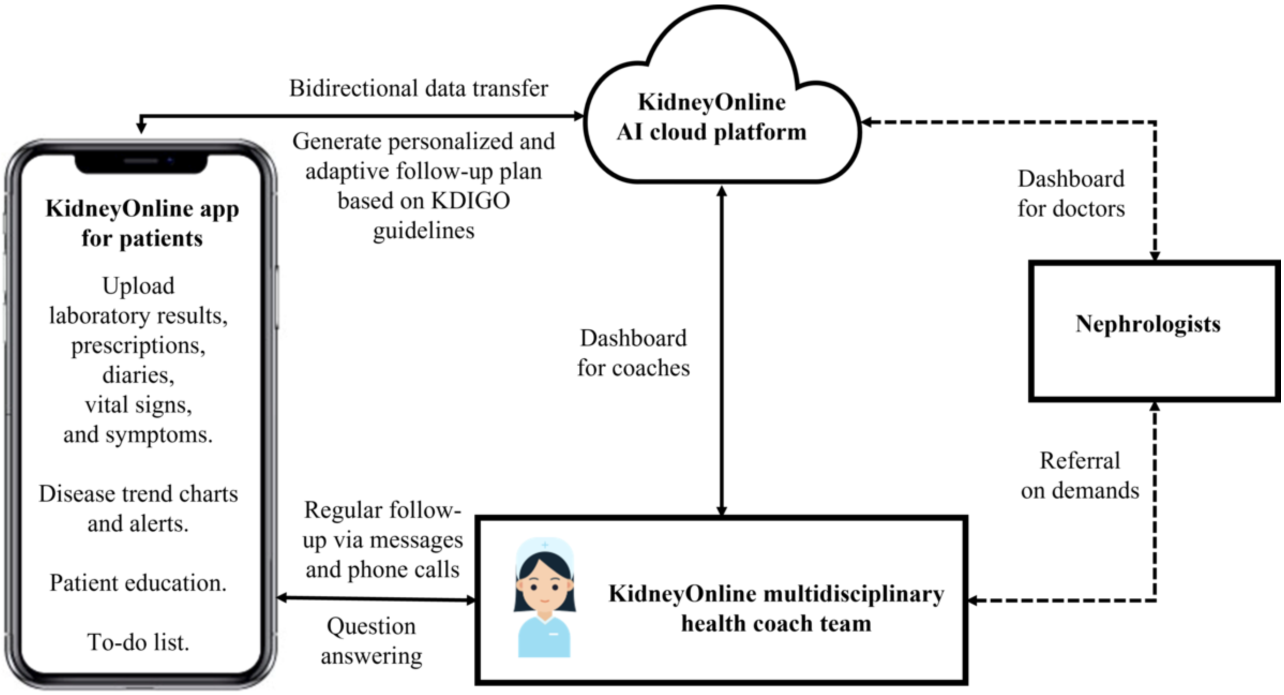
Intervention

The KidneyOnline intelligent patient care system was a nurse-led, patient-oriented collaborative management system as an adjunct to regular clinic visits for patients with CKD (Figure 1). The intelligent system was empowered by artificial intelligence (AI) and a health coach team, which consisted of a group of experienced nurses trained by nephrologists, dieticians, and social workers. The system consisted of a smartphone app for patients, a web-based clinical dashboard app for health care providers, and a data server for information management (Figure 1).

Participants in the KidneyOnline intelligent patient care group were able to experience at least 5 aspects of service provided by the system, which are detailed in Textbox 1 (Figure 2).

Patients in the conventional care group only had access to the health-related educational materials provided by the app. They could get professional suggestions from the nephrologists through regular clinical visits, but they did not have access to the service provided by AI and the health coach team.

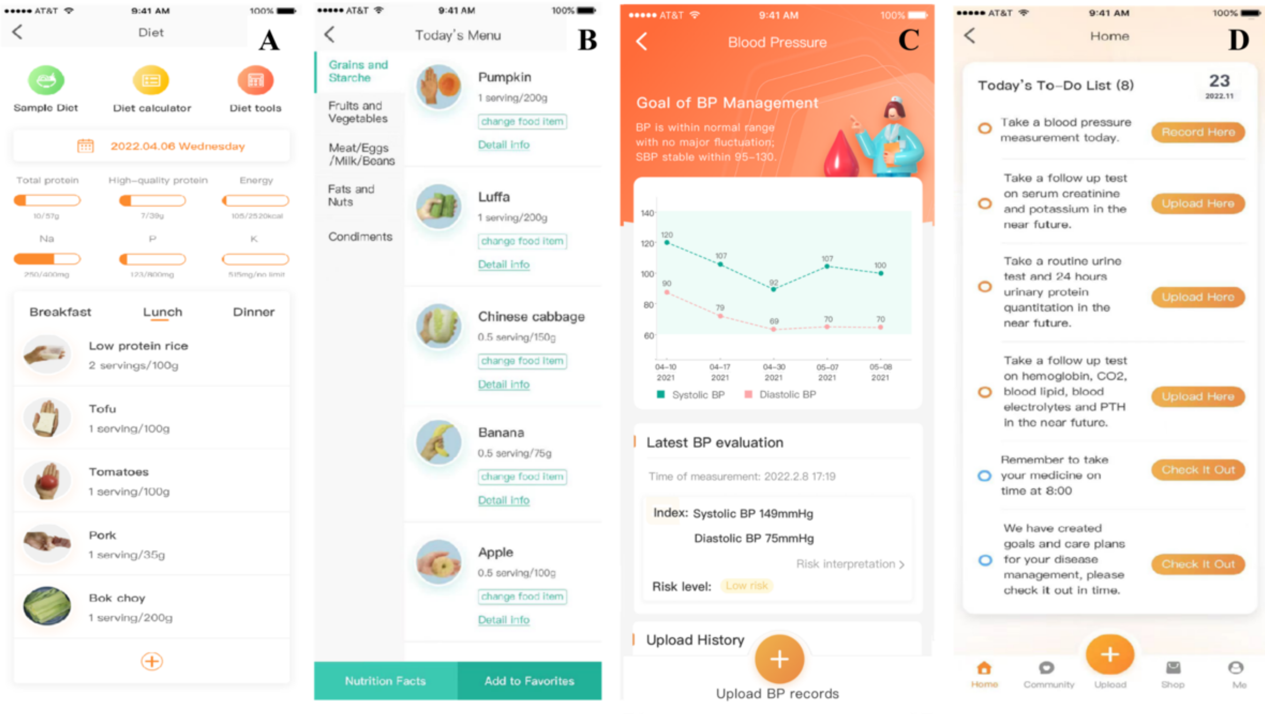
Figure 1. Framework of KidneyOnline intelligent care system. AI: artificial intelligence; KDIGO: Kidney Disease: Improving Global Outcomes (clinical practice guidelines).



Textbox 1. The 5 aspects of service provided to patients by the KidneyOnline system.

1. Interpretation of disease condition and corresponding guidance. The intelligent care system helped explain patients' diagnoses and medication prescriptions, analyze and interpret lab results, and remind patients about medication precautions. The system also provided guidance on lifestyle interventions, including diet, exercise, and sleep. Patients could receive artificial intelligence (AI)–generated personalized recipes based on their disease condition and food preferences, and they were able to make food diaries to see if the nutrition requirements were met.
2. Regular check-ups. The intelligent care system followed each patient regularly to check if the patient was in good condition, assure if the recent BP was well-controlled, evaluate whether the patient's current medication was reasonable, and assess whether the patient's dietary intake was reasonable.
3. Early warnings. Through the algorithm, the intelligent system was able to identify risks associated with lab results and certain medications and make early warnings to minimize these risks. If abnormal lab results requiring immediate attention were identified, the system would send a medical alert to both the patient and the health coach. The health coach would take the initiative to contact the patient and make suggestions, including reminding the patient about medications, providing guidance on diet, exercise, and other lifestyle aspects, and reminding the patient to consult with doctors.
4. Real-time question-and-answer fields empowered by knowledge graphs. We established a renal knowledge graph according to the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines for chronic kidney disease (CKD), which provided intelligent search and human-computer dialogues to enhance the efficiency and professional competency of the health coach team. Empowered by the graphs, the health coaches promptly provided real-time answers to any questions the patients had (Multimedia Appendix 1). These include laboratory results analysis, medication precautions, dietary guidance, exercise guidance, advice on coping with unexpected situations, advice on the correct way to take medicine to avoid kidney injuries, and so forth.
5. Clinical reminders. According to the patient's overall condition, the intelligent system could generate a personalized follow-up plan, reminding the patient to go to the doctor at regular intervals, helping them sort out test items they need to finish, and checking follow-up results to confirm whether the follow-up visit was completed.

Figure 2. Artificial intelligence (AI)–generated personalized recipes, blood pressure (BP) monitoring. and personal checklist dashboards of the KidneyOnline app.



Data Collection

The foundation of this intelligent system was built upon the structurization of patients' health-related data using deep learning optical character recognition (OCR). Conventionally, patients' health data came from a variety of sources, including (1) patients' self-reported signs and symptoms; (2) data from intelligent home devices such as sphygmomanometer, and (3) patients' past medical history, clinical notes, drug prescriptions, lab results, pathological reports, imagological exams, and so on. In the KidneyOnline intelligent system, patients uploaded those data simply by taking photos, and the intelligent system extracted the data efficiently by utilizing deep learning OCR. Combined with manual verification, a structured database was built so that patients' health-related data could be integrated and analyzed quantitatively.

The participants took photos of their medical records, laboratory test results, and clinical prescriptions and uploaded them on the mobile app. All electronic data and photographs were uploaded instantly to a secure, cloud-based server. To comply with security and privacy regulations, patients' smartphones were password-protected, and data were encrypted. Only the researchers were able to access the data on a cloud storage platform.

Outcomes

The primary outcome of our study was the first occurrence of either a 30% decrease in eGFR or an incidence of ESKD. Secondary end points included changes in 24-hour proteinuria and changes in mean arterial pressure (MAP). eGFR results were calculated by using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula. Blood pressure data were collected through the app based on home blood pressure measurements uploaded by the patients, using either mercury or an electronic sphygmomanometer.

Ethics Approval

This study was approved by the local research ethics board from Anqing Municipal Hospital (2022-033) and was conducted according to the Declaration of Helsinki.

Statistical Analysis

Reporting Descriptive Data

The distributional properties of data were expressed as mean (SD) for continuous variables with a normal distribution or median (IQR) for variables with a skewed distribution. For continuous data, the independent or paired Student *t* test was used for within-group and between-group comparisons; for categorical variables with percentages, the chi-square or McNemar test was used. Clinical parameters including 24-hour proteinuria and MAP during the follow-up were compared using Student *t* tests (for normally distributed continuous variables) and Wilcoxon signed-rank tests (for nonnormally distributed continuous variables).

Propensity Score Matching

To balance the confounding factors between the KidneyOnline intelligent patient care system group and the conventional CKD care group, 1:1 propensity score matching (PSM) was performed using nearest neighbor algorithms with a 0.9 caliper width of 0.02 pooled standard deviations. Matching was based on baseline characteristics of age, sex, BMI, baseline eGFR, MAP, and proteinuria levels.

Assessment of the Benefits of KidneyOnline

For survival analysis, both the Kaplan-Meier method and Cox proportional hazards model were used. The Kaplan-Meier method was applied to evaluate the cumulative incidence of primary outcomes in both groups following PSM. Additionally, the Cox proportional hazards model was utilized to identify predictive factors associated with the outcomes. For matched data after PSM, the Cox proportional hazards model with gamma frailty was used. All missing information was treated as missing data without imputation. Subgroup analyses were performed after stratifying according to the median age (<33 years vs ≥33 years), sex, kidney function (eGFR <60 mL/min/1.73 m^2 vs ≥60 mL/min/1.73 m^2), and proteinuria (<1 g/24 h vs ≥1g/24 h, <3 g/24 h vs ≥3 g/24 h), as well as the median value of baseline MAP (<88.7 mmHg vs ≥88.7 mmHg). Statistical analyses were performed using Python and Lifelines, an open-access survival analysis library written in Python. A 2-sided P<.05 was considered statistically significant.

Results

Baseline Characteristics

Between January 2017 and January 2021, 78,007 potentially eligible patients were screened. Among them, 2060 (2.6%) were eligible for our analysis according to the inclusion and exclusion criteria. There were 1600 (77.7%) patients in the KidneyOnline intelligent patient care system group and 460 (22.3%) patients in the conventional care group (Figure 3). Among the 2060 patients enrolled in our study, the mean age was 35.6 (SD 9.5) years, 1175 (57%) were female, and they had an average BMI of 22.7 (SD 4.2) kg/m^2 (Table 1). Upon registration, the mean eGFR was 88.6 (SD 31.3) mL/min/1.73 m^2, and the mean proteinuria level was 1.4 (SD 1.8) g/24 h. The average MAP was 89 (SD 11) mmHg. During the follow-up period, 1539 (75%) patients were treated with a renin-angiotensin-aldosterone system blocker (RASB), 524 (25%) were treated with steroids, and 418 (20%) were treated with immunosuppressants. Compared with patients in the conventional care group, patients in the KidneyOnline intelligent patient care system group had a more favorable BMI (mean 22.5, SD 4 kg/m^2 vs mean 23.2, SD 4.8 kg/m^2; P=.004) and lighter 24-hour proteinuria (median 0.7, IQR 0.3-1.6 g/24 h vs median 0.9, IQR 0.3-1.9 g/24 h; P=.033) but worse baseline kidney function (mean 87.8, SD 31.3 mL/min/1.73 m^2 vs mean 91.5, SD 31.4 mL/min/1.73 m^2; P=.024). There were no significant differences in the context of RASB and steroids or immunosuppressants usage between the 2 groups (Table 1).

Figure 3. Flow diagram of patient enrollment for the analysis. eGFR: estimated glomerular filtration rate.

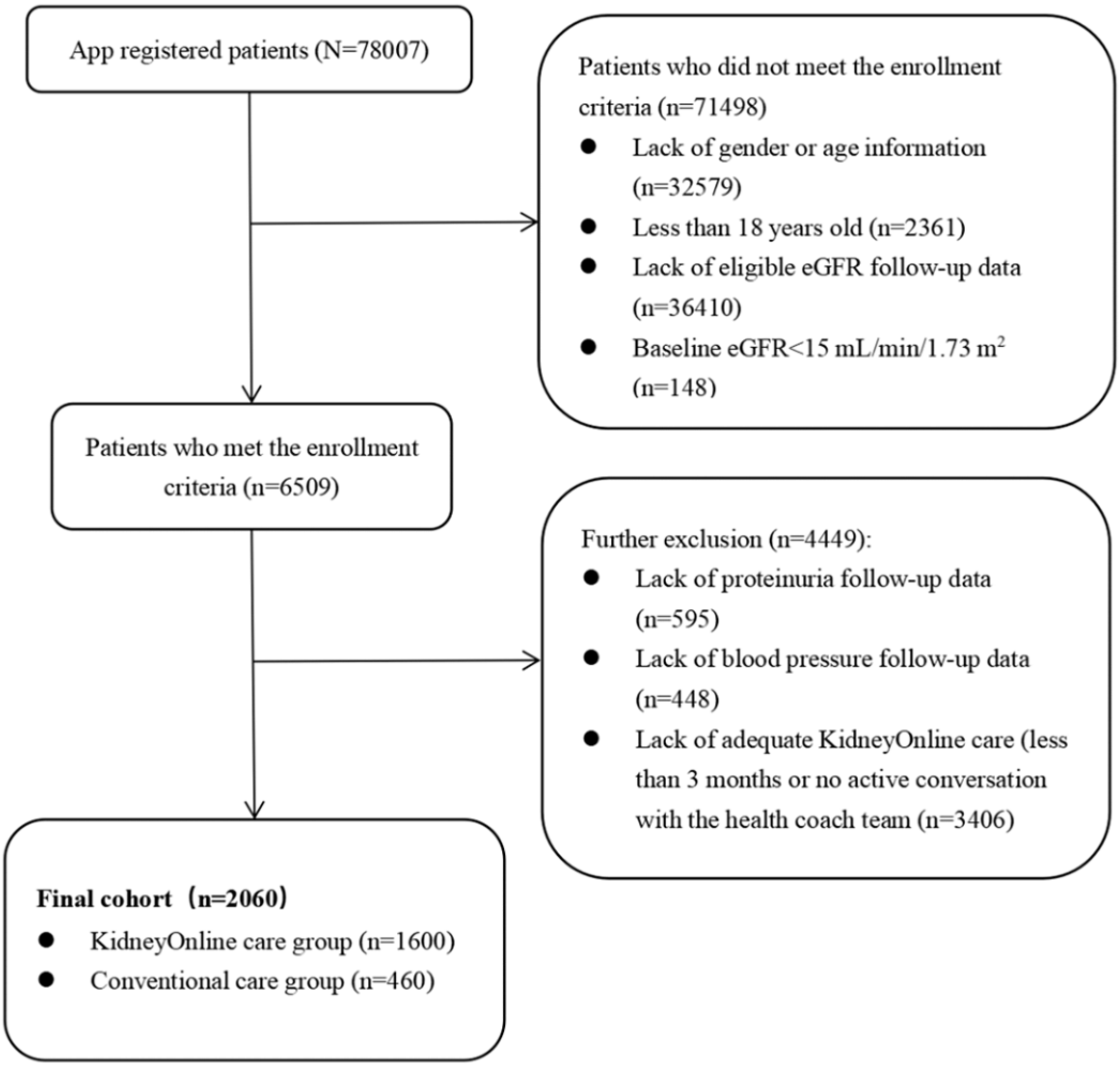


Table 1. Baseline and follow-up characteristics of patients in the whole cohort.

Demographics	Total (N=2060)	Conventional care group (n=460)	KidneyOnline care group (n=1600)	*P* value
Age (years), mean (SD)	35.6 (9.5)	35.3 (9.6)	35.8 (9.5)	.32
Sex, n (%)				
Female	1175 (57)	248 (53.9)	927 (57.9)	.12
Male	885 (43)	212 (46.1)	673 (42.1)	.12
BMI (kg/m^2), mean (SD)	22.7 (4.2)	23.2 (4.8)	22.5 (4)	.004
Etiology of CKD[a], n (%)				
IgA[b] nephropathy or IgA vasculitis	945 (45.9)	182 (39.6)	763 (47.7)	.002
Membranous nephropathy	185 (9)	49 (10.7)	136 (8.5)	.18
Focal segmental glomerular sclerosis	63 (3.1)	13 (2.8)	50 (3.1)	.86
Hypertensive nephropathy	53 (2.6)	23 (5)	30 (1.9)	<.001
Diabetic nephropathy	7 (0.3)	1 (0.2)	6 (0.4)	.95
Other types of nephritis	246 (11.9)	50 (10.9)	196 (12.2)	.47
Unknown etiology	559(27.1)	141(30.7)	418(26.1)	.062
Kidney transplant recipient	2 (0.1)	1 (0.2)	1 (0.1)	.93
Laboratory results				
Baseline eGFR[c] (mL/min/1.73 m^2), mean (SD)	88.6 (31.3)	91.5 (31.4)	87.8 (31.3)	.024
CKD stage 3-4, n (%)	443 (21.5)	88 (19.1)	355 (22.2)	.18
Baseline proteinuria (g/24 h), median (IQR)	0.8 (0.3-1.7)	0.9 (0.3-1.9)	0.7 (0.3-1.6)	.033
Baseline MAP[d] (mmHg), mean (SD)	88.6 (10.9)	89.5 (11.1)	88.4 (10.8)	.051
Drug therapies, n (%)				
RASB[e]	1539 (74.7)	333 (72.4)	1206 (75.4)	.19
Steroids	524 (25.4)	127 (27.6)	397 (24.8)	.23
Immunosuppressants	418 (20.3)	94 (20.4)	324 (20.3)	.93

[a]CKD: chronic kidney disease.

[b]IgA: immunoglobulin A.

[c]eGFR: estimated glomerular filtration rate.

[d]MAP: mean arterial pressure.

[e]RASB: renin-angiotensin-aldosterone system blocker.

Benefits of KidneyOnline Care System for the Whole Population

After a mean follow-up of 18.1 (SD 9.5) months, the primary composite kidney outcome occurred in 121 (8%) participants in the KidneyOnline intelligent patient care system group and 33 (7%) in the conventional care group, with 86 (5%) versus 21 (5%) and 35 (2%) versus 12 (3%) participants reaching a 30% eGFR decrease and ESKD, respectively.

Multivariate Cox regression with stepwise procedures was used to analyze the risk factors of composite kidney outcome. After adjustments for age, gender, baseline eGFR, proteinuria, MAP, and use of RASB, steroids, and immunosuppressants, individuals in the KidneyOnline care group were less likely to progress to ESKD compared with the conventional group, with a hazard ratio of 0.375 (95% CI 0.221-0.638; *P*<.001) (Multimedia Appendix 2).

Characteristics of the Individuals After PSM

A 1:1 PSM analysis was performed to balance the selection bias between the 2 groups. A total of 902 patients with 451 (50%) patients in each group were successfully matched. In total, the average age was 35.5 (SD 9.4) years, and 54% (n=487) were female. The average BMI was 22.9 (SD 3.8) kg/m^2. The baseline MAP was 89.1 (SD 10.8) mmHg, and the 24-hour proteinuria was 0.8 (IQR 0.3-1.8) g/24 h. The baseline eGFR of the 902 patients was 91.9 (SD 30.7) mL/min/1.73 m^2 (Table 2). Overall, there were 672 (75%) patients who received RASB, 232 (26%) who received steroids, and 197 (22%) who received immunosuppressant therapy. There were no significant differences in laboratory results and treatments between the KidneyOnline intelligent patient care system and the conventional care group at baseline (Table 2).

Table 2. Baseline and follow-up characteristics of patients after propensity score matching (PSM).

Demographics	Total (N=902)	Conventional care group (n=451)	KidneyOnline care group (n=451)	*P* value
Age (years), mean (SD)	35.5 (9.4)	35.27 (9.6)	35.67 (9.2)	.51
Sex, n (%)				
Female	487 (54)	244 (54.1)	243 (53.9)	>.99
Male	415 (46)	207 (45.9)	208 (46.1)	>.99
BMI (kg/m^2), mean (SD)	22.9 (3.8)	22.9 (4.0)	23.0 (3.6)	.90
Etiology of CKD[a], n (%)				
IgA[b] nephropathy or IgA vasculitis	386 (42.8)	181 (40.1)	205 (45.5)	.13
Membranous nephropathy	95 (10.5)	45 (10)	50 (11.1)	.64
Focal segmental glomerular sclerosis	23 (2.5)	12 (2.7)	11 (2.4)	>.99
Hypertensive nephropathy	29 (3.2)	22 (4.9)	7 (1.6)	.006
Diabetic nephropathy	4 (0.4)	1 (0.2)	3 (0.7)	.63
Other types of nephritis	99 (11)	48 (10.6)	51 (11.3)	.84
Unknown etiology	265 (29.4)	141 (31.3)	124 (27.5)	.24
Kidney transplant recipient	1 (0.1)	1 (0.2)	0 (0)	>.99
Laboratory results				
Baseline eGFR[c] (mL/min/1.73 m^2), mean (SD)	91.9 (30.7)	91.5 (31.3)	92.3 (30.2)	.63
CKD stage 3-4, n (%)	163 (18.1)	86 (19.1)	77 (17.1)	.49
Baseline proteinuria (g/24 h), median (IQR)	0.8 (0.3-1.8)	0.8 (0.3-1.8)	0.8 (0.3-1.6)	.29
Baseline MAP[d] (mmHg), mean (SD)	89.1 (10.8)	89.3 (11.0)	88.9 (10.5)	.56
Drug therapies				
RASB[e]	672 (74.5)	327 (72.5)	345 (76.5)	.20
Steroids	232 (25.7)	123 (27.3)	109 (24.2)	.32
Immunosuppressants	197 (21.8)	93 (20.6)	104 (23.1)	.41

[a]CKD: chronic kidney disease.

[b]IgA: immunoglobulin A.

[c]eGFR: estimated glomerular filtration rate.

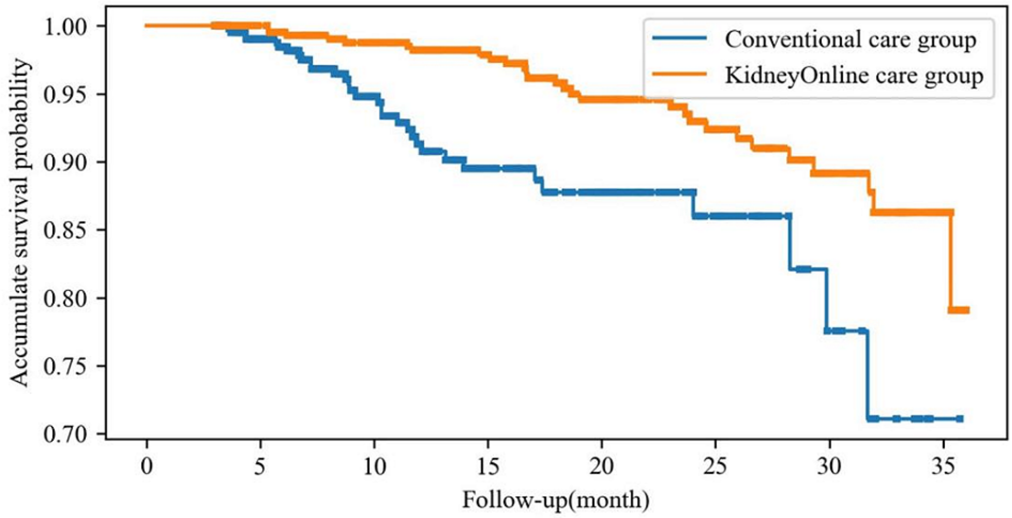
[d]MAP: mean arterial pressure.

[e]RASB: renin-angiotensin-aldosterone system blocker.

Benefits of the KidneyOnline Care System for Propensity-Matched Individuals

After a mean follow-up period of 15.8 (SD 9.5) months, the primary composite kidney outcome occurred in 28 (6%) participants in the KidneyOnline intelligent patient care system group and in 32 (7%) in the conventional care group, with a hazard ratio (HR) of 0.391 (95% CI 0.231-0.661; *P*<.001) (Figure 4), with 20 (4%) versus 20 (4%) and 8 (2%) versus 12 (3%) participants reaching a 30% decrease in eGFR and ESKD, respectively. A subgroup analysis showed that the KidneyOnline intelligent patient care system significantly reduced the risk of composite kidney outcome in all subgroups stratified by age (<33 vs ≥33 years), sex, kidney function (eGFR <60 vs ≥60 mL/min/1.73 m^2), and proteinuria (<1 vs ≥ g/24 h, <3 vs ≥3 g/24 h). However, the KidneyOnline intelligent patient care system improved the composite kidney outcomes in patients with elevated levels of MAP, but it did not improve composite kidney outcomes in the normal MAP group (Multimedia Appendix 3).

Figure 4. Kaplan-Meier curve of primary composite kidney outcomes.



	0	5	10	15	20	25	30	35
Conventional care group								
At risk	451	368	205	128	79	43	16	1
Censored	0	79	230	297	344	379	404	418
Events	0	4	16	26	28	29	31	32
KidneyOnline care group								
At risk	451	424	362	299	217	148	79	19
Censored	0	27	84	144	217	282	347	405
Events	0	0	5	8	17	21	25	27

Changes in MAP During Follow-up

Since the KidneyOnline intelligent patient care system showed different effects in different MAP groups, we further analyzed the MAP during follow-up between the KidneyOnline group and the conventional CKD care group. The baseline MAP values were similar between the 2 groups. The MAP significantly decreased from 88.9 (SD 10.5) mmHg at baseline to 85.6 (SD 7.9) mmHg at 6 months (*P*<.001) in the KidneyOnline group and from 89.3 (SD 11.1) mmHg to 87.5 (SD 8.2) mmHg (*P*=.002) in the conventional CKD care group. After 6 months, the MAP in the KidneyOnline group participants remained at a lower level compared with that in the conventional CKD care group (Multimedia Appendix 4A). This trend was also observed when participants were stratified according to the median value of MAP (88.7 mmHg) (Multimedia Appendix 4B,C).

Changes in Proteinuria During Follow-up

The mean level of 24-hour proteinuria in the KidneyOnline intelligent patient care system group was not significantly different from the conventional CKD care group at baseline (1.4 vs 1.5 g/24 h, *P*=.36). During the follow-up, the mean level of 24-hour proteinuria decreased to 0.8 g/24 h in the KidneyOnline intelligent patient care system group and 0.8 g/24 h in the conventional care group at end of 12 months. There was no significant difference in the mean level of 24-hour proteinuria between the 2 groups by the end of 24 months (0.7 vs 0.7 g/24 h, *P*=.92) (Multimedia Appendix 5).

Discussion

Principal Findings

In this study, we described a nurse-led, smartphone-based patient-oriented system designed to help disease management in patients living with stages 1 to 4 of CKD. Our findings demonstrate that this intelligent care system was associated with better blood pressure control and a reduced risk of kidney failure. This provided a novel strategy for promoting a healthy lifestyle and improving kidney prognosis in patients with CKD, regardless of their scheduled consultations.

Comparisons to Prior Work

The incidence and prevalence of CKD have been persistently increasing because of the aging population, who experience a high incidence of metabolic disorders such as hypertension, diabetes, and obesity [8-10]. There have been efforts to find innovative and efficient ways to improve patient outcomes, but the results have been conflicting due to varied intervention facets and intensities. Few studies showed improvement in hard kidney end points or eGFR change. The intervention conducted in the Multifactorial Approach and Superior Treatment Efficacy in Renal Patients With the Aid of Nurse Practitioners (MASTERPLAN) study [11] initially reported negative results of the strict implementation of CKD guidelines to patients with CKD in decreasing the risk of ESKD. Nevertheless, after a prolonged follow-up, Peeters et al [12] demonstrated that additional support by nurse practitioners, including lifestyle intervention, use of mandatory medication, and implementation of CKD guidelines reduced the risk of composite renal endpoints (death, ESKD and 50% increase in serum creatinine) by 20%. The care management performed by nurse practitioners at least quarterly resulted in a decrease in eGFR of 0.45 mL/min/1.73

m^2 per 1.73 m^2 per year in the intervention group compared with the control group [12]. In a study conducted in Taiwan, Barrett et al [13] showed that multidisciplinary education according to the National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (NKF/KDOQI) guidelines decreased the incidence of dialysis and reduced mortality in predialysis patients with CKD after a mean follow-up of 11.7 months. Recently, in their 3-month-prospective study, Li et al [14] reported that patients with CKD who received diet, exercise, and self-management education delivered via a wearable device and smartphone app experienced a slower rate of eGFR decline than those given only conventional care. These studies suggested that beyond traditional nephrologist-centered clinical consultations, proper education in combination with the careful management of patients with CKD could potentially improve kidney outcomes even in the short term. In the KidneyOnline intelligent patient care system, patients were educated by nurses who received training according to the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines. Other interventions including lifestyle intervention and education were all integrated into the KidneyOnline care system. Our study confirmed previous findings that nurse care could improve the renal outcomes of patients with CKD via patient education, disease interpretation, and lifestyle intervention.

Analysis of Findings

Our study demonstrated that the KidneyOnline care system reduced the risk of composite kidney end points, irrespective of age, baseline eGFR, and proteinuria. This effect was multifactorial. One of the main underlying reasons for this effect was the well-controlled blood pressure, which was significantly lower at 6 months compared with that at the time of enrollment in the KidneyOnline care group, and it remained at a lower level throughout the follow-up period. In China, the prevalence of hypertension is higher in patients with CKD compared with the general population, but awareness of hypertension in patients with CKD was reported to be 80.7% in 2018[15]. Previous studies have shown that adequate control of blood pressure was suboptimal in patients with CKD [16,17]. In fact, refractory hypertension in patients was largely attributed to inadequate adherence to prescribed medication [18]. Low medication adherence was associated with CKD progression [19,20]. It has been established that well-designed mobile apps could effectively improve medication adherence in cardiovascular disease [21,22]. In the KidneyOnline care system group, the trend graph and brief report about blood pressure, real-time online questions and answers, and medication intake suggestions were all potential factors that might have contributed to patients' good adherence. Moreover, constant online communication and laboratory test reminders possibly prompted participants' clinical visits to adjust their medication regimens in time, thus reducing the risk of adverse drug reactions. Finally, healthy lifestyle modifications, including sufficient physical activity, proper BMI control, smoking cessation, and healthy diet habits, were reported to help slow the progression of CKD in patients with preserved kidney function [23]. Participants joining the KidneyOnline care group often gave positive feedback on the smart reminder for helping them organize their diets and make healthy lifestyle choices.

Interestingly, there was an improvement in proteinuria in both the KidneyOnline intelligent care group and the conventional care group. This may be attributed to the KidneyOnline app's educational materials on the benefits of a low-salt diet. In a typical Chinese diet, the salt intake is often much higher than recommended. The KidneyOnline app enabled patients from both groups to access free educational materials on the importance of a low-salt diet, as well as tips on limiting salt intake. Moreover, the KidneyOnline app offered a variety of low-salt diet recipes, which might have helped increase patients' compliance with a low-salt diet. Additionally, there was a high percentage of patients with glomerulonephritis, with 25.7% (n=232) being treated with steroids. The use of steroids and immunosuppressants is another important factor leading to the improvement in proteinuria.

The rate of smartphone ownership has exploded in China during the last 10 years. Telehealth apps provide new opportunities to enhance self-management, behavior change, and medication adherence not only for people living in urban areas but also for those in rural areas. As Duan et al [24] have shown, the prevalence of CKD in China's rural areas was 16.4% between 2015 and 2017. Education level, personal income, alcohol consumption, and hypertension were all risk factors associated with insufficient kidney function. The KidneyOnline care system provides a web-based solution for patient-centered care and helps reduce the time and cost associated with long travels to seek medical advice. Thus, our mode of care could facilitate patients' self-management in a cost-effective manner, especially in areas facing a shortage of medical resources.

Strengths and Limitations

To our knowledge, this study was the first to assess the efficacy of a smartphone-based, nurse-led, patient-oriented management system for patients with CKD across China using hard kidney end points. The 4-year observational data from the real world demonstrated the efficacy of this newly developed system. Inevitably, our study had several limitations. First, there were over 70,000 user registrations on our app. Only 2060 patients' records fulfilled our criteria and were eligible for analysis. This could have resulted in selection bias in both groups. Patients with a stronger sense of self-management and motivation were more likely to be included in this study. Second, patients who had severe symptoms or advanced pathological lesions were less likely to choose our patient care system; instead, they received treatments in large hospitals on their first visits. Moreover, patients with minor renal impairment could potentially have a lower probability of joining the KidneyOnline care system. In addition, due to the cost associated with the KidneyOnline care system, patients who joined the system probably had a higher economic status compared with patients in the control group. Nonetheless, we have shown that patients in the KidneyOnline care group experienced better prognoses in terms of composite renal outcomes, irrespective of their age, baseline eGFR, and proteinuria. More data are needed to illustrate the efficacy of our mode of care in those with more impaired kidney function, as well as those with very slight renal impairment. Third, we did not analyze other possible end points related to the utilization of the KidneyOnline care system, such as cardiovascular comorbidities and hospitalization frequencies, which would help us better evaluate our mode of management for patients with CKD. Additionally, the participants were young and normotensive; there was a high prevalence of glomerulonephritis but a limited number of diabetes cases, and these factors may have limited the feasibility of the KidneyOnline care system for patients with CKD caused by diabetic nephropathy. It is also important to note that our study was conducted in China; because the education and economic levels of patients with CKD vary across countries, a new mode of patient care based on mobile phones in other countries is anticipated.

Conclusion

We developed a smartphone-based, nurse-led, patient-oriented management system to facilitate health care for patients with nondiabetic CKD in China. Through multifaceted patient care, our mode of patient management was associated with a reduced risk of composite kidney outcomes. This strengthened the evidence of telehealth interventions to promote kidney health and long-term management for patients with CKD.

Acknowledgments

We are grateful to all the patients and controls for their participation in this study.

Data Availability

The data sets analyzed in this study are not publicly available due to business confidentiality reasons but are available from the corresponding author upon reasonable request.

Conflicts of Interest

Authors JW, TZ, XC, SX, and FH are employed by the Beijing Kidney Health Technology Co Ltd. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Demonstration of KidneyOnline dashboard for patients and health coach team.
[PNG File , 279 KB - mhealth_v11i1e45531_app1.png]

Multimedia Appendix 2
Multivariate Cox regression analysis of composite kidney outcome (before patient matching). eGFR: estimated glomerular filtration rate; MAP: mean arterial pressure; RASB: renin-angiotensin-aldosterone system blocker.
[PNG File , 344 KB - mhealth_v11i1e45531_app2.png]

Multimedia Appendix 3
Forest plot of the comparison between the 2 different modes of care in different subgroups (after patient matching). eGFR: estimated glomerular filtration rate; MAP: mean arterial pressure.
[PNG File , 137 KB - mhealth_v11i1e45531_app3.png]

Multimedia Appendix 4
Changes in mean arterial pressure (MAP) during follow-up in the whole population and subgroups.
[PPTX File , 285 KB - mhealth_v11i1e45531_app4.pptx]

Multimedia Appendix 5
Changes in 24-hour proteinuria in the 2 matched groups of patients after propensity score matching (PSM) during follow-up.
[PNG File , 124 KB - mhealth_v11i1e45531_app5.png]

References

1. Ayodele OE, Alebiosu CO. Burden of chronic kidney disease: an international perspective. Adv Chronic Kidney Dis 2010 May;17(3):215-224. [doi: 10.1053/j.ackd.2010.02.001] [Medline: 20439090]
2. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004 Sep 23;351(13):1296-1305. [doi: 10.1056/NEJMoa041031] [Medline: 15385656]
3. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet 2012 Mar 03;379(9818):815-822. [doi: 10.1016/S0140-6736(12)60033-6] [Medline: 22386035]
4. Forbes A, Gallagher H. Chronic kidney disease in adults: assessment and management. Clin Med (Lond) 2020 Mar 12;20(2):128-132 [FREE Full text] [doi: 10.7861/clinmed.cg.20.2] [Medline: 32165439]

5. Tuot DS, Boulware LE. Telehealth applications to enhance CKD knowledge and awareness among patients and providers. Adv Chronic Kidney Dis 2017 Jan;24(1):39-45 [FREE Full text] [doi: 10.1053/j.ackd.2016.11.017] [Medline: 28224941]
6. Lewis RA, Lunney M, Chong C, Tonelli M. Identifying Mobile Applications Aimed at Self-Management in People With Chronic Kidney Disease. Can J Kidney Health Dis 2019;6:2054358119834283 [FREE Full text] [doi: 10.1177/2054358119834283] [Medline: 30899533]
7. Stevenson JK, Campbell ZC, Webster AC, Chow CK, Tong A, Craig JC, et al. eHealth interventions for people with chronic kidney disease. Cochrane Database Syst Rev 2019 Aug 06;8:CD012379. [doi: 10.1002/14651858.CD012379.pub2] [Medline: 31425608]
8. Wang F, He K, Wang J, Zhao M, Li Y, Zhang L, et al. Prevalence and risk factors for CKD: a comparison between the adult populations in China and the United States. Kidney Int Rep 2018 Sep;3(5):1135-1143 [FREE Full text] [doi: 10.1016/j.ekir.2018.05.011] [Medline: 30197980]
9. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. Lancet 2013 Jul 20;382(9888):260-272. [doi: 10.1016/S0140-6736(13)60687-X] [Medline: 23727169]
10. Yang C, Wang H, Zhao X, Matsushita K, Coresh J, Zhang L, et al. CKD in China: evolving spectrum and public health implications. Am J Kidney Dis 2020 Aug;76(2):258-264. [doi: 10.1053/j.ajkd.2019.05.032] [Medline: 31492486]
11. van Zuilen AD, Bots ML, Dulger A, van der Tweel I, van Buren M, Ten Dam MAGJ, et al. Multifactorial intervention with nurse practitioners does not change cardiovascular outcomes in patients with chronic kidney disease. Kidney Int 2012 Sep;82(6):710-717 [FREE Full text] [doi: 10.1038/ki.2012.137] [Medline: 22739979]
12. Peeters MJ, van Zuilen AD, van den Brand JA, Bots ML, van Buren M, Ten Dam MAGJ, et al. Nurse practitioner care improves renal outcome in patients with CKD. J Am Soc Nephrol 2014 Feb;25(2):390-398 [FREE Full text] [doi: 10.1681/ASN.2012121222] [Medline: 24158983]
13. Wu I, Wang S, Hsu K, Lee C, Sun C, Tsai C, et al. Multidisciplinary predialysis education decreases the incidence of dialysis and reduces mortality--a controlled cohort study based on the NKF/DOQI guidelines. Nephrol Dial Transplant 2009 Nov;24(11):3426-3433. [doi: 10.1093/ndt/gfp259] [Medline: 19491379]
14. Li W, Chiu F, Zeng J, Li Y, Huang S, Yeh H, et al. Mobile health app with social media to support self-management for patients with chronic kidney disease: prospective randomized controlled study. J Med Internet Res 2020 Dec 15;22(12):e19452 [FREE Full text] [doi: 10.2196/19452] [Medline: 33320101]
15. Yan Z, Wang Y, Li S, Wang J, Zhang L, Tan H, China National Survey of Chronic Kidney Disease Working Group. Hypertension control in adults with CKD in China: baseline results from the Chinese Cohort Study of Chronic Kidney Disease (C-STRIDE). Am J Hypertens 2018 Mar 10;31(4):486-494. [doi: 10.1093/ajh/hpx222] [Medline: 29304216]
16. Peralta CA, Hicks LS, Chertow GM, Ayanian JZ, Vittinghoff E, Lin F, et al. Control of hypertension in adults with chronic kidney disease in the United States. Hypertension 2005 Jun;45(6):1119-1124. [doi: 10.1161/01.HYP.0000164577.81087.70] [Medline: 15851626]
17. Plantinga LC, Miller ER, Stevens LA, Saran R, Messer K, Flowers N, Centers for Disease ControlPrevention Chronic Kidney Disease Surveillance Team. Blood pressure control among persons without and with chronic kidney disease: US trends and risk factors 1999-2006. Hypertension 2009 Jul;54(1):47-56 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.109.129841] [Medline: 19470881]
18. Burnier M, Santschi V, Favrat B, Brunner HR. Monitoring compliance in resistant hypertension: an important step in patient management. J Hypertens Suppl 2003 May;21(2):S37-S42. [doi: 10.1097/00004872-200305002-00007] [Medline: 12929906]
19. Tangkiatkumjai M, Walker D, Praditpornsilpa K, Boardman H. Association between medication adherence and clinical outcomes in patients with chronic kidney disease: a prospective cohort study. Clin Exp Nephrol 2017 Jun;21(3):504-512. [doi: 10.1007/s10157-016-1312-6] [Medline: 27438073]
20. Cedillo-Couvert EA, Ricardo AC, Chen J, Cohan J, Fischer MJ, Krousel-Wood M, CRIC Study Investigators. Self-reported medication adherence and CKD progression. Kidney Int Rep 2018 May;3(3):645-651 [FREE Full text] [doi: 10.1016/j.ekir.2018.01.007] [Medline: 29854972]
21. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
22. Li Y, Gong Y, Zheng B, Fan F, Yi T, Zheng Y, et al. Effects on adherence to a mobile app-based self-management digital therapeutics among patients with coronary heart disease: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 15;10(2):e32251 [FREE Full text] [doi: 10.2196/32251] [Medline: 34906924]
23. Ricardo AC, Anderson CA, Yang W, Zhang X, Fischer MJ, Dember LM, CRIC Study Investigators. Healthy lifestyle and risk of kidney disease progression, atherosclerotic events, and death in CKD: findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2015 Mar;65(3):412-424 [FREE Full text] [doi: 10.1053/j.ajkd.2014.09.016] [Medline: 25458663]
24. Duan J, Wang C, Liu D, Qiao Y, Pan S, Jiang D, et al. Prevalence and risk factors of chronic kidney disease and diabetic kidney disease in Chinese rural residents: a cross-sectional survey. Sci Rep 2019 Jul 18;9(1):10408 [FREE Full text] [doi: 10.1038/s41598-019-46857-7] [Medline: 31320683]

Abbreviations

AI: artificial intelligence
CKD: chronic kidney disease
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
eGFR: estimated glomerular filtration rate
ESKD: end-stage kidney disease
HR: hazard ratio
KDIGO: Kidney Disease: Improving Global Outcomes
MAP: mean arterial pressure
NKF/KDOQI: National Kidney Foundation's Kidney Disease Outcomes Quality Initiative
OCR: optical character recognition
PSM: propensity score matching
RASB: renin-angiotensin-aldosterone system blocker

Edited by L Buis; submitted 05.01.23; peer-reviewed by WY Li, C Yang; comments to author 21.02.23; revised version received 12.03.23; accepted 24.04.23; published 01.06.23.

Please cite as:
Liu W, Yu X, Wang J, Zhou T, Yu T, Chen X, Xie S, Han F, Wang Z
Improving Kidney Outcomes in Patients With Nondiabetic Chronic Kidney Disease Through an Artificial Intelligence–Based Health Coaching Mobile App: Retrospective Cohort Study
JMIR Mhealth Uhealth 2023;11:e45531
URL: https://mhealth.jmir.org/2023/1/e45531
doi:10.2196/45531
PMID:37261895

Original Paper

Impact of Mobile Apps in Conjunction With Percutaneous Endoscopic Gastrostomy on Patients' Complications, Quality of Life, and Health-Related Self-Care Behaviors: Randomized Clinical Trial

Bi-Lian Chen[1,2], MSc; Han-Chung Lien[3], MD, PhD; Shyh-Sheng Yang[4], MD; Shiao-Chi Wu[5], PhD; Hsien-Hsien Chiang[1], MSc; Li-Chan Lin[6], PhD

[1]Department of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[2]Department of Nursing, Taichung Veterans General Hospital, Taichung, Taiwan
[3]Division of Gastroenterology, Taichung Veterans General Hospital, Taichung, Taiwan
[4]The Division of Thoracic Surgery, Taichung Veterans General Hospital, Taichung, Taiwan
[5]Institute of Health and Welfare Policy, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[6]Department of Nursing, Asia University, Taichung, Taiwan

Corresponding Author:
Li-Chan Lin, PhD
Department of Nursing
Asia University
500 Lioufeng Road
Wufeng
Taichung, 41354
Taiwan
Phone: 886 4 23323456 ext 1827
Fax: 886 4 23390193
Email: lichan2009@gmail.com

Abstract

Background: Percutaneous endoscopic gastrostomy (PEG) is commonly chosen for long-term enteral nutrition support. However, common complications of PEG include wound infection, leakage, obstruction, bleeding, dislodgement, pneumonia, peritonitis, and more. The anticipation of these complications by both patients and their family caregivers underscores the essential requirement of ongoing technical guidance for the daily care of PEG and the adoption of preventative strategies.

Objective: This study aimed to establish and compare a health education program utilizing a tracking system for PEG using a mobile app (PEG app) and instant messaging software versus a paper-based health education program with instant messaging software. Their effectiveness in preventing complications, avoiding hospital readmissions, improving self-care practices, and enhancing quality of life outcomes was assessed.

Methods: A randomized controlled trial design was used, and the study sample consisted of patients from a medical center in central Taiwan who underwent thoracic surgery or gastroenterology procedures. Inclusion criteria were being a new case undergoing his or her first gastric tube insertion and having the ability to operate a smartphone. Exclusion criteria were cases requiring tube replacement or nasogastric tubes. A total of 74 participants were enrolled, with 37 participants in the experimental group and 37 participants in the control group. Data collection took place from hospitalization until 1 month after discharge. The experimental group received care using the gastric tube tracking system (PEG app) and the Line app that included phone, text, and photo capture capabilities, while the control group received routine nursing care and used the Line app.

Results: The experimental group demonstrated a significant reduction in the occurrence of complications compared with the control group (χ^2_1=12.087, P=.001). Specifically, the occurrence of leakage events was significantly lower in the experimental group than in the control group (χ^2_1=12.906, P=.001). However, the experimental group exhibited superior self-care ability compared with the control group (t_{72}=2.203, P=.03). There was no significant difference in overall quality of life scores between the experimental and control groups (t_{72}=1.603, P=.11). However, the experimental group showed better social aspects of quality of life than the control group (t_{72}=2.164, P=.03).

Conclusions: Integration of the PEG app with instant messaging can enhance self-care ability, improve social aspects of quality of life, and reduce complications. The study results suggest that the PEG app could be used as an adjunct tool to promote patients' self-directed management of their gastric tube at home, particularly for patients who have undergone their first PEG placement and are being discharged from the hospital.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300071271; https://tinyurl.com/4vvy584e

(JMIR Mhealth Uhealth 2023;11:e48970) doi:10.2196/48970

KEYWORDS

percutaneous endoscopic gastrostomy; mobile applications tracking system; self-care; complications; quality of life; mobile application; mHealth app; mHealth intervention; health promotion; health education; endoscopy; application; education; gastrostomy; care; prevention; behavior; tracking; utilization

Introduction

Percutaneous endoscopic gastrostomy (PEG) is a preferred method for long-term enteral nutrition support. Studies indicate that 5% to 7.8% of nursing home residents receive gastrostomy feeding [1-3]. Wong et al [4] found that 2.5% of residents in Singapore's long-term care facilities were using a gastrostomy tube. In Taiwan, cultural factors and the lack of reimbursement for PEG procedures by the National Health Insurance before 2009 contributed to a reported prevalence rate of 0.4% for PEG use, with percutaneous endoscopic jejunostomy use in long-term care facilities ranging from 0.1% to 0.3% [5]. Although these figures suggest a lower prevalence in Asian countries compared with Western countries, it is conceivable that there will be an increased use of PEG for long-term enteral feeding in the near future. This prediction is based on the availability of health education and coverage for PEG insertion by the National Health Insurance in Taiwan as of 2009.

Common complications of PEG include wound infection (3%-50%), leakage (10%-42.3%), obstruction (8%-35%), bleeding (32%), dislodgement (14.3%), pneumonia, peritonitis (1%-18%), and buried bumper syndrome (1.5%-0.8%) [6-10]. A study by Ang et al [11] that involved interviews of 18 patients and caregivers about their experience with PEG tube use revealed that home-based patients with PEG and their caregivers often feel anxious about insufficient self-care knowledge and information about what to expect after tube insertion. They particularly fear the occurrence of complications. Farrag et al [12] also noted that handling the complexities of enteral feeding techniques requires ongoing technical guidance from someone to explain the daily care of PEG and how to avoid complications, as well as to observe signs of complications and promptly report them to health care personnel for appropriate management.

With the advancement of technology and the widespread use of smartphones, instant messaging apps, which can be downloaded for free on smartphones or computers via the internet, have transformed health education from static paper-based methods to dynamic audiovisual and video formats. These apps provide real-time voice and visual information, allowing one-on-one or group video chats; the exchange of photos, audio, and video; and real-time online question and answer functionality. However, despite the rapidity and convenience of instant messaging as a means of communication, when applied to disease prevention and health education campaigns, there is often a lack of standardized guidelines within the medical field [13-17].

As a result, medical apps, including medical informatics apps, have emerged as a new educational approach. These apps offer advantages of standardization, privacy, data confidentiality, and systematic management [18]. The National Institute for Health and Care Excellence in the United Kingdom defines mobile health (mHealth) as the use of software downloaded on smartphones or computers via the internet for the exchange of calls as well as voice, text, and image messages [19,20]. Health care providers can use mHealth to track, monitor, and analyze patients' health data in real time, providing feedback on their health status [21]. This approach can improve patient compliance, strengthen patient-centered self-management of health [22], reduce emergency room visits [23], and decrease health care expenses resulting from complications [24,25].

These effects result from integrating smartphone apps with technology that changes people's lifestyles and behavior [26]. The support by Wong et al [27-29] for the development of mHealth apps, particularly those grounded in a health-social approach model, underscores their potential to improve self-care, quality of life, and health outcomes for community-dwelling older adults and patients with chronic diseases. These apps may achieve this by enhancing self-efficacy, promoting sustained behavior change, encouraging regular app use, leveraging visual communication, and reducing reliance on traditional health care services.

Medical information apps have demonstrated effectiveness for privacy and confidentiality, self-health management, and enhancing self-care capabilities. We conducted a systematic literature review with specific search criteria combining the keywords "percutaneous endoscopic gastrostomy" and "mHealth application" using the Boolean operator "AND" and restricting the publication year range from 1995 to 2023. These criteria were applied to the following 4 databases: CINAHL, ProQuest, PubMed, and the Cochrane Library. Among the databases scrutinized, only ProQuest yielded 2 conference abstracts that were relevant to the convergence of PEG and mHealth apps within the specified publication year range. The literature search failed to uncover any articles specifically addressing the use of mHealth apps in the context of PEG procedures.

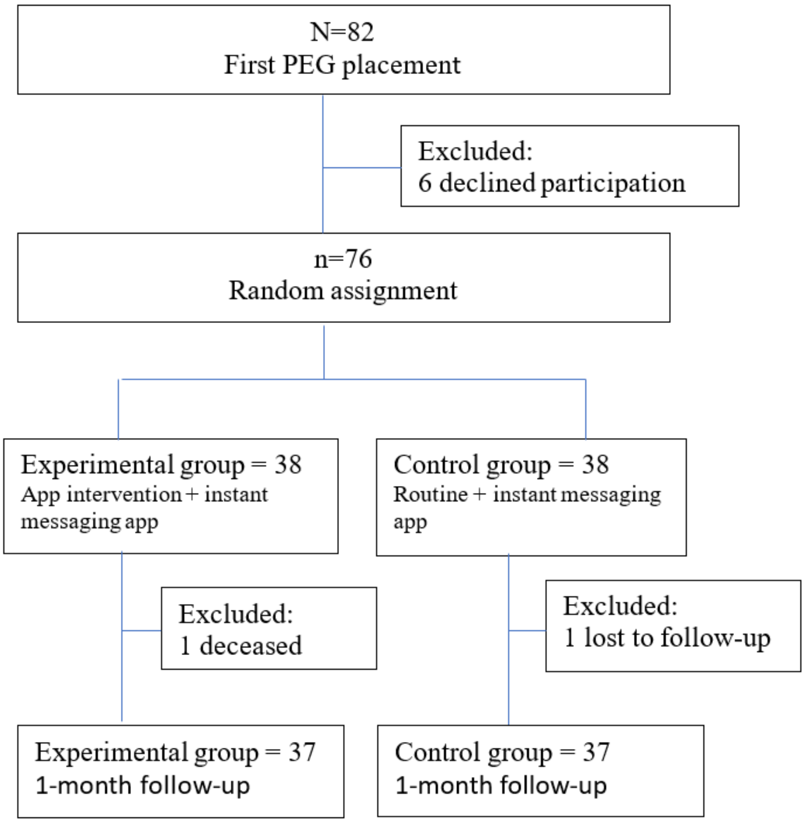
As such, the aim of this study was to establish and compare the effectiveness of a mobile app for PEG integrated with real-time communication software with that of a paper-based health education program also integrated with real-time communication software. This study evaluated the implications on several key health-related outcomes, including complications and self-care, as well as perceived well-being in terms of quality of life. Additionally, the study assessed the impact on the use of health care services, specifically looking at hospital or emergency readmissions.

Methods

Study Design

This study used a randomized controlled trial design with a study period spanning from August 7, 2018, to December 31, 2022. Random allocation was achieved using computer-generated grouping, with numbers assigned and placed in sealed envelopes. Envelopes were drawn by cases at the time of enrollment to determine group allocation, continuing until the estimated sample size was reached, as illustrated in Figure 1.

Figure 1. Flow of participants for the study of an app intervention for PEG. PEG: percutaneous endoscopic gastrostomy.



Recruitment

A total of 82 patients underwent initial catheter placement. After excluding 6 patients who refused participation, 76 patients were randomized and assigned to respective groups, with 38 patients in each group for the study. During the 1-month follow-up period, there was 1 death owing to complications of chemotherapy in the experimental group, and 1 patient was lost to follow-up in the control group, resulting in 37 patients in each group who completed the study.

Research Site and Participants

The study was conducted at a prestigious medical center located in the central region of Taiwan. Inclusion criteria for the study were as follows: (1) patients who were undergoing their first PEG placement; (2) patients who had access to mobile devices such as smartphones, tablets, or computers; (3) patients or their caregivers who were able to operate mobile devices; (4) patients who could upload data according to the designated schedule; and (5) patients who could communicate in English or Taiwanese. Exclusion criteria were as follows: (1) patients with nasogastric tubes and (2) patients who required replacement of gastrostomy or nasogastric tubes during the study period.

Sample Size Estimation for Power Calculation

The sample size estimation for power calculation was grounded in the primary outcome measure, which focused on the most common complication—wound infection. This approach was inspired by the work of Pattison and Young [30], who indicated that the infection rates were 31% in the context of an educational booklet intervention and 58% in the control group. In the Department of General Surgery at the University of Washington, the readmission rate because of postoperative wound infection within 1 month after an intervention with a mobile app was 38% (5/13), according to Sanger et al [31]. According to the Power and Sample Size Program, sample size estimation was performed using the occurrence of wound infection complications as the outcome indicator for both the experimental and control groups treated as independent samples. With expected rates of 38% in the experimental group and 58% in the control group, an α level of .05, and a β level of .9, the calculated sample size was 32 patients in each group, for a total of 64 patients. Considering an estimated loss to follow-up rate of 10 patients (15%), a total of 74 patients were planned to be enrolled in the study.

Randomization and Groups

The research tool involves the completion of a “Participant Consent Form” by the patient or their legal representative, followed by random allocation to either the intervention group or the control group through a drawing of lots. Patients in the intervention group had the mobile device app “Percutaneous Endoscopic Gastrostomy Tracking System (HTML 5)“ (PEG app) installed during their hospitalization.

The intervention group received the PEG app and Line app (for instant messaging) for guidance. The protocol encompassed a teaching and feedback process administered by the principal investigator. This process involved daily changes of the wound dressing and applying sterile gauze, which were conducted by the patient or their caregiver postoperation. Additionally, the process included ongoing monitoring of the wound's condition and regular measurement of body temperature and body weight for a duration of 1 week. Patients were instructed to self-report any symptoms related to their wound, including but not limited to redness, swelling, heat, pain, discharge, and odor, using the designated mobile app. They were encouraged to upload daily wound photos until the completion of 1 week after the procedure. In addition to the aforementioned wound symptom reporting, patients were asked to complete a self-care knowledge questionnaire as a pretest. Furthermore, in the fourth week following tube placement, patients were instructed to complete questionnaires related to self-care knowledge, self-care ability, and their quality of life through the designated mobile app. In the event of an emergency, such as visits to the emergency department within 3 days or hospital readmissions within 14 days, patients or their family members were instructed to use the mobile app to report the situation on a daily basis. The data collected through the app were then accessed by the principal investigator through a designated website. The Line app (instant messaging) provided real-time communication for patients or family members to address any caregiving issues, including the ability to conduct voice or video calls and upload wound photos for immediate response.

The control group received traditional paper-based gastric tube education handouts, which include information on the purpose, indications, daily care instructions, and complications. Additionally, for ethical consideration, the Line app (instant messaging) was used for real-time communication with patients or family members to address any caregiving issues, including immediate response through voice or video calls and uploading wound photos for prompt evaluation.

Research Instruments and Assessment of Reliability and Validity

Regarding hardware, the mobile devices included smartphones, tablets, computers, a server, a thermometer, and a scale. Regarding software, the PEG tracking system was compatible with Android or iOS mobile devices or desktop computers with the Windows operating system, and the Line app allowed real-time communication. The research team installed the ”Active Server Pages (the Gastric Tube Monitoring Platform)“ through the Google browser, which automatically sent alerts to the platform regarding patients’ ”elevated body temperature“ and ”signs of wound infection.“ The assessment tools included basic patient information, physiological monitoring forms, wound assessment forms and complication reporting, self-care knowledge, self-care ability, and quality of life measures.

The app interface (prototype) was designed based on a needs analysis of patients, family members, and clinical nurses. The researchers used draw.io software to create the interface, which included the login page, the home page, basic information, wound inspection form, temperature trend chart, gastric tube wound photography, recognition of complications, complications reporting, question-and-answer functionality, quality of life questionnaire, self-care knowledge questionnaire, and self-care ability questionnaire.

Outcome Measures

Data collection took place at 2 specific time points: initially, before the intervention began to gather demographic information and assess self-care knowledge. Subsequently, data collection occurred after the intervention, within 1 week of the completion of the 1-month study period, to assess complications, quality of life, self-care knowledge, and self-care ability.

Primary Outcomes

Following the health-social approach model, complications and quality of life were the primary outcome measures.

Complications

Complications including wound infection, leakage, migration, pneumonia, granulation, bleeding, admission to the emergency department within 3 days, and hospital readmission within 14 days were assessed.

World Health Organization Quality of Life Brief Version

In 1995, the World Health Organization (WHO) established a quality of life questionnaire (WHOQOL) that includes the following 4 major domains: psychological, physiological, social, and environmental. The domains use a 5-point Likert scale. A score of 1 represents ”very dissatisfied,“ 2 represents ”dissatisfied,“ 3 represents ”neutral,“ 4 represents ”satisfied,“ and 5 represents ”very satisfied“ [32]. The Taiwan version of the WHOQOL questionnaire is a concise version, with 26 items and 2 additional local items, totaling 28 items (WHOQOL-BREF). Translated by Yao et al [33], the questionnaire has shown good internal consistency and reliability, with a Cronbach α of .91 and content validity ranging from 0.53 to 0.78.

Secondary Outcomes

Self-Care Knowledge Assessment

The self-care knowledge assessment primarily evaluated the level of understanding by individuals or primary caregivers about self-care for gastric tubes, based on the following 4 dimensions: wound care, tube feeding, complications, and emergency medical care. A self-designed questionnaire with 10 items assessed self-care knowledge, with correct answers given a score of 1 and incorrect or unknown answers given a score of 0. A higher score indicates better self-care knowledge, with a total possible score of 10 and a minimum score of 0. The content validity index (CVI) during the first round of expert review was 0.58. Following semantic modifications to the eighth item, “emergency situations related to gastric tubes,” to remove compound options, as suggested by the experts, the second round of expert review yielded a CVI of 1. The internal consistency of the ”self-care knowledge scale“ was determined to be Cronbach α=.66.

Self-Care Ability

The self-designed questionnaire for self-care ability primarily assessed the level of competence in performing gastric tube care behaviors by patients or primary caregivers. It covers the following 3 dimensions: measures for tube feeding, prevention of complications, and observation of symptoms of complications. The self-designed questionnaire consists of 11 items, which are rated using a 5-point Likert scale, with response options of ”always“ (5 points), ”often“ (4 points), ”sometimes“ (3 points), ”rarely“ (2 points), and ”never“ (1 point). After revisions, the second round of expert review yielded a CVI of 0.98. The internal consistency reliability of the ”self-care ability scale“ was Cronbach α=.78.

Ethical Considerations

This study was approved by the Human Research Ethics Review Committee of Taichung Veterans General Hospital (on August 7, 2018; institutional review board number: CF18189A-1).

Data Processing and Statistical Analysis

Descriptive analysis was performed using SPSS 26.0, while inferential statistical analysis involved chi-square tests, independent *t* tests, and paired *t* tests to analyze categorical or continuous variables to examine whether there were differences between the 2 groups being compared.

Results

The mean ages of the experimental and control groups were 60.63 (SD 10.29) years and 60.79 (SD 9.63) years, respectively. In the experimental group, there were 35 men (35/38, 92%); in the control group, there were 31 men (31/38, 82%). There were no significant differences between the 2 groups regarding basic demographic characteristics, as shown in Table 1.

In addition to demographic data, self-care knowledge was the sole research variable that had a pretest assessment. During the study, 1 participant passed away, and another was lost to follow-up, accounting for a small proportion of missing data (2/76, 3%). To address this limited missing data, imputation was conducted by utilizing the mean self-care knowledge in the posttest for both the experimental and control groups. Table 2 shows that the posttest knowledge scores in the experimental group and control group were significantly higher than the pretest scores, with significant differences observed in the paired *t* tests (experimental group: t_{37}=3.99, *P*=.001; control group: t_{37}=4.75, *P*<.001). Upon further analysis of the change in self-care knowledge scores from pre to postintervention, the experimental group exhibited a larger change in scores than the control group. However, it is important to note that this difference did not reach statistical significance.

Table 3 shows that 5 individuals in the experimental group and 19 individuals in the control group experienced complications. The incidence of complications was significantly lower in the experimental group than in the control group (χ^2_1=12.087, *P*=.001). Further analysis revealed that the control group had a higher incidence of complications, including infection, leakage, displacement, pneumonia, bleeding, and granulation tissue formation, than the experimental group. However, only the incidence of leakage was significantly lower in the experimental group than in the control group (χ^2_1=12.906, *P*=.001). In addition, there were no significant differences between the 2 groups regarding visits to the emergency department within 3 days and hospital readmission within 14 days.

In terms of self-care ability, the experimental group demonstrated significantly better self-care ability than the control group (t_{72}=2.203, *P*=.03), as indicated by higher scores on the self-care ability scale. Higher scores on the scale represent better self-care ability.

In terms of quality of life, there were no significant differences in total scores between the experimental and control groups. However, upon further examination of the quality of life in terms of physiological, psychological, social, and environmental aspects, the experimental group showed significantly higher scores in the social domain than the control group (t_{72}=2.164, *P*=.03). At the same time, there were no differences in other domains, as shown in Table 4.

Table 1. Participant characteristics (N=76).

Characteristics	Control group (n=38)	Experimental group (n=38)	Statistical test (*t* test or chi-square test; *df*)	*P* value
Age (years), mean (SD)	60.79 (9.63)	60.63 (10.29)	0.06 (74)[a]	.95
BMI (kg/m^2), mean (SD)	21.04 (4.19)	21.93 (4.24)	0.92 (74)[a]	.36
Sex, n (%)			1.84 (1)[b]	.18
Female	7 (18)	3 (8)		
Male	31 (82)	35 (92)		
Chemotherapy, n (%)	33 (87)	34 (90)	0.126 (1)[b]	.72
Radiotherapy, n (%)	34 (90)	34 (90)	0 (1)[b]	>.99
Caregiving responsibility, n (%)			4.41 (2)[b]	.11
Spouse	24 (63)	15 (40)		
Child or grandchild	10 (26)	15 (40)		
Other	4 (11)	8 (21)		
Patient education, n (%)			2.65 (4)[b]	.62
Illiterate	2 (5)	1 (3)		
Primary school	5 (13)	7 (18)		
Junior high school	11 (29)	7 (18)		
Senior high school	20 (53)	22 (58)		
Undergraduate or higher	0 (0)	1 (3)		
Caregiver education (n=71), n (%)			6.62 (5)[b]	.20
Illiterate	1 (3)	0 (0)		
Primary school	2 (5)	0 (0)		
Junior high school	5 (14)	2 (6)		
Senior high school	17 (46)	14 (41)		
Undergraduate	11 (30)	13 (38)		
Graduate school or higher	1 (3)	5 (15)		
NG[c] used, n (%)	12 (32)	12 (32)	0 (1)[b]	>.99

[a]*t* test.
[b]Chi-square test.
[c]NG: nasogastric.

Table 2. Comparison of knowledge between pre and posttests for the experimental and control groups (N=76).

Group[a]	Pretest score	Posttest score	MD[b]	Paired *t* test (*df*)	*P* value
Control group	7.03 (1.61)	8.29 (1.22)	1.26	4.75 (37)	<.001
Experimental group	7.08 (1.53)	8.61 (1.79)	1.52	3.99 (37)	.001

[a]Comparison between the groups: t_{74}=0.81, *P*=.42.
[b]MD: mean difference.

Table 3. Comparison of complications between the experimental and control groups (N=74).

Complications	Control group (n=37), n (%)	Experimental group (n=37), n (%)	Chi-square (*df*)	*P* value
Complication of any type (n=24)	19 (51)	5 (14)	12.09 (1)	.001
Infection	7 (19)	3 (8)	1.85 (1)	.17
Leakage	15 (41)	2 (5)	12.91 (1)	.001
Migration	1 (3)	0 (0)	—[a]	>.99
Pneumonia	1 (3)	0 (0)	—[a]	>.99
Granulation	3 (8)	0 (0)	—[a]	>.99
Bleeding	1 (3)	0 (0)	—[a]	>.99
Admission to the emergency department (3 days)	1 (3)	0 (0)	—[a]	>.99
Hospital readmission (14 days)	2 (5)	2 (5)	—[a]	>.99

[a]Not applicable because Fisher exact tests were conducted.

Table 4. Comparison of self-care ability and quality of life (QOL) between the experimental and control groups (N=74), assessed using independent *t* tests.

Self-care ability and quality of life	Control group (n=37), mean (SD)	Experimental group (n=37), mean (SD)	*t* test (*df*)	*P* value
Ability for self-care of the PEG[a]	47.49 (5.63)	50.30 (5.33)	2.200 (72)	.03
QOL				
Physical	25.27 (5.69)	26.27 (4.50)	0.837 (72)	.41
Psychological	19.51 (3.71)	20.57 (3.67)	1.220 (72)	.22
Social	33.05 (6.70)	35.97 (4.72)	2.164 (72)	.03
Environment	14.11 (2.72)	14.76 (2.37)	0.567 (72)	.28
Total QOL score	91.95 (16.81)	97.57 (13.12)	1.603 (72)	.11

[a]PEG: percutaneous endoscopic gastrostomy.

Discussion

Principal Findings

The research findings demonstrated that using the PEG and Line apps can improve self-care behaviors, reduce complications, and enhance quality of life at the social level. The experimental group demonstrated significantly better self-care abilities than the control group. In terms of behavior change and habit formation, the experimental group utilized the app platform to proactively capture and report abnormal wound information, such as redness, swelling, heat, and pain, through daily wound photos, wound infection checks, and vital sign data feedback. Researchers received email notifications and were able to immediately observe changes in the patients' recent wound condition, particularly the quantity, color, and location of secretions, as well as slight fever symptoms indicated by body temperature, on the app platform. This early warning system enabled patients to increase the frequency of wound dressing changes and minimize further deterioration of wound infections. Both groups were knowledgeable about daily care and complication reporting, which is consistent with the findings by Ang et al [11] and Farrag et al [12] from interviews with 29 patients and their families on their experience with gastrostomy tubes, in which red flags for complications and seeking medical resources for tube-related issues were reported.

The experimental group experienced significantly fewer complications than the control group. Specifically, in the experimental group, there were 5 cases of complications (3 cases of infection and 2 cases of leakage) reported through daily wound checks and vital sign measurements. In comparison, the control group had 19 cases of complications (15 cases of leakage and 7 cases of infection, with 1 individual possibly experiencing 2 complications concurrently). The significant reduction in leakage complications in the experimental group can be attributed to the gastrostomy tube app platform, which primarily focused on medical care and allowed continuous uploading of daily data (wound symptoms, vital signs, complication reporting, and health care personnel's analysis of tube changes). In contrast, the control group relied on ad hoc communication to identify and report abnormalities, leading to intermittent photo uploads. The continuous uploading feature for the experimental group enabled a comparison of wound progression over time, facilitating timely detection of differences in care and adjustment of dressing change methods. Patients gained confidence in self-care through social support, changed their health behaviors, and developed a habit of daily reporting on their gastrostomy tube care. This early detection of leakage and prompt intervention helped reduce the risk of skin breakdown and physiological discomfort, such as pain. Furthermore, the lack of difference in wound infection rates between the 2 groups can be attributed to the small number of cases and the consistent use of antibiotics before and after surgery in both groups. Similarly, the lack of significant differences in emergency room visits and hospital readmissions in both groups can be attributed to the small number of cases, making it difficult to achieve statistically significant results. It is recommended that future research studies consider increasing the sample size to address this limitation.

Regarding quality of life, there was no significant difference in total scores between the experimental and control groups. This is primarily because both groups had a comparable number of patients with esophageal cancer who underwent chemotherapy or radiotherapy, which has a significant physiological and psychological impact, as highlighted by Farrag et al [12]. A series of chemotherapy and radiotherapy treatments has similar physiological effects and disease progression, with 50% of patients with head and neck tumors experiencing weight loss of 6.5% to 15% and unexpected pain after chemotherapy [12]. This is consistent with findings by Schneider et al [34], and both groups of patients experienced similar physiological challenges.

However, the experimental group had significantly higher scores in terms of quality of life at the social level than the control group. This could be attributed to the ”cost-benefit decision-making and self-efficacy in adopting healthy behaviors through the introduction of technologies such as health app usage, resulting in individual health and societal behavioral changes,” as indicated by Wong et al [27-29] with the health-social approach model and benefit of video communication. Regarding the patients’ family member experiences, they expressed that “using the gastric tube app to obtain daily information about the complications and questions and answers related to the gastric tube has been helpful and provides a sense of security.” The advantages and benefits of acquiring knowledge about gastric tube care through the app outweigh the disadvantages. Patients' changes in awareness can lead to progression to a point at which they enhance self-efficacy in gastric tube care by consistently implementing gastric tube care behaviors through regular app usage. A patient described the following:

> *When the tumor (cancer) was discovered, it was a sudden shock, with limited knowledge about health education and future treatment directions. Fortunately, the app helped record and take photos of the wound, etc. It helped me get through the toughest days. Recording on the app every day filled me with hope and confidence. With app management, we can discover overlooked steps and compare daily records.*

This implies that, with guidance from health care professionals and through recording and quality control, patients regain confidence and self-efficacy. This is consistent with the findings by Singh et al [35] from a semistructured interview with 22 patients with spinal cord injuries, which identified the following 3 characteristics: a desire to attain health, efforts to learn and use apps, and acceptance of social support [35,36].

This study confirms the behavior change theory of the PEG app, in which patients initially set the goal of wound healing and learn about gastric tube care knowledge through the app. On a daily basis, they self-report their gastric tube health, pay attention to complications, compare changes in their gastric tube, and make timely corrections for any oversights. The social resources of feedback, guidance, and clarification from health care providers facilitate changes in patients’ self-directed health behaviors.

The lack of a significant difference in PEG knowledge between the experimental and control groups may be attributed to a couple of possible reasons. (1) Both groups had smartphones and could access PEG knowledge online and through patient support group chats, resulting in improved knowledge after intervention for both groups, and (2) both groups had instant messaging apps for immediate access to answers to questions, which could also be a contributing factor to the lack of a significant difference in knowledge. In terms of nursing application, the research findings demonstrate that the gastric tube action tracking system dynamically improved patients’ self-care knowledge, skills, and quality of life related to gastric tube management while reducing the incidence of complications and emergency department visits. Following the principle of a one-stop service, it is recommended to integrate the PEG app into the case management workflow of the hospital system, from inpatient to discharge, and expand it to cancer centers in parallel, achieving seamless integration of mHealth app information across hospital platforms, which brings substantial benefits to patients in terms of continuity of care, resulting in a win-win-win outcome. In terms of data collection, we suggest using big data approaches to collect wound evolution, vital signs, infection indicators, wound photos, and other features for early prediction of gastric tube wound infection through artificial intelligence algorithms.

Limitations

In terms of research limitations, the participants knew to which group they belonged, which might have introduced potential biases and limitations to the objectivity of data collection, as blinding was not possible. Therefore, for future studies, it is recommended to involve personnel other than the researchers themselves in data collection to minimize researcher-related effects. With respect to research tools, considering the time and cost constraints, we suggest following the approach proposed by Cheng et al [37] to enable data exchange between the app care messages and hospital electronic medical record to efficiently transmit patients’ home-based information to the hospital, reduce unnecessary physician inquiries and ineffective examination costs, and promote the concept of remote care in smart health care. Additionally, the feature of sending reminder messages through the app required payment of SMS text messaging fees. However, due to insufficient funds, this feature did not function as intended. Instead, we resorted to using a direct phone line for notifications.

Conclusions

The group using the PEG app indicated significant improvements in self-care abilities, reduced complications, and better social support in terms of quality of life compared with the control group. These findings highlight the benefits of the PEG app for home-based postoperative care of patients with gastrostomy tubes. This study suggests that patients discharged after their first PEG placement should be assisted in using the PEG app to facilitate independent self-management of their gastrostomy tubes at home.

Acknowledgments

The authors express our gratitude to Taichung Veterans General Hospital for permitting us to recruit research participants and conduct the study on their premises, as well as to the 5 clinical experts who offered their feedback on the instrument.

Authors' Contributions

LCL and BLC contributed to the study design and implementation, data analysis, interpretation of the findings, and preparation of the manuscript. HCL and SSY contributed to the study design and identified the clinical cases. SCW contributed to data analysis and interpretation of the findings. HHC contributed to the study design.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 81 KB - mhealth_v11i1e48970_app1.pdf]

References

1. Law A, Stevens J, Walkey AJ. Gastrostomy tube use in the critically ill, 1994–2014. Annals ATS 2019 Jun;16(6):724-730. [doi: 10.1513/annalsats.201809-638oc]
2. Becker W, Hilbert T. [Gastrostromy tube feeding of elderly inpatients in Bremen]. Gesundheitswesen 2004 Dec;66(12):806-811. [doi: 10.1055/s-2004-813838] [Medline: 15609216]
3. Wade DT. How many patients in a prolonged disorder of consciousness might need a best interests meeting about starting or continuing gastrostomy feeding? Clin Rehabil 2018 Nov 21;32(11):1551-1564. [doi: 10.1177/0269215518777285] [Medline: 29779405]
4. Wong A, Sowa P, Banks M, Bauer JD. Home enteral nutrition in Singapore's long-term care homes-incidence, prevalence, cost, and staffing. Nutrients 2019 Oct 17;11(10):2492 [FREE Full text] [doi: 10.3390/nu11102492] [Medline: 31627289]
5. Lin L, Wu S, Chen HS, Wang TG, Chen MY. Prevalence of impaired swallowing in institutionalized older people in taiwan. J Am Geriatr Soc 2002 Jun;50(6):1118-1123. [doi: 10.1046/j.1532-5415.2002.50270.x] [Medline: 12110075]
6. Anderloni A, Di Leo M, Barzaghi F, Semeraro R, Meucci G, Marino R, et al. Complications and early mortality in percutaneous endoscopic gastrostomy placement in lombardy: a multicenter prospective cohort study. Dig Liver Dis 2019 Oct;51(10):1380-1387. [doi: 10.1016/j.dld.2019.03.024] [Medline: 31010743]
7. Blumenstein I, Shastri Y, Stein J. Gastroenteric tube feeding: techniques, problems and solutions. World J Gastroenterol 2014 Jul 14;20(26):8505-8524 [FREE Full text] [doi: 10.3748/wjg.v20.i26.8505] [Medline: 25024606]
8. Boeykens K, Duysburgh I, Verlinden W. Prevention and management of minor complications in percutaneous endoscopic gastrostomy. BMJ Open Gastroenterol 2022 Jul;9(1):e000975 [FREE Full text] [doi: 10.1136/bmjgast-2022-000975] [Medline: 35851280]
9. Gang M, Kim JY. Short-term complications of percutaneous endoscopic gastrostomy according to the type of technique. Pediatr Gastroenterol Hepatol Nutr 2014 Dec;17(4):214-222 [FREE Full text] [doi: 10.5223/pghn.2014.17.4.214] [Medline: 25587521]
10. Mobily M, Patel JA. Palliative percutaneous endoscopic gastrostomy placement for gastrointestinal cancer: Roles, goals, and complications. World J Gastrointest Endosc 2015 Apr 16;7(4):364-369 [FREE Full text] [doi: 10.4253/wjge.v7.i4.364] [Medline: 25901215]
11. Ang S, Lim S, Lim ML, Ng XP, Madeleine L, Chan MM, et al. Health care professionals' perceptions and experience of initiating different modalities for home enteral feeding. Clin Nutr ESPEN 2019 Apr;30:67-72. [doi: 10.1016/j.clnesp.2019.02.005] [Medline: 30904231]
12. Farrag K, Shastri Y, Beilenhoff U, Aksan A, Stein J. Percutaneous endoscopic gastrostomy (PEG): a practical approach for long term management. BMJ 2019 Jan 22;364:k5311. [doi: 10.1136/bmj.k5311] [Medline: 30670385]
13. Chang S, Wu W, Hu YJ, Lai HY, Wong TC. Quasi-experimental design for using an interactive social media intervention program to improve truck drivers' health beliefs and eating behaviors. BMC Public Health 2022 Aug 04;22(1):1486 [FREE Full text] [doi: 10.1186/s12889-022-13883-6] [Medline: 35927651]

14. Chiang K, Wang HH. Nurses' experiences of using a smart mobile device application to assist home care for patients with chronic disease: a qualitative study. J Clin Nurs 2016 Jul;25(13-14):2008-2017. [doi: 10.1111/jocn.13231] [Medline: 27136280]
15. Fu MR, Axelrod D, Guth A, Scagliola J, Rampertaap K, El-Shammaa N, et al. A web- and mobile-based intervention for women treated for breast cancer to manage chronic pain and symptoms related to lymphedema: randomized clinical trial rationale and protocol. JMIR Res Protoc 2016 Jan 21;5(1):e7 [FREE Full text] [doi: 10.2196/resprot.5104] [Medline: 26795447]
16. Kobayashi T, Nishina Y, Tomoi H, Harada K, Tanaka K, Matsumoto E, et al. Corowa-kun: A messenger app chatbot delivers COVID-19 vaccine information, Japan 2021. Vaccine 2022 Jul 30;40(32):4654-4662 [FREE Full text] [doi: 10.1016/j.vaccine.2022.06.007] [Medline: 35750541]
17. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 01;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
18. Jahn H, Jahn I, Behringer W, Lyttle MD, Roland D, Paediatric Emergency Research United Kingdom and Ireland (PERUKI). A survey of mHealth use from a physician perspective in paediatric emergency care in the UK and Ireland. Eur J Pediatr 2021 Aug;180(8):2409-2418 [FREE Full text] [doi: 10.1007/s00431-021-04023-0] [Medline: 33763717]
19. Jolly K, Sidhu M, PSM COPD Group, Bower P, Madurasinghe V, MRC START Group. Improving recruitment to a study of telehealth management for COPD: a cluster randomised controlled 'study within a trial' (SWAT) of a multimedia information resource. Trials 2019 Jul 24;20(1):453 [FREE Full text] [doi: 10.1186/s13063-019-3496-z] [Medline: 31340853]
20. Nwe K, Larsen M, Nelissen N, Wong DCW. Medical mobile app classification using the National Institute for Health and Care Excellence evidence standards framework for digital health technologies: interrater reliability study. J Med Internet Res 2020 Jun 05;22(6):e17457 [FREE Full text] [doi: 10.2196/17457] [Medline: 32501271]
21. Shuwandy M, Zaidan B, Zaidan A, Albahri AS. Sensor-based mHealth authentication for real-time remote healthcare monitoring system: a multilayer systematic review. J Med Syst 2019 Jan 06;43(2):33 [FREE Full text] [doi: 10.1007/s10916-018-1149-5] [Medline: 30612191]
22. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Mar 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
23. Atreja A, Khan S, Rogers JD, Otobo E, Patel NP, Ullman T, HealthPROMISE Consortium Group. Impact of the mobile HealthPROMISE platform on the quality of care and quality of life in patients with inflammatory bowel disease: study protocol of a pragmatic randomized controlled trial. JMIR Res Protoc 2015 Mar 18;4(1):e23 [FREE Full text] [doi: 10.2196/resprot.4042] [Medline: 25693610]
24. Callahan C, Buchanan N, Stump TE. Healthcare costs associated with percutaneous endoscopic gastrostomy among older adults in a defined community. J Am Geriatr Soc 2001 Nov;49(11):1525-1529. [doi: 10.1046/j.1532-5415.2001.4911248.x] [Medline: 11890593]
25. Merrick S, Farrell D. Head and neck cancer patients' experiences of percutaneous endoscopic gastrostomy feeding: a Q-methodology study. Eur J Cancer Care (Engl) 2012 Jul;21(4):493-504. [doi: 10.1111/j.1365-2354.2012.01326.x] [Medline: 22329827]
26. Ye J, Ma Q. The effects and patterns among mobile health, social determinants, and physical activity: a nationally representative cross-sectional study. AMIA Jt Summits Transl Sci Proc 2021;2021:653-662 [FREE Full text] [Medline: 34457181]
27. Wong A, Wong F, Chow KKS, Wong SM, Lee PH. Effect of a telecare case management program for older adults who are homebound during the COVID-19 pandemic: a pilot randomized clinical trial. JAMA Netw Open 2021 Sep 01;4(9):e2123453 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.23453] [Medline: 34499135]
28. Wong A, Bayuo J, Wong FKY, Chow KKS, Wong SM, Lau ACK. The synergistic effect of nurse proactive phone calls with an mHealth app program on sustaining app usage: 3-arm randomized controlled trial. J Med Internet Res 2023 May 01;25:e43678 [FREE Full text] [doi: 10.2196/43678] [Medline: 37126378]
29. Wong A, Wong F, Bayuo J, Chow KKS, Wong SM, Lee AYL. A randomized controlled trial of an mHealth application with nursing interaction to promote quality of life among community-dwelling older adults. Front Psychiatry 2022;13:978416 [FREE Full text] [doi: 10.3389/fpsyt.2022.978416] [Medline: 36329920]
30. Pattison D, Young A. Effect of a multi‐disciplinary care team on the management of gastrostomy feeding. J Human Nutrition Diet 2003 Oct 31;10(2):103-109 [FREE Full text] [doi: 10.1046/j.1365-277x.1997.00044.x]
31. Sanger PC, Hartzler A, Han SM, Armstrong CAL, Stewart MR, Lordon RJ, et al. Patient perspectives on post-discharge surgical site infections: towards a patient-centered mobile health solution. PLoS One 2014;9(12):e114016 [FREE Full text] [doi: 10.1371/journal.pone.0114016] [Medline: 25436912]

32. Snell D, Siegert R, Surgenor LJ, Dunn JA, Hooper GJ. Evaluating quality of life outcomes following joint replacement: psychometric evaluation of a short form of the WHOQOL-Bref. Qual Life Res 2016 Jan;25(1):51-61. [doi: 10.1007/s11136-015-1044-1] [Medline: 26068734]
33. Yao G, Chung CW, Yu CF, Wang JD. Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. J Formos Med Assoc 2002 May;101(5):342-351. [Medline: 12101852]
34. Schneider SM, Pouget I, Staccini P, Rampal P, Hebuterne X. Quality of life in long-term home enteral nutrition patients. Clin Nutr 2000 Mar;19(1):23-28 [FREE Full text] [doi: 10.1054/clnu.1999.0068] [Medline: 10700530]
35. Singh G, Simpson E, MacGillivray MK, Sawatzky B, Adams J, Ben Mortenson W. Expectations of a health-related mobile self-management app intervention among individuals with spinal cord injury. Top Spinal Cord Inj Rehabil 2022;28(2):205-215 [FREE Full text] [doi: 10.46292/sci21-00022] [Medline: 35521055]
36. Jeon E, Park HA. Development of the IMB model and an evidence-based diabetes self-management mobile application. Healthc Inform Res 2018 Apr;24(2):125-138 [FREE Full text] [doi: 10.4258/hir.2018.24.2.125] [Medline: 29770246]
37. Cheng C, Werner N, Doutcheva N, CONSORT Group. Codesign and usability testing of a mobile application to support family-delivered enteral tube care. Hosp Pediatr 2020 Aug;10(8):641-650 [FREE Full text] [doi: 10.1542/hpeds.2020-0076] [Medline: 32616602]

Abbreviations

CVI: content validity index
mHealth: mobile health
PEG: percutaneous endoscopic gastrostomy
WHO: World Health Organization
WHOQOL-BREF: World Health Organization Quality of Life Brief Version

Edited by L Buis; submitted 13.05.23; peer-reviewed by J Volkman, AKC Wong; comments to author 22.08.23; revised version received 20.09.23; accepted 22.09.23; published 20.10.23.

Please cite as:
Chen BL, Lien HC, Yang SS, Wu SC, Chiang HH, Lin LC
Impact of Mobile Apps in Conjunction With Percutaneous Endoscopic Gastrostomy on Patients' Complications, Quality of Life, and Health-Related Self-Care Behaviors: Randomized Clinical Trial
JMIR Mhealth Uhealth 2023;11:e48970
URL: https://mhealth.jmir.org/2023/1/e48970
doi:10.2196/48970
PMID:37862072

Original Paper

Older Adults’ Perceptions About Using Intelligent Toilet Seats Beyond Traditional Care: Web-Based Interview Survey

Pouyan Esmaeilzadeh[1], PhD

[1]Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics
College of Business
Florida International University
Modesto A. Maidique Campus
11200 S.W. 8th St, RB 261B
Miami, FL, 33199
United States
Phone: 1 (305) 348 3302
Email: pesmaeil@fiu.edu

Abstract

Background: In contemporary society, age tech (age technology) represents a significant advancement in health care aimed at enhancing patient engagement, ensuring sustained independence, and promoting quality of life for older people. One innovative form of age tech is the intelligent toilet seat, which is designed to collect, analyze, and provide insights based on toileting logs and excreta data. Understanding how older people perceive and interact with such technology can offer invaluable insights to researchers, technology developers, and vendors.

Objective: This study examined older adults’ perspectives regarding the use of intelligent toilet seats. Through a qualitative methodology, this research aims to unearth the nuances of older people’s opinions, shedding light on their preferences, concerns, and potential barriers to adoption.

Methods: Data were collected using a web-based interview survey distributed on Amazon Mechanical Turk. The analyzed data set comprised 174 US-based individuals aged ≥65 years who voluntarily participated in this study. The qualitative data were carefully analyzed using NVivo (Lumivero) based on detailed content analysis, ensuring that emerging themes were coded and classified based on the conceptual similarities in the respondents’ narratives.

Results: The analysis revealed 5 dominant themes encompassing the opinions of aging adults. The perceived benefits and advantages of using the intelligent toilet seat were grouped into 3 primary themes: health-related benefits including the potential for early disease detection, continuous health monitoring, and seamless connection to health care insights. Technology-related advantages include the noninvasive nature of smart toilet seats and leveraging unique and innovative data collection and analysis technology. Use-related benefits include ease of use, potential for multiple users, and cost reduction owing to the reduced need for frequent clinical visits. Conversely, the concerns and perceived risks were classified into 2 significant themes: psychological concerns, which included concerns about embarrassment and aging-related stereotypes, and the potential emotional impact of constant health monitoring. Technical performance risks include concerns centered on privacy and security, device reliability, data accuracy, potential malfunctions, and the implications of false positives or negatives.

Conclusions: The decision of older adults to incorporate intelligent toilet seats into their daily lives depends on myriad factors. Although the potential health and technological benefits are evident, valid concerns that need to be addressed remain. To foster widespread adoption, it is imperative to enhance the advantages while simultaneously addressing and mitigating the identified risks. This balanced approach will pave the way for a more holistic integration of smart health care devices into the routines of the older population, ensuring that they reap the full benefits of age tech advancements.

(JMIR Mhealth Uhealth 2023;11:e46430) doi:10.2196/46430

KEYWORDS

older adults; age tech; intelligent toilet seat; survey interview; qualitative study; mobile phone

Introduction

Age Technology

Worldwide demographic trends have witnessed an increase in the proportion of the older adult population. Many older adults have physical conditions or health issues that must be constantly monitored. However, mobility limitations, cognitive impairment, chronic pain, and the costs of being monitored through frequent appointments and admissions may challenge older adults to be engaged in their health care [1]. These issues may negatively affect the functional abilities, physical activities, social activities, and quality of life of older adults. Moreover, aging-related problems may significantly contribute to social isolation and depression [2]. Thus, these physical (eg, declining health and increased vulnerability to diseases), emotional (eg, increased risk of depression and loneliness), cognitive (eg, decline in memory and cognitive abilities), social (eg, reduced social interactions), and economic (eg, difficulty in paying for long-term care) issues may impair independence and the ability to perform daily activities [3].

It is critical to address the combinations of these aging-related issues and find strategies and ways to support older adults as they age to help maintain their independence and overall well-being. Previous studies have highlighted that technology may improve physical, cognitive, emotional, social, and economic issues and help older adults perform core daily activities [4,5]. The development of new technologies and technological devices enables older adults to manage their health conditions effectively, efficiently, and independently. Age technology (age tech) is designed to meet the needs of older adults and those caring for them [6]. Various technologies and devices are devoted to supporting and improving the quality of life for older adults as they age. Age tech can potentially improve the lives of older adults in different ways. First, age tech for physical issues is an assistive technology that helps with mobility and independence [7]. Second, age tech is a health care technology for managing diseases and well-being that provides remote monitoring systems through wearables and devices [8]. Third, age tech aids in daily tasks such as smart home technologies that facilitate activities of daily living using voice commands or mobile devices. Fourth, age tech is used for cognitive functions, such as cognitive aids that help with memory and cognitive abilities, including virtual assistants and reminder apps [9]. Fifth, age tech for emotional and social issues is a communication technology that reduces feelings of loneliness and isolation through videoconferencing and messaging applications [10].

Moreover, studies show that using age tech could potentially help reduce health care costs (to address economic issues) in some ways [11]. For instance, age tech can allow older adults to receive medical care without leaving their homes to visit health care professionals in person or potentially prevent unnecessary hospitalization or readmissions, leading to cost savings. Age tech could potentially lower overall health care costs by improving medication management and helping older adults better manage chronic conditions.

The key areas examined in previous studies on aging technology can be categorized into 6 groups. The first category focuses on demographic trends related to the aging population trends, such as increasing life expectancy and the growing number of older adults worldwide [12]. The second category is related to the impact of technology on aging and how technology can support older people in improving their overall quality of life by facilitating day-to-day activities and ensuring that aging individuals can lead independent and comfortable lives (such as using tools to adjust lighting, room temperature, or lock doors using voice commands) [13]. The third refers to the design and usability of age tech, such as user-centered design, accessibility, ease of use, and usability of technology for older adults [14]. The fourth group considers the impact of age tech on the health and well-being of older people, such as the technology used to manage existing health conditions, prevent potential diseases, or even boost mental health [15]. The fifth category examines the ethics of collecting biological and personal habit data in residences and ethical considerations of using age tech for older adults, such as privacy, security, confidentiality, and potential exploitation [16]. The last emphasizes the future directions of age tech, such as the potential for new personal health devices, artificial intelligence (AI)–powered applications, and emerging technologies to support older adults. This aligns with previous studies highlighting the disruptive types of AI systems being implemented in routine care [17].

Smart Toilet Seats as a Subset of Age Tech

Using age tech can have several advantages for older adults, such as improved access to care, better health management, increased independence, and connected care [18,19]. However, the older population may not be likely to use age tech because of several potential technological challenges, such as cost, physical challenges, lack of confidence and support, complexity, inadequate infrastructure, and privacy and security concerns [20-22]. A study showed that older people are less willing to modify their housing using technology to meet their needs [23]. As a result, older adults may resist using age tech, preferring traditional methods of receiving health care or communicating with health care givers and professionals [24-26]. Previous studies have examined the adoption of various age tech solutions (eg, personal health devices, wearable technology, telemedicine, health care robots, and social robots) with different functionalities. The adoption of smart technology in the home is on the rise, with internet-connected devices such as smart speakers, appliances, and home automation systems becoming increasingly common [27]. Smart toilets are one area of smart home technology that promises to improve health outcomes. Smart toilets are integrated with sensors and analytical capabilities that can track various health and wellness metrics, from heart rate to weight changes. For aging and disabled populations, smart toilets present an opportunity to monitor health changes unobtrusively and share important data with medical professionals through telehealth platforms.

Smart toilet technology is rapidly gaining popularity, with the global smart toilet market projected to grow significantly in the coming years. The market was valued at US $3.6 billion in 2022 and is expected to reach US $5.5 billion by 2029, witnessing a compound annual growth rate of 6.2% [28]. Another report suggested that the market could reach a value of US $22.2 billion by 2030, with a compound annual growth rate of 15.12% [29]. The increasing demand for advanced bathroom solutions, driven by growing consumer awareness about hygiene and sanitation, is a key factor contributing to this growth [30].

Smart options can reduce strain and effort for caregivers assisting people who require help in using the toilet. With remote operation through a smartphone app, caregivers do not have to be directly present to assist older adults. Beyond basic accessibility features, smart toilets can monitor vital signs, track changes, and share data with medical teams via telehealth platforms. Sensors in smart toilets can measure heart rate and blood pressure when a person sits. Optical sensors can evaluate a person's vein health to screen for conditions such as congestive heart failure. Smart toilets can also track urine and stool properties such as color, volume, concentration, and frequency to detect possible urinary tract infections, dehydration issues, and gastrointestinal problems. Trends and concerns regarding changes can automatically send alerts to a person's physician or care team to enable rapid intervention and treatment. Smart toilet seats can detect multiple signs of illness through automated urine and stool analyses. This early detection can help users address health issues before they become more severe, potentially preventing hospitalizations and the spread of infectious diseases among older adult communities [31]. Smart toilet seats have been shown to be beneficial in older adults living in residences. For example, the TrueLoo toilet helped reduce patient falls by 11% in a study involving memory care patients in California [32]. By tracking changes in urine and providing daily wellness reports, smart toilets can aid in the early detection of health issues and provide faster and more efficient care [33].

Adoption by Older People

The global impact of the COVID-19 pandemic has underscored the critical importance of age tech for the future, especially as it relates to the health care and well-being of the aging population [34]. During the pandemic, many older adults found themselves isolated from their families and usual support systems, leading to an increased reliance on technology. Out of necessity, many older adults adapted to new digital platforms, including telehealth services and communication applications, to stay connected with their health care providers, families, and communities. This unexpected shift brought about a transformation in the way older adults perceive and interact with technology [6]. Their newfound comfort with these platforms may potentially translate into a more receptive attitude toward other innovative solutions and digital services [35]. By drawing parallels with their recent embrace of digital health tools during the pandemic, it becomes evident that there could be growth in acceptance and trust in age tech among the older adult community. This evolving landscape presents a promising opportunity for the wider adoption of smart health technologies tailored to the unique needs of older adults.

The aging population faces unique health challenges and monitoring their well-being is paramount. The human body, through its waste products, provides valuable insights into health and well-being that can be critical for health care. Although this information is rich in diagnostic potential, there has traditionally been a reluctance, especially among older people, to harness these data. This hesitance might stem from cultural taboos, privacy concerns, or simply the lack of user-friendly and noninvasive tools to analyze waste matter [21,22,36]. Understanding this context is imperative for innovative solutions tailored to older adults. Intelligent toilet seats, designed with the consideration of older people in mind, present a groundbreaking approach. By seamlessly integrating technology with daily utility, these devices provide older adults with a nonintrusive method to monitor their health indicators continuously. However, to realize their full potential, it is crucial to address the acceptance of such technology among older adult demographics. It is not merely about proving the technological efficacy of the intelligent toilet seat; it is equally vital to demonstrate its value in a manner that resonates with older people. In a landscape with various health monitoring tools, the justification for an intelligent toilet seat lies in its unique proposition of unobtrusive, continuous health monitoring that respects the comfort and privacy of older people. Addressing this acceptance challenge is vital for the broader adoption and success of such innovations in older adults' health care.

Study Objectives

This research focuses on an AI-powered device that has been limitedly studied in the age tech literature. An intelligent toilet seat combines an internet-connected toilet seat and continuous monitoring using AI to generate valuable insights from waste matter discharged from the human body. This advanced technology, mainly used in homes or nursing facilities, seeks to analyze the excreta of individuals to derive valuable health insights from their toileting patterns. We used a qualitative approach in this study as the mentioned age tech is designed based on purposes different from a traditional toilet seat and is still relatively new to the public. Moreover, little is known about older adults' adoption patterns related to this system in previous studies. Thus, we aimed to gain a deeper understanding of older adults' perspectives and attitudes regarding a particular age tech (ie, intelligent toilet seats).

The main objectives of this study are as follows:

1. Explore older adults' perceptions and understanding of intelligent toilet seats and their attitudes toward using them in their daily lives
2. Discover the barriers and challenges older adults may encounter when using intelligent toilet seats
3. Uncover the factors that encourage older adults to adopt intelligent toilet seats
4. Examine the effectiveness of interventions aimed at promoting the adoption of intelligent toilet seats among older adults

This qualitative study could allow researchers to gain a rich and nuanced understanding of the expectations, perceptions, and attitudes regarding intelligent toilet seat adoption among older adults. This information could inform the development of interventions to improve age tech integration into health care services, support user adoption based on the potential benefits and risks of the technology, and identify areas where additional research is needed.

Literature Review

According to World Health Organization reports, the global population is aging rapidly, with the number of adults aged ≥60 years expected to more than double between 2015 and 2050 [37]. This demographic shift comes with various health challenges, including increased risk of falls, incontinence, and other conditions affecting quality of life and independence. Technology integration into health care, particularly for older people, has been growing. Among these technological innovations, smart toilet seats have gained attention for their potential health monitoring capabilities. Intelligent or smart toilet seats have emerged as an assistive technology that may help address some of these age-related bathroom challenges. This literature review synthesizes the current research on the use and perceptions of intelligent toilet seats among older adults.

The literature review on using smart toilet seats by older adults revealed several key findings. First, the study by Simpson [38] highlighted the use of technologies originally developed for mobile robots, such as smart wheelchairs, to accommodate the population of older people. This suggests that similar technological advancements can be applied to intelligent toilet seats to enhance the experience for aging adults. Pal et al [39] also discussed the negative perception modeling of older adults toward embracing the smart home revolution. This study emphasizes the importance of understanding the attitudes and preferences of older adults when introducing new technologies, including smart toilet seats. Furthermore, using a fuzzy inference system, Kurnianingsih et al [40] proposed a personalized adaptive system for older adult care in smart homes. This research demonstrates the potential for intelligent technologies such as smart toilet seats to be integrated into a comprehensive ecosystem that caters to the specific needs of older adults.

Moreover, Borelli et al [41] present HABITAT, as Internet of Things (IoT) solution for independent older adults. This study highlights the importance of interoperability among different smart devices, which can be applied to designing and implementing smart toilet seats to ensure seamless integration within a larger smart home environment. Finally, Marques et al [42] emphasized the significance of innovative and assistive eHealth technologies for the older adult demographic. This study underscores the potential benefits of incorporating smart toilet seats as part of a broader ecosystem of eHealth solutions to improve the overall well-being of aging adults.

Smart toilet seats are a relatively new technology that integrates sensors, AI, and automation to add advanced features to a standard toilet seat [43]. Key functions of smart toilet seats include the ability to sense the presence of a user, adjust the seat temperature, offer remote control options, analyze toilet use patterns, and, in some cases, monitor basic health data such as heart rate and blood pressure [44]. Along with recent market research reports, major manufacturers such as Toto, Kohler, and Brondell have recently introduced smart toilet models; however, their adoption remains limited [45].

According to Huang et al [46], researchers plan to apply funding to a smart toilet seat model embedded with sensors that can collect vital signs. A recent study explored the development of a smart toilet system for aging people and persons with disabilities [47]. Respondents felt that a smart toilet seat could be beneficial, highlighting its potential for improving comfort, accessibility, and dignity for older adults. Another study concluded that respondents had positive perceptions of smart toilet seats, emphasizing their potential benefits [48]. The literature review in this study also discusses the effects of extended exposure to smart toilet systems. A study presented a smart AI-based toilet concept that uses 3D depth data to automatically preconfigure the height and tilt of a motorized toilet seat [49]. Although not specifically addressing the needs of older adults, such features can enhance accessibility and autonomy for users with mobility issues. An article discussed a disease-detecting “precision health” toilet that can sense multiple signs of illness through automated urine and stool analysis [31]. Although not focused on older adults, this technology has the potential to benefit users of all ages by providing early detection of health issues. Some smart toilet seat models also analyze urine and stool to screen for possible health issues, such as urinary tract infections, kidney disease, and colorectal cancer [50]. Although still in the early stages, such capabilities could provide convenience and peace of mind for older users and caregivers.

Although intelligent toilet seats show promise for older users, research points to several limitations and concerns. Privacy and hacking risks are frequently raised as smart seats collect personal use data [33]. The seats’ high costs, complex interfaces, and need for regular cleaning may also deter adoption, especially among older people without caregiver support. Studies further indicate reluctance among older adults to use new bathroom technologies. Yeo et al [51] conducted focus groups with adults aged >65 years and found broad skepticism; respondents felt that smart toilets were unnecessary, embarrassing, and too complicated to use. This highlights the need for better education and training regarding the benefits of smart seats.

Methods

Research Context

This study focuses on a particular age tech—intelligent toilet seats. This age tech combines an internet-connected toilet seat and a continuous monitoring system to generate important insights from excrement and urine. The in-toilet technology measures various parameters of excreta, such as volume, consistency, frequency, color, and chemical composition. The data collected from these toilet seats can be used for various purposes, such as tracking bowel movements for medical diagnosis, monitoring hydration levels, and detecting changes in health status. The key idea is to monitor the early signs and signals of chronic issues that are very common in the older population (or even younger people) and be able to prompt corrective actions before they become more problematic. It is also equipped with sensor technology to determine who the user is and provide results for each user. Scanning the toilet bowl to determine the excreta size and shape, the system can indicate potential health issues such as dehydration, dietary imbalances, gastrointestinal diseases, infections, or even early signs of colorectal cancer. Regular monitoring and analysis of stool and urine can provide insights into an individual’s microbiome health, hydration levels, and kidney function. In addition, any sudden or drastic change in excretion patterns can serve as an alert for the need for medical consultation or intervention.

The smart toilet can report the collected data from toileting logs and send data or analysis of wellness parameters to the designated care team's dashboard. The data are analyzed using AI, and if abnormalities are detected, the system will send an alert to health care providers' systems. Smart toilets generally connect with the broader network of interconnected devices, known as IoT, allowing access to the gathered data and additional functionalities. By checking health updates from the system, health care teams can directly inform the user, family, or caregivers of any anomalies or deviations detected in the health data collected by the smart toilet seats. On the basis of the results, health care professionals can provide real-time feedback and suggestions to improve personal health and wellness. The smart toilet can also directly send health data or analysis to users, enabling older people to continuously monitor wellness parameters and health conditions and track trends and insights.

Type of Smart Toilet Seats

In the rapidly evolving field of smart health monitoring, myriad brands and types of intelligent toilet seats have emerged, each with distinct capabilities and features tailored to specific health requirements. Several basic seats provide comfort, such as warm seats and cleaning functions. Some seats might focus on advanced sensors to detect hydration levels, whereas others can prioritize temperature monitoring or even microbiome analysis. However, our study, in particular, focuses on a specific category of smart toilet seats that primarily analyze excreta, both stools and urine. The rationale behind this focus is the rich array of health insights that can be gained from such waste matter. For instance, variations in stool consistency can indicate digestive health, potential infections, or even chronic conditions such as irritable bowel syndrome. Similarly, urine can offer clues regarding hydration, kidney function, and the presence of specific compounds or infections.

To illustrate, brand A's intelligent toilet might emphasize pH monitoring in urine to identify potential urinary tract infections, whereas brand B's product could prioritize stool consistency analysis to detect digestive anomalies. In contrast, the type of intelligent toilet seat we focus on combines both these functionalities, offering a holistic overview of an individual's health through dual analysis. Although it would be beneficial to provide a comprehensive comparison of all available smart toilet seats in the market, such a broad scope could dilute the primary objective of our study. Our focus remains on understanding older adults' perception and acceptance of technology that analyzes both stool and urine to detect potential health issues and monitor critical health parameters.

Study Design

This study used a qualitative research approach via a survey interview. Our study aimed to explore and understand older people's views about intelligent toilet seats and the factors influencing their feelings, perceptions, and attitudes. As the use of this technology is increasing, there is a need for a solid adoption pattern. Our research's present state is exploratory and discovery to uncover factors affecting older people adopting smart toilet seats. Thus, leveraging a qualitative approach is appropriate to effectively capture the influential factors that may facilitate or deter adoption in the future [52]. Survey interviews were used as the data collection method in this study. This approach allows researchers to ask a set of standardized questions to a sample of respondents to gather information and collect insights into their opinions, attitudes, behaviors, or experiences [53]. We administered a web-based survey questionnaire to the respondents to obtain more information from the participants by capturing their words, ideas, and expressions. The reason for choosing web-based versus in-person data collection was 2-fold: (1) having in-person interviews with older people could be difficult and inconvenient for them (eg, owing to health conditions) and might interrupt data-collection flows; (2) the web-based survey interview is helpful for conducting social science studies to gather data from larger samples [54].

This research methodology was designed based on content analysis and thematic analysis. In the context of our study on older adults' perceptions of intelligent toilet seats, both content and thematic analyses offer valuable frameworks. Although content analysis helps structurally categorize and quantify responses, thematic analysis delves deeper into the nuances and underlying patterns within those responses. Both methods, when combined, provide a comprehensive understanding of older adults' attitudes, beliefs, and potential barriers or facilitators to adopting such technology. This process involves collecting data, coding the data to detect themes and patterns, and using the coded data to develop a theory that explains the relationships between concepts. The baseline questions used in this study focused on 5 categories. The first area was general familiarity with AI and awareness of smart devices. The remaining 4 questions were mainly asked to discover the respondents' perceptions of the specific age tech being studied: perceived benefits and advantages, perceived concerns and risks, overall opinions, and willingness to use. Multimedia Appendix 1 presents the survey questions used in this study.

At the beginning of data collection, the system was clearly defined, and the features were explained using words and terms that someone not specialized in this field or unfamiliar with this technology could easily comprehend. We also consulted 2 well-published scholars in the field of AI in health care to ensure that the description of the technology being studied was readable and understandable. We made minor changes to the definition and feature description according to the feedback received. We refrained from using phrases or expressions that may convey positive or negative connotations to minimize possible leading effects on the answers. Finally, we conducted a pilot test involving 25 randomly selected older adults to verify the clarity of the survey language. The feedback from these respondents confirmed that they comprehended the study's aims and the definitions and questions presented to them.

Ethical Considerations

Beyond obtaining ethical clearance from the institutional review board of Florida International University (Reference #112755), several steps were ensured to maintain the research's integrity, confidentiality, and ethical soundness. First, participants were provided with a clear and detailed informed consent form. Second, participation was voluntary, and respondents could opt out at any stage without consequences. Third, data anonymity and confidentiality were upheld as no personal identifiers were collected, and data were stored securely with access restricted to research team members. Fourth, the potential risks and benefits of participation were clearly outlined in the consent form.

Respondents, Data Collection, and Sample Representativeness

Individuals (as potential users of intelligent toilet seats) are the unit of analysis in this study. However, this study focused on older adults who may need to monitor their health and wellness parameters more frequently than other age groups. Older adults may experience various changes that increase the risk of health problems such as diabetes and chronic kidney disease. Monitoring health parameters (such as infection and kidney function) can help detect these issues early and prevent complications. Early detection and treatment can improve health outcomes and quality of life in older people and help physicians identify any changes in health status and adjust treatment plans accordingly. Thus, older people can be the best target market for this age tech.

Two inclusion criteria were set that were consistent with the study objective. The first was age, including older adults aged ≥65 years, and the second was the location, limiting respondents to individuals in the United States. Data collection was performed in January 2023 through Amazon's Mechanical Turk (MTurk). Previous studies have proven that MTurk is a suitable survey tool for collecting individual-level data [55,56]. Researchers, as requesters, can use this crowdsourcing website to reach out to potential participants (ie, MTurk workers) in numerous countries to conduct a survey. MTurk workers, often referred to as Turkers, are a diverse group of people from across the globe who engage in the tasks known as human intelligence tasks. Several studies compare MTurk to conventional data collection methods in health and medical literature and support using this platform for various academic purposes (eg, in health care research) [57]. Existing literature in clinical research highlights that because of a larger network of people, the MTurk population is more representative of the US population than other web-based surveys [58]. Moreover, some studies validated crowdsourcing platforms (such as MTurk) for collecting data from older adult respondents [59,60].

Although MTurk provides a diverse pool of respondents, it is essential to acknowledge the potential limitations concerning its representativeness. Our sample from MTurk comprises a subset of the older adult population, particularly those familiar with and who have access to technology. This may mean that they are more technology-savvy and possibly more open to adopting new technologies than the general older adult population. Moreover, MTurk workers might have a different socioeconomic status [61]. Participation in MTurk often serves as a supplementary income source, which might indicate a different economic stratum compared with other older adults. Finally, although our study was limited to respondents in the United States, MTurk workers can come from diverse geographical and urban or rural backgrounds. Their distribution might not align perfectly with the broader distribution of older adults in the United States. Thus, in our analysis, we were cautious about overgeneralizing our findings, emphasizing that they primarily apply to older adults familiar with web-based platforms.

All questions were structured and open ended, allowing the respondents to express their in-depth opinions and ideas to generate a more profound understanding of the phenomenon. It should be highlighted that the survey was anonymous, and no personally identifiable information was collected from respondents. We also did not define character limits for answers to avoid fake or false responses owing to the imposed pressure of reaching a data-entry limit. However, we discarded records with no or irrelevant responses. A total of 202 individuals who met the inclusion criteria attempted the survey interview. We found that the responses of 28 individuals were not satisfactory (either no or unrelated responses were provided). Therefore, the final data set included responses from 174 older adults living in the United States. All the respondents received US $5 as an incentive for completing the survey.

Data Analysis

Following our qualitative method, we used content analysis to interpret and analyze the response transcripts to identify and induce patterns, themes, meaningful structures, and trends, which can help researchers make inferences about the studied population. Content analysis is mainly used to draw inferences about the attitudes, beliefs, or behaviors of the people who produced the content [62]. Two researchers with experience in age tech and older adults' adoption of personal health devices coded the responses to perform explorative content analysis. The coding procedures used in our qualitative content analysis were open, axial, and selective coding [63]. In the initial coding stage (ie, open coding), researchers independently examined the answers and generated a list of codes or labels that described the content based on their interpretation of the data. The relationships and connections between the previously generated codes were explored using axial coding. At this stage, the codes were grouped into categories and subcategories to explain the relationships between the codes. In selective coding, researchers focused on the most important or relevant categories or conceptual themes that had emerged from the open and axial coding stages.

The 2 coders independently coded line-by-line answers from 174 surveys using the NVivo software tool. With NVivo, researchers imported the data, applied codes, and categorized data into themes. As 2 coders coded the same data, we used interrater reliability (IRR) to measure the degree to which different coders agreed on data coding [64]. We used 3 measures to calculate IRR: percentage agreement, κ coefficient, and intraclass correlation coefficient (ICC) [65]. As discrepancies in coding interpretations arose, it was crucial to ensure that any biases or misunderstandings were addressed. For conflicts in coding, the 2 coders initiated in-depth discussions to explore the roots of their discrepancies, share insights from their individual interpretations, and converge on a shared understanding. This iterative process ensured that both coders fully understood each other's perspectives and was essential for refining and solidifying the coding framework. Although incorporating a third reviewer might have added another layer of arbitration, we found that our systematic discussions and the use of IRR measures, especially after repeated coding, provided robust validity. The coders' consensus meetings were comprehensive, focusing on understanding the essence of each response and reconciling the differences. We finalized our coding after reaching a mutual agreement and achieving significantly improved IRR metrics. It is worth noting that the achieved IRR values, after discussion, were well within the acceptable range, demonstrating the reliability of our dual-coder system. We obtained a percentage agreement of 72%, a κ coefficient of 0.69, and an ICC of 0.67. To attain a higher threshold and ensure the validity of the results, we repeated the coding process until the 3 metrics fell within the acceptable range. To achieve better consistency and agreement between ≥2 raters, the coders met to resolve coding disagreements and reach a coding consensus. Then, the coders achieved a percentage agreement of 91%, a κ score of 0.80, and an ICC score of 0.83, representing the validity of the coding process and the results.

Results

Demographic Characteristics

The web-based survey was completed by 174 respondents aged ≥65 years. Of the respondents, 73 (41.9%) were aged between 65 and 70 years, 59 (33.9%) were aged between 70 and 75 years, 30 (17.2%) were aged between 75 and 80 years, and 12 (6.9%) were aged between 80 and 85 years. A small majority of the respondents were male, with 92 (52.9%) male and 82 (47.1%) female respondents. A significant proportion of the respondents had undergraduate or graduate education (120/174, 68.9%), indicating that the sample is relatively educated compared with the general population. This is consistent with previous research indicating that individuals with higher education levels are more likely to seek web-based health information and are more aware of technological changes [66,67]. Chronic health conditions were reported by 37.9% (66/174) of the respondents, with diabetes, arthritis, and chronic kidney disease being the most common issues cited. Most of the respondents (127/174, 73%) lived with family or others, whereas the remaining (47/174, 27%) lived alone.

At the beginning of the survey, the respondents were asked about their overall familiarity with AI-powered and smart devices. Overall, 13.8% (24/174) of the respondents indicated that they were not familiar with AI and smart devices, 37.2% (64/174) reported being aware of AI-based devices, though they had never personally used them, and 23% (40/174) were familiar with and had used some general AI-based devices (such as smart thermostats, smart cameras, and robot vacuums). Finally, 46% (80/174) of the respondents indicated familiarity with AI-powered health care tools or devices (such as AI for cancer diagnosis, tumor detection, or managing diabetes). Moreover, 25.3% (44/174) of the respondents reported using AI-enhanced applications to manage chronic health issues and dietary preferences or to receive medication reminders.

Moreover, before sharing the description of smart toilet seats, we asked the respondents whether they knew anything about this technology. Overall, 60.1% (106/174) of the respondents said they had heard about it or read (at least) an article about this topic on social media, newspapers, or websites. After showing the description, all respondents reported that they reasonably understood what a smart toilet seat looked like and how it worked. It is worth noting that none of the respondents directly had used an intelligent toilet seat. This evidence is expected because understudied technology is still relatively new to the older population.

When asked for their overall opinion on smart toilet seats, a large majority, 83.9% (146/174), exhibited positive attitudes, whereas 16.1% (28/174) held negative views toward this age tech innovation. As for their willingness to use this technology, 64.9% (113/174) of the respondents stated that they were willing to use it, 28.2% (49/174) of the respondents considered future use, and 6.9% (12/174) of the respondents were not inclined to use it. Regarding the best use case for adopting this technology, 63.8% (111/174) of the respondents suggested that nursing homes were the most appropriate setting, followed by rehabilitation centers (44/174, 25.3%) and personal use in homes (19/174, 10.9%).

Positive Opinions (Benefits and Advantages)

Respondents were asked to describe their opinions on the benefits and advantages of using intelligent toilet seats. Textbox 1 shows open codes and common concepts related to the use of age tech. Open codes were derived directly from the raw data, capturing specific details and nuances from respondents' responses. During open coding, researchers label and categorize phenomena found in the text based on their properties and dimensions. These codes were then grouped by researchers based on shared themes or ideas, capturing the overarching essence or underlying patterns of the data presented. Through this process, the intricate details of individual responses are distilled into broader concepts that capture the collective sentiment and insights of the respondents.

Textbox 1. Open codes for perceived benefits and advantages related to intelligent toilet seats.

Early detection of diseases

- Detection of infectious diseases, anomaly detection, preventive care, timely discovery of serious issues, identification of diseases, preventing serious illness

Multiple users

- Many users, usable to several users, supporting multiple users, accessible to different users, multiple individuals

Easy to use

- User-friendliness, easy-to-use technology, simple, intuitive controls, no complicated steps, no learning required, no instructions

Health care cost reduction

- Lowering the overall cost of health care, reducing costs of travel, lowering the number of appointments, minimizing unnecessary visits, cost of extra treatments

Seamless connection to clinical reports

- Interactive communication, real-time data sharing, better connection, sharable reports, better collaboration with caregivers

Constant monitoring

- Regular check-ups, ongoing screening, continuing surveillance of health status, monitoring symptoms, constant observation of signs, being controlled

Safe technology

- No harm, not dangerous to individuals, safe function, reliable technology, no injuries, no side effects

Unique function

- Innovative technology, new functionality, unique attributes, competitive advantage

Table 1 shows 8 categories of constructs. Table 1 also exhibits example quotes selected from the respondents' answers to the survey questions. These quotes can support the constructs (related to perceived benefits and advantages) extracted from the answers. The 8 constructs are as follows:

1. *Early detection of diseases*: Respondents expressed the view that smart toilet seat technology might offer potential benefits in identifying health issues such as urinary tract infections. They believed that the early identification of these problems would enable users to address them promptly. Some respondents felt that this could prevent more severe health complications such as the spread of infectious diseases within older adult communities. Several respondents highlighted the potential of such technology in facilitating timely and preventive care, thereby possibly enhancing the quality of health care they receive.
2. *Multiple users*: Respondents mentioned an appreciation for the device's sensor technology, which they understood could identify individual users. They highlighted that, according to their understanding, the system could analyze excreta by assessing attributes such as size, color, consistency, frequency, and shape, allowing for individualized results. This feature, they felt, ensured that multiple residents could use the seat without data overlap. In addition, some respondents noted the potential convenience of suspending data collection, especially during visits from guests, by simply pressing a button on the seat.
3. *Easy to use*: Respondents perceived that the smart toilet could be user-friendly. They believed that it required minimal setup, had an uncomplicated design, and would not require intricate instruction manuals or guidelines. Some respondents understood that the toilet used computer vision technology to gather health data autonomously and to forward it for assessment. As such, many felt relieved that they would not have to engage in maintenance, manually transmit data, or decipher analytical reports. The overarching sentiment was that adopting this intelligent device would not necessitate intricate technological training or education.
4. *Health care cost reduction*: Respondents expressed the belief that maintaining an automated electronic record of bowel movements and urination could potentially lead to financial savings. They felt that such records could minimize the need for regular physician visits, associated transportation costs, and subsequent treatments. Many respondents perceived that smart toilet seats might diminish the expenses tied to health monitoring by curtailing the frequency of physician visits and unwarranted medical examinations. There was a shared sentiment that this technology could be cost-effective in the long term, particularly for older adults grappling with chronic health issues.
5. *Seamless connection to clinical reports*: Respondents noted the potential of the smart toilet to relay data from toileting logs directly to the dashboards of designated care teams. They mentioned that in the event of detected anomalies,

direct communication could be initiated with the patient, their family, or even their physician. Furthermore, the option to provide printouts of the data to both the patient's family and medical professionals was considered valuable. Respondents felt that such features could foster improved collaboration and interaction between caregivers, health care providers, and family members.

6. *Constant monitoring*: Respondents believed that smart toilet seats might offer them a means to consistently oversee wellness indicators and health states, allowing for the identification of trends and insights. They noted the potential for the technology, in collaboration with health care teams, to consistently monitor for any significant health deviations or irregularities without requiring their active participation. Many respondents shared a sentiment of reassurance, feeling that their clinical data would be regularly assessed and any chronic health concerns would remain under the watchful eye of their health care professionals.
7. *Safe technology design*: Respondents highlighted the appeal of smart toilet seats owing to their lack of need for battery charging or wearing additional devices such as pendants. They also emphasized the absence of harmful radiation or potential safety hazards associated with their use. The simplicity of the device, with fewer components that might malfunction, was observed to reduce apprehensions about possible user errors. Many respondents believed that smart toilet seats presented a low-risk tool for older adults in terms of physical safety and potential software issues, which they perceived would not compromise their health.
8. *Unique function*: Respondents were intrigued by the distinctive functionality of the technology, which integrates an internet-enabled toilet seat with continuous monitoring capabilities to derive insights from fecal and urinary outputs. They noted that age tech taps into an emerging scientific domain that blends optics, AI, sensors, IoT, and cloud computing to craft a personalized health portrait based on toileting data. Many respondents felt that such technology shifts the health care paradigm, encouraging a more proactive stance than a reactive one. The bespoke features of this age tech seemed particularly suited to the needs of older adults, especially those managing chronic health conditions.

It should be noted that although both early detection of diseases and constant monitoring aim to optimize health management, they operate on distinct principles. Early detection focuses on the proactive identification of specific health conditions, often before symptoms manifest, aiming for timely intervention when diseases are most treatable. It is rooted in intentionality, targeting diseases known to benefit from early intervention through periodic screenings. In contrast, constant monitoring is an ongoing surveillance of an individual's overall health status. It encompasses a broader spectrum of health metrics, capturing data continuously, often through technological aids. Its primary purpose is to track and respond to any health changes, whether indicative of a disease or a temporary fluctuation. Although both are crucial, they offer unique approaches and benefits in health care.

Table 1. List of constructs, definitions, anecdotal evidence, and count for perceived benefits and advantages related to intelligent toilet seats.

Axial codes (constructs)	Definition	Sample quotes	Respondents, n
Early detection of diseases	The extent to which intelligent toilet seats can improve the chance of early detection of serious health conditions and diseases	• "This technology is good for general health checking and early disease detection." • "At my age, any head start on detecting a health problem can make all the difference. If this seat can give me a heads-up about potential issues, then I'm all for it."	102
Multiple users	The extent to which different individuals can use intelligent toilet seats without possible disruption or confusion	• "Me, my wife, or even our kids can use the same smart seat without worrying about result confusion." • "Our household has six people; it'd be great if the intelligent toilet could differentiate between all of us and still function effectively."	63
Easy to use	The extent to which using intelligent toilet seats is free of effort and does not need training or complicated instructions	• "It seems very easy to use this seat because there is no training or complicated instruction on how to use it." • "If it's called 'intelligent,' I assume it'll be smart enough to make things easy for me."	81
Health care cost reduction	The extent to which using intelligent toilet seats can reduce overall health care costs.	• "I am fed up with many appointments that I need to have to just check my urination and digestion issues. They are unnecessary money going out of my pocket. This seat can solve cost-related issues." • "Preventive care is always cheaper than treating advanced conditions. I see the potential in these toilets to keep me ahead of any health troubles."	46
Seamless connection to clinical reports	The extent to which using intelligent toilet seats can help users access their clinical data and improve communication and collaboration with caregivers	• "I can see my health conditions and clinical reports. Also, my doctors can check out the results promptly. So we can have a better dialogue about my conditions." • "The ability to link my daily health metrics from the toilet to my medical records? Now, that's smart technology."	78
Constant monitoring	The extent to which using intelligent toilet seats can enhance the monitoring and screening of wellness parameters	• "I believe continuous monitoring of toileting will be useful in the sense that I am sure my wellness is monitored and in case of immediate treatment, I am in good hands. • "I think the beauty of this technology is its consistency. The constant monitoring means no anomalies go unnoticed."	83
Safe technology design	The extent to which intelligent toilet seats have safe design and harmless functionality, and using intelligent toilet seats will not impose physical harm on individuals	• "No parts or functionality of these seats can harm my health." • "It appears that the design is foolproof and there's no risk of any harm, then why not? I'm on board."	56
Unique function	The extent to which intelligent toilet seats leverage unique and innovative technology for data collection and data analysis	• "This technology is very new. Using computer algorithms and AI to detect diseases from bowel movement is very innovative." • "It's not just about collecting data but how it's analyzed. The technology in this toilet seems to think out of the box."	39

Table 2 presents the key themes representing the main perceived benefits and advantages of using intelligent toilet seats. The core themes are health care-related benefits, technology-related advantages, and use-related benefits. The theme of health care-related benefits refers to the contributions of intelligent toilet seats to the early detection of diseases, interactive communication with caregivers, and monitoring of wellness parameters. Technology-related advantages reflect the safe functionality and innovative features of this age tech. The use-related benefits theme describes how smart toilet seats are easy to use, straightforward, efficient, and usable by multiple individuals.

Table 2. Selective codes representing perceived benefits and advantages related to intelligent toilet seats.

Selective codes (themes)	Constructs involved	Definition
Health care-related benefits	Early detection of diseases plus seamless connection to clinical reports plus constant monitoring	The extent to which older people may believe that intelligent toilet seats can help detect diseases in their early stages, strengthen interactive connections to clinical data, and maintain constant monitoring of health parameters
Technology-related advantages	Safe technology design plus unique function	The extent to which older people may believe that intelligent toilet seats are safe technology with innovative features
Utilization-related benefits	Easy to use plus multiple users plus health care cost reduction	The extent to which older people may believe that intelligent toilet seats are easy to use, usable by several users, and efficient

Negative Opinions (Concerns and Risks)

Respondents were asked to describe their perceptions and opinions regarding the concerns and risks of using intelligent toilet seats. Textbox 2 depicts open codes and common concepts related to using this technology.

Textbox 2. Open codes for perceived concerns and risks related to intelligent toilet seats.

Embarrassment

- Feeling embarrassed, uncomfortable, ashamed, unpleasant, discomfort

Unnecessary concerns

- Consistent stress, overthinking, obsession, feeling phantom alerts, persistent notifications

Aging-related stereotypes

- Negative images, untrue ideas, stereotypes, beliefs not based in fact, negative attitudes

Privacy and security risks

- Collection practices, security measures, sharing mechanisms, leaked data, data breaches, hackers, vulnerable systems, data protection

Data quality risks

- Invalid and unreliable wellness parameters, malfunction, false negative or false positive, biased results, data accuracy

Internet outage risks

- Down network, internet connection problems, Wi-Fi issues, disrupted communication, connectivity issues, slow network, equipment failure, outage

Table 3 demonstrates 6 categories of the constructs. Some sample narrations chosen from the respondents' answers are presented in Table 3. These quotes can support the constructs (related to perceived concerns and risks) taken from the answers. The 6 constructs are as follows:

1. *Embarrassment*: Respondents expressed potential feelings of discomfort associated with the intimate nature of data collection using the intelligent toilet seat. The primary function of the tool is to analyze fecal and urinary outputs to provide health insights, and this direct focus on personal waste might evoke feelings of embarrassment among older adults. Although the technology operates discreetly, several respondents indicated a preference for a device that monitors wellness parameters in a less intrusive manner.
2. *Unnecessary concerns*: Respondents voiced concerns about the potential for heightened anxiety stemming from continuous monitoring by intelligent toilet seats. Although the technology aims to provide proactive health insights, some older adults felt that being constantly monitored might lead them to become overly preoccupied with minor health anomalies that might not necessarily require medical intervention. The constant anticipation of alerts or notifications about their health metrics might inadvertently induce stress, which could have negative implications for their overall well-being, both mentally and physically.
3. *Aging-related stereotypes*: Some respondents expressed concerns about potential age-related perceptions surrounding the use of smart toilet seats. They feared that adopting such technology might inadvertently signal to others that they are managing severe chronic health conditions such as chronic kidney disease, urinary tract infections, gastrointestinal bleeding, constipation, or prostate cancer. This association, whether real or perceived, could reinforce negative stereotypes about the inevitable decline of health with age. However, it is essential to note that many older adults may embrace this age tech not necessarily because they manage specific health conditions, but as a proactive measure to maintain overall wellness and preemptively address potential health concerns.
4. *Privacy and security issues*: Respondents expressed reservations about data privacy and security associated with

using intelligent toilet seats. They voiced their apprehensions regarding the nature of the data gathered, its sharing mechanisms, and potential access by unauthorized individuals. The inherent risks of internet-connected devices, such as potential data breaches and cyberattacks, further heightened their concerns. Given the sensitive nature of the data collected, many respondents indicated a preference for using devices that strictly adhere to Health Insurance Portability and Accountability Act guidelines. They emphasized the importance of ensuring that data are anonymized, not linked with personal identifiers or comprehensive health records, and not shared with third parties for nonconsented uses, such as research, to safeguard their privacy.

5. *Data quality*: Respondents expressed concerns about the accuracy and reliability of data gathered by the intelligent toilet seats. Although the device uses high-resolution optical scanning through lasers to identify potential health markers, the respondents questioned the consistent accuracy of these measurements. They voiced skepticism regarding the system's ability to always capture valid wellness indicators, especially when foreign objects might be in the toilet. The potential for technical glitches that could lead to inaccurate readings, either false negatives or false positives, has also emerged as a concern. Furthermore, the inherent limitations and potential biases of AI and machine learning algorithms raised ethical apprehensions. Respondents emphasized that inaccuracies or biases in data might influence clinical decisions or treatments, underscoring the criticality of data precision in their willingness to adopt such technology.
6. *Internet outage risks*: Respondents highlighted concerns about potential internet disruptions and their impact on the functionality of intelligent toilet seats. Given that these devices rely on IoT technology for real-time data transmission, any lapse in internet connectivity could hinder the timely sharing of health data with medical professionals. They perceived that such interruptions could also impair communication and coordination between users and their health care teams, leading to the risk of conveying outdated or stale health parameters. As these sensing devices are dependent on an active internet connection to function effectively, any downtime could pause their key operations. Therefore, some respondents believed that the presence of a stable internet infrastructure, possibly supported by advancements such as 5G and robust Wi-Fi networks, would be crucial for the uninterrupted operation of such devices.

Table 3. List of constructs, definitions, anecdotal evidence, and count for perceived concerns and risks related to intelligent toilet seats.

Axial codes (constructs)	Definition	Sample quotes	Respondents, n
Embarrassment	The extent to which older people may feel uncomfortable and embarrassed by using intelligent toilet seats	• "I will not feel OK when I am supposed to use it. It is embarrassing that a device is checking your private stuff to collect data." • "I get that it's advanced technology, but there's a certain level of privacy and dignity you want in the bathroom. I'd feel a bit exposed."	83
Unnecessary concerns	The extent to which using intelligent toilet seats can raise excessive mental obsession	• "I will feel obsessed all time since I will wait for a new alert saying that I am diagnosed with a serious infection that should be addressed ASAP." • "Every time it would analyze something, I'd be worried if I'm healthy or if there's something wrong. Do I really need that daily stress?"	41
Aging-related stereotypes	The extent to which using intelligent toilet seats may create unfair and inaccurate attitudes that others hold about older people	• "If I use it at home, my friends and guests will think I am using this because I am not feeling good, because I am old, and I cannot control myself." • "Just because we're aging doesn't mean we need a device for every little thing. I don't like to become a symbol of dependency."	38
Privacy and security risks	The extent to which using intelligent toilet seats can cause data security and privacy invasion concerns	• "With the introduction of any new technology, there is always the risk of data breaches and malicious attacks that could invade data privacy." • "Based on all these hacking stories in the news, how can I be sure the information from the toilet seat remains confidential?"	124
Data quality risks	The extent to which using intelligent toilet seats can cause wrong, biased, and invalid clinical reports	• "Human oversight and technical bias. AI systems can be subject to human bias and errors, which can lead to incorrect or unethical decisions." • "What if the toilet malfunctions and gives a false reading? A false alarm can be just as harmful."	79
Internet outage risks	The extent to which the functionality of intelligent toilet seats can be disrupted due to network problems and internet outage	• "This technology will not work if the internet connection is not stable. No internet means interruption, and the toilet seat system will be down." • "I've had my Wi-Fi go out plenty of times. If the toilet relies on that, it sounds like a struggle waiting to happen."	45

Table 4 depicts the key themes signifying the main perceived concerns and risks of using intelligent toilet seats. The core themes were psychological concerns and technical performance risks. The psychological concern theme refers to uncomfortable feelings, unnecessary stress, and unfair and inaccurate attitudes of others toward aging owing to the use of smart toilet seats. Technical performance risks reflect possible technical malfunctions, inaccurate medical data, unsecured networks, and the potential threat of data breaches and unauthorized access.

Table 4. Selective codes representing perceived concerns and risks related to intelligent toilet seats.

Selective codes (themes)	Constructs involved	Definition
Psychological concerns	Embarrassment plus unnecessary concerns plus aging-related stereotypes	The extent to which older people may believe that using intelligent toilet seats may cause uncomfortable feelings, raise some unnecessary concerns, and develop stereotypes around them about aging
Technical performance risks	Privacy and security risks plus data quality concerns plus internet outage risks	The extent to which older people may believe that using intelligent toilet seats may cause possible technical risks, such as network vulnerability, unauthorized access, erroneous data, and malfunctions

Discussion

Theoretical Contributions

This qualitative study attempts to gain a deeper understanding of older adults' adoption of a particular age tech. The research methodology adopted in this study focused on exploring older adults' attitudes and perceptions regarding the use of intelligent toilet seats. Choosing a qualitative method to investigate older adults' perceptions of intelligent toilet seats, despite their lack of direct experience, is crucial for capturing in-depth insights. Even if they have not used such technology, understanding older adults' initial perceptions is invaluable; their viewpoints can highlight potential barriers to adoption, misconceptions, and design or educational improvements that vendors and developers might need to consider. Qualitative research delves deeper into perceptions, allowing for an open-ended exploration of initial reactions, apprehensions, and expectations. Given that this is potentially a first encounter with such technology for many older adults, this method captures the nuanced views shaped by cultural, generational, or personal beliefs. Although questionnaires provide quantifiable data, they might not grasp the breadth and depth of understanding qualitative approaches offer, especially when examining novel products or concepts.

Through survey interviews, the research aims to uncover the factors influencing older adults' decision to adopt or reject this technology and the perceived benefits and challenges they might encounter with its use. This section explains the adoption facilitators and barriers that may shape opinions about this AI-powered technology and affect adoption decisions among older adults. Moreover, according to the results, we propose a guiding framework for future studies followed by practical implications and study limitations. It is worth noting that these theoretical implications and practical recommendations can be extended to a large extent to the broader realm of age tech, as smart toilet seats represent just one subset of this larger technological category.

Adoption Facilitators

Health Care–Related Benefits

These were considered the most favorable advantages of using an intelligent toilet seat in this study. This result confirms that the key reason for using this intelligent system among older people is to improve health outcomes and enhance the quality of life. This is in line with previous studies on age tech adoption among the older population [68]. Building a smart environment through age tech and AI-powered devices can improve the quality of life, enhance safety, and provide timely health monitoring for older adults [69]. Similarly, AI-enabled screening tools can analyze toileting data to identify potential health issues, including conditions that may not be easily detectable through physical examinations or frequent visits. This technology can monitor excreta data, which holds valuable information about an individual's health and alerts the individual and their health care provider to potential health concerns. Early detection of diseases leads to more prompt and effective treatment, improving health outcomes and quality of life for older adults [70]. It also increases the chances of survival and recovery for serious conditions such as cancer and allows more effective management of chronic conditions, such as diabetes.

Previous studies have indicated that older adults and their caregivers can efficiently obtain and integrate medical records and test results through seamless access to clinical reports [71]. Having complete and current clinical information enables health care providers to make informed decisions and provide better care to older adults [72]. Furthermore, seamless access to clinical reports facilitates improved communication and coordination among health care providers, ensuring that older adults receive consistent and effective treatment. With electronic access to real-time clinical reports, health care providers can save time and reduce the need for manual recordkeeping and duplicative tests, improving efficiency in managing chronic conditions in older adults.

Toilet-based health monitoring tools using smart toilets could offer preventive home-based continuous health monitoring for older adults experiencing chronic issues such as diabetes, liver disorders, and kidney diseases [44]. Regular monitoring of health parameters, such as frequency and volume of excreta, can provide older adults with the information needed to effectively manage chronic conditions, such as diabetes and kidney problems. Continuous monitoring allows health care providers to quickly detect changes in older peoples' health, leading to improved outcomes [73]. This proactive approach to health can increase older adults' engagement in managing their chronic conditions. In addition, the large amount of health data generated from continuous monitoring can be analyzed to provide valuable data-driven insights into health trends and patterns.

Technology-Related Advantages

One common concern related to age tech among older adults is the potential risk it may pose to users, possibly leading to injuries [74]. For instance, previous studies have highlighted that technology may endanger users' health owing to malfunction and unsafe parts. Thus, safety is an important consideration when designing age tech for older adults, particularly for those with physical challenges [18]. The whole concept of age tech is to empower older adults to maintain their independence and remain actively engaged in their communities while ensuring their safety, health, and well-being through technology-driven solutions [75]. Safe technology aims to minimize harm to individuals and be reasonably free from causing damage. Intelligent toilet seats are designed in a way that does not cause harm to older adults or other users. Even if the AI-based algorithm does not work properly, the system will not physically injure older people because of improper use or dangerous functionality. As this age tech is not complicated to use and no specific training is required, the risk of unintended consequences and unforeseen circumstances that may threaten the health of older adults is lower.

Previous studies have indicated that aging management technology can help older adults monitor their physical activity levels and health markers using innovative features [76]. Advanced sensors and scanners enable intelligent toilet seats to accurately check bowel movements, track toileting data, and generate personalized and sophisticated analyses of an individual's health risk factors. AI algorithms can predict future trends and patterns based on the toileting habits of older adults, allowing for proactive and preventive measures to be taken [77]. These patterns and relationships may not be easily noticeable by humans. Moreover, continuous data analysis provides real-time feedback. These unique features (ie, pattern recognition, predictive analysis, AI algorithm, machine learning, big data analytics, and real-time feedback) can offer new insights and opportunities to improve care for older adults [78].

Use-Related Benefits

As mentioned in previous studies, various types of age tech should be easy to use and require little to no training or instructions to operate [79]. The goal of ease of use is to minimize the frustration and confusion that older adults may experience when using age tech and to help them achieve their goals quickly and efficiently. The accessibility and usability of age tech can promote inclusivity and help to bridge the digital divide [12]. Smart toilets equipped with AI technology seemed to be free of effort and self-explanatory to the respondents of this study. This study also confirms the importance of the property of a technology or system that makes it simple and intuitive to operate and understand. The higher the user-friendliness level and accessibility of technology, the more likely it is that older adults would adopt it. Moreover, another aspect of usability is that different individuals or multiple people (of a family or in a nursing home) can share the same AI-based toilet seat. This technology is sharable but can be customized to meet the needs of different users. The data were collected from the toileting logs of each user and analyzed without interfering with the results of other users.

As highlighted in prior research, a common reason older adults may incorporate age tech in their daily activities is health care cost saving [75]. Costs because of medication mismanagement (especially for older adults with chronic health problems) are an integral part of their monthly expenditure [80]. Older people use age tech to reduce travel costs to and from physician's appointments and costs related to receiving care planning and various treatment options. Using AI-based toilet seats can reduce the need for more expensive treatments, unnecessary medical tests, additional medications, and hospitalization through early detection and treatment of diseases that can prevent or delay the condition's progression. Thus, older adults can save money on health care costs, and the health care system can also reduce the burden of more costly treatments and hospital stays.

Adoption Barriers

Psychological Concerns

Emotional, ethical, and mental health issues arising from using technology have been highlighted in previous studies [81]. Age tech use can contribute to feelings of anxiety and stress, particularly when it involves detecting abnormalities in health parameters and the pressure to receive undesirable reports. The shadow of receiving alerts or alarms implying that older adults have been diagnosed with a difficult health issue can make them stressed, decrease their attention span, and generate unnecessary concerns [82]. Older adults may also feel embossed owing to the manner of data collection by AI-based seats as they monitor older people through excreta data. Consistent with previous studies, respondents repeatedly highlighted the concept of unobtrusively monitoring wellness parameters [76], as they believed that continuous monitoring of an older adult's physical and emotional well-being should be conducted without causing discomfort or disruption. Thus, age tech collection practices, monitoring, and providing early warning signs of possible health issues should be in a manner that respects ethics, human rights, and the dignity of older adults.

Another source of stress and discomfort from using AI-based toilet seats is age-related stereotypes, which have been examined in previous studies [83]. Older adults may feel the stigma associated with using intelligent toilet seats as others (family or friends) would relate this need to aging and infirm health conditions. They may feel that using this technology is seen as a sign of weakness and dependence, or that it is only for older people with serious health problems. However, many older adults may use this technology only for prevention or early detection of diseases, not because they currently experience a serious infection. Previous studies indicate that perceived stigma can harm the dignity and self-esteem of older adults by making them feel ashamed of their use of technology [84]. It can also prevent older adults from embracing age tech and taking advantage of its benefits.

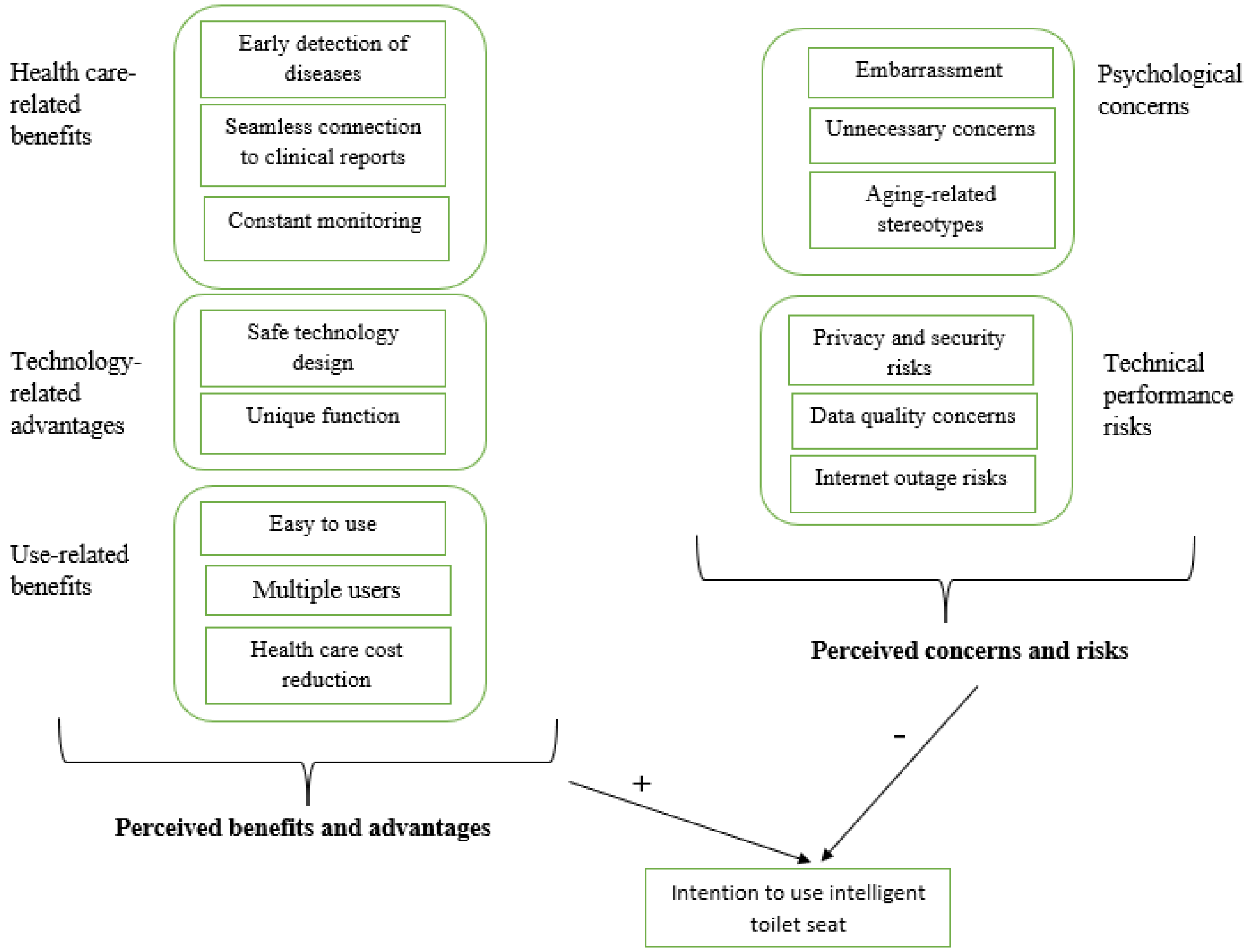
Technical Performance Risks

As highlighted in previous studies, concerns about potential technical issues are integral to using AI-based devices in health care [85]. Technical concerns can arise because AI algorithms embedded in age tech may produce inaccurate results or make incorrect decisions, primarily if they are based on faulty data or trained on limited, outdated, or biased data sets [86]. Older adults might be concerned about the quality and reliability of collected data, as sensors and AI algorithms mismeasure wellness parameters; all further analyses will also be flawed. As intelligent toilet seats send toileting logs over the internet to a central database, similar to other AI-based devices, they can be susceptible to hacking or other security threats, compromising sensitive information or disrupting their operation. Intelligent toilet seats function over the internet, and the probability of transmitting real-time data is low when the internet connection is poor. In this case, it is plausible that the system misses part of the data that could be very valuable for diagnosis and choosing treatment options. In addition, similar to other AI-powered tools, using AI algorithms in intelligent toilet seats can raise concerns about privacy and the protection of personal data [87], as AI systems may collect, store, and analyze vast amounts of information about older adults.

A Proposed Guiding Framework

The results indicate that older adults expect several benefits from using intelligent toilet seats; however, concerns and risks are not negligible. The intention to adopt these AI-based toilet seats among older adults depends on comparing perceived positive and negative opinions associated with the system. Thus, if AI-powered toilet seats deliver less added value to older adults, they will negatively favor integrating the system into their daily lives. The guiding framework outlines the key variables, relationships, and mechanisms that are expected to exist in the research phenomenon being studied. This study contributes to the current discussion on using AI-powered systems in health care by providing a better picture of the potential benefits and possible concerns that may shape older adults' perceptions about using intelligent toilet seats. Figure 1 presents a guiding framework illustrating the primary themes, categories, and concepts derived from the survey interviews.

Figure 1. Guiding framework.



Practical Suggestions

On the basis of the current state of research in age tech and our findings of older adults' perceptions and adoption of intelligent toilet seats, this study offers several practical recommendations and effective interventions that address both the facilitators and barriers to adoption, aiming to enhance the acceptance and use of AI-powered tools within the aging population.

1. *Increasing awareness and understanding*: Developers of AI-based toilet seats need to increase awareness and understanding of age technology through education and training programs for potential users focusing on its benefits and safe use. Educating older adults on the potential benefits of age tech can provide open and honest conversations about it on vendors' websites, such as transparent, frequently asked questions, or web-based forums.
2. *Keeping technology accessible and easy to use*: Many older adults have physical conditions or health issues that make it difficult to use new technologies. Developers need to make intelligent toilet seats accessible and user-friendly, for example, by offering simple-to-use interfaces with little or no instructions, uncomplicated interactions with technology, and safe design. Moreover, integrating this age tech into health care services, such as telehealth and remote monitoring, can increase its accessibility and perceived value and make it more acceptable and accessible to older adults.
3. *Encouraging social support and reducing stigma*: To protect the dignity of older adults, it is essential to address the stigma surrounding technology use. One plausible way is to promote social support networks, such as family and friends, to encourage and assist older adults in adopting intelligent toilet seats. Educating older adults on the benefits of this technology is necessary via marketing efforts to alleviate the stigma associated with intelligent toilet seats. Encouraging them to try it and see how it can improve their daily life can also help to reduce stigma. It is also important to highlight that using this technology does not necessarily mean an older adult is dependent or weak. Instead, it can be seen as proactive and responsible, allowing individuals to care for their health and well-being.
4. *Offering tailored interventions and policies*: Developing tailored interventions specific to the needs and experiences of different older adult populations, such as those with chronic conditions or low income. Older adults may have

different preferences and comfort levels with technology; therefore, assessing their needs and finding a solution that works best for them is crucial. For example, older adults with chronic issues may feel stressed, and receiving frequent warnings may exacerbate their uncomfortable experience with the system. Frequent alerts can be stressful and overwhelming for older adults with chronic illnesses. It is essential to find a balance between providing helpful reminders and notifications and avoiding adding to the stress and anxiety that older individuals may already be experiencing. One possible solution is to allow older adults to customize their notification settings and to choose the frequency and type of reminders they would like to receive. For example, some individuals may prefer to receive a daily summary of their health information. In contrast, others may choose to receive real-time notifications for critical events or changes in their condition. Moreover, using gentle and nonintrusive reminders could help AI-based toilet seat adoption be approached thoughtfully and compassionately. Another tailored example is the feasibility of using an automated bidet intervention to decrease physical assistance required from caregivers for toileting and toilet hygiene [88].

5. *Reassurance on system uptime*: Developers need to provide information about how intelligent toilet seats have been tested and validated and the measures in place to ensure their uptime. Some steps can be taken to mitigate the impact of internet outages. For instance, designing the seats with an offline mode so that they can continue to function even if the internet connection is lost. This could include storing data locally and syncing it with a remote server when the connection is restored. Moreover, seats can regularly back up the collected data to ensure that they are not lost during an internet outage. Thus, developing emergency protocols for internet outages should assure users that system downtime does not disrupt patient care.
6. *Monitoring and evaluating AI algorithms*: AI systems can sometimes make decisions that result in unintended consequences. Biases, discriminatory components, and errors may influence their mechanism in the data on which they are trained. The development of AI systems needs to prioritize safety, security, and ethical considerations. In addition, it is important to continuously monitor and evaluate their algorithms to identify and address potential risks. Thus, system designers must regularly check the effectiveness and safety of AI components embedded in intelligent toilet seats to promote positive health outcomes among older adults.
7. *Addressing privacy and security issues*: System designers need to be mindful of privacy and security concerns and the potential for misinterpretation of data. It is important to use AI algorithms that are based on reliable and unbiased data. In addition, developers must design AI systems that are secure, interoperable, and scalable to ensure that appropriate privacy and security measures are in place.
8. *Involving older adults*: Giving roles to older adults in selecting and setting up their age tech can help them embrace it faster. By providing older people with control over the system and helping them understand how intelligent toilet seats work to provide seamless connections to health care data, they are more likely to feel comfortable and confident using it. This policy can also help to build trust and increase technology adoption. Our results show that the best use case for implementing intelligent toilet seats seems to be in nursing homes. Nevertheless, it is worth noting that using this age tech is a personal choice and that older adults should not feel pressured to use it if uncomfortable. Everyone has the right to make decisions about their health and well-being, which should be respected and supported.

Limitations and Directions for Future Studies

This qualitative research on older adults' perceptions about toilet seats equipped with AI can provide valuable insights into their attitudes and perceptions, but it also has some limitations. First, the sample size of older adults in the study (n=174) may not be representative of the larger population of older adults in the United States, limiting the generalizability of the findings. It would be interesting for future studies to examine the perceptions of a larger sample of older adults. Second, older adults may have limited literacy on emerging technologies used in health care, which could impact their understanding of AI-powered technology and their perceptions of its use in health care. Future studies should focus on aging adults who are fairly familiar with age tech and AI-based systems in health care. Third, we used a web-based survey using MTurk as a recruitment method. This data collection method may introduce bias into the sample, as older adults who are more open to technology might be more likely to participate in the study. Thus, future research can extend this study by using other recruitment methods (such as in-person interviews) to examine a more representative sample. Fourth, although the insights from this study offer valuable perspectives on the acceptance and potential use of smart toilet seats by older adults, it is crucial to note that none of the respondents had first-hand experience with these devices before the study. As such, their feedback was based on perceptions rather than on actual use. For a more comprehensive understanding, future research should consider examining the experiences of individuals who have actively used smart toilet seats. This would directly assess the utility, challenges, and benefits associated with these innovative devices.

Fifth, the process of analyzing qualitative data can be challenging because it often involves coding and categorizing large amounts of unstructured data. In this study, 2 researchers familiar with age tech analyzed and coded the data. As qualitative studies rely heavily on the researcher's interpretation of data, the researchers' personal experiences with AI technology could influence their interpretation of the data and the findings. Future researchers can build upon this study by incorporating a wider range of perspectives and reducing researcher bias through collaboration with larger and more diverse research teams. This could result in identifying additional concepts, themes, or quotes that may have been overlooked in this study. Sixth, the respondents in this study have not raised the cost of using smart toilet seats. However, smart toilet seats with advanced biometric monitoring capabilities could cost approximately US $1000 plus installation fees, presenting a major financial barrier for most older adults. At this stage, there is little evidence that health insurance providers or medical practices are promoting smart toilets for in-home use. The high costs raise questions of how average consumers could realistically afford these devices. Further research is needed on potential subsidization or creative financing models to improve access to the potential benefits of smart toilets. Seventh, we did not focus on a particular smart toilet brand to gain a broader understanding of general user perceptions and attitudes toward the technology. Nonetheless, this could be an interesting area of research for future studies to delve deeper into brand-specific comparisons and their varied functionalities. Finally, this qualitative study identified key factors (perceived benefits and risks) influencing older adults' intentions to use intelligent toilet seats. However, further quantitative research is required to fully understand the impact of these factors. This could involve examining the significance and importance of these constructs in implementing this technology in nursing homes. In addition, more studies are required to test and validate the relationships between the variables identified in the guiding model to determine the proposed model's predictive power. However, despite these limitations, this qualitative research can provide important insights into the attitudes and adoption of intelligent toilet seats among the aging population and help inform the design and implementation of AI-based devices in health care.

Conclusions

Previous studies indicate that age tech has the potential to greatly advance the lives of older adults by providing support, increasing independence, improving health care outcomes, and enhancing quality of life. For aging individuals and those with disabilities or chronic illnesses, smart toilets present an opportunity to live more independently while maintaining close monitoring of their health status. By sharing data and alerts with medical teams through telehealth platforms, smart toilets enable safer at-home living and can capture potential health issues early on. Their adoption stands to meaningfully improve the quality of life and health outcomes for susceptible populations. This study is an attempt to deeply explore, identify, and categorize aging adults' perceptions of using a particular age tech, namely intelligent toilet seats. As the use of AI-based toilet seats for monitoring toileting logs is still a relatively new area of technology, more research is needed to determine the full range of benefits and potential privacy concerns. This qualitative study provides valuable insights into the opinions and attitudes of older adults toward the use of an intelligent toilet seat. The findings indicate a positive attitude toward using this age tech owing to health care benefits, technology-related advantages, and use benefits. However, the respondents also raised concerns about psychological distress and technological performance risks. It is crucial to identify ways to maximize the benefits of the technology while minimizing the risks it may pose to promote the widespread adoption of intelligent toilet seats among older adults. Addressing the barriers and concerns when developing technological solutions for older adults and designing technology based on their needs, expectations, and abilities can help them embrace toilet seats equipped with AI solutions. The results suggest a need for more education and awareness about the benefits (ie, early detection of diseases) and limitations of AI-based toilet seats, as well as increased efforts to address mental stress, concerns about the reliability of AI algorithms, and privacy and security concerns. Further research is needed to understand the broader impact of technology on older adults and to develop solutions that are both accessible and beneficial to this population. This research contributes to the field of age tech and assists in developing effective and user-friendly technological solutions for older adults.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based survey.
[DOCX File , 19 KB - mhealth_v11i1e46430_app1.docx]

References

1. Jakobsson E, Nygård L, Kottorp A, Malinowsky C. Experiences from using eHealth in contact with health care among older adults with cognitive impairment. Scand J Caring Sci 2019 Jun;33(2):380-389. [doi: 10.1111/scs.12634] [Medline: 30628736]
2. Perissinotto C, Holt-Lunstad J, Periyakoil VS, Covinsky K. A practical approach to assessing and mitigating loneliness and isolation in older adults. J Am Geriatr Soc 2019 Apr;67(4):657-662. [doi: 10.1111/jgs.15746] [Medline: 30762228]
3. Etxeberria I, Etxebarria I, Urdaneta E. Profiles in emotional aging: does age matter? Aging Ment Health 2018 Oct;22(10):1304-1312. [doi: 10.1080/13607863.2017.1286450] [Medline: 28282728]
4. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J Soc Robot 2018 Mar 1;10:643-668. [doi: 10.1007/s12369-018-0471-x]
5. Isabet B, Pino M, Lewis M, Benveniste S, Rigaud AS. Social telepresence robots: a narrative review of experiments involving older adults before and during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Mar 30;18(7):3597 [FREE Full text] [doi: 10.3390/ijerph18073597] [Medline: 33808457]

6. Sixsmith A. COVID-19 and AgeTech. Qual Ageing Older Adults 2020 Nov 11;21(4):247-252. [doi: 10.1108/qaoa-07-2020-0029]
7. Hakobyan L, Lumsden J, O'Sullivan D, Bartlett H. Mobile assistive technologies for the visually impaired. Surv Ophthalmol 2013;58(6):513-528. [doi: 10.1016/j.survophthal.2012.10.004] [Medline: 24054999]
8. Dias D, Paulo Silva Cunha J. Wearable health devices-vital sign monitoring, systems and technologies. Sensors (Basel) 2018 Jul 25;18(8):2414 [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]
9. Yamagata C, Coppola JF, Kowtko M, Joyce S. Mobile app development and usability research to help dementia and Alzheimer patients. In: Proceedings of the IEEE Long Island Systems, Applications and Technology Conference (LISAT). 2013 Presented at: IEEE Long Island Systems, Applications and Technology Conference (LISAT); May 3, 2013; Farmingdale, NY. [doi: 10.1109/lisat.2013.6578252]
10. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
11. Malwade S, Abdul SS, Uddin M, Nursetyo AA, Fernandez-Luque L, Zhu XK, et al. Mobile and wearable technologies in healthcare for the ageing population. Comput Methods Programs Biomed 2018 Jul;161:233-237. [doi: 10.1016/j.cmpb.2018.04.026] [Medline: 29852964]
12. Shestakova NN, Djanelidze MG, Skvortsova MB. AgeTech innovations as an instrument for older population inclusion. In: Proceedings of the 3rd International Scientific Conference on Sustainable Development (ESG 2022). 2022 Presented at: 3rd International Scientific Conference on Sustainable Development (ESG 2022); 2022; St. Petersburg, FL. [doi: 10.1007/978-3-031-30498-9_11]
13. Sixsmith A, Sixsmith J, Fang ML, Horst B. AgeTech for cognitive health and dementia. In: AgeTech, Cognitive Health, and Dementia. Cham, Switzerland: Springer; 2020.
14. Chu CH, Biss RK, Cooper L, Quan AM, Matulis H. Exergaming platform for older adults residing in long-term care homes: user-centered design, development, and usability study. JMIR Serious Games 2021 Mar 09;9(1):e22370 [FREE Full text] [doi: 10.2196/22370] [Medline: 33687337]
15. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
16. Schicktanz S, Schweda M. Aging 4.0? Rethinking the ethical framing of technology-assisted eldercare. Hist Philos Life Sci 2021 Aug 03;43(3):93 [FREE Full text] [doi: 10.1007/s40656-021-00447-x] [Medline: 34342739]
17. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022 Oct 05;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
18. Sixsmith A. AgeTech: technology-based solutions for aging societies. In: Rootman I, Edwards P, Levasseur M, Grunberg F, editors. Promoting the Health of Older Adults: The Canadian Experience. Toronto, ON: Canadian Scholars; Jul 2021.
19. Liu L, Daum C, Miguel Cruz A, Neubauer N, Perez H, Ríos Rincón A. Ageing, technology, and health: advancing the concepts of autonomy and independence. Healthc Manage Forum 2022 Sep;35(5):296-300 [FREE Full text] [doi: 10.1177/08404704221110734] [Medline: 35924794]
20. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Oct;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
21. Abadir PM, Chellappa R, Choudhry N, Demiris G, Ganesan D, Karlawish J, et al. The promise of AI and technology to improve quality of life and care for older adults. Nat Aging 2023 Jun;3(6):629-631. [doi: 10.1038/s43587-023-00430-0] [Medline: 37231197]
22. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/#:~:text=Some%2034%25%20of%20older%20internet,me%20how%20to%20use%20it.%E2%80%9D [accessed 2023-02-19]
23. Wang X, He Y, Zhang H. How to influence behavioral intention toward age-friendly home modifications in urban older people aged 70. Gerontol Geriatr Med 2023 Feb 25;9:23337214231152697 [FREE Full text] [doi: 10.1177/23337214231152697] [Medline: 36860701]
24. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
25. Molino M, Cortese CG, Ghislieri C. Technology acceptance and leadership 4.0: a quali-quantitative study. Int J Environ Res Public Health 2021 Oct 15;18(20):10845 [FREE Full text] [doi: 10.3390/ijerph182010845] [Medline: 34682588]
26. Winson A. The Industrial Diet: The Degradation of Food and the Struggle for Healthy Eating. Vancouver, BC: University of British Columbia Press; 2013.
27. Kaushik K, Bhardwaj A, Dahiya S. Smart home IoT forensics: current status, challenges, and future directions. In: Proceedings of the International Conference on Advancement in Computation & Computer Technologies (InCACCT). 2023 Presented at: International Conference on Advancement in Computation & Computer Technologies (InCACCT); May 5-6, 2023; Gharuan, India. [doi: 10.1109/incacct57535.2023.10141730]

28. Global smart toilet market research report 2023. Valuates Reports. URL: https://reports.valuates.com/market-reports/QYRE-Othe-3U404/smart-toilet [accessed 2023-10-20]
29. Smart toilet market size to reach USD 22.20 billion at a CAGR of 15.12% by 2030 - report by market research future (MRFR). Globe Newswire. 2023 Feb 28. URL: https://www.globenewswire.com/news-release/2023/02/28/2617058/0/en/Smart-Toilet-Market-Size-To-Reach-USD-22-20-Billion-at-a-CAGR-of-15-12-by-2030-Report-by-Market-Research-Future-MRFR.html [accessed 2023-10-20]
30. Global smart toilet market size, share and industry trends analysis report by application (commercial and residential), by distribution channel (offline and online), by regional outlook and forecast, 2022 - 2028. Research and Markets. 2022 Jul. URL: https://www.researchandmarkets.com/reports/5645775/global-smart-toilet-market-size-share-and [accessed 2023-10-14]
31. Armitage H. ‘Smart toilet’ monitors for signs of disease. Stanford Medicine. 2020 Apr 6. URL: https://med.stanford.edu/news/all-news/2020/04/smart-toilet-monitors-for-signs-of-disease.html#:~:text=A%20disease%2Ddetecting%20%E2%80%9Cprecision%20health,a%20new%20Stanford%20study%20reports.&text=The%20smart%20toilet%20automatically%20sends,cloud%2Dbased%20system%20for%20safekeeping [accessed 2023-02-19]
32. Dorman A. ‘Smart’ toilet seat tracks patients’ health, aids in diagnoses. McKnight's Senior Living. 2023 Jun 26. URL: https://www.mcknightsseniorliving.com/home/news/tech-daily-news/smart-toilet-seat-tracks-patients-health-aids-in-diagnoses/ [accessed 2023-10-20]
33. Ge TJ, Rahimzadeh VN, Mintz K, Park WG, Martinez-Martin N, Liao JC, et al. Passive monitoring by smart toilets for precision health. Sci Transl Med 2023 Feb;15(681):eabk3489 [FREE Full text] [doi: 10.1126/scitranslmed.abk3489] [Medline: 36724240]
34. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105 [FREE Full text] [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
35. Perdana A, Mokhtar IA. Seniors' adoption of digital devices and virtual event platforms in Singapore during Covid-19. Technol Soc 2022 Feb;68:101817 [FREE Full text] [doi: 10.1016/j.techsoc.2021.101817] [Medline: 34785827]
36. Orlov L. 10 barriers to boosting tech adoption by older adults in 2021. Aging and Health Technology Watch. 2021 Aug 18. URL: https://www.ageinplacetech.com/blog/barriers-boosting-tech-adoption-older-adults-2021 [accessed 2023-02-18]
37. Ageing and health. World Health Organization. 2022 Oct 1. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2023-10-14]
38. Simpson RC. Smart wheelchairs: a literature review. J Rehabil Res Dev 2005;42(4):423-436 [FREE Full text] [doi: 10.1682/jrrd.2004.08.0101] [Medline: 16320139]
39. Pal D, Papasratorn B, Chutimaskul W, Funilkul S. Embracing the smart-home revolution in Asia by the elderly: an end-user negative perception modeling. IEEE Access 2019;7:38535-38549. [doi: 10.1109/access.2019.2906346]
40. Kurnianingsih K, Nugroho L, Widyawan W, Lazuardi L, Prabuwono AS, Mantoro T. Personalized adaptive system for elderly care in smart home using fuzzy inference system. Int J Pervasive Comput Commun 2018 Sep 03;14(3/4):210-232. [doi: 10.1108/IJPCC-D-18-00002]
41. Borelli E, Paolini G, Antoniazzi F, Barbiroli M, Benassi F, Chesani F, et al. HABITAT: an IoT solution for independent elderly. Sensors (Basel) 2019 Mar 12;19(5):1258 [FREE Full text] [doi: 10.3390/s19051258] [Medline: 30871107]
42. Marques B, McIntosh J, Valera A, Gaddam A. Innovative and assistive eHealth technologies for smart therapeutic and rehabilitation outdoor spaces for the elderly demographic. Multimodal Technol Interact 2020 Oct 22;4(4):76. [doi: 10.3390/mti4040076]
43. Park SM, Won DD, Lee BJ, Escobedo D, Esteva A, Aalipour A, et al. A mountable toilet system for personalized health monitoring via the analysis of excreta. Nat Biomed Eng 2020 Jun;4(6):624-635 [FREE Full text] [doi: 10.1038/s41551-020-0534-9] [Medline: 32251391]
44. Tasoglu S. Toilet-based continuous health monitoring using urine. Nat Rev Urol 2022 Apr;19(4):219-230. [doi: 10.1038/s41585-021-00558-x] [Medline: 35064251]
45. Smart toilet market size by application (residential and commercial), by distribution channel (offline and online), regions, global industry analysis, share, growth, trends, and forecast 2023 to 2032. The Brainy Insights. URL: https://www.thebrainyinsights.com/report/smart-toilet-market-13543 [accessed 2023-10-24]
46. Huang JJ, Yu SI, Syu HY. Development of the smart toilet equipment with measurements of physiological parameters. In: Proceedings of the 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing. 2012 Presented at: 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing; September 4-7, 2012; Fukuoka, Japan. [doi: 10.1109/uic-atc.2012.143]
47. Hermsen S, Verbiest V, Buijs M, Wentink E. Perceived use cases, barriers, and requirements for a smart health-tracking toilet seat: qualitative focus group study. JMIR Hum Factors 2023 Aug 11;10:e44850 [FREE Full text] [doi: 10.2196/44850] [Medline: 37566450]
48. Balaceanu C, Marcu I, Suciu G, Dantas C, Mayer P. Developing a smart toilet system for ageing people and persons with disabilities. In: Proceedings of the 6th Conference on the Engineering of Computer Based Systems. 2019 Presented at:

ECBS '19: 6th Conference on the Engineering of Computer Based Systems; September 2-3, 2019; Bucharest, Romania. [doi: 10.1145/3352700.3352716]
49. Lumetzberger J, Mayer P, Kampel M, Panek P. Smart toilet seat configuration for more autonomy using an AI-based 3D depth sensor. In: Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference. 2021 Presented at: PETRA '21: The 14th PErvasive Technologies Related to Assistive Environments Conference; June 29-July 2, 2021; Corfu, Greece. [doi: 10.1145/3453892.3464895]
50. Bae JH, Lee HK. User health information analysis with a urine and feces separable smart toilet system. IEEE Access 2018 Dec 05;6:78751-78765. [doi: 10.1109/access.2018.2885234]
51. Caldeira C, Nurain N, Heintzman AA, Molchan H, Caine K, Demiris G, et al. How do I compare to the other people?": older adults' perspectives on personal smart home data for self-management". Proc ACM Hum Comput Interact 2023 Oct 04;7(CSCW2):1-32. [doi: 10.1145/3610029]
52. Berg BL, Lune H. Qualitative Research Methods for the Social Sciences. London, UK: Pearson Education; Nov 15, 2011.
53. DiStaso MW, Bortree DS. Multi-method analysis of transparency in social media practices: survey, interviews and content analysis. Public Relat Rev 2012 Sep;38(3):511-514. [doi: 10.1016/j.pubrev.2012.01.003]
54. Al-Salom P, Miller CJ. The problem with online data collection: predicting invalid responding in undergraduate samples. Curr Psychol 2017 Sep 9;38(5):1258-1264. [doi: 10.1007/s12144-017-9674-9]
55. Marge M, Banerjee S, Rudnicky AI. Using the Amazon Mechanical Turk for transcription of spoken language. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2010 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; March 14-19, 2010; Dallas, TX URL: https://ieeexplore.ieee.org/abstract/document/5494979 [doi: 10.1109/icassp.2010.5494979]
56. Mellis AM, Bickel WK. Mechanical Turk data collection in addiction research: utility, concerns and best practices. Addiction 2020 Oct 24;115(10):1960-1968 [FREE Full text] [doi: 10.1111/add.15032] [Medline: 32135574]
57. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018 Apr 4;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
58. Zhang B, Gearhart S. Collecting online survey data: a comparison of data quality among a commercial panel and MTurk. Surv Pract 2020 Dec 03;13(1):1-10. [doi: 10.29115/SP-2020-0015]
59. Turner AM, Engelsma T, Taylor JO, Sharma RK, Demiris G. Recruiting older adult participants through crowdsourcing platforms: Mechanical Turk versus prolific academic. AMIA Annu Symp Proc 2020;2020:1230-1238 [FREE Full text] [Medline: 33936499]
60. Ogletree AM, Katz B. How do older adults recruited using MTurk differ from those in a national probability sample? Int J Aging Hum Dev 2021 Sep 20;93(2):700-721. [doi: 10.1177/0091415020940197] [Medline: 32683886]
61. Shaw A, Hargittai E. Do the online activities of Amazon Mechanical Turk workers mirror those of the general population? A comparison of two survey samples. Int J Commun 2021;15 [FREE Full text]
62. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: SAGE Publications; 2019.
63. Blair E. A reflexive exploration of two qualitative data coding techniques. J Methods Measure Soc Sci 2015;6(1):14-29. [doi: 10.2458/v6i1.18772]
64. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research: norms and guidelines for CSCW and HCI practice. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-23. [doi: 10.1145/3359174]
65. Bornmann L, Mutz R, Daniel HD. A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants. PLoS One 2010 Dec 14;5(12):e14331 [FREE Full text] [doi: 10.1371/journal.pone.0014331] [Medline: 21179459]
66. Kwan G, Shaw JA, Murnane L. Internet usage within healthcare: how college students use the internet to obtain health information. J Consum Health Internet 2019 Dec 02;23(4):366-377. [doi: 10.1080/15398285.2019.1681247]
67. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online health information seeking among US adults: measuring progress toward a healthy people 2020 objective. Public Health Rep 2019 Sep 12;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]
68. Demiris G, Hensel BK. Technologies for an aging society: a systematic review of "smart home" applications. Yearb Med Inform 2008:33-40. [Medline: 18660873]
69. Bail K, Gibson D, Acharya P, Blackburn J, Kaak V, Kozlovskaia M, et al. Using health information technology in residential aged care homes: an integrative review to identify service and quality outcomes. Int J Med Inform 2022 Sep;165:104824 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104824] [Medline: 35792376]
70. Ijaz N, Buta B, Xue QL, Mohess DT, Bushan A, Tran H, et al. Interventions for frailty among older adults with cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 2022 Feb 08;79(5):482-503 [FREE Full text] [doi: 10.1016/j.jacc.2021.11.029] [Medline: 35115105]
71. Lawless MT, Marshall A, Mittinty MM, Harvey G. What does integrated care mean from an older person's perspective? A scoping review. BMJ Open 2020 Jan 22;10(1):e035157 [FREE Full text] [doi: 10.1136/bmjopen-2019-035157] [Medline: 31974092]

72. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc Manage Forum 2020 Jan 24;33(1):10-18. [doi: 10.1177/0840470419873123] [Medline: 31550922]
73. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
74. Eby DW, Molnar LJ, Zhang L, St Louis RM, Zanier N, Kostyniuk LP, et al. Use, perceptions, and benefits of automotive technologies among aging drivers. Inj Epidemiol 2016 Dec 19;3(1):28 [FREE Full text] [doi: 10.1186/s40621-016-0093-4] [Medline: 27868168]
75. Rubeis G, Fang ML, Sixsmith A. Equity in AgeTech for ageing well in technology-driven places: the role of social determinants in designing AI-based assistive technologies. Sci Eng Ethics 2022 Oct 27;28(6):49 [FREE Full text] [doi: 10.1007/s11948-022-00397-y] [Medline: 36301408]
76. Czaja SJ, Boot WR, Charness N, Rogers WA. Designing for Older Adults: Principles and Creative Human Factors Approaches, Third Edition. Boca Raton, FL: CRC Press; 2019.
77. Fritz RL, Dermody G. A nurse-driven method for developing artificial intelligence in "smart" homes for aging-in-place. Nurs Outlook 2019 Mar;67(2):140-153 [FREE Full text] [doi: 10.1016/j.outlook.2018.11.004] [Medline: 30551883]
78. Li W, Chai Y, Khan F, Jan SR, Verma S, Menon VG, Kavita, et al. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mobile Netw Appl 2021 Jan 06;26(1):234-252. [doi: 10.1007/s11036-020-01700-6]
79. Wang S, Bolling K, Mao W, Reichstadt J, Jeste D, Kim HC, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019 Apr 10;7(2):60 [FREE Full text] [doi: 10.3390/healthcare7020060] [Medline: 30974780]
80. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 01;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
81. Panico F, Cordasco G, Vogel C, Trojano L, Esposito A. Ethical issues in assistive ambient living technologies for ageing well. Multimed Tools Appl 2020 Jul 17;79(47-48):36077-36089. [doi: 10.1007/s11042-020-09313-7]
82. Abouzahra M, Ghasemaghaei M. The antecedents and results of seniors’ use of activity tracking wearable devices. Health Policy Technol 2020 Jun;9(2):213-217. [doi: 10.1016/j.hlpt.2019.11.002]
83. Mariano J, Marques S, Ramos MR, Gerardo F, Cunha CL, Girenko A, et al. Too old for technology? Stereotype threat and technology use by older adults. Behav Inf Technol 2021 Feb 07;41(7):1503-1514. [doi: 10.1080/0144929x.2021.1882577]
84. Coghlan S, Waycott J, Lazar A, Neves BB. Dignity, autonomy, and style of company: dimensions older adults consider for robot companions. Proc ACM Hum Comput Interact 2021 Apr 22;5(CSCW1):1-25 [FREE Full text] [doi: 10.1145/3449178] [Medline: 34308262]
85. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020 Jul 22;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
86. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
87. Zhang Z, Shi Q, He T, Guo X, Dong B, Lee J, et al. Artificial intelligence of toilet (AI-Toilet) for an integrated health monitoring system (IHMS) using smart triboelectric pressure sensors and image sensor. Nano Energy 2021 Dec;90:106517. [doi: 10.1016/j.nanoen.2021.106517]
88. Bollinger R, Somerville E, Keglovits M, Hu YL, Stark S. Feasibility of an automated bidet intervention to decrease caregiver burden. Am J Occup Ther 2021 Sep 01;75(5):5020. [doi: 10.5014/ajot.2021.043919] [Medline: 34780632]

Abbreviations

Age tech: age technology
AI: artificial intelligence
ICC: intraclass correlation coefficient
IoT: Internet of Things
IRR: interrater reliability
MTurk: Mechanical Turk

Edited by L Buis; submitted 11.02.23; peer-reviewed by M Jordan-Marsh, J Sanford, A Peeples; comments to author 13.10.23; revised version received 19.10.23; accepted 06.11.23; published 01.12.23.

Please cite as:
Esmaeilzadeh P
Older Adults' Perceptions About Using Intelligent Toilet Seats Beyond Traditional Care: Web-Based Interview Survey
JMIR Mhealth Uhealth 2023;11:e46430
URL: https://mhealth.jmir.org/2023/1/e46430
doi:10.2196/46430
PMID:38039065

Original Paper

Effects of User-Reported Risk Factors and Follow-Up Care Activities on Satisfaction With a COVID-19 Chatbot: Cross-Sectional Study

Akanksha Singh[1*], MS, PhD; Benjamin Schooley[2*], MBA, PhD; Nitin Patel[3], MD

[1]Department of Health Services Administration, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[2]IT & Cybersecurity, Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT, United States

[3]Hackensack Meridian Health, Hackensack, NJ, United States

[*]these authors contributed equally

Corresponding Author:
Benjamin Schooley, MBA, PhD
IT & Cybersecurity
Department of Electrical and Computer Engineering
Brigham Young University
240 Engineering Building
Provo, UT, 84602
United States
Phone: 1 8014220027
Email: Ben_Schooley@byu.edu

Abstract

Background: The COVID-19 pandemic influenced many to consider methods to reduce human contact and ease the burden placed on health care workers. Conversational agents or chatbots are a set of technologies that may aid with these challenges. They may provide useful interactions for users, potentially reducing the health care worker burden while increasing user satisfaction. Research aims to understand these potential impacts of chatbots and conversational recommender systems and their associated design features.

Objective: The objective of this study was to evaluate user perceptions of the helpfulness of an artificial intelligence chatbot that was offered free to the public in response to COVID-19. The chatbot engaged patients and provided educational information and the opportunity to report symptoms, understand personal risks, and receive referrals for care.

Methods: A cross-sectional study design was used to analyze 82,222 chats collected from patients in South Carolina seeking services from the Prisma Health system. Chi-square tests and multinomial logistic regression analyses were conducted to assess the relationship between reported risk factors and perceived chat helpfulness using chats started between April 24, 2020, and April 21, 2022.

Results: A total of 82,222 chat series were started with at least one question or response on record; 53,805 symptom checker questions with at least one COVID-19–related activity series were completed, with 5191 individuals clicking further to receive a virtual video visit and 2215 clicking further to make an appointment with a local physician. Patients who were aged >65 years (P<.001), reported comorbidities (P<.001), had been in contact with a person with COVID-19 in the last 14 days (P<.001), and responded to symptom checker questions that placed them at a higher risk of COVID-19 (P<.001) were 1.8 times more likely to report the chat as helpful than those who reported lower risk factors. Users who engaged with the chatbot to conduct a series of activities were more likely to find the chat helpful (P<.001), including seeking COVID-19 information (3.97-4.07 times), in-person appointments (2.46-1.99 times), telehealth appointments with a nearby provider (2.48-1.9 times), or vaccination (2.9-3.85 times) compared with those who did not perform any of these activities.

Conclusions: Chatbots that are designed to target high-risk user groups and provide relevant actionable items may be perceived as a helpful approach to early contact with the health system for assessing communicable disease symptoms and follow-up care options at home before virtual or in-person contact with health care providers. The results identified and validated significant design factors for conversational recommender systems, including triangulating a high-risk target user population and providing relevant actionable items for users to choose from as part of user engagement.

(JMIR Mhealth Uhealth 2023;11:e43105) doi:10.2196/43105

KEYWORDS

patient engagement; chatbot; population health; health recommender systems; conversational recommender systems; design factors; COVID-19

Introduction

Background and Significance

Since the discovery of SARS-CoV-2 and subsequent COVID-19 disease outbreak began in December 2019, the pandemic has challenged health care systems and societies worldwide to respond in unprecedented ways to protect both patients and staff and prepare for surges of patients who are critically ill. Health care institutions have revamped services and delivery, pushing digital transformations and innovative models for crisis management and health care delivery at an unforeseen scale [1-3]. An important goal has been to manage patient load, directing a large number of patients and the general public to an appropriate level of care [4]. Indeed, a variety of patient support pathways for COVID-19 have become more commonplace [5], including those facilitated by telehealth, telephony [6], email, and interactive chat [7,8]. These technologies have been applied during the pandemic, for example, in primary care visits [9], patient-family communications [10], postdischarge follow-ups, and palliative care [11-13]. A range of benefits have been reported. For example, robust, physician-directed 24/7 COVID-19 response hotlines [14] can help meet increased health care service demands during the acute phase of a pandemic, conserving scarce resources such as personal protective equipment and testing supplies and preventing the spread of infections to patients and health care workers. These trends have motivated investigation into new and better ways to provide remote care facilitated by technology, such as remote patient monitoring, e-visits, e-consults, mobile health, and tele-education [15].

This study aimed to evaluate the impacts of an interactive conversational recommender telehealth system (ie, chatbot) for engaging users during the COVID-19 pandemic in the state of South Carolina to address the health care needs of users while also trying to reduce exposure to the virus for users and providers alike. We assessed the impact of its use to inform the design considerations of future conversational recommender systems.

During the period following the outbreak of COVID-19, health care delivery was dramatically pushed toward telehealth and telemedicine [16-19]. Telehealth technologies have provided an effective and efficient way to deliver health care during the pandemic [20,21] for older populations [22] and patients with chronic conditions [23] and to reduce burnout for physicians working under epidemic conditions [24]. Telehealth virtual visits have reduced in-person visits by 33% [25] and are believed to have significant potential for positive socioeconomic impact on health care delivery [26-28]. Telehealth has rapidly expanded to include low-acuity conditions (including acute respiratory infections) [29] and infectious disease consultation, diagnosis, and monitoring [30].

Virtual visits have taken different forms, including synchronous video visits either prescheduled or on demand [31], telephony-based care, and chat-based virtual visits with providers, and more recently, chatbots have been introduced [32,33]. Artificial intelligence (AI) chatbots have also made their way into remote care practice [34-38]. Although physicians have reported mixed views on the value of chatbots, some studies have found that providers acknowledge cost savings and benefits when the functionality of the chatbot plays a clear strategic role for an organization [39,40].

Since the COVID-19 outbreak, chatbots have been used to provide remote assessments to triage potential patients [41,42], expand access to health care education, and try to ease supply and demand challenges for human health care providers [43]. Challenges to wider adoption include the need for more humanlike conversations and social and ethical considerations [13,34,44,45].

Chatbots for Patient Engagement and Satisfaction

Personalization and evaluation of chatbots are needed to enhance patient engagement and satisfaction. Direct-to-consumer telehealth satisfaction has generally been high among patients [46], with one study reporting satisfaction with 85% of virtual visits [47]. However, significant variation in quality has been found across virtual visit providers for the management of common acute illnesses [48]. In one study, the quality of telehealth was lacking in terms of appropriateness of ordering tests and scheduling follow-up visits for patients with acute respiratory infections [49].

Although satisfaction levels have been high for video-based direct-to-consumer virtual visits, little is known about satisfaction and patient engagement with emerging technologies such as asynchronous chatbots. Benefits have been found for the management of breast cancer [50] and the postoperative management of patients of orthopedics [51]. Other studies have reported low levels of patient engagement during ureteroscopic follow-up [52]. Factors affecting chatbot acceptance, encompassing engagement and satisfaction, include their perceived utility and trustworthiness. In contrast, factors such as poor patient computer skills and dislike for talking to computers have negatively affected patient satisfaction [32]. In contrast, the ability to personalize a chatbot experience, such as the selection of a preferred language, positively affects user trust and engagement [53]. Chatbots represent an interactive technology that requires assessment of its ability to engage users in a helpful and appealing way while reducing exposure and health care use.

Patient Engagement During Pandemics

We found very limited research on patient engagement strategies during an epidemic using tools such as chatbots. A recent study reported on a self-triage and self-scheduling tool to stratify and recommend appropriate care for patients with COVID-19 [54]. A similar triage tool was recently implemented for patients with multiple sclerosis to identify those who are at high risk for the purpose of reducing COVID-19 spread at multiple sclerosis centers [55].

As with many new technologies, chatbots carry the inherent risks of possible user resistance and technology immaturity. Dependence on emergent and complex technologies such as AI and natural language processing represents a significant risk for health care organizations as AI and natural language processing are still far from providing humanlike experiences inclusive of human cognition and empathy [32,56]. In contrast, chatbots offer advantages and benefits, such as the possibility of simplifying information search processes, the availability of easy-to-use instant messaging–like user interfaces, the capability to leverage previous user interactions for sentiment analysis, and the ability to provide personalized responses [56], while also reducing health care provider workload [57]. Understanding the benefits and risks of these technologies as they mature is important for assessing their impacts and future directions. In this study, we assessed the user-perceived helpfulness of a chatbot implemented at a large health care system.

Satisfaction with technology is a key metric to assess the impact of patient engagement strategies and interventions such as the use of chatbots for providing telehealth services during the COVID-19 pandemic. Recent studies on advances in clinical recommender systems and chatbots point to a need to explore evaluation paradigms for chatbots [58,59], including necessary constructs such as user satisfaction and effectiveness [60,61]. In this study, we used these evaluation paradigms to identify factors that inform the design of chatbots targeted to increase user satisfaction and effectiveness. We analytically triangulated the target population for a COVID-19 chatbot designed to engage patients who are at risk or feel that they are at risk of the disease and correlated the reported risk factors with chat helpfulness. On the basis of guidance from the Centers for Disease Control and Prevention (CDC), users at higher risk include those who are aged ≥65 years, report cough or breathing difficulty in the past few weeks, report any of the CDC-specified relevant comorbidities, have been in close contact with anyone who has tested positive for COVID-19 within the past 14 days, report CDC-specified COVID-19 symptoms in the past week, and are identified by the health system as being in the high-risk symptom category. Thus, users who are at a higher risk of COVID-19 will report the chat to be helpful more frequently than the group of users who are at a lower risk of COVID-19 (hypothesis 1).

We propose that users who find a relevant actionable item to engage with during their chat session will report the chat to be more helpful. Engagement means that the user finds and clicks on a hyperlink that navigates them to the described item. There may be many ways to help users engage with an action item regarding COVID-19, such as directing them to a CDC-specified COVID-19 information page, relevant links or telephone numbers to seek care, in-person provider appointments in their geographic area, telehealth appointments, vaccination information, information and telephone numbers to seek vaccination appointments in their area, and CDC travel guidelines to help them prepare for potential exposure. We posit that the use of user-relevant actionable items during a chat session increases user reporting of chat helpfulness. Thus, we analytically identified the relevant actionable items and links that are associated with user satisfaction with the chatbot. Hence, users who engage with an actionable item during their chat session will report the chat to be more helpful (hypothesis 2).

Methods

Overview

A case study approach was used to analyze a chatbot implementation by a large health system in the Southeastern United States. The health system aimed to provide free chatbot access to COVID-19 information resources, symptom checking, and referrals for additional in-person or virtual care to the public. Chatbot responses were collected by the case study organization and then analyzed by researchers using chi-square tests and multinomial logistic regression models to find significant differences between groups in which categorical responses were provided.

Study Setting

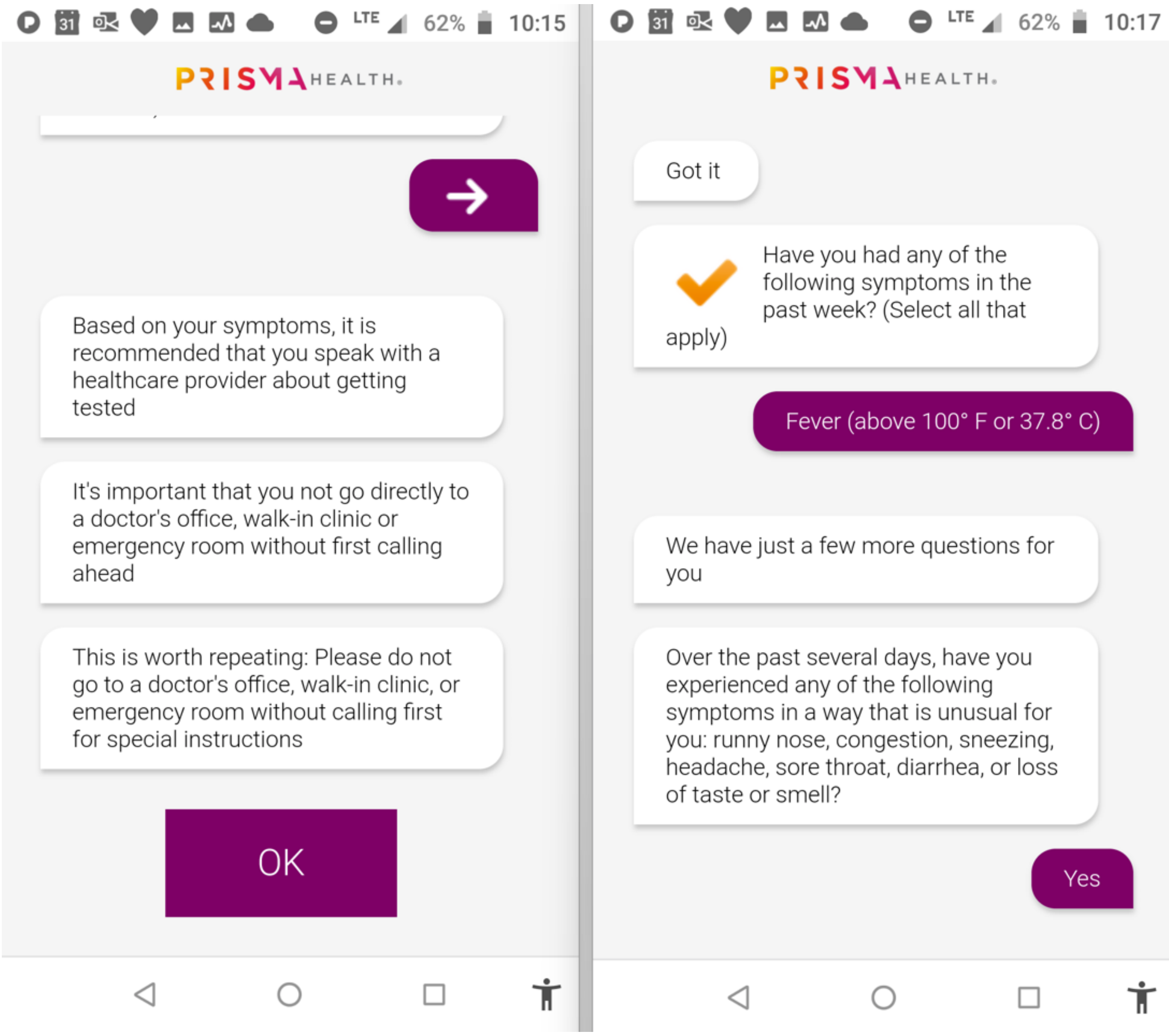
Prisma Health is the largest and most comprehensive health care system in the state of South Carolina, treating 1.2 million unique patients in the 2019 financial year with 330 physician practice sites and 18 hospitals inclusive of 2984 beds, 30,000 team members, 2 level-1 trauma centers, 2 comprehensive stroke centers, 2 affiliated medical schools and nursing schools, 50 residency and fellowship programs, and 560 residents or fellows. The health system serves a geographic area where 45% of South Carolina's 5.2 million residents live within a 15-minute drive [62]. As with most health care organizations grappling with COVID-19, Prisma Health was challenged with how best to provide access to care while protecting patients and employees from infection during the pandemic. There are significant population health challenges in South Carolina, including a high prevalence of chronic conditions. South Carolina was ranked 42 of all US states in terms of health determinants and outcomes in 2019 [63]. To address the needs of their patient population, Prisma Health developed a strategy for virtual care to serve as an important entry point into the health care system for both existing patients and the inquiring public. Easy-to-find connection points were constructed between asynchronous and synchronous virtual visits and scheduling of in-person primary care and specialty care services.

Chatbot Pathway

In response to COVID-19, Prisma Health moved its multiyear strategy to rapidly implement a chatbot in early 2020. Telehealth offerings were expanded to include virtual primary care visits, a free COVID-19 telehealth screening service, and a 1-800 telephony-based COVID-19 hotline. In April 2020, a chatbot was implemented to provide COVID-19 education; symptom screening; and referrals for follow-up care to the telephony service, video telehealth screening, and physicians' offices across the Prisma Health system. Chatbot technology is an essential component of Prisma Health's overall virtual care strategy to engage users with initial questions, guide users to appropriate levels of care, and provide a helpful experience to users while reducing staff time involved in discussing issues that can be more effectively addressed with interactive technologies. The chatbot asked a range of questions to individuals who freely and voluntarily used the system. The range and types of questions asked depended on how each individual user responded to the previous chatbot question, formulated as a decision tree. The chatbot placed each user in a symptom risk zone based on their responses. For the symptom zone assessment, red-yellow-green categories were predetermined by a clinician committee. “Red” was assigned to a chatbot user who reported any symptom or combination of symptoms with or without any other reported risk factor. For example, a person reporting “fever” (question 7) and a comorbidity of diabetes (question 4) was assigned the “Red” risk level. “Yellow” was assigned to any chatbot user reporting any comorbidity or close contact with a patient with COVID-19 in the last 14 days or who reported being aged >65 years. For example, a person not reporting any symptoms (question 7) but having a comorbidity (question 4) of diabetes was assigned “Yellow.” “Green” was assigned to users who did not report risk factors. The effective start date for the symptom zone assessment was September 16, 2020.

Figure 1 shows screenshots of a partial chatbot conversation checking for COVID-19 symptoms in a patient. As shown in these figures, relevant instructions or suggestions were provided to the user following each patient interaction with the symptom checker. Assessment of the Prisma Health COVID-19 chatbot for its helpfulness to consumers was core and central to this study. As noted previously, it was believed that the chatbot would be viewed as helpful by those who reported risk factors associated with the disease compared with those who reported no or fewer risk factors.

Figure 1. Screenshot of a partial Prisma Health chatbot COVID-19 symptom checker conversation.



Data Collection and Analysis

Data Collection and Coding

We collected data from the Prisma Health chatbot (powered by Conversa) between April 24, 2020, and April 21, 2022. Data were received including user chatbot responses and user follow-up actions (ie, activities). Secondary analysis of data included analysis of individual chat responses to COVID-19 symptom questions; education resources that were consumed about COVID-19; and user activities pertaining to seeking an in-person or telehealth appointment, seeking a COVID-19 vaccination, and responses to a question regarding the helpfulness of the chat.

We used Python with the Pandas library (Python Software Foundation) to transform data from chat user symptom responses and chat user action responses to create one response and activity series for each user in each row. We then coded each response variable (ie, question 1, question 2, and so on). We coded all missing responses as "SYSMIS" and excluded them from the analysis. "Yes" and "No" responses were coded as 0 and 1, respectively. Questions that included "No," "Somewhat," and "Yes" responses were coded as 0, 0.5, and 1, respectively (eg, question 2). Question 6 was coded as "No"=0, "Yes"=1, and "Not sure"=2. Symptoms in question 7 were divided into 4 categories: "No symptom" (coded as 0), "1 symptom" (coded as 1), "2 symptoms" (coded as 2), and "3 symptoms" (coded as 3). Similarly, symptom zone assignment by the chatbot in question 12 was classified as "Green" (0), "Yellow" (1), and "Red" (2).

User activities were coded into 7 categories: "Seeking COVID-19 information" (A1), "Contact Prisma Health" (A2), "Seeking in-person appointment" (A3), "Seeking telehealth appointment" (A4), "Seeking vaccination" (A5), "Seeking travel guidelines" (A6), and "Seeking vaccination information" (A7). Multimedia Appendix 1 illustrates how coded action categories (eg, seeking COVID-19 information) were mapped to the actual hyperlinks located in the chatbot. Any user who recorded one or more activities in a chat session was coded as 1, whereas users who did not engage in a chatbot activity were coded as 0. Each of the activity items was represented in the chatbot as internal Prisma Health hyperlinks or external hyperlinks for users to click to engage in actions outside the chatbot, including seeking additional information, calling a phone number, finding local physicians, or making a web-based appointment.

Data Analysis

We analyzed the data using the SPSS Statistics software (version 27; IBM Corp). First, we analyzed the frequencies and proportions of categorical variables to describe the characteristics of the participants and describe user responses and activities. As all the variables being analyzed were categorical, missing data for chat responses were labeled as missing (SYSMIS) and excluded from the analysis rather than applying data imputation. Second, we conducted chi-square tests to compare the differences in the distribution of risk factor responses and chat helpfulness. Third, we conducted multinomial logistic regression to estimate the odds ratios (ORs) with 95% CIs of various categorical responses for several COVID-19 risk factors and activity variables (independent variables) on the categorical dependent variable (ie, chat helpfulness). Chatbot user responses including "Yes" or "Somewhat" were controlled to examine the robustness of the associations. We used the multinomial logistic regression method for OR analysis as it is helpful in determining associations between nominal (categorical) variables that are not ordinal [64]. In the basic model, we examined the association between chat helpfulness and symptom zone. In model 1, we included the responses to the question on age-related risk. In model 2, we included the responses to the remaining risk questions. While analyzing the impact of user-selected activities on chat helpfulness, we used a basic model to estimate the ORs with 95% CIs of various activity groups that were found to significantly affect the chat helpfulness responses in step 3.

We further explored models to enhance the basic model for chat helpfulness and activities by adding responses as covariates. We stopped when there were insufficient data to establish any further improvement in significance and OR.

Ethics Approval

This study was conducted under the University of South Carolina institutional review board–approved protocol (Pro00101062).

Results

Overview

Between April 24, 2020, and April 21, 2022, a total of 82,222 chat series were started with at least one question or response on record. A total of 53,805 symptom checker questions were accompanied by at least one COVID-19–related activity by users. Among those activities, 5191 chat users clicked further to receive a virtual video visit, and 2215 clicked to make an appointment with a local physician. Chatbot users reported using the tool primarily for the following reasons: checking symptoms, learning about the virus, learning about treatment, learning how to avoid infection, signing up for SMS text message alerts, or accomplishing something else. A total of 9931 users answered the following question: "Did you find this chat helpful?" Similarly structured and phrased questions and response choices about app [65], provider [66-68], and health care program helpfulness [69] have been used in previous research. Chat participants were unique users defined by email addresses, and their responses were analyzed further. Table 1 shows the frequency distributions of user responses to the chatbot questions.

Table 1. Chatbot response and activity frequencies (N=53,805).

COVID-19 risk question or user activity and response classification and coding of categories	Response	Total response series, n (%)
Are you aged ≥65 years? (question 1)		
0	No (0)	43,135 (80.17)
1	Yes (1)	3817 (7.09)
SYSMIS	Missing (system)	6853 (12.74)
Did you find this chat helpful? (question 2)		
0	No (0)	3346 (6.22)
1	Yes (1)	3971 (7.38)
0.5	Somewhat (0.5)	2614 (4.86)
SYSMIS	Missing (system)	43,874 (81.54)
Did your cough or breathing difficulty start or become significantly worse sometime in the past few weeks? (question 3)		
0	No (0)	9113 (16.94)
1	Yes (1)	16,922 (31.45)
SYSMIS	Missing (system)	27,770 (51.61)
Do you have any of the following conditions: diabetes, heart disease, chronic lung disease, or any condition that lowers your body's ability to fight infection (pregnancy or chemotherapy or steroids for cancer or sickle cell disease)? (question 4)		
0	No (0)	37,934 (70.5)
1	Yes (1)	8805 (16.36)
SYSMIS	Missing (system)	7066 (13.13)
Have you been in close contact with anyone who has tested positive for COVID-19 within the past 14 days? (question 6)		
0	No (0)	10,986 (20.42)
2	Not sure (2)	22,061 (41)
1	Yes (1)	13,584 (25.25)
SYSMIS	Missing (system)	7174 (13.3)
Have you had any of the following symptoms in the past week? (select all that apply; question 7)		
No symptom (0)	None of these (0)	18,570 (34.51)
1 symptom (1)	Cough (1)	13,328 (24.77)
1 symptom (1)	Shortness of breath (1)	3357 (6.24)
1 symptom (1)	Fever (>100 °F or >37.8 °C; 1)	4088 (7.6)
2 symptoms (2)	Fever (>100 °F or >37.8 °C); cough (2)	2345 (4.36)
2 symptoms (2)	Fever (>100 °F or >37.8 °C); shortness of breath (2)	427 (0.79)
2 symptoms (2)	Cough; shortness of breath (2)	4833 (8.98)
3 symptoms (3)	Fever (>100 °F or >37.8 °C); cough; shortness of breath (3)	1490 (2.77)
SYSMIS	Missing (system)	5367 (9.97)
Over the past several days, have you experienced any of the following symptoms in a way that is unusual for you: runny nose, congestion, sneezing, headache, sore throat, diarrhea, or loss of taste or smell? (question 11)		
0	No (0)	10,408 (19.34)
1	Yes (1)	36,772 (68.34)
SYSMIS	Missing (system)	6625 (12.31)
Symptom zone (system assigned; question 12)		
Green (0)	Green (0)	2493 (4.63)

COVID-19 risk question or user activity and response classification and coding of categories	Response	Total response series, n (%)
Yellow (1)	Yellow or yellow plus (1)	15,519 (28.84)
Red (2)	Red (2)	13,601 (25.28)
SYSMIS	Missing (system)	22,192 (41.25)
Seeking COVID-19 information (A1[a])		
1	Seeking COVID-19 information	2718 (5.05)
0	Missing (system)	51,087 (94.95)
Contact Prisma Health (A2)		
1	Contact Prisma Health	2 (0)
0	Missing (system)	53,803 (100)
Seeking in-person appointment (A3)		
1	Seeking in-person appointment	2215 (4.12)
0	Missing (system)	51,590 (95.88)
Seeking telehealth appointment (A4)		
1	Seeking telehealth appointment	5191 (9.65)
0	Missing (system)	48,614 (90.35)
Seeking vaccination (A5)		
1	Seeking vaccination	4477 (8.32)
0	Missing (system)	49,328 (91.68)
Seeking travel guidelines (A6)		
1	Seeking travel guidelines	29 (0.05)
0	Missing (system)	53,776 (99.95)
Seeking vaccination information (A7)		
1	Seeking vaccination information	3177 (5.9)
0	Missing (system)	50,628 (94.1)

[a]A1-A7: user-selected activities.

Significant COVID-19 Risk Factors

The Prisma Health Conversa chatbot decision tree uses the selections from a user to determine the symptom zone. User responses to COVID-19 risk factors such as age of >65 years in question 1, cough and breathing difficulty in question 3, comorbidities in question 4, close contact with a person with COVID-19 in question 6, symptom list in question 7, and additional recent general symptoms in question 11 were calculated to determine a symptom zone in question 12 for each series of questions answered. Figure 2 represents the frequency distributions of chat helpfulness responses in relation to these risk factor questions. Table 2 displays the results of the chi-square tests. Each symptom question (1, 3, 4, 6, 7, 11, and 12) was individually assessed for significance in determining chat helpfulness (Table 2). Further analyses using multinomial logistic regression are presented in the following section.

Figure 2. Frequency of chat helpfulness responses as they correlate with significant COVID-19 risk assessment questions. Q1: question 1; Q2: question 2; Q3: question 3; Q4: question 4; Q6: question 6; Q7: question 7; Q11: question 11; Q12: question 12.

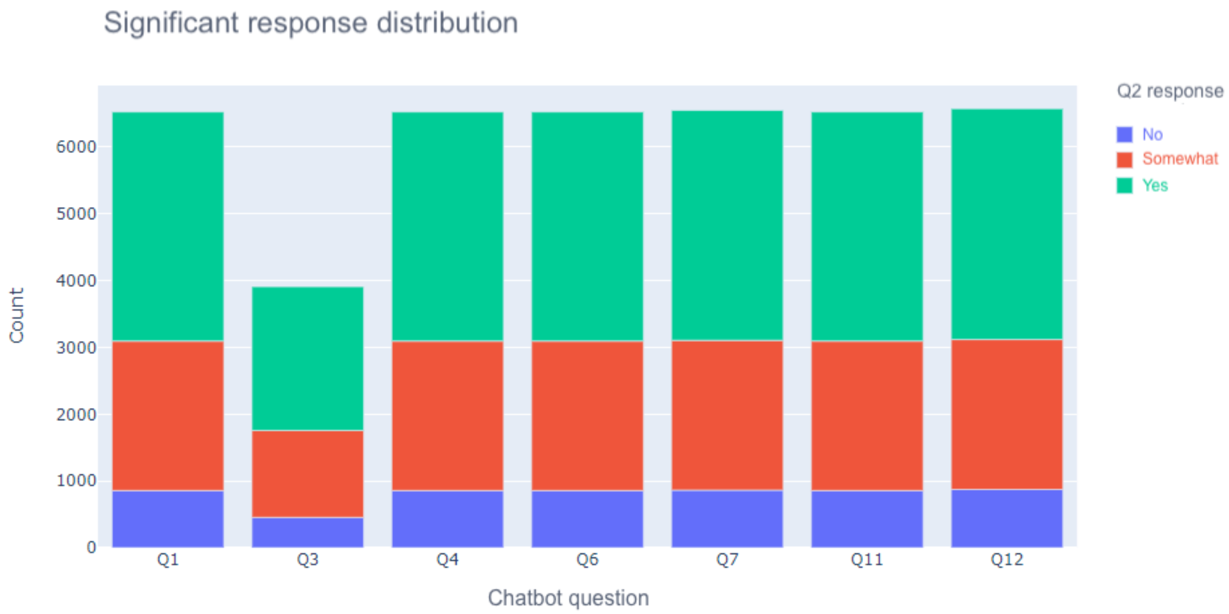


Table 2. Chatbot user risk factor questionnaire responses and COVID-19–related activity characteristics in various chat helpfulness response groups.

COVID-19 risk question or user activity and response category	Responses to “Did you find this chat helpful?” (question 2), n (%)			Pearson chi-square *P* value
	No	Somewhat	Yes	
Are you aged ≥65 years? (question 1)				<.001
No (n=5820)	799 (13.7)	2025 (34.8)	2996 (51.5)	
Yes (n=708)	58 (8.2)	214 (30.2)	436 (61.6)	
Total (n=6528)	857 (13.1)	2239 (34.3)	3432 (52.6)	
Did your cough or breathing difficulty start or become significantly worse sometime in the past few weeks? (question 3)				<.001
No (n=1141)	176 (15.4)	415 (36.4)	550 (48.2)	
Yes (n=2770)	279 (10.1)	887 (32)	1604 (57.9)	
Total (n=3911)	455 (11.6)	1302 (33.3)	2154 (55.1)	
Do you have any of the following conditions: diabetes, heart disease, chronic lung disease, or any condition that lowers your body’s ability to fight infection (pregnancy or chemotherapy or steroids for cancer or sickle cell disease)? (question 4)				<.001
No (n=4994)	697 (14)	1722 (34.5)	2575 (51.6)	
Yes (n=1534)	160 (10.4)	517 (33.7)	857 (55.9)	
Total (n=6528)	857 (13.1)	2239 (34.3)	3432 (52.6)	
Have you been in close contact with anyone who has tested positive for COVID-19 within the past 14 days? (question 6)				<.001
No (n=1519)	183 (12.1)	441 (29)	895 (58.9)	
Yes (n=1789)	317 (17.7)	598 (33.4)	874 (48.9)	
Not sure (n=3222)	357 (11.1)	1200 (37.2)	1665 (51.7)	
Total (n=6530)	857 (13.1)	2239 (34.3)	3434 (52.6)	
Have you had any of the following symptoms in the past week? (select all that apply; question 7)				<.001
No symptom (n=2180)	359 (16.5)	772 (35.4)	1049 (48.1)	
1 symptom (n=2587)	293 (11.3)	906 (35)	1388 (53.7)	
2 symptoms (n=1516)	181 (11.9)	479 (31.6)	856 (56.5)	
3 symptoms (n=269)	30 (11.2)	87 (32.3)	152 (56.5)	
Total (n=6552)	863 (13.2)	2244 (34.3)	3445 (52.6)	
Over the past several days, have you experienced any of the following symptoms in a way that is unusual for you: runny nose, congestion, sneezing, headache, sore throat, diarrhea, or loss of taste or smell? (question 11)				.001
No (n=1301)	184 (14.1)	391 (30)	726 (55.8)	
Yes (n=5226)	673 (12.9)	1848 (35.4)	2705 (51.8)	
Total (n=6527)	857 (13.1)	2239 (34.3)	3431 (52.6)	
Symptom zone (system assigned; question 12)				<.001
Green (n=440)	82 (18.6)	110 (25)	248 (56.4)	
Yellow (n=2781)	454 (16.3)	1034 (37.2)	1293 (46.5)	
Red (n=3356)	337 (10)	1102 (32.8)	1917 (57.1)	
Total (n=6577)	873 (13.3)	2246 (34.1)	3458 (52.6)	
Seeking COVID-19 information (A1[a])				<.001
Yes (n=338)	39 (11.5)	117 (34.6)	182 (53.8)	
Contact Prisma Health (A2)				N/A[b]
Yes (n=0)	0 (0)	0 (0)	0 (0)	
Seeking in-person appointment (A3)				<.001

COVID-19 risk question or user activity and response category	Responses to “Did you find this chat helpful?” (question 2), n (%)			Pearson chi-square *P* value
	No	Somewhat	Yes	
Yes (n=188)	36 (19.1)	68 (36.2)	84 (44.7)	
Seeking telehealth appointment (A4)				<.001
Yes (n=246)	48 (19.5)	91 (37)	107 (43.5)	
Seeking vaccination (A5)				<.001
Yes (n=502)	67 (13.3)	146 (29.1)	289 (57.6)	
Seeking travel guidelines (A6)				.96
Yes (n=3)	1 (33.3)	1 (33.3)	1 (33.3)	
Seeking vaccination information (A7)				.87
Yes (n=656)	225 (34.3)	167 (25.5)	264 (40.2)	

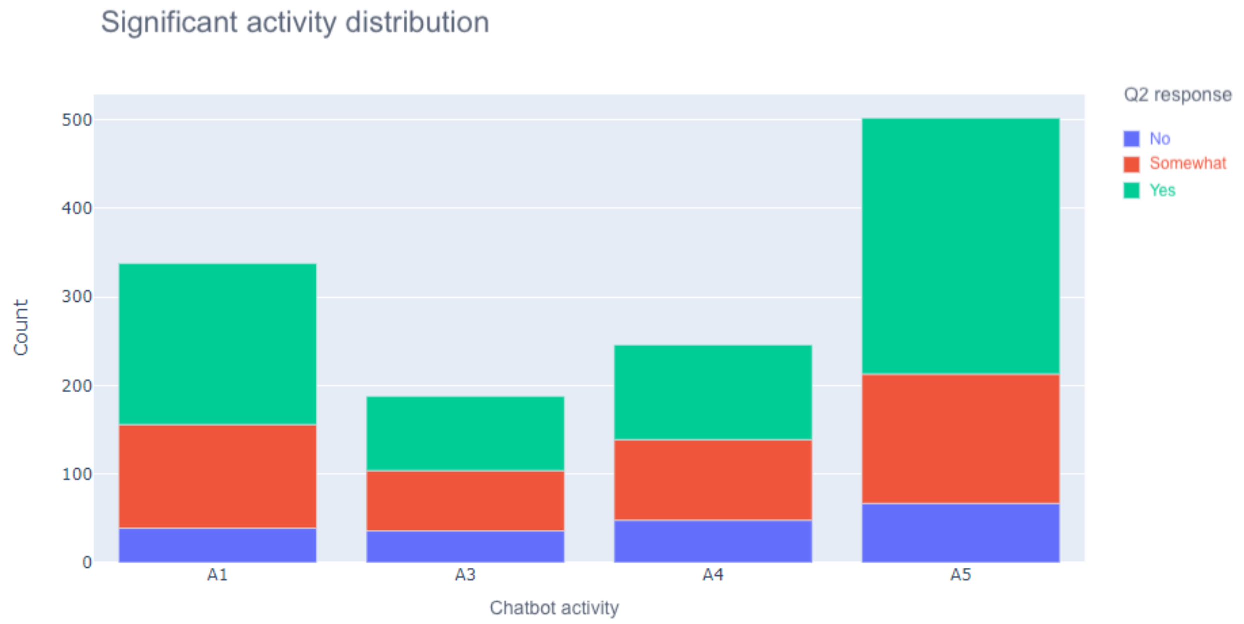
[a]A1-A7: user-selected activities.

[b]N/A: not applicable; no statistics could be computed.

Significant User Activities

We conducted chi-square tests to determine the significance of various user-selected activities in relation to chat helpfulness (Table 2). Activities such as seeking travel guidelines and vaccination-related educational information were not found to be significant factors in determining chat helpfulness. Statistically significant relationships were found (P<.001) between chat helpfulness and several activities, including seeking COVID-19 information (A1), in-person appointments (A3), telehealth appointments (A4), and vaccination (A5). Figure 3 represents the frequency distributions of chat helpfulness responses in relation to chatbot-prompted activities that were found to be significant. Additional multinomial logistic regression analyses are reported in the following section.

Figure 3. Frequency of chat helpfulness responses in relation to significant chatbot-prompted activities. Q2: question 2.



Relationship Between Persons at High Risk of COVID-19 and Positive Chat Helpfulness

To analyze the effect of chat responses and COVID-19 risk factors on chat helpfulness, we applied a multinomial logistic regression (Table 3). To test hypothesis 1, we examined the relationship between chat helpfulness and COVID-19 risk factor symptom zone. We tested 3 models, including a basic model 0, and then subsequent models in which we adjusted for age risk factor (question 1; model 1) and then for all risk factors (questions 1, 3, 4, 6, 7, and 11; model 2).

Table 3. Multinomial logistic regression of chat helpfulness for COVID-19 risk factor variables.

Question (variable) and response category	Odds ratio (95% CI)[a]		
	Basic model[b]	Model 1[c]	Model 2[d]
Chat "somewhat" helpful (dependent category)			
Symptom zone (question 12)			
Green	0.410 (0.301-0.560)[e]	0.471 (0.341-0.651)[e]	0.339 (0.123-0.938)[f]
Yellow	0.696 (0.591-0.821)[e]	0.674 (0.571-0.795)[e]	0.550 (0.345-0.878)[f]
Red	Reference	Reference	Reference
Are you aged ≥65 years? (question 1)			
No	N/A[g]	0.668 (0.493-0.906)[h]	0.877 (0.562-1.368)
Yes	N/A	Reference	Reference
Have you had any of the following symptoms in the past week? (select all that apply; question 3)			
1 symptom	N/A	N/A	1.257 (0.788-2.006)
2 symptoms	N/A	N/A	0.948 (0.599-1.501)
3 symptoms	N/A	N/A	Reference
Did your cough or breathing difficulty start or become significantly worse sometime in the past few weeks? (question 4)			
No	N/A	N/A	1.097 (0.717-1.677)
Yes	N/A	N/A	Reference
Over the past several days, have you experienced any of the following symptoms in a way that is unusual for you: runny nose, congestion, sneezing, headache, sore throat, diarrhea, or loss of taste or smell? (question 6)			
No	N/A	N/A	1.166 (0.770-1.764)
Yes	N/A	N/A	Reference
Do you have any of the following conditions: diabetes, heart disease, chronic lung disease, or any condition that lowers your body's ability to fight infection (pregnancy or chemotherapy or steroids for cancer or sickle cell disease)? (question 7)			
No	N/A	N/A	0.881 (0.676-1.148)
Yes	N/A	N/A	Reference
Have you been in close contact with anyone who has tested positive for COVID-19 within the past 14 days? (question 11)			
No	N/A	N/A	1.135 (0.820-1.571)
Yes	N/A	N/A	0.569 (0.448-0.723)[e]
Not sure	N/A	N/A	Reference
"Yes" chat helpful (dependent category)			
Symptom zone (question 12)			
Green	0.532 (0.404-0.700)[e]	0.606 (0.453-0.810)[e]	0.555 (0.218-1.409)
Yellow	0.501 (0.428-0.586)[e]	0.471 (0.402-0.553)[e]	0.416 (0.266-0.650)[e]
Red	Reference	Reference	Reference
Are you aged ≥65 years? (question 1)			
No	N/A	0.445 (0.333-0.593)[e]	0.680 (0.446-1.037)
Yes	N/A	Reference	Reference
Have you had any of the following symptoms in the past week? (select all that apply; question 3)			
1 symptom	N/A	N/A	1.208 (0.776-1.880)
2 symptoms	N/A	N/A	0.993 (0.643-1.533)
3 symptoms	N/A	N/A	Reference
Did your cough or breathing difficulty start or become significantly worse sometime in the past few weeks? (question 4)			

Question (variable) and response category	Odds ratio (95% CI)[a]		
	Basic model[b]	Model 1[c]	Model 2[d]
No	N/A	N/A	0.929 (0.619-1.394)
Yes	N/A	N/A	Reference
Over the past several days, have you experienced any of the following symptoms in a way that is unusual for you: runny nose, congestion, sneezing, headache, sore throat, diarrhea, or loss of taste or smell? (question 6)			
No	N/A	N/A	1.301 (0.877-1.931)
Yes	N/A	N/A	Reference
Do you have any of the following conditions: diabetes, heart disease, chronic lung disease, or any condition that lowers your body's ability to fight infection (pregnancy or chemotherapy or steroids for cancer or sickle cell disease)? (question 7)			
No	N/A	N/A	0.780 (0.606-1.003)
Yes	N/A	N/A	Reference
Have you been in close contact with anyone who has tested positive for COVID-19 within the past 14 days? (question 11)			
No	N/A	N/A	1.342 (0.984-1.832)
Yes	N/A	N/A	0.591 (0.471-0.742)[e]
Not sure	N/A	N/A	Reference

[a]The odds ratios and 95% CIs for the "Somewhat" and "Yes" responses for chat helpfulness were calculated, with "No" as the reference sample.

[b]In the basic model, we examined the association between symptom zone (question 12) and chat helpfulness (question 2).

[c]In model 1, we also adjusted for COVID-19 age risk factor (question 1) over the basic model.

[d]In model 2, we also adjusted for all COVID-19 risk factors (questions 1, 3, 4, 6, 7, and 11) over the basic model.

[e]*P*<.001.

[f]*P*<.05.

[g]N/A: not applicable; this variable was adjusted in subsequent models.

[h]*P*<.01.

From the 2 years of data on COVID-19 chatbot use, in 18.46% (9931/53,805) of chats, the question regarding chat helpfulness (question 2) was answered. In 26.32% (2614/9931) of these chats, the chat was reported to be "Somewhat" helpful, and in 39.99% (3971/9931), the chat was reported to be helpful ("Yes"; Table 1). Of the 6585 chats in which chat helpfulness was reported, 440 (6.68%) were assigned the "Green" symptom zone, 2781 (42.23%) were assigned the "Yellow" symptom zone, and 3356 (50.96%) were assigned the "Red" symptom zone.

As noted in Table 3, we found that the chat series that were assigned the "Green" symptom zone were 0.410 (95% CI 0.301-0.560) times less likely to find the chat somewhat helpful compared with the reference category of "Red" symptom zone. Similarly, chat series that were assigned the "Yellow" symptom zone were 0.696 (95% CI 0.591-0.821) times less likely to find the chat somewhat helpful compared with the reference category of "Red" symptom zone.

When adjusted for the age risk variable (question 1) in *model 1*, the adjusted OR (aOR) of finding the chat somewhat helpful slightly decreased in the "Yellow" symptom zone (aOR 0.674, 95% CI 0.571-0.795) compared with the "Red" symptom zone (reference category) but slightly increased for the "Green" symptom zone (aOR 0.471, 95% CI 0.341-0.651) compared with the "Red" symptom zone. In *model 2*, when adjusted over model 1 for the rest of the risk assessment variables (questions 3, 4, 6, 7, and 11), the aORs of both the "Green" (aOR 0.339, 95% CI 0.123-0.938) and "Yellow" (aOR 0.550, 95% CI 0.345-0.878) symptom zones decreased significantly compared with that of the "Red" symptom zone.

In summary, the chat series that were assigned a "Red" (higher) symptom zone were approximately 1.8 times more likely to find the chat somewhat helpful compared with the chat series assigned a "Yellow" symptom zone and approximately 2.95 times more likely to find the chat somewhat helpful compared with chat series assigned a "Green" (lower) symptom zone. Thus, we conclude that the chat series that were assigned a higher risk of COVID-19 were more likely to find the chat somewhat helpful compared with those that were assigned a lower COVID-19 risk level. The OR increased when adjusted for age and other risk categories.

Similarly, in the basic model, chat series that were assigned a "Green" symptom zone were 0.532 times (95% CI 0.404-0.700) less likely to answer "Yes" than those assigned a "Red" symptom zone. Those with a "Yellow" symptom zone were 0.501 times (95% CI 0.428-0.586) less likely to answer "Yes" than those assigned a "Red" symptom zone. When adjusted for the age risk factor (question 1) in model 1, the likelihood of the chat series in the "Green" symptom zone answering "yes" slightly increased (aOR 0.606, 95% CI 0.453-0.810) compared with chat series in the "Red" symptom zone but slightly decreased for chat series in the "Yellow" symptom zone (aOR

0.471, 95% CI 0.402-0.553). In model 2, when adjusted for the remaining risk factors (questions 3, 4, 6, 7, and 11), the aOR of the "Yellow" symptom zone (aOR 0.416, 95% CI 0.266-0.650) decreased compared with that of the basic model, whereas for the "Green" symptom zone (aOR 0.555, 95% CI 0.218-1.409), the aOR increased slightly when compared with that of the basic model.

To summarize, after adjusting for all COVID-19 risk factor questions in model 2, the chat series that were assigned the "Red" symptom zone were approximately 1.8 times more likely to respond to the chat helpfulness question (question 2) with a "Yes" compared with the chat series that were assigned the "Green" symptom zone. Similarly, the chat series that were assigned the "Red" symptom zone were approximately 2.4 times more likely to answer the chat helpfulness question (question 2) with a "Yes" compared with the chat series that were assigned the "Yellow" symptom zone.

To conclude, higher COVID-19 risk associations in the chat series resulted in a greater likelihood of positive chat helpfulness from users. The results indicate that people in higher COVID-19 risk categories may find the chat more positively helpful than those in lower COVID-19 risk categories, supporting hypothesis 1.

Relationship Between Chat User Activity and Positive Chat Helpfulness

We further analyzed chat activities in association with chat helpfulness using multinomial regression (Table 4). As the activities were mutually exclusive for each chat response, we did not analyze the interassociations of activities with each other. Thus, we used a basic model to determine the ORs of each relevant activity and chat helpfulness.

Table 4. Multinomial logistic regression for chat helpfulness and activities.

Chat helpfulness question and response category	Odds ratio (95% CI)[a]	
	Chat "somewhat" helpful	"Yes" chat helpful
Seeking COVID-19 information[b] (A1[c])		
No (0)	0.252 (0.175-0.363)[d]	0.246 (0.173-0.348)[d]
Yes (1)	Reference	Reference
Seeking in-person appointment[e] (A3)		
No (0)	0.407 (0.271-0.612)[d]	0.503 (0.340-0.746)[d]
Yes (1)	Reference	Reference
Seeking telehealth appointment[f] (A4)		
No (0)	0.404 (0.283-0.575)[c]	0.526 (0.373-0.741)[c]
Yes (1)	Reference	Reference
Seeking vaccination[g] (A5)		
No (0)	0.345 (0.257-0.463)[c]	0.260 (0.199-0.341)[c]
Yes (1)	Reference	Reference

[a]The odds ratios and 95% CIs for the "Somewhat" and "Yes" responses for chat helpfulness were calculated, with "No" as the reference sample.

[b]In this category, we examined the association between seeking COVID-19 information and chat helpfulness (question 2).

[c]A1-A7: user-selected activities.

[d]*P*<.001.

[e]In this category, we examined the association between seeking an in-person appointment and chat helpfulness (question 2).

[f]In this category, we examined the association between seeking a telehealth appointment and chat helpfulness (question 2).

[g]In this category, we examined the association between seeking vaccination and chat helpfulness (question 2).

In the 9931 chat series in which users answered the chat helpfulness question (question 2), some users carried out a variety of activities after selecting symptoms and receiving a symptom zone assignment. The activities with significant associations with positive chat helpfulness responses included seeking COVID-19 information (A1; 338/9931, 3.4%); seeking in-person appointments with Prisma Health network general practitioners (A3; 188/9931, 1.89%); seeking telehealth services (A4; 246/9931, 2.48%); and, finally, seeking vaccination (A5; 502/9931, 5.05%).

When analyzing these activities for association with chat helpfulness, we compared the ORs of the "Somewhat" and "Yes" chat helpfulness categories with the "No" chat helpfulness response as a reference category. Table 4 presents the ORs of the chat series for these activities, with the type of activity as the reference category. For this analysis, the chat series that sought these activities were in reference categories; hence, the ORs were inverted, and the β values were negative.

The user chat series that did not seek COVID-19 information were 0.252 times (95% CI 0.175-0.363) less likely to find the chat somewhat helpful and 0.246 times (95% CI 0.173-0.348) less likely to respond to the chat helpfulness question (question 2) with a "Yes" (reference category) than those that sought COVID-19 information. The user chat series that did not seek in-person appointments were 0.407 times (95% CI 0.271-0.612) less likely to find the chat somewhat helpful and 0.503 times (95% CI 0.340-0.746) less likely to respond to the chat helpfulness question (question 2) with a "Yes" than those that sought in-person appointments. The user chat series that did not seek telehealth appointments were 0.404 times (95% CI 0.283-0.575) less likely to find the chat somewhat helpful and 0.526 times (95% CI 0.373-0.741) less likely to respond to the chat helpfulness question (question 2) with a "Yes" than those that sought a telehealth appointment. Finally, user chat series that did not seek vaccination were 0.345 times (95% CI 0.257-0.463) less likely to find the chat somewhat helpful and 0.260 times (95% CI 0.199-0.341) less likely to respond to the chat helpfulness question (question 2) with a "Yes" than those that sought vaccination.

In summary, the chat series that sought additional follow-up activities (A1, A3, A4, and A5) found the chat to be more positively helpful than those that did not seek these activities. Users seeking COVID-19 information were approximately 3.97 times more likely to find the chat somewhat helpful and 4.07 times more likely to respond to the chat helpfulness question (question 2) with a "Yes" than those who did not seek COVID-19 information. Users seeking an in-person appointment were approximately 2.46 times more likely to find the chat somewhat helpful and 1.99 times more likely to respond to the chat helpfulness question (question 2) with a "Yes" than those who did not seek an in-person appointment. Users seeking telehealth services were approximately 2.48 times more likely to find the chat "Somewhat" helpful and 1.9 times more likely to respond to the chat helpfulness question (question 2) with a "Yes" than those who did not seek telehealth services. Users seeking vaccination were approximately 2.9 times more likely to find the chat "Somewhat" helpful and 3.85 times more likely to respond to the chat helpfulness question (question 2) with a "Yes" than those who did not seek COVID-19 vaccination.

We can conclude that our second hypothesis is supported regarding the fact that providing an actionable solution to chatbot users positively increases user perceptions of chat helpfulness.

Discussion

Principal Findings

The primary contribution of this study was to identify two main factors for designing chatbots for increased helpfulness: (1) identification of a target user group, in this case, a higher-risk user population for COVID-19; and (2) providing relevant actionable items for the target population. Extending access to care while also limiting patient and provider exposure to COVID-19 was a core motivation for implementing the Prisma Health chatbot. This concept is supported by previous research, with one early study reporting that telehealth availability expands access to care [70]. That same study indicated that telehealth users were younger and healthier than those who visited physicians' offices or the emergency department for similar conditions. Our study results showed that chatbot users who believed themselves to be at risk of COVID-19, those with comorbidities, those aged >65 years, those exposed to other people with the virus, and those reporting COVID-19 symptoms found the chatbot service more helpful than those who may be younger, healthier, or at a lower risk of having the disease.

The results also showed that users who engaged with a relevant actionable item as a result of the chatbot conversation found the chat more helpful. The actionable items selected at the end of the chat conversation may have increased user perceptions of the helpfulness of the chat by assisting the user in taking meaningful action after better understanding their COVID-19 condition.

Helpfulness can be considered a component of a larger grouping of constructs aimed at assessing system effectiveness, utility, or beneficence and can be a key measure of program quality and a direct indicator of satisfaction [71] and overall success. Thus, these results may indicate that the chatbot in this study helps the health care organization achieve its objective of providing helpful technology-enabled engagement for people who are at higher risk of having COVID-19. The results suggest that the chatbot in this study was helpful for the health care organization to identify the higher-risk target audience as well as provide them with an actionable item resulting from the conversation with the chatbot. Additional studies are needed to generalize the findings to the broader context of health care systems and chatbot experiences.

An important purpose of the chatbot from an organizational perspective was to take advantage of accessible, lower-cost, virtual visit technologies. As noted in a recent study, telehealth visit costs are lower than in-person alternatives at retail health clinics, urgent care centers, emergency departments, and primary care physician offices for acute, nonurgent conditions [72]. In contrast, direct-to-consumer telehealth may increase access by making care more convenient for certain users, but it may also increase use and health care spending [73]. In line with this previous study, the chatbot in our study was designed to provide patients and the general public in the geographic region with a low-cost, easy-to-use, nonintrusive entry into the health care system before using a video-based telehealth option. The long-term downstream cost impacts of chatbots should continue to be studied [74].

There are several limitations to this study. We used convenience sampling in our methods by conducting a secondary analysis of real-world observational chatbot data, thus lacking sociogeographic and demographic controls. We do not know if the same results will be true of chatbot users in other locations or with conditions outside this case study and outside the COVID-19 pandemic context. The observational data lacked a deeper understanding of the users, including their socioeconomic status, educational level, ethnicity, and other data points common in purposively controlled studies. We also could not determine whether the users were using the chatbot for themselves or as caregivers. The question used in this study to assess helpfulness contained a positive-leaning bias. The question and response options were kept simple to maintain ease of use for users. However, the assessment of the results should take this into account. Furthermore, the symptoms were purely self-reported in the chatbot, with no way to clinically validate the symptom responses before a provider interaction. Users were people from 2 major regions of 1 state, all of whom sought information from the website of a single health care enterprise. Thus, only people with a computing device and internet access were able to use the system. The chatbot used in this study was offered in both English and Spanish. No other languages were supported. There were too few responses in Spanish to conduct a reliable analysis.

Further research is needed to understand individuals' expectations, needs, perceptions, and experiences relative to the use of AI chatbots, especially by certain demographics such as older age groups. Qualitative studies could be useful in this regard, for example, to understand what factors lead a user to abandon the chat and to further understand the wide range of missing chat response data. Further research is needed to understand the effects that the use of chatbot technologies will have on health care use. Future research should assess chatbot satisfaction across a range of measures beyond helpfulness, such as trustworthiness, by assessing a wider range of user demographics beyond age.

Conclusions

Users at higher risk of COVID-19 found the chatbot technology in this study to be more helpful, indicating that the use of remote patient engagement tools offered by the local health care system may provide value to the local community as an important response to the pandemic. In addition, this study demonstrated that many users intend to engage further with the health care system beyond their chatbot experience to schedule virtual visits, speak with a COVID-19 consultant over the phone, and schedule an appointment with a local provider. These findings may aid health care systems in their current and future chatbot implementations for COVID-19 or for other conditions that greatly affect the operations of health care systems. The helpfulness of interactive technology is an important way to measure the overall effectiveness of a conversational recommender system and strategy. We conclude that users of the system engaged in a manner that seemed to be focused on their personal well-being and that the virtual care strategy used was designed to engage patients and the public and to manage scarce resources as effectively as possible. We also conclude that chatbot design considerations should focus on using just those features that are most beneficial to the target audience and affected providers. Thus, identifying high-impact features, conversational pathways, and recommendations can be very helpful to inform the design of future conversational recommender systems or chatbots.

Acknowledgments

This study was funded by the Office of the Vice President for Research, University of South Carolina, under the COVID-19 Research Initiative (grant 80003688). Furthermore, the authors thank Prisma Health for providing access to their Conversa chatbot data to conduct this study.

Conflicts of Interest

BS, AS, and NP report grants from the University of South Carolina during the conduct of the study. NP reports personal fees from Conversa outside the submitted work. NP was not an advisor before this study and was added as an advisor only after Prisma Health began using the Conversa technology.

Multimedia Appendix 1

Mapping for activity categories and chatbot activity links.

[DOCX File , 15 KB - mhealth_v11i1e43105_app1.docx]

References

1. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May 01;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
2. Krausz M, Westenberg JN, Vigo D, Spence RT, Ramsey D. Emergency response to COVID-19 in Canada: platform development and implementation for eHealth in crisis management. JMIR Public Health Surveill 2020 May 15;6(2):e18995 [FREE Full text] [doi: 10.2196/18995] [Medline: 32401218]
3. Watson AR, Wah R, Thamman R. The value of remote monitoring for the COVID-19 pandemic. Telemed J E Health 2020 Sep 01;26(9):1110-1112. [doi: 10.1089/tmj.2020.0134] [Medline: 32384251]
4. Sensmeier J. Enhancing the patient and provider experience through telehealth. Nurs Manage 2020 Nov;51(11):8-15. [doi: 10.1097/01.NUMA.0000719376.64125.a0] [Medline: 33086365]
5. Wright JH, Caudill R. Remote treatment delivery in response to the COVID-19 pandemic. Psychother Psychosom 2020;89(3):130-132 [FREE Full text] [doi: 10.1159/000507376] [Medline: 32213775]
6. Sacco G, Lléonart S, Simon R, Noublanche F, Annweiler C, TOVID Study Group. Communication technology preferences of hospitalized and institutionalized frail older adults during COVID-19 confinement: cross-sectional survey study. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e21845 [FREE Full text] [doi: 10.2196/21845] [Medline: 32896832]

7. Volpp K, Meer E, Rosin R, Lee V, Washington V, Volpp KG. Asked and answered: building a chatbot to address COVID-19-related concerns. NEJM Catalyst in care delivery. 2020 Jun 18. URL: https://catalyst.nejm.org/doi/pdf/10.1056/CAT.20.0230 [accessed 2021-02-09]
8. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
9. Gadzinski AJ, Gore JL, Ellimoottil C, Odisho AY, Watts KL. Implementing telemedicine in response to the COVID-19 pandemic. J Urol 2020 Jul;204(1):14-16 [FREE Full text] [doi: 10.1097/JU.0000000000001033] [Medline: 32249678]
10. Wittenberg E, Goldsmith JV, Chen C, Prince-Paul M, Johnson RR. Opportunities to improve COVID-19 provider communication resources: a systematic review. Patient Educ Couns 2021 Mar;104(3):438-451 [FREE Full text] [doi: 10.1016/j.pec.2020.12.031] [Medline: 33455825]
11. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying connected in the COVID-19 pandemic: telehealth at the largest safety-net system in the united states. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]
12. Ersek M, Smith D, Griffin H, Carpenter JG, Feder SL, Shreve ST, et al. End-of-life care in the time of COVID-19: communication matters more than ever. J Pain Symptom Manage 2021 Aug;62(2):213-22.e2 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.12.024] [Medline: 33412269]
13. Begun J, Jiang H. Health care management during COVID-19: insights from complexity science. NEJM Catalyst Innovations in Care Delivery. 2020 Oct 09. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0541 [accessed 2023-11-14]
14. Margolius D, Hennekes M, Yaho J, Einstadter D, Gunzler D, Chehade N, et al. On the front (phone) lines: results of a COVID-19 hotline in Northeast Ohio. medRxiv. Preprint posted online May 13, 2020 2020 [FREE Full text] [doi: 10.1101/2020.05.08.20095745]
15. Wijesooriya NR, Mishra V, Brand PL, Rubin BK. COVID-19 and telehealth, education, and research adaptations. Paediatr Respir Rev 2020 Sep;35:38-42 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.009] [Medline: 32653468]
16. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
17. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
18. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
19. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810. [doi: 10.2196/18810] [Medline: 32238336]
20. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
21. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
22. Zubatsky M, Berg‐Weger M, Morley J. Using telehealth groups to combat loneliness in older adults through ‐19. J Am Geriatr Soc 2020 May 22;68(8):1678-1679. [doi: 10.1111/JGS.16553] [Medline: 28498166]
23. Totten AM, Womack DM, Eden KB, McDonagh MS, Griffin JC, Grusing S, et al. Telehealth: mapping the evidence for patient outcomes from systematic reviews. Agency for Healthcare Research and Quality. URL: https://pubmed.ncbi.nlm.nih.gov/27536752/ [accessed 2023-12-05]
24. Moazzami B, Razavi-Khorasani N, Dooghaie Moghadam A, Farokhi E, Rezaei N. COVID-19 and telemedicine: immediate action required for maintaining healthcare providers well-being. J Clin Virol 2020 May;126:104345 [FREE Full text] [doi: 10.1016/j.jcv.2020.104345] [Medline: 32278298]
25. Shah SJ, Schwamm LH, Cohen AB, Simoni MR, Estrada J, Matiello M, et al. Virtual visits partially replaced in-person visits in an ACO-based medical specialty practice. Health Aff (Millwood) 2018 Dec 01;37(12):2045-2051 [FREE Full text] [doi: 10.1377/hlthaff.2018.05105] [Medline: 30633681]
26. Ray KN, Kahn JM. Connected subspecialty care: applying telehealth strategies to specific referral barriers. Acad Pediatr 2020;20(1):16-22 [FREE Full text] [doi: 10.1016/j.acap.2019.08.002] [Medline: 31404707]
27. Jennett PA, Affleck Hall L, Hailey D, Ohinmaa A, Anderson C, Thomas R, et al. The socio-economic impact of telehealth: a systematic review. J Telemed Telecare 2003;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]
28. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]

29. Poon SJ, Schuur JD, Mehrotra A. Trends in visits to acute care venues for treatment of low-acuity conditions in the United States from 2008 to 2015. JAMA Intern Med 2018 Oct 01;178(10):1342-1349 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3205] [Medline: 30193357]
30. Coombes CE, Gregory ME. The current and future use of telemedicine in infectious diseases practice. Curr Infect Dis Rep 2019 Oct 19;21(11):41-27 [FREE Full text] [doi: 10.1007/s11908-019-0697-2] [Medline: 31630276]
31. Sterling R, LeRouge C. On-demand telemedicine as a disruptive health technology: qualitative study exploring emerging business models and strategies among early adopter organizations in the united states. J Med Internet Res 2019 Nov 15;21(11):e14304 [FREE Full text] [doi: 10.2196/14304] [Medline: 31730038]
32. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 May;5(3):2055207619871808-2055207619871809 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
33. Karp S, Kaushik D. Systems for transitioning telephony-based and in-person servicing interactions to and from an artificial intelligence (AI) chat session U.S. Patent 10,630,840,. US Patent and Trademark Office. Washington DC; 2020 Apr 21. URL: https://patents.google.com/patent/US10630840B1/en [accessed 2023-12-04]
34. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of artificial intelligence within the telehealth domain. Yearb Med Inform 2019 Aug 9;28(1):35-40 [FREE Full text] [doi: 10.1055/s-0039-1677897] [Medline: 31022750]
35. Khan R, Das A. Introduction to chatbots. In: Build Better Chatbots. Berkeley, CA: Apress; Dec 15, 2015:1-11.
36. Somvanshi M, Chavan P, Tambade S, Shinde SV. A review of machine learning techniques using decision tree and support vector machine. In: Proceedings of the 2nd International Conference on Computing, Communication, Control, and Automation. 2017 Presented at: ICCUBEA '17; August 12-13, 2016; Pune, India p. 95-100. [doi: 10.1109/ICCUBEA.2016.7860040]
37. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
38. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 04;43(5):135-193. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
39. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
40. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
41. Reid V, Hurst T. Q and A: why providence health wants you to use its COVID-19 triage chatbot. Advisory Board: IT Forefront. 2020 Mar 26. URL: https://digitalcommons.providence.org/publications/2993 [accessed 2023-12-04]
42. Espinoza J, Crown K, Kulkarni O. A guide to chatbots for COVID-19 screening at pediatric health care facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
43. Bharti U, Bajaj D, Batra H, Lalit S, Gangwani A. Medbot: conversational artificial intelligence powered chatbot for delivering telehealth after COVID-19. In: Proceedings of the 5th International Conference on Communication and Electronics Systems. 2020 Presented at: ICCES '20; June 10-12, 2020; Coimbatore, India p. e18898. [doi: 10.1109/ICCES48766.2020.9137944]
44. Contreras CM, Metzger GA, Beane JD, Dedhia PH, Ejaz A, Pawlik TM. Telemedicine: patient-provider clinical engagement during the COVID-19 pandemic and beyond. J Gastrointest Surg 2020 Jul;24(7):1692-1697 [FREE Full text] [doi: 10.1007/s11605-020-04623-5] [Medline: 32385614]
45. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying connected in the COVID-19 pandemic: telehealth at the largest safety-net system in the United States. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]
46. Jannach D, Manzoor A, Cai W, Chen L. A survey on conversational recommender systems. ACM Comput Surv 2021 May 25;54(5):1-36 [FREE Full text] [doi: 10.1145/3453154] [Medline: 28213342]
47. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar 01;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
48. Martinez KA, Rood M, Jhangiani N, Kou L, Rose S, Boissy A, et al. Patterns of use and correlates of patient satisfaction with a large nationwide direct to consumer telemedicine service. J Gen Intern Med 2018 Oct 20;33(10):1768-1773 [FREE Full text] [doi: 10.1007/s11606-018-4621-5] [Medline: 30112737]
49. Schoenfeld AJ, Davies JM, Marafino BJ, Dean M, DeJong C, Bardach NS, et al. Variation in quality of urgent health care provided during commercial virtual visits. JAMA Intern Med 2016 May 01;176(5):635-642 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8248] [Medline: 27042813]
50. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023 [FREE Full text] [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]

51. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
52. Campbell K, Louie P, Levine B, Gililland J. Using patient engagement platforms in the postoperative management of patients. Curr Rev Musculoskelet Med 2020 Aug 01;13(4):479-484 [FREE Full text] [doi: 10.1007/s12178-020-09638-8] [Medline: 32388724]
53. Goldenthal SB, Portney D, Steppe E, Ghani K, Ellimoottil C. Assessing the feasibility of a chatbot after ureteroscopy. Mhealth 2019 Jul 01;5(7):8-5 [FREE Full text] [doi: 10.21037/mhealth.2019.03.01] [Medline: 31019961]
54. Linder C. The effects of a healthcare Chatbots' language and persona on user trust, satisfaction, and chatbot effectiveness. Proquest Dissertations 2020 May 01:1-91.
55. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
56. Bonavita S, Tedeschi G, Atreja A, Lavorgna L. Digital triage for people with multiple sclerosis in the age of COVID-19 pandemic. Neurol Sci 2020 May;41(5):1007-1009 [FREE Full text] [doi: 10.1007/s10072-020-04391-9] [Medline: 32303856]
57. Valtolina S, Barricelli B, Di Gaetano S. Communicability of traditional interfaces VS chatbots in healthcare and smart home domains. Behav Inf Technol 2019 Jun 30;39(1):108-132 [FREE Full text] [doi: 10.1080/0144929X.2019.1637025] [Medline: 31404707]
58. Reis L, Maier C, Mattke J, Weitzel T. Chatbots in healthcare: status QUO, application scenarios for physicians and patients and future directions. In: Proceedings of the 28th European Conference on Information Systems. 2020 Presented at: ECIS '20; June 15-17, 2020; Virtual Event URL: https://aisel.aisnet.org/ecis2020_rp/163 [doi: 10.1258/135763303771005207]
59. Gao C, Lei W, He X, de Rijke M, Chua T. Advances and challenges in conversational recommender systems: a survey. AI Open 2021 Oct 19;2(16):100-126. [doi: 10.1016/J.AIOPEN.2021.06.002]
60. Rafailidis D, Manolopoulos Y. The Technological Gap Between Virtual Assistants and Recommendation Systems. arXiv. Preprint posted online December 21, 2018 2018. [doi: 10.48550/arxiv.1901.00431] [Medline: 30193357]
61. Tran TN, Felfernig A, Trattner C, Holzinger A. Recommender systems in the healthcare domain: state-of-the-art and research issues. J Intell Inf Syst 2020 Dec 17;57(1):171-201. [doi: 10.1007/S10844-020-00633-6] [Medline: 31630276]
62. De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K. Health recommender systems: systematic review. J Med Internet Res 2021 Jun 29;23(6):e18035 [FREE Full text] [doi: 10.2196/18035] [Medline: 34185014]
63. FY19 statistical profile: Prisma health at a glance. Prisma Health Report. Greenville, South Carolina; 2019 Sep 30. URL: https://prismahealth.org/pdfs/fy19-statistical-profile [accessed 2023-12-03]
64. United Health Foundation. Analysis of America's Health Rankings composite measure. America's Health Rankings. URL: https://www.americashealthrankings.org [accessed 2021-02-11]
65. Zhu S, Zhuang Y, Lee P, Li JC, Wong PW. Leisure and problem gaming behaviors among children and adolescents during school closures caused by COVID-19 in Hong Kong: quantitative cross-sectional survey study. JMIR Serious Games 2021 May 07;9(2):e26808-e26840 [FREE Full text] [doi: 10.2196/26808] [Medline: 33960954]
66. CAHPS clinician and group survey. Agency for Healthcare Research and Quality. 2018. URL: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html [accessed 2023-06-17]
67. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
68. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):e87-e64 [FREE Full text] [doi: 10.2196/RESPROT.3852]
69. Puzia M, Laird B, Green J, Huberty J. Parents' perceptions of their children's engagement in a consumer-based meditation mobile app: cross-sectional survey study. JMIR Pediatr Parent 2020 Nov 13;3(2):e24536 [FREE Full text] [doi: 10.2196/24536] [Medline: 33185564]
70. Pagoto S, Tulu B, Agu E, Waring M, Oleski J, Jake-Schoffman D. Using the habit app for weight loss problem solving: development and feasibility study. JMIR Mhealth Uhealth 2018 Jun 20;6(6):e145 [FREE Full text] [doi: 10.2196/MHEALTH.9801] [Medline: 30950796]
71. Uscher-Pines L, Mehrotra A. Analysis of Teladoc use seems to indicate expanded access to care for patients without prior connection to a provider. Health Aff (Millwood) 2014 Feb 07;33(2):258-264 [FREE Full text] [doi: 10.1377/hlthaff.2013.0989] [Medline: 24493769]
72. Sanny L, Susastra AC, Yusramdaleni R. The analysis of customer satisfaction factors which influence chatbot acceptance in Indonesia. Manag Sci Lett 2020;10:1225-1232. [doi: 10.5267/J.MSL.2019.11.036]
73. Gordon AS, Adamson WC, DeVries AR. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. J Med Internet Res 2017 Feb 17;19(2):e35 [FREE Full text] [doi: 10.2196/jmir.6783] [Medline: 28213342]

74. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491 [FREE Full text] [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]

Abbreviations

AI: artificial intelligence
aOR: adjusted odds ratio
CDC: Centers for Disease Control and Prevention
OR: odds ratio

Edited by L Buis; submitted 29.09.22; peer-reviewed by S Thirunavukkarasu, J Luo, K Wac; comments to author 28.04.23; revised version received 19.06.23; accepted 03.11.23; published 14.12.23.

Please cite as:
Singh A, Schooley B, Patel N
Effects of User-Reported Risk Factors and Follow-Up Care Activities on Satisfaction With a COVID-19 Chatbot: Cross-Sectional Study
JMIR Mhealth Uhealth 2023;11:e43105
URL: https://mhealth.jmir.org/2023/1/e43105
doi:10.2196/43105
PMID:38096007

Short-Form Video Exposure and Its Two-Sided Effect on the Physical Activity of Older Community Women in China: Secondary Data Analysis

Chen Wu[1], PhD; Si Chen[1], PhD; Shan Wang[1], MSc; Sijing Peng[1], PhD; Jiepin Cao[2], PhD

1

2

Corresponding Author:
Chen Wu, PhD

Abstract

Background: There is a tendency for older adults to become more physically inactive, especially older women. Physical inactivity has been exacerbated since the COVID-19 pandemic. Lockdowns and information-based preventive measures for COVID-19 increased the number of short-form video app users and short-form video exposure, including content exposure and the duration of exposure, which has demonstrated important effects on youths' health and health-related behaviors. Despite more older adults viewing short-form videos, less is known about the status of their short-form video exposure or the impacts of the exposure on their physical activity.

Objective: This study aims to describe physical activity–related content exposure among older adults and to quantify its impacts along with the duration of short-form video exposure on step counts, low-intensity physical activity (LPA), and moderate-to-vigorous physical activity (MVPA).

Methods: We analyzed a subsample (N=476) of older women who used smartphones and installed short-form video apps, using the baseline data collected from an ongoing cohort study named the Physical Activity and Health in Older Women Study (PAHIOWS) launched from March to June 2021 in Yantai, Shandong Province, China. The information on short-form video exposure was collected by unstructured questions; physical activity–related content exposure was finalized by professionals using the Q-methodology, and the duration of exposure was transformed into hours per day. Step counts, LPA, and MVPA were assessed with ActiGraph wGT3X-BT accelerometers. Multiple subjective and objective covariates were assessed. Linear regression models were used to test the effects of short-form video exposure on step counts, LPA, and MVPA. MVPA was dichotomized into less than 150 minutes per week and 150 minutes or more per week, and the binary logistic regression model was run to test the effects of short-form video exposure on the achievement of spending 150 minutes or more on MVPA.

Results: Of 476 older women (mean age 64.63, SD 2.90 years), 23.7% (113/476) were exposed to physical activity–related short-form videos, and their daily exposure to short-form videos was 1.5 hours. Physical activity–related content exposure increased the minutes spent on MVPA by older women (B=4.14, 95% CI 0.13-8.15); the longer duration of short-form video exposure was associated with a reduced step count (B=−322.58, 95% CI −500.24 to −144.92) and minutes engaged in LPA (B=−6.95, 95% CI −12.19 to −1.71) and MVPA (B=−1.56, 95% CI −2.82 to −0.29). Neither content exposure nor the duration of exposure significantly increased or decreased the odds of older women engaging in MVPA for 150 minutes or more per week.

Conclusions: Short-form video exposure has both positive and negative impacts on the physical activity of older adults. Efforts are needed to develop strategies to leverage the benefits while avoiding the harms of short-form videos.

(JMIR Mhealth Uhealth 2023;11:e45091) doi:10.2196/45091

KEYWORDS

short-form video; media exposure; physical activity; step counts; older adults; apps

Introduction

Globally, populations are aging, and physical activity (PA) is an essential parameter of healthy aging in societies worldwide, as being physically active can reduce the risk of all-cause mortality, cognition decline, functional limitation, and psychological problems of older adults [1,2]. Although the World Health Organization recommends at least 150 minutes of moderate-intensity PA per week or at least 75 minutes of vigorous-intensity PA for older adults, few of them could meet the guidelines. Most older adults engage in low-intensity PA (LPA) in their daily lives [3], and the most common type of PA is step based [4]. In recent years, studies have found that engaging in LPA provides health benefits, such as improved sleep quality [5] and reduced incidence of heart failure rehospitalization [6]. Additionally, taking more steps per day could decrease depression and all-cause mortality in later life regardless of the stepping intensity [7,8]. However, there is a tendency for older adults to become more physically inactive [9], and compared to older men, older women are generally more sedentary and less active [10,11]. This phenomenon has been exacerbated during the COVID-19 pandemic [12]. Short-form video consumption may be the reason underlying this phenomenon.

To keep entertained while staying safe indoors due to lockdowns or quarantines for COVID-19, people increased the use of digital media [13], and the short-form video exploded in popularity. Statistics reported in December 2020 showed that more than two-thirds of consumers in the United States spent between 30 minutes to 3 hours daily watching short-form videos [14]. In China, according to the Chinese Statistical Report on Internet Development by the China Internet Network Information Center, the number of short-form video users has increased by 100 million and reached 873 million from March to December 2020 [15], and individuals typically spent more than 18 hours per week watching short-form videos [16]. Older adults prefer to have in-person interactions; COVID-19 serves as a reminder of the existing "digital divide" between younger and older adults and an opportunity to bring all generations together by helping bridge digital divides [17,18]. In China, a lot of older adults who are not netizens (ie, internet users) gained access to the internet and learned skills to use smartphones with the help of their children, relatives, and community volunteers during the pandemic motivated by the lockdown and adhering to an information-based preventive measure (ie, a color-based QR code on contact-tracing apps to determine an individual's exposure risk to COVID-19) [19,20]. The number of older netizens doubled and reached 110.8 million by December 2020 [15], and 31.3% of new older netizens were short-form video users [16]. Short-form video platforms have higher user stickiness, and short-form video exposure may continue in the daily lives of older adults beyond the types of lockdown measures from COVID-19.

Much research has been done on media exposure in youth cohorts, yet the term media exposure is not explicitly defined in the literature. Most studies in youth cohorts used content exposure or the duration of exposure as proxies of media exposure, which can guide the operationalization of the short-form video exposure concept. Social contagion theory proposes that emotions or behaviors would be rapidly spread from one individual to another, sometimes without rational thought and reason [21]. This notion is empirically supported by the content exposure literature. For example, Ngqangashe and Backer [22] found that adolescents exposed to sweet snack videos were less likely to choose fruits and vegetables, and those exposed to fruit and vegetable videos were less likely to select sweet snacks. As to the duration of exposure, the displacement hypothesis proposed that web-based communication would reduce digital media users' psychological well-being by replacing time spent with strong ties or close relationship partners [23], and this hypothesis was corroborated by a lot of evidence. In addition to psychological impacts, increased media exposure was found to be associated with behavioral problems in youth [24]. With the rise of older netizens in the internet world, the challenges of short-form video exposure on older adults' health also emerge. One recent study found that exposure to food videos and longer duration of exposure to short-form videos had significant effects on the overweight and obesity of older adults [25]. However, to our knowledge, less is known about PA-related short-form video exposure in older adults, and the impacts of PA-related short-form video exposure and the duration of exposure on the PA of older adults remain unclear.

To address these research gaps and advance science in older adults' PA, especially the population less likely to engage in PA (ie, older women [10,11]), this exploratory study was designed to describe PA-related short-form video exposure among older adults and test the impacts of content exposure and the duration of exposure on their step counts, LPA, and moderate-to-vigorous PA (MVPA), respectively.

Methods

Participants

This secondary analysis used baseline data collected from an ongoing cohort study named the Physical Activity and Health in Older Women Study (PAHIOWS) launched from March to June 2021 in Yantai, Shandong Province, China. All investigators in PAHIOWS were trained to collect self-reported and objective data in advance by the principal investigator. Following the gatekeeper's (ie, the director of the community center) approval, the recruitment of participants was launched. Traditional approaches including poster campaigns and person-to-person interactions were used to recruit participants. To entice participation, both monetary (gifts) and nonmonetary incentives were used. The nonmonetary incentive was a printed health report including the information on body composition assessed by Tanita MC-180 (Bailida Co, Tokyo, Japan) and arterial stiffness and vascular obstruction assessed by an automatic oscillometer device (VaSera VS-1500AE, Fukuda-Denshi, Tokyo, Japan). The inclusion criteria set for the parent study were community women ages 60-70 years who could communicate, had no cognitive impairment (the Mini-Mental State Examination score: >17 for illiteracy, >20 for primary school, and >24 for middle school and above), and provided written informed consent. There were 1370 older women enrolled in PAHIOWS; all data were collected in person, and a double-checking strategy was used to eliminate data entry mistakes. In our study, we solely used data from participants using smartphones with short-form video apps Tiktok/Douyin or Kuaishou, Toutiao, and Haokan, as they were apps older adults often used in China.

Ethics Approval

The ethical oversight of the parent study was obtained from the institutional review board in the School of Nursing and Rehabilitation, Shandong University, China (2020-R001).

Measurements

Response Variable: Physical Activity

Physical activity including step counts, LPA, and MVPA was assessed by the ActiGraph wGT3X-BT accelerometers

(ActiGraph, Pensacola, FL), which is a valid tool to assess the intensity of PA [26] and step counts [27]. The location of the accelerometer may influence its validity, and wearing accelerometers on the hip is recommended by the manufacturer for free-living adults. Trained investigators used consistent instructions to guide participants to wear the accelerometers on the hip using a hip-worn belt on the spot and to wear the device all day except when sleeping, swimming, or showering. We collected all devices 7 days later, and within this time frame, two reminders were sent out by phone calls to enhance participants' adherence to wearing the accelerometer and wearing it as instructed by investigators. As such, the random error was also reduced, and therefore, the reliability of this tool was maintained. Raw data were documented 30 times per second and were transformed into counts of movement in an epoch length of 60 seconds by ActiLife software 6.13.4 (ActiGraph). Nonwearing time was defined as an interval of at least 90 consecutive minutes of zero counts with an allowance of up to 2 minutes of 0-100 counts. A valid day is defined as a minimum wearing time of 10 hours, and the data of participants with at least 4 valid days were included for analyses, following recent recommendations [28]. Daily LPA (100-1951 counts per minute) and MVPA (1952 counts and above per minute) were classified based on the Freedson criterion [29], which is widely used in the literature and applicable to older adults, and were exported along with daily step counts.

Predicting Variable: Short-Form Video Exposure

Short-form video exposure was evaluated by two unstructured questions, covering content exposure (ie, what kinds of short-form videos do you often view?) and the duration of exposure per day (ie, how much time do you spend on watching short-form videos per day?). Participants' responses to the former question constituted the Q sample; 14 professionals with expertise in kinesiology and exercise science evaluated the relevance of content with PA following the steps of the Q-methodology [30], and participants with any response rated as most relevant to PA were coded as PA-related content exposure. The duration of exposure was not PA specific, and participants' responses to the duration of exposure were all transformed into hours per day.

Covariates

Sociodemographic characteristics measured included variables of age, education background (elementary school, middle and high school, or college and university), and living alone (yes or no).

The health characteristics assessed included BMI (BMI<18.5 kg/m², 18.5 kg/m² to <28 kg/m², or BMI≥28 kg/m²), history of falls (yes or no), history of falls with injuries (yes or no), history of heart diseases or otolithiasis (yes or no), balance index level, gait speed, musculoskeletal pain, and forced vital capacity.

The balance index level was tested with the Super Balance III Static Balance Test System (Acmeway, Beijing, China). Participants were instructed to complete a couple of static moves and hold each position for 30 seconds on the force platform, and the trajectory and velocity data of the center of gravity movement was documented by the system and automatically transformed into three balance index levels of low, medium, and high.

Gait speed was tested with a 5-meter walk. We measured the time walked over 5 meters at a normal pace from a moving start with a manual stopwatch and calculated the speed. Gait speed was dichotomized into two groups (<1.0 m/s and ≥1.0 m/s), with reference to the Asian Working Group for Sarcopenia consensus in 2019 [31].

We used self-reported questions to ascertain participants' musculoskeletal pain at seven sites: knees (×2), foot (×2), spine (×1), and hips (×2). Participants reporting yes to any pain site were classified into the musculoskeletal pain group, and those without pain at any site were classified as the no musculoskeletal pain group.

Forced vital capacity was tested by a spirometer (CMCS-FHL, Beijing, China) and was measured from the end of inspiration to the end of expiration while participants were in the standing position. Forced vital capacity was dichotomized into <2500 ml and ≥2500 ml based on the clinical criterion.

Statistical Analysis

No outlier was detected by visualizing the data with a box plot, and no strategy was used to handle the missing values in this study as the number of missing values was extremely small (around 1%). Sample characteristics were presented as means (SDs) or medians (IQRs) for continuous variables and presented as numbers (percentages) for categorical variables. The content exposure and the duration of exposure were regressed on step counts, LPA, and MVPA, respectively. Covariates including age, education background, living alone, BMI, history of falls, history of heart diseases or otolithiasis, balance index level, gait speed, musculoskeletal pain, and forced vital capacity were controlled for each regression model. BMI, education background, and the balance index level were transformed into dummy variables before entering each linear regression model. For MVPA, we did an additional analysis by dichotomizing MVPA into below 150 minutes per week and 150 minutes and above per week, and regressed content exposure and the duration of exposure on it while controlling for all covariates. We also did sensitivity analyses by replacing the history of falls with the history of falls with injuries in all models. The unstandardized regression coefficient (B), odds ratio (OR) and its 95% CI were calculated. All analyses were performed with SPSS (version 26; IBM Corp), and *P*<.05 was considered statistically significant.

Results

Sample Characteristics

Table 1 summarizes the characteristics of the sample. Older women in this secondary analysis had a mean age of 64.63 (SD 2.90) years, with the majority receiving education at the middle and high school level (n=351, 73.7%), living with others (n=426, 89.5%), and having abnormal BMI (n=377, 79.2%). Around one-third of participants had a history of falling (n=133, 27.9%) or a history of heart disease or otolithiasis (n=157, 33%), and 24.6% (n=117) of participants reported a history of falling with injuries. Most of the 476 participants were graded as low on the balance index level (n=383, 80.5%) and had forced vital capacity lower than 2500 ml (n=392, 82.4%). Approximately two-thirds of participants had musculoskeletal pain (n=310, 65.1%), and as many as 97.1% (n=462) of participants in this study demonstrated satisfactory gait speed (ie, ≥1 m/s). There were 23.7% (n=113) of participants exposed to PA-related short-form videos, and the median duration of exposure to short-form videos was 1.5 (IQR 1-2.5) hours per day. Participants walked an average of 8186.84 steps per day monitored by ActiGraph wGT3X-BT, and the mean (SD) time spent on LPA and MVPA was 303.15 (76.60) and 32.11 (18.83) minutes per day, respectively. Only a few participants (n=71, 14.9%) in this study spent 150 minutes or more on MVPA weekly.

Table . Sample characteristics of older women (N=476).

Variables			Values
Age (years), mean (SD)			64.63 (2.90)
Education background, n (%)			
	Elementary school		53 (11.1)
	Middle and high school		351 (73.7)
	College and university		72 (15.1)
Living alone, n (%) [a]			
	Yes		48 (10.1)
	No		426 (89.5)
BMI (kg/m²), n (%)			
	<18.5		0 (0.0)
	18.5 to <24		99 (20.8)
	24 to <28		232 (48.7)
	≥28		145 (30.5)
History of falling, n (%)			
	Yes		133 (27.9)
	No		343 (72.1)
History of falling with injuries, n (%)			
	Yes		117 (24.6)
	No		359 (75.4)
History of heart diseases or otolithiasis, n (%)			
	Yes		157 (33.0)
	No		319 (67.0)
Balance index level, n (%)			
	Low		383 (80.5)
	Medium		78 (16.4)
	High		15 (3.2)
Gait speed (m/s), n (%) [b]			
	<1		11 (2.3)
	≥1		462 (97.1)
Musculoskeletal pain, n (%)			
	Yes		310 (65.1)
	No		166 (34.9)
Forced vital capacity (ml), n (%)			
	<2500		392 (82.4)
	≥2500		84 (17.6)
Short-form video exposure			
	Physical activity–related content exposure, n (%)		
		Yes	113 (23.7)
		No	363 (76.3)
	The duration of exposure (hours/day), median (IQR)[c]		1.5 (1-2.5)
Physical activity, mean (SD)			

Variables		Values
	Step counts (steps/day)	8186.84 (2621.88)
	Low-intensity physical activity (minutes/day)	303.15 (76.60)
	Moderate-to-vigorous physical activity (minutes/day)	32.11 (18.83)

[a]Missing values: n=2.
[b]Missing values: n=3.
[c]Missing value: n=1.

Associations of Short-Form Video Exposure With the Step Counts, LPA, and MVPA of Older Women

Table 2 presents the results of linear regression models on step counts, LPA, and MVPA. When holding other variables as constant, a 1-unit increase in the duration of exposure to short-form videos significantly decreased 322.58 walk steps per day for older women. We also found significantly negative impacts of the duration of exposure to short-form videos on the minutes spent on the LPA (B=−6.95, 95% CI −12.19 to −1.71) and MVPA (B=−1.56, 95% CI −2.82 to −0.29) of older women. PA-related content exposure was only found to be significantly associated with increases in the minutes spent on MVPA (B=4.14, 95% CI 0.13-8.15). In the binary logistic regression model, the content exposure failed to significantly increase the odds of spending 150 minutes or more on MVPA (OR 1.313, 95% CI 0.723-2.381), and the duration of exposure failed to significantly reduce the odds (OR 0.949, 95% CI 0.772-1.167). The results remained stable when replacing the variable of history of fall with history of fall with injuries in each model.

Table . Factors associated with step counts, lower-intensity physical activity, and moderate-to-vigorous physical activity of community older women (N=476).

	Step counts, B (95% CI)				Low-intensity physical activity, B (95% CI)				Moderate-to-vigorous physical activity, B (95% CI)			
	Model 1	*P* value	Model 2	*P* value	Model 1	*P* value	Model 2	*P* value	Model 1	*P* value	Model 2	*P* value
Age	–50.27 (–135.61 to 35.07)	.25	–50.17 (–135.53 to 35.18)	.25	–3.70 (–6.20 to –1.19)	.004	–3.69 (–6.20 to –1.20)	.004	–0.34 (–0.94 to 0.27)	.28	–0.33 (–0.94 to 0.27)	.28
Education background (reference: elementary school)												
Middle and high school	–38.62 (–811.08 to 733.83)	.92	–36.61 (–808.85 to 735.74)	.93	–16.39 (–39.15 to 6.38)	.16	–16.39 (–39.16 to 6.37)	.16	3.51 (–1.99 to 9.01)	.21	3.51 (–1.99 to 9.00)	.21
College and university	–450.93 (–1418.28 to 516.43)	.36	–456.87 (–1424.63 to 510.89)	.35	–15.53 (–43.98 to 12.91)	.28	–15.57 (–44.03 to 12.88)	.28	–1.98 (–8.85 to 4.89)	.57	–2.03 (–8.90 to 4.84)	.56
Living alone (reference: no)												
Yes	–210.57 (–993.93 to 572.79)	.60	–213.71 (–997.11 to 569.69)	.60	4.55 (–18.55 to 27.65)	.70	4.53 (–18.57 to 27.63)	.70	–3.08 (–8.66 to 2.50)	.28	–3.10 (–8.68 to 2.47)	.28
BMI (kg/m^2; reference: 18.5 to <24)												
24 to <28	99.55 (519.14 to 718.22)	.75	72.69 (–544.84 to 690.24)	.82	17.24 (–1.01 to 35.48)	.06	17.11 (–1.10 to 35.32)	.07	–2.09 (–6.50 to 2.31)	.35	–2.25 (–6.64 to 2.15)	.32
≥28	–67.94 (–749.45 to 613.58)	.85	–90.54 (–772.51 to 591.44)	.79	27.58 (7.55 to 47.61)	.007	27.47 (7.43 to 47.51)	.007	–3.99 (–8.83 to 0.84)	.11	–4.13 (–8.97 to 0.70)	.09
History of fall (reference: no)												
Yes	–564.57 (–1102.76 to –26.39)	.04	—[a]	—	–2.71 (–18.51 to 13.10)	.74	—	—	–3.39 (–7.21 to 0.43)	.08	—	—
History of fall with injuries (reference: no)												
Yes	—	—	–586.44 (–1146.69 to –26.20)	.04	—	—	–2.96 (–19.39 to 13.48)	.72	—	—	–3.66 (–7.63 to 0.31)	.07
History of heart diseases or otolithiasis (reference: yes)												
No	45.55 (–476.29 to 567.39)	.86	42.27 (–480.03 to 564.57)	.87	0.44 (–14.87 to 15.76)	.96	0.41 (–14.92 to 15.74)	.96	0.12 (–3.58 to 3.82)	.95	0.088 (–3.61 to 3.79)	.96
Balance index level (reference: low)												
Medium	159.20 (–498.41 to 816.81)	.63	184.73 (–472.39 to 841.86)	.58	2.69 (–16.61 to 21.98)	.79	2.80 (–16.47 to 22.08)	.78	–1.77 (–6.43 to 2.89)	.46	–1.62 (–6.27 to 3.03)	.49

	Step counts, B (95% CI)				Low-intensity physical activity, B (95% CI)				Moderate-to-vigorous physical activity, B (95% CI)			
	Model 1	*P* value	Model 2	*P* value	Model 1	*P* value	Model 2	*P* value	Model 1	*P* value	Model 2	*P* value
High	–492.97 (–1884.22 to 898.27)	.49	–474.54 (–1865.52 to 916.44)	.50	–1.93 (–42.96 to 39.09)	.93	–1.85 (–42.87 to 39.17)	.93	–7.87 (–17.78 to 2.03)	.12	–7.77 (–17.67 to 2.13)	.12
Gait speed (m/s) (reference: <1)												
≥1	1252.66 (–307.40 to 2812.71)	.12	1224.28 (–338.31 to 2786.86)	.12	–4.73 (–50.73 to 41.27)	.84	–4.91 (–50.97 to 41.16)	.83	9.76 (–1.35 to 20.86)	.09	9.55 (–1.57 to 20.67)	.09
Musculoskeletal pain (reference: no)												
Yes	–441.71 (–950.59 to 67.16)	.09	–447.47 (–955.72 to 60.78)	.08	–6.70 (–21.67 to 8.27)	.38	–6.71 (–21.66 to 8.24)	.38	–2.80 (–6.41 to 0.82)	.13	–2.82 (–6.43 to 0.79)	.13
Forced vital capacity (ml; reference: <2500)												
≥2500	131.63 (–490.57 to 753.83)	.68	122.54 (–500.04 to 745.12)	.70	12.78 (–5.49 to 31.04)	.17	12.73 (–5.54 to 31.00)	.17	5.29 (0.88 to 9.70)	.02	5.23 (0.82 to 9.64)	.02
Short-form video exposure												
Content exposure (reference: no)												
Yes	370.94 (–194.07 to 935.94)	.20	390.43 (–174.15 to 955.00)	.18	2.49 (–14.10 to 19.09)	.77	2.58 (–13.99 to 19.16)	.76	4.14 (0.13 to 8.15)	.04	4.25 (0.25 to 8.25)	.04
Duration of exposure (h)	–322.58 (–500.24 to –144.92)	<.001	–322.05 (–499.76 to –144.35)	<.001	–6.95 (–12.19 to –1.71)	.01	–6.95 (–12.18 to –1.71)	.01	–1.56 (–2.82 to –0.29)	.02	–1.55 (–2.81 to –0.28)	.02
Adjusted R^2	0.085	—	0.085	—	0.070	—	0.070	—	0.102	—	0.103	—

[a]Not applicable.

Associations of Sample Characteristics With the Step Counts, LPA, and MVPA of Older Women

As depicted in Table 2, we found LPA reduced with advancing age (B=−3.70, 95% CI −6.20 to −1.19). A history of falling significantly reduced daily steps (B=564.57, 95% CI −1102.76 to −26.39) for older women, but failed to significantly reduce the minutes spent on LPA and MVPA. When replacing the history of fall with the history of fall with injuries, we found the history of fall with injuries significantly reduced daily steps (B=−586.44, 95% CI −1146.69 to −26.20). A BMI of 28 and above was significantly associated with increases in the minutes older women spent on LPA (B=27.58, 95% CI 7.55-47.61), and forced vital capacity of 2500 ml and above was found to be significantly associated with increases in the minutes spent on MVPA (B=5.29, 95% CI 0.88-9.70). In the binary logistic regression model, forced vital capacity of 2500 ml and above significantly increased the odds of spending 150 minutes or more on MVPA (OR 2.247, 95% CI 1.224-4.125), while a history of falling significantly reduced the odds (OR 0.423, 95% CI 0.204-0.877). Significant results remained stable when replacing the history of falls with the history of falls with injuries in each model.

Discussion

Principal Findings

Quarantine, limited in-person gatherings, and other lockdown measures as a result of COVID-19 have restricted PA [32,33] and simultaneously increased the number of older adult netizens [15]. Approximately one-third of new older adult netizens were short-form video users, and short-form video apps have demonstrated higher user stickiness, indicating that users would stay longer and frequently revisit [16]. Furthermore, empirical evidence from younger adults demonstrated that social media exposure influences their health behaviors such as food choice and smoking behavior [25,34]. However, less is known about the status of short-form video exposure and its association with PA among older adults. By analyzing 476 older women who were users of short-form video apps, we found that 23.7% (n=113) of them were exposed to PA-related short-form videos, and their daily exposure to short-form videos was 1.5 hours. In addition, we identified that the duration of exposure was significantly associated with reduced step counts and minutes engaged in LPA and MVPA. PA-related content exposure was significantly associated with an increase in minutes spent on MVPA but failed to significantly increase the odds of engaging in MVPA for 150 minutes or more per week among older adults.

Evidence has demonstrated that taking at least 7500 steps daily generates health benefits. For example, Saint-Maurice et al [35] analyzed data collected from 4840 adults (mean age 56.8 years; n=2435, 54% women) and found that taking 8000 steps per day was associated with significantly lower all-cause mortality compared with taking 4000 steps per day. Lee et al [7] analyzed data collected from 16,741 older women (mean age 72 years) and found that mortality rates progressively decreased before leveling at approximately 7500 steps per day. In this study, older women took an average of 8186.84 steps daily (ie, exceeding 7500 steps), and we found that a 1-hour increase in short-form video exposure significantly reduced 322.8 steps per day. This finding indicates that research is needed to explore the appropriate time spent on short-form videos by older women, and such efforts might inform strategies for app developers and community health workers to avoid the hazardous impacts of short-form videos on the health of older women.

We found that women spent an average of 303.15 minutes per day on LPA and 32.11 minutes per day on MVPA, and approximately 15% of older women would achieve the goal of spending 150 minutes or more on MVPA per week, which was slightly higher than the rates reported in studies from high-income countries [36,37]. The longer duration of exposure was significantly associated with a reduction in the minutes spent on LPA and MVPA, and the PA-related content exposure was significantly associated with an increase in the minutes spent on MVPA. Neither the duration of exposure nor PA-related content exposure was found to be significantly associated with older adults' engagement in MVPA for 150 minutes or more per week. Being physically active is beneficial for healthy aging [1,2]; our findings strengthen the necessity of investigating the recommended time spent on short-form video apps. In addition, we found that less than one-quarter of older women viewed short-form videos related to PA. Short-form video apps often recommend content starting from the interests or preferences of the users; our findings indicate that short-form video apps may optimize the recommendation system such as passively pushing PA-related short-form videos, and such efforts may increase the content-related exposure of older women and promote engaging in more MVPA.

Except for media exposure, we found several sample characteristics that were associated with PA in older women. In this study, 27.9% (133/476) of older women reported a history of falls, which was slightly lower than those reported in older adults of both genders (ie, 30%-50%) [38], and women with a history of falls would take 564.57 fewer steps per day than those without, reducing the odds of achieving the goal of spending 150 minutes or more per week on MVPA. We also found that older women with a BMI of 28 kg/m^2 would spend an extra 27.58 minutes per day on LPA compared with those with normal BMI, and this is inconsistent with the evidence in the literature that PA declines with an increase in weight [39]. This inconsistency can be explained by the historical effect of COVID-19 (ie, obesity is a well-known risk factor for COVID-19, severe COVID-19, and its complications, and when sparked by the pandemic, older women with a BMI above 28 kg/m^2 may have become more sensitive to their health and formed new behaviors such as being physically active to protect or enhance their well-being [40,41]). Last, forced vital capacity declined by 20-30 ml per year starting at the age of 30 years [42], and in this study, only 17.6% (84/476) of older women had forced vital capacity of 2500 ml or more. Exercise limitation is a well-known consequence of respiratory conditions [43], our study corroborated this notion and found that compared with women with unsatisfactory forced vital capacity, women with a forced vital capacity ≥2500 ml were more likely to have spent more minutes on MVPA and to achieve the goal of engaging in MVPA for 150 minutes or more per week.

Limitations and Future Work

This study is not free of limitations. First, limited by the secondary data analysis, we failed to assess and control variables that might influence the inference of our study. For example, older adults are accustomed to using traditional media [44], which may also transfer PA-related information and influence the outcome of interest in this study. Additionally, the built environment is particularly relevant to the PA of older adults [45,46], and this variable was not included in PAHIOWS either. Future studies may want to corroborate findings from this study by assessing and controlling these variables. Second, we failed to document the time older women spent specifically watching PA-related short-form videos. The duration of time older women spent watching PA-related short-form videos may provide a better understanding of the impact of content exposure on PA. However, collecting and calculating PA-related short-form video exposure would be challenging; future studies may want to develop strategies to collect these data and that may help give more specific recommendations for the PA of older adult netizens. Third, we found opposite conclusions on relationships between content exposure with PA and the duration of exposure with PA. Moreover, empirical evidence has demonstrated that media exposure would bring mental health benefits to older adults [47]. More studies are needed to explore the balance point to leverage the benefit while avoiding the negative effects of short-form videos. Lastly, PAHIOWS recruited only community older women aged 60-70 years; extrapolating the findings from this study to a sample of male and female older adults or different age cohorts should be done cautiously.

Conclusions

Our study found that short-form videos, which are being increasingly viewed by older adults, introduced both positive and negative impacts on their PA. Exposure to content related to PA would increase the minutes older adults spent on MVPA per day, while the duration of exposure would decrease their steps counts and minutes spent on LPA and MVPA. However, neither of the exposures were able to predict the achievement of spending 150 minutes or more per week on MVPA among older women.

Acknowledgments

We appreciate the contributions of Dr Mary H Palmer (School of Nursing, University of North Carolina at Chapel Hill) and two master's degree students, Li Li and Xinning Peng (School of Nursing and Rehabilitation, Shandong University), to this manuscript.

Data Availability

The data set analyzed during this study is available from the corresponding author upon reasonable request.

Authors' Contributions

CW designed the study, analyzed and interpreted the data, and drafted and revised the manuscript. SC collected the data, designed the study, and revised the manuscript for important intellectual content. SW, SP, and JC interpreted the data and revised the manuscript for important intellectual content. All authors made substantial contributions to the study and approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

References

1. Cunningham C, O'Sullivan R, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: a systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020 May;30(5):816-827. [doi: 10.1111/sms.13616] [Medline: 32020713]
2. Sawa R, Asai T, Doi T, Misu S, Murata S, Ono R. The association between physical activity, including physical activity intensity, and fear of falling differs by fear severity in older adults living in the community. J Gerontol B Psychol Sci Soc Sci 2020 Apr 16;75(5):953-960. [doi: 10.1093/geronb/gby103] [Medline: 30219902]
3. Rees-Punia E, Deubler E, Campbell P, Gapstur SM, Patel A. Light-intensity physical activity in a large prospective cohort of older US adults: a 21-year follow-up of mortality. Gerontology 2020 May;66(3):259-265. [doi: 10.1159/000502860] [Medline: 31600755]
4. Kaleth AS, Slaven JE, Ang DC. Does increasing steps per day predict improvement in physical function and pain interference in adults with fibromyalgia? Arthritis Care Res (Hoboken) 2014 Dec;66(12):1887-1894. [doi: 10.1002/acr.22398] [Medline: 25049001]
5. Seol J, Park I, Kokudo C, Zhang S, Suzuki C, Yajima K, et al. Distinct effects of low-intensity physical activity in the evening on sleep quality in older women: a comparison of exercise and housework. Exp Gerontol 2021 Jan;143:111165. [doi: 10.1016/j.exger.2020.111165] [Medline: 33232794]
6. Miyahara S, Fujimoto N, Dohi K, Sugiura E, Moriwaki K, Omori T, et al. Postdischarge light-intensity physical activity predicts rehospitalization of older Japanese patients with heart failure. J Cardiopulm Rehabil Prev 2018 May;38(3):182-186. [doi: 10.1097/HCR.0000000000000296] [Medline: 29251652]
7. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 1;179(8):1105-1112. [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
8. Hsueh MC, Stubbs B, Lai YJ, Sun CK, Chen LJ, Ku PW. A dose response relationship between accelerometer assessed daily steps and depressive symptoms in older adults: a two-year cohort study. Age Ageing 2021 Feb 26;50(2):519-526. [doi: 10.1093/ageing/afaa162] [Medline: 32980870]
9. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
10. Lee YS. Gender differences in physical activity and walking among older adults. J Women Aging 2005;17(1-2):55-70. [doi: 10.1300/J074v17n01_05] [Medline: 15914419]
11. China Health and Retirement Longitudinal Study. China health and pension report. 2019. URL: charls.pku.edu.cn/info/1010/1009.htm [accessed 2023-07-05]
12. Yamada M, Kimura Y, Ishiyama D, Otobe Y, Suzuki M, Koyama S, et al. Effect of the COVID-19 epidemic on physical activity in community-dwelling older adults in Japan: a cross-sectional online survey. J Nutr Health Aging 2020 Sep;24(9):948-950. [doi: 10.1007/s12603-020-1424-2] [Medline: 33155619]
13. Olvera C, Stebbins GT, Goetz CG, Kompoliti K. Tiktok tics: a pandemic within a pandemic. Mov Disord Clin Pract 2021 Aug 9;8(8):1200-1205. [doi: 10.1002/mdc3.13316] [Medline: 34765687]
14. Advanced Television. Research: pandemic drives short-form video viewing. 2020 Dec 7. URL: advanced-television.com/2020/12/07/research-pandemic-drives-short-form-video-viewing/ [accessed 2022-11-3]

15. China Internet Network Information Center. The 47th statistical report on internet development in China. 2021. URL: www.cnnic.net.cn/n4/2022/0401/c88-1125.html [accessed 2022-10-05]
16. China Internet Network Information Center. The 49th statistical report on internet development in China. 2022. URL: www.cnnic.net.cn/n4/2022/0401/c88-1131.html [accessed 2022-10-07]
17. Seifert A, Cotten SR, Xie B, Carr D. A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021 Feb 17;76(3):e99-e103. [doi: 10.1093/geronb/gbaa098]
18. Yao Y, Zhang H, Liu X, Liu X, Chu T, Zeng Y. Bridging the digital divide between old and young people in China: challenges and opportunities. Lancet Healthy Longev 2021 Mar;2(3):e125-e126. [doi: 10.1016/S2666-7568(21)00032-5] [Medline: 36098110]
19. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
20. Liu S, Yang L, Zhang C, Xiang YT, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18. [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
21. Redl F. The phenomenon of contagion and “shock effect” in group therapy. In: Eissler KR, editor. Searchlights on Delinquency; New Psychoanalytic Studies. Madison, CT: International Universities Press; 1949.
22. Ngqangashe Y, Backer C. The differential effects of viewing short-form online culinary videos of fruits and vegetables versus sweet snacks on adolescents’ appetites. Appetite 2021 Nov 1;166:105436. [doi: 10.1016/j.appet.2021.105436] [Medline: 34119561]
23. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T, Scherlis W. Internet paradox. A social technology that reduces social involvement and psychological well-being? Am Psychol 1998 Sep;53(9):1017-1031. [doi: 10.1037//0003-066x.53.9.1017] [Medline: 9841579]
24. Guerrero MD, Barnes JD, Chaput JP, Tremblay MS. Screen time and problem behaviors in children: exploring the mediating role of sleep duration. Int J Behav Nutr Phys Act 2019 Nov 14;16(1):105. [doi: 10.1186/s12966-019-0862-x] [Medline: 31727084]
25. Chen K, He Q, Pan Y, Kumagai S, Chen S, Zhang X. Short video viewing, and not sedentary time, is associated with overweightness/obesity among Chinese women. Nutrients 2022 Mar 21;14(6):1309. [doi: 10.3390/nu14061309] [Medline: 35334966]
26. Xie HD, Shang Y, Ouyang YY, Luo J. Tri-axial accelerometer placed at different locations of the body to assess the variation of energy expenditure under different exercise conditions. Chin J Tissue Eng Res 2023 Jan 16;27(23):3707-3713. [doi: 10.12307/2023.512]
27. Ngueleu AM, Barthod C, Best KL, Routhier F, Otis M, Batcho CS. Criterion validity of ActiGraph monitoring devices for step counting and distance measurement in adults and older adults: a systematic review. J Neuroeng Rehabil 2022 Oct 17;19(1):112. [doi: 10.1186/s12984-022-01085-5] [Medline: 36253787]
28. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
29. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
30. Churruca K, Ludlow K, Wu W, Gibbons K, Nguyen HM, Ellis LA, et al. A scoping review of Q-methodology in healthcare research. BMC Med Res Methodol 2021 Jun 21;21(1):125. [doi: 10.1186/s12874-021-01309-7] [Medline: 34154566]
31. Chen LK, Woo J, Assantachai P, Auyeung TW, Chou MY, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]
32. Constandt B, Thibaut E, De Bosscher V, Scheerder J, Ricour M, Willem A. Exercising in times of lockdown: an analysis of the impact of COVID-19 on levels and patterns of exercise among adults in Belgium. Int J Environ Res Public Health 2020 Jun 10;17(11):4144. [doi: 10.3390/ijerph17114144] [Medline: 32532013]
33. Füzéki E, Groneberg DA, Banzer W. Physical activity during COVID-19 induced lockdown: recommendations. J Occup Med Toxicol 2020 Aug 12;15:25. [doi: 10.1186/s12995-020-00278-9] [Medline: 32817753]
34. Roby NU, Hasan MT, Hossain S, Christopher E, Ahmed MK, Chowdhury AB, et al. Puff or pass: do social media and social interactions influence smoking behaviour of university students? A cross-sectional mixed methods study from Dhaka, Bangladesh. BMJ Open 2020 Nov 3;10(11):e038372. [doi: 10.1136/bmjopen-2020-038372] [Medline: 33148734]
35. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160. [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
36. O’Brien MW, Kimmerly DS, Mekari S. Greater habitual moderate-to-vigorous physical activity is associated with better executive function and higher prefrontal oxygenation in older adults. Geroscience 2021 Dec;43(6):2707-2718. [doi: 10.1007/s11357-021-00391-5] [Medline: 34081258]

37. Jefferis BJ, Sartini C, Lee IM, Choi M, Amuzu A, Gutierrez C, et al. Adherence to physical activity guidelines in older adults, using objectively measured physical activity in a population-based study. BMC Public Health 2014 Apr 19;14:382. [doi: 10.1186/1471-2458-14-382] [Medline: 24745369]
38. Vaishya R, Vaish A. Falls in older adults are serious. Indian J Orthop 2020 Jan 24;54(1):69-74. [doi: 10.1007/s43465-019-00037-x] [Medline: 32257019]
39. Pietiläinen KH, Kaprio J, Borg P, Plasqui G, Yki-Järvinen H, Kujala UM, et al. Physical inactivity and obesity: a vicious circle. Obesity (Silver Spring) 2008 Feb;16(2):409-414. [doi: 10.1038/oby.2007.72] [Medline: 18239652]
40. Demeulemeester F, de Punder K, van Heijningen M, van Doesburg F. Obesity as a risk factor for severe COVID-19 and complications: a review. Cells 2021 Apr 17;10(4):933. [doi: 10.3390/cells10040933] [Medline: 33920604]
41. Mohammad S, Aziz R, Al Mahri S, Malik SS, Haji E, Khan AH, et al. Obesity and COVID-19: what makes obese host so vulnerable? Immun Ageing 2021 Jan 4;18(1):1. [doi: 10.1186/s12979-020-00212-x] [Medline: 33390183]
42. Bruel Tronchon N, Frappé E, Chomette Ballereau S, Barthélémy JC, Costes F. Prediction of pulmonary restriction from forced vital capacity in elderly is similar using GLI and ERS equations. Lung 2014 Oct;192(5):775-779. [doi: 10.1007/s00408-014-9627-0] [Medline: 25064631]
43. Bédard A, Carsin AE, Fuertes E, Accordini S, Dharmage SC, Garcia-Larsen V, et al. Physical activity and lung function-cause or consequence? PLoS One 2020 Aug 20;15(8):e0237769. [doi: 10.1371/journal.pone.0237769] [Medline: 32817718]
44. Wang H, Sun X, Wang R, Yang Y, Wang Y. The impact of media use on disparities in physical and mental health among the older people: an empirical analysis from China. Front Public Health 2022 Sep 26;10:949062. [doi: 10.3389/fpubh.2022.949062] [Medline: 36225780]
45. Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt M, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet 2016 May 28;387(10034):2207-2217. [doi: 10.1016/S0140-6736(15)01284-2] [Medline: 27045735]
46. Zang P, Xian F, Qiu H, Ma S, Guo H, Wang M, et al. Differences in the correlation between the built environment and walking, moderate, and vigorous physical activity among the elderly in low- and high-income areas. Int J Environ Res Public Health 2022 Feb 8;19(3):1894. [doi: 10.3390/ijerph19031894] [Medline: 35162915]
47. Han M, Tan XY, Lee R, Lee JK, Mahendran R. Impact of social media on health-related outcomes among older adults in Singapore: qualitative study. JMIR Aging 2021 Feb 17;4(1):e23826. [doi: 10.2196/23826] [Medline: 33595437]

Abbreviations

LPA: low-intensity physical activity
MVPA: moderate-to-vigorous physical activity
OR: odds ratio
PA: physical activity
PAHIOWS: Physical Activity and Health in Older Women Study

Edited by A Mavragani; submitted 15.12.22; peer-reviewed by H Wang, S Kim; revised version received 10.07.23; accepted 18.08.23; published 13.09.23.

Please cite as:
Wu C, Chen S, Wang S, Peng S, Cao J
Short-Form Video Exposure and Its Two-Sided Effect on the Physical Activity of Older Community Women in China: Secondary Data Analysis
JMIR Mhealth Uhealth 2023;11:e45091
URL: https://mhealth.jmir.org/2023/1/e45091
doi:10.2196/45091

Original Paper

mHealth Apps Targeting Obesity and Overweight in Young People: App Review and Analysis

Elena Vlahu-Gjorgievska[1], PhD; Andrea Burazor[1], BSc; Khin Than Win[1], PhD; Vladimir Trajkovik[2], PhD

[1]School of Computing and Information Technology, University of Wollongong, Wollongong, Australia

[2]Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Skopje, Republic of North Macedonia

Corresponding Author:
Elena Vlahu-Gjorgievska, PhD
School of Computing and Information Technology
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 42214606
Email: elenavg@uow.edu.au

Abstract

Background: Overweight and obesity have been linked to several serious health problems and medical conditions. With more than a quarter of the young population having weight problems, the impacts of overweight and obesity on this age group are particularly critical. Mobile health (mHealth) apps that support and encourage positive health behaviors have the potential to achieve better health outcomes. These apps represent a unique opportunity for young people (age range 10-24 years), for whom mobile phones are an indispensable part of their everyday living. However, despite the potential of mHealth apps for improved engagement in health interventions, user adherence to these health interventions in the long term is low.

Objective: The aims of this research were to (1) review and analyze mHealth apps targeting obesity and overweight and (2) propose guidelines for the inclusion of user interface design patterns (UIDPs) in the development of mHealth apps for obese young people that maximizes the impact and retention of behavior change techniques (BCTs).

Methods: A search for apps was conducted in Google Play Store using the following search string: [“best weight loss app for obese teens 2020”] OR [“obesity applications for teens”] OR [“popular weight loss applications”]. The most popular apps available in both Google Play and Apple App Store that fulfilled the requirements within the inclusion criteria were selected for further analysis. The designs of 17 mHealth apps were analyzed for the inclusion of BCTs supported by various UIDPs. Based on the results of the analysis, BCT-UI design guidelines were developed. The usability of the guidelines was presented using a prototype app.

Results: The results of our analysis showed that only half of the BCTs are implemented in the reviewed apps, with a subset of those BCTs being supported by UIDPs. Based on these findings, we propose design guidelines that associate the BCTs with UIDPs. The focus of our guidelines is the implementation of BCTs using design patterns that are impactful for the young people demographics. The UIDPs are classified into 6 categories, with each BCT having one or more design patterns appropriate for its implementation. The applicability of the proposed guidelines is presented by mock-ups of the mHealth app “Morphe,” intended for young people (age range 10-24 years). The presented use cases showcase the 5 main functionalities of Morphe: learn, challenge, statistics, social interaction, and settings.

Conclusions: The app analysis results showed that the implementation of BCTs using UIDPs is underutilized. The purposed guidelines will help developers in designing mHealth apps for young people that are easy to use and support behavior change. Future steps involve the development and deployment of the Morphe app and the validation of its usability and effectiveness.

(JMIR Mhealth Uhealth 2023;11:e37716) doi:10.2196/37716

KEYWORDS

behavior change techniques; user interface design patterns; mHealth apps; obesity; lifestyle; mobile app; mobile health; mobile phone

Introduction

Obesity has nearly tripled in the last 30 years, with the World Health Organization estimating that around 340 million or 27% of the world's children and adolescents are overweight or obese [1]. Overweight and obesity have been linked to several serious health problems and medical conditions, including an increase in the risk for noncommunicable diseases such as cardiovascular diseases, diabetes, musculoskeletal disorders, endometrial cancers as well as other types of cancers [1]. Excessive weight and obesity can lead to not only physiological medical complications but also severe psychological effects [2]. The social and emotional well-being and self-esteem of young people are especially impacted during this important developmental phase of life, with these negative consequences tracking well into an individual's later life [3,4]. Further, there is a general reduction in the intake of certain food groups and nutrients and an increase in the consumption of junk food and sugary drinks [5,6], as well as a significant decrease in engagement in moderate-to-vigorous physical exercises during this transition period between adolescence and adulthood [7]. Therefore, targeting young people (age range 10-24 years) is very important [8].

The assumption that nutrition and physical activity behaviors are mediators of body weight provides the basis for behavioral interventions for obesity, which are largely derived from the principles of classical conditioning and social theories [9]. A person's behavior is predominantly responsible for maintaining health and plays an important role in the prevention, management, and treatment of overweight and obesity. Behavior change techniques (BCTs) are descriptors (replicable components of an intervention) designed to enable behavior change by addressing important targets of capability, opportunity, and motivation. The refined taxonomy of BCT—Coventry, Aberdeen, and London-Refined (CALO-RE)—is specifically tailored toward the change of physical activity and promotion of healthy eating behaviors [10].

Mobile phone ownership is ubiquitous, especially among young people. Based on the media use report, 91% of youth between 12 and 15 years of age own a mobile device [11]. The mobile devices are carried by their owners most of the time and are rarely switched off [12]; therefore, they can provide notifications to the users at particular moments, thereby enhancing the engagement and adoption of certain behaviors. These devices can also be used for collecting and analyzing user data, which facilitate the capability to automate certain processes, consequently reducing a user's cognitive load in navigation and selection activities [13]. These characteristics make mobile phones good candidates for delivering digitally supported obesity interventions.

Mobile health (mHealth) apps present a unique opportunity, particularly for young people, to revolutionize the way health behavior change interventions are delivered [14]. However, despite the potential for improved engagement in long-term interventions [15], health interventions delivered by these devices are short-lived. Literature shows that most users cease mHealth app activity within a few uses, and a quarter of mHealth apps are found to be used only 1 time after installation [16]. The factors that impact the adoption of mHealth apps are well-researched, and there is no significant evidence to suggest that adoption alone can improve an individual's health [16]. The continuation of use where technology supports user engagement in behavior change is the area that can enhance positive outcomes [15]. Thus, the continuation of use of mHealth apps greatly impacts their overall efficacy and potential for success.

The user interface and experience of mobile apps strongly influence users' perception and satisfaction and have a strong impact on the app's adoption and continuation of use [17]. The user interface design patterns (UIDPs) are descriptions of the best practices within user interface design. They are general reusable solutions to commonly occurring problems and can ensure that user interfaces flow well and are easy and enjoyable to use. In addition, UIDPs can reduce the cognitive load and improve the overall performance of the app. Furthermore, literature [18] suggests that the overall "look and feel" of apps impacts the adoption by young people, while the perception that health apps were designed primarily for adults was found to be a barrier in using the app. In this context, applying well-known user interface design principles and patterns [19] can improve the efficacy of mHealth apps and contribute toward its continuation of use.

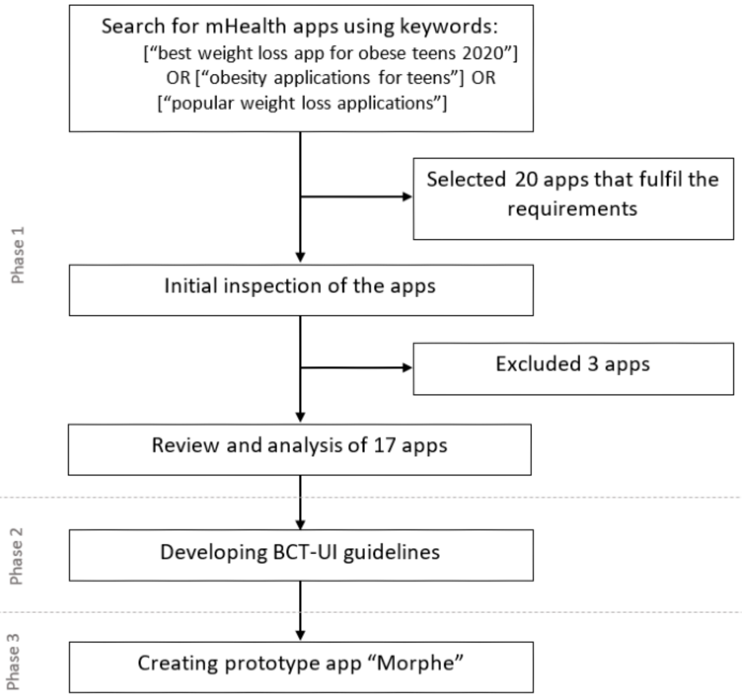
Previous research in this field is primarily focused on mHealth interventions for the adult population without a specific view of the young people demographics [15]. However, evidence suggests that mHealth interventions may be viable in effecting positive health changes in young people as well [20]. The variable results for using mHealth apps by young people could be also explained by the lack of available apps specifically tailored to offer weight management for this group [14]. Additionally, there is a scarcity of research on the impact of UIDPs on the efficacy of BCTs in mHealth apps for obese young people.

The aims of this research were to (1) review and analyze mHealth apps targeting obesity and overweight and (2) propose guidelines for the inclusion of UIDPs in the development of mHealth apps for obese young people, which maximizes the impact and retention of BCTs.

Methods

The overview of the study methodology is presented in Figure 1. A search for apps was conducted in Google Play Store by using a combination of the following keywords: ["best weight loss app for obese teens 2020"] OR ["obesity applications for teens"] OR ["popular weight loss applications"]. The inclusion criteria for the apps were as follows: free; available in both Google Play and Apple App Store; do not require the use of external devices; appropriate for individuals between 10 and 24 years of age; and app's primary purpose (app category, tags, and description) is stated as health, nutrition, physical activity improvement, targeting obesity, or specified as a tool for obesity intervention.

Figure 1. Study methodology diagram. BCT-UI: behavior change technique–user interface; mHealth: mobile health.



The first 20 most popular apps that met the above described criteria were selected for further analysis. During the app inspection, 3 of the selected apps were found to require in-app purchases to access the key functionality and thus were excluded from the study, leaving 17 apps for analysis. Two researchers installed the apps on separate devices and analyzed the features of the apps independently. When there was a discrepancy in the opinions, all the authors discussed them until a consensus was reached.

The selected mHealth apps were analyzed for the inclusion of BCTs from the CALO-RE taxonomy [10] (Multimedia Appendix 1) in their design, taking into account the UIDPs (selected from [19] and [21] Multimedia Appendix 2) used to support those BCTs. UIDPs were noted only in cases where utilized to implement BCTs or some other key functionalities of the app. The generalist UIDPs such as affordance for tap or swipe or key input patterns were not considered in this analysis. However, any glaring problems with an app's interface that had the potential to disrupt the user's experience were noted. Additionally, any features that were not available for free (but only as in-app purchases) were not considered in the analysis.

Based on the results of the review and analysis of the apps, BCT-user interface (BCT-UI) design guidelines classifying UIDPs into 6 categories were developed. Furthermore, using the proposed guidelines, a prototype app called "Morphe" was designed. The purpose of Morphe was to showcase the applicability of the BCT-UI guidelines in the development of mHealth apps targeting young people (age range 10-24 years) with obesity.

Ethical Considerations

As this study does not include experiments on human subjects, no ethical approval was sought.

Results

Review of Apps

Based on the app category, descriptions, and offered features, selected apps were classified into 2 groups: (1) apps focused on physical activity (n=12) and (2) apps focused on nutrition (n=5). Our results (shown in the tables below) and Multimedia Appendix 3 indicate that 20 (out of 40) BCTs listed in the CALO-RE taxonomy were present to some degree in the analyzed apps. The most frequently employed BCTs were related to self-monitoring of behavior (15 apps), followed by prompt practice (9 apps), providing feedback on the performance of the behavior (9 apps), goal setting (behavior) (8 apps), and successful behavior contingent rewards (8 apps). The goal setting of behavior and self-monitoring of behavior were mostly implemented in combination (8 apps) as well as by self-monitoring of behavior with prompt practice (9 apps). The goal setting of behavior and goal setting of behavioral outcome were combined in 5 apps. Six of the BCTs implemented in the apps focused on physical activities were not implemented in the nutrition-focused apps (Table 1).

Table 1. Number of apps implementing each behavior change technique or user interface design pattern.

Characteristics	Apps (n)
Behavior change technique[a]	
#16: Self-monitoring of behavior	15
#19: Provide feedback on performance	9
#26: Prompt practice	9
#5: Goal setting (behavior)	8
#13: Successful behavior contingent rewards	8
#6: Goal setting (outcome)	7
#17: Self-monitoring of behavioral outcome	7
#21: Instruction on how to perform behavior	7[b]
#1: Information provision (general)	6
#28: Facilitate social comparison	6
#22: Demonstrate behavior	6[b]
#9: Setting graded tasks	5[b]
#12: Effort or progress contingent rewards	4
#14: Shaping	4
#29: Plan social support	4
#40: Stimulate anticipation of future rewards	4
#4: Information provision (others' behavior)	3
#3: Information provision (others' approval)	1[b]
#7: Action planning	1[b]
#36: Stress management	1[b]
User interface design pattern	
Content: Cards	14
Charts: Sparklines	13
Content: Filters	11
Gamification-Rewards: Collectibles	10
Notifications: Triggers	10
Charts: Drilldown	9
Content: Search	8
Charts: Dashboard	5
Form: Calculator	5
Social: Profile	5
Content: Article list	5[b]
Gamification-Rewards: Praise	5[b]
Social: Connecting	5[b]
Charts: Threshold	4
Social: Activity Stream	4
Form: Multistep	3
Social: Comments	3
Social: Groups	3

Characteristics	Apps (n)
Gamification-Rewards: Points	3[b]
Gamification-Rewards: Unlock features	3[b]
Gamification: Leaderboard	3[b]
Content: Favorites	2
Form: Registration with Personalization	2
Personalization	2
Gamification: Levels	2[b]
Social: Reactions	2[b]
Charts: Overview plus Data	1[b]
Gamification: Appropriate Challenge	1[b]
Onboarding: Tutorials	1[b]

[a]Behavior change techniques are listed as per the numbering in the Coventry, Aberdeen, and London-Refined taxonomy.

[b]Behavior change techniques and user interface design patterns not implemented in apps focused on nutrition.

Our analysis identified several UIDPs used in the implementation of BCTs (Table 1 and Table 2). Information provision was often implemented using cards (14 apps) and complemented with search and filter functionalities. Pattern filters was implemented in 11 apps, and search was included in 8 apps. Favorites was underutilized (only in 2 apps) besides the many obvious opportunities for its implementation. For successful behavior contingent rewards or effort or progress contingent rewards, most apps utilized collectibles (10 apps). Providing feedback was mostly implemented using charts, with sparklines (13 apps) and drilldown (9 apps) included in apps to provide feedback about the performance of the behavior. None of the apps implemented UIDPs such as scarcity, social proof, Kairos, and interactive preview, while tutorials (1 app) and personalization (2 apps) were underutilized. It needs to be noted that 11 of the UIDPs implemented in the apps focused on physical activities were not implemented in the nutrition-focused apps (Table 1).

Table 2. Overall characteristics of the apps in this review.

Focus, app name	Type/target group	Behavior change techniques	Total	User interface design patterns
Physical activity focus				
NFL Play 60	Type: Exergaming; Educational target group: age range 6-8/9-12 years	#1-Information provision (general); #9-Set graded tasks; #12-Effort or progress contingent rewards; #13-Successful behavior contingent rewards; #14-Shaping	5	Onboarding: Tutorials; Content: Cards; Charts: Sparklines; Gamification–Rewards: Collectibles; Gamification–Rewards: unlock features; Gamification–Rewards: Points; Gamification–Rewards: Praise; Triggers; Gamification: Levels
Runtastic (Adidas Running)	Type: Fitness/physical activity tracking; Educational target group: General	#1-Information provision (general); #3-Information provision (others' approval); #5-Goal setting (behavior); #16-Self-monitoring of behavior; #28-Facilitate social comparison	5	Content: Filters; Content: Cards; Content: Search; Form: Multistep; Form: Calculator; Form: Registration with Personalization; Social: Connecting; Social: Profile; Social: Activity Stream; Social: Groups; Gamification: Leaderboard
7-Minute Workout	Type: Physical activity; Target group: General	#13-Successful behavior contingent rewards; #16-Self-monitoring of behavior; #17-Self-monitoring of behavioral outcome; #19-Provide feedback on performance; #22-Demonstrate behavior; #26-Prompt practice	6	Charts: Overview plus Data; Charts: Threshold; Content: Filters; Content: Cards; Form: Calculator; Gamification-Rewards: Collectibles; Gamification–Rewards: Praise; Triggers
Sworkit	Type: Fitness/physical activity tracking; Educational target group: General, children aged 12 years and younger	#9-Set graded tasks; #16-Self-monitoring of behavior; #21-Instruction on how to perform behavior; #22-Demonstrate behavior; #26-Prompt practice	5	Charts: Sparklines; Content: Cards; Content: Search; Content: Filters; Content: Article list; Triggers
Endomondo	Type: Fitness/physical activity tracking; Educational target group: General	#5-Goal setting (behavior); #12-Effort or progress contingent rewards; #16-Self-monitoring of behavior; #19-Provide feedback on performance; #28-Facilitate social comparison; #29-Plan social support	6	Charts: Sparklines; Charts: Drilldown; Content: Cards; Content: Filters; Gamification–Reward: Praise; Gamification-Reward: Collectibles; Gamification: Leaderboard; Social: Connecting; Social: Profile; Social: Comments; Social: Reactions; Social: Activity stream

Focus, app name	Type/target group	Behavior change techniques	Total	User interface design patterns
Couch to 5K (C25K)	Type: Fitness training Target group: General	#9-Setting graded tasks, #16-Self-monitoring of behavior #21-Instruction on how to perform behavior #26-Prompt practice	4	Trigger
Fitify	Type: Fitness/physical activity tracking Educational target group: General	#13-Successful behavior contingent rewards #14–Shaping #16-Self-monitoring of behavior #19-Provide feedback on performance #21-Instruction on how to perform behavior #22-Demonstrate behavior #26-Prompt practice	7	Content: Cards Content: Article list Content: Filters Content: Search Charts: Sparklines Charts: Drilldown Charts: Dashboard Gamification-Reward: Praise Gamification-Reward: Collectibles Triggers
Fitness Bud-dy	Type: Fitness/physical activity tracking Calorie tracking Target group: General	#5-Goal setting (behavior) #6-Goal setting (outcome) #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #19-Provide feedback on performance #21-Instruction on how to perform behavior #22-Demonstrate behavior	7	Content: Filters Content: Search Content: Cards Content: Article list Charts: Drilldown Charts: Sparklines Form: Calculator Form: Multistep
FitOn	Type: Fitness/physical activity tracking Educational target group: General	#1-Information provision (general) #4-Information provision (others' behavior) #5-Goal setting (behavior) #6-Goal setting (outcome) #7-Action Planning #13-Successful behavior contingent rewards #16-Self-monitoring of behavior #21-Instruction on how to perform behavior #22-Demonstrate behavior #26-Prompt practice #28-Facilitate social comparison #29-Plan social support #40-Stimulate anticipation of future rewards	13	Content: Search Content: Filters Content: Article List Content: Cards Content: Favorites Social: Connecting Social: Profile Gamification-Reward: Collectibles Gamification-Reward: Praise Triggers Personalization
FitBit	Type: Fitness/physical activity tracking Educational target group: General	#1-Information provision (general) #4-Information provision (others' behavior) #5-Goal setting (behavior) #6-Goal setting (outcome) #12-Effort or progress contingent rewards #13-Successful behavior contingent rewards #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #19-Provide feedback on performance #21-Instruction on how to perform behavior #22-Demonstrate behavior #28-Facilitate social comparison #29-Plan social support #36-Stress management #40-Stimulate anticipation of future rewards	15	Content: Cards Content: Filters Charts: Dashboard Charts: Sparklines Charts: Drilldown Charts: Threshold Social: Activity Stream Social: Comments Social: Reactions Social: Groups Social: Connecting Social: Profile Gamification: Leaderboard Gamification-Rewards: Collectibles Triggers

Focus, app name	Type/target group	Behavior change techniques	Total	User interface design patterns
Zombies, Run!	Type: Exergaming Physical activity Target group: General	#9-Setting graded tasks #13-Successful behavior contingent rewards #14–Shaping #19-Provide feedback on performance #21-Instruction on how to perform behavior #40-Simulate anticipation of future rewards	6	Content: Cards Content: Article List Charts: Sparklines Gamification-Rewards: Collectibles Gamification-Rewards: Unlock features Gamification–Rewards: Points Gamification: Appropriate Challenge
Walkr	Type: Exergaming Physical activity Target group: General, kid-friendly	#9-Setting graded tasks #13-Successful behavior contingent rewards #16-Self-monitoring of behavior #26-Prompt practice #28-Facilitate social comparison	5	Gamification-Reward: Collectibles Gamification-Reward: Points Gamification-Reward: Unlock features Gamification: Levels Charts: Sparklines Social: Connecting Triggers
Nutrition focus				
Cron-O-Meter	Type: Calorie/nutrition tracker Target group: General	#5-Goal setting (behavior) #6-Goal setting (outcome) #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #19-Provide feedback on performance	5	Chart: Dashboard Chart: Sparklines Chart: Drilldown Charts: Threshold Content: Search Content: Filters Content: Cards Form: Registration with Personalization Form: Multistep Form: Calculator
Eat the Rainbow Food Journal	Type: Food journal Educational target group: General	#1-Information provision (general) #5-Goal setting (behavior) #12-Effort or progress contingent rewards #14–Shaping #16-Self-monitoring of behavior #40-Stimulate anticipation of future rewards	6	Chart: Drilldown Charts: Sparklines Charts: Dashboard Content: Cards Content: Filters Content: Search Content: Favorites Gamification-Reward: Collectibles Personalization
MyFitnessPal	Type: Calorie/nutrition tracker Physical activity tracker Target group: General	#1-Information provision (general) #4-Information provision (others' behavior) #5-Goal setting (behavior) #6-Goal setting (outcome) #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #19-Provide feedback on performance #26-Prompt practice	8	Content: Cards Forms: Calculator Charts: Sparklines Charts: Drilldown Triggers
Fitatu	Type: Food tracker Weight tracker Target group: General	#6-Goal setting (outcome) #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #26-Prompt practice	5	Charts: Sparklines Charts: Drilldown Content: Search Content: Filters

Focus, app name	Type/target group	Behavior change techniques	Total	User interface design patterns
Lose It! Calorie Counter	Type: Calorie/nutrition tracker Physical activity tracker Target group: General	#6-Goal setting (outcome) #13-Successful behavior contingent rewards #16-Self-monitoring of behavior #17-Self-monitoring of behavioral outcome #19-Provide feedback on performance #26-Prompt practice #28-Facilitate social comparison #29-Plan social support	8	Content: Cards Chart: Sparklines Chart: Threshold Chart: Dashboard Chart: Drilldown Social: Comments Social: Activity Stream Social: Groups Social: Profile Triggers Gamification-Rewards: Collectibles

BCT-UI Design Guidelines

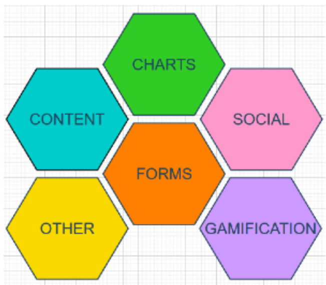
The app analysis results clearly show that implementing BCTs by using UIDPs is underutilized. Only 12 BCTs and 13 UIDPs are implemented in 5 or more apps, of which 3 BCTs and 3 UIDPs are not implemented in the nutrition-focused apps. To overcome this gap and help developers create mHealth apps for obese young people that are easy to use and effective in motivating users to engage in behavior change, we propose BCT-UI design guidelines (presented in Figure 2). These guidelines focus on implementing BCTs using UIDPs that are impactful for the young people demographics. The guidelines use the BCTs from Michie's CALO-RE framework [10] with selected UIDPs [19,21,22].

Each BCT may have one or more UIDPs appropriate for its implementation. UIDPs are classified into 6 categories (Figure 3). The content represents any display of the information, where the information can be textual or graphical. The presentation of the content can include patterns such as article list, cards, option to mark items as favorites, filter and search to refine or locate content of interest, as well as social proof (that can be presented by textual references). The charts represent how the data can be visually presented, including different types of charts such as dashboard, drilldown, interactive preview, overview plus data, sparklines, or threshold. These user interface patterns can help in visualizing users' behavioral and outcome goals, action planning, self-monitoring, or present feedback on the performance. The forms category represents umbrella patterns that focus on structure or feature rather than a specific form implementation. In this context, the forms are components that support input from the user. The gamification elements represent the items that deliver rewards or stimulate challenge and competition. In these UIDPs, the rewards can be implemented as collectibles, points, praise, or unlocking specific features and can be used to introduce different challenge levels or create appropriate challenges based on user preferences. The leaderboards as well can be used to facilitate social comparison. The social elements are patterns (activity stream, groups, comments, reactions) that allow users to connect with their peers for social support and comparison. The connection can be anonymous or delivered through social media integration. The "other" category includes patterns that can support delivering prompts of various kinds (such as Kairos and Triggers), tutorials, and general design considerations such as the incorporation of personalization and customization. These UIDPs can be used as nudges to intervene at specific times when the user will be open to receiving advice or performing the goal behavior.

Figure 2. Proposed design guidelines for the use of each user interface pattern in the context of the behavior change techniques.

	Content: Article List	Content: Cards	Content: Favorites	Content: Filters	Content: Search	Content: Social Proof	Charts: Dashboard	Charts: Drilldown	Charts: Interactive Preview	Charts: Overview plus Data	Charts: Sparklines	Charts: Threshold	Form: Calculator	Form: Multistep	Form: Registration with Personalization	Gamification - Rewards: Collectibles	Gamification - Rewards: Points	Gamification - Rewards: Praise	Gamification - Rewards: Unlock features	Gamification: Appropriate Challenge	Gamification: Leaderboard	Gamification: Levels	Social: Activity Stream	Social: Comments	Social: Connecting	Social: Groups	Social: Reactions	Social: Profile	Notifications: Kairos	Notifications: Triggers	Onboarding: Tutorials	Personalization	Customization	Scarcity
#1-Information provision (general)	√	√	√	√	√																									√		√		
#2-Information provision (to the individual)	√	√	√	√	√	√																								√		√		
#3-Information provision (others' approval)																								√			√							
#4-Information provision (others' behavior)	√	√	√	√	√	√																	√			√								
#5-Goal setting (behavior)														√	√					√											√			
#6-Goal setting (outcome)									√				√	√	√					√											√			
#7-Action planning									√				√	√	√														√	√	√			
#8-Identifying barriers/problem resolution		√		√	√									√	√														√	√	√			
#9-Setting graded tasks											√									√		√										√		√
#10-Review of behavioral goals				√			√			√		√		√						√									√	√				
#11-Review of outcome goals				√			√		√	√		√	√	√																				
#12-Effort or progress contingent rewards																√	√	√	√													√		√
#13-Successful behavior contingent rewards																√	√	√	√													√		√
#14-Shaping																√	√	√	√	√		√										√		√
#15-Generalization of target behavior	√	√	√	√	√																								√	√				
#16-Self-monitoring of behavior									√				√	√																		√		
#17-Self-monitoring of behavioral outcome									√				√	√																		√		
#18-Focus on past success	√	√		√			√			√	√	√		√															√	√				
#19-Provide feedback on performance		√		√			√	√		√	√	√																						
#20-Informing when and where to perform behavior	√	√	√	√	√																								√	√		√		
#21-Instruction on how to perform behavior	√	√	√	√	√																													
#22-Demonstrate behavior																															√			
#23-Training to use prompts															√																√			
#24-Environmental restructuring	√	√																		√									√	√				
#25-Agreement of behavioral contract		√												√	√																√			
#26-Prompt practice																													√	√		√		
#27-Use of follow-up prompts																													√	√		√		
#28-Facilitate social comparison																					√		√		√	√		√						
#29-Plan social support																								√	√	√	√							
#30-Prompt identification as a role model																					√				√	√		√	√	√				
#31-Prompt anticipated regret																														√				
#32-Fear arousal	√	√	√	√	√																									√		√		
#33-Prompt self talk																				√									√	√		√		
#34-Prompt use of imagery																				√									√	√		√		
#35-Relapse prevention	√	√		√	√				√				√	√															√	√				
#36-Stress management	√	√	√	√	√															√						√			√	√		√		
#40-Stimulate anticipation of future rewards																													√	√	√			√

Figure 3. User interface design pattern categories.



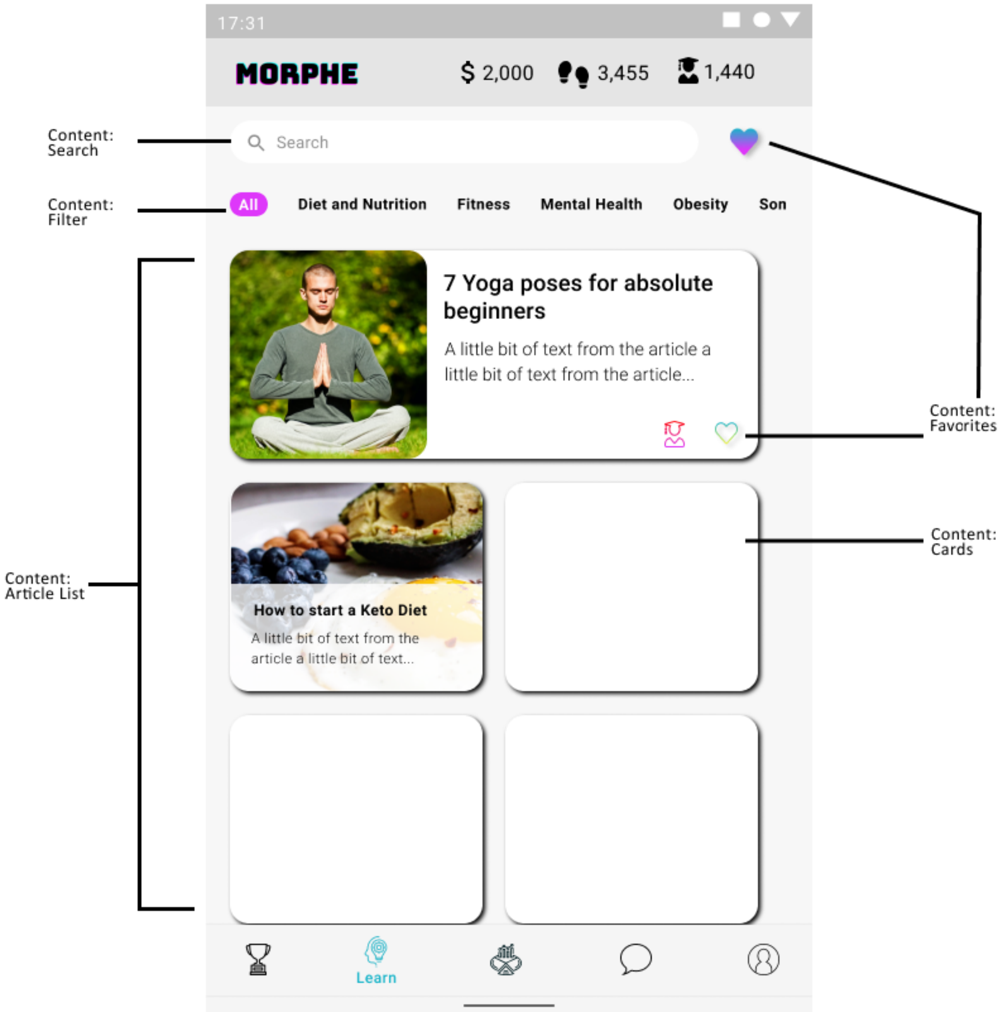
Use Cases: Using UIDPs to Implement BCTs

The proposed design guidelines offer the utilization of different UIDPs in the implementation of various BCTs. Besides the use cases presented in this section, miscellaneous examples of each pattern in the context of their BCTs are presented in Multimedia Appendix 4. In the given use cases, we have used mock-ups of the Morphe app developed by the authors to showcase the applicability of the proposed BCT-UI guidelines. This app is intended for young people (age 10-24 years) and has 5 main functionalities: learn, challenge, statistics, social interaction, and (user's) settings. Additionally, the Morphe app uses notifications to nudge the user toward the desired behavior.

Use Case 1: Content Presentation

The presentation of the content in the app is very important for the usability of the app. Content UIDP dictates how app users can access, refine, and interact with app information and experiences. In our Morphe app (Figure 4), the educational content in the "Learn" functionality is presented using the article list pattern. In general, article lists tend to contain multiple rectangular cards used to store and deliver content. Each card contains an image, a title, and a brief description to allow the user to understand what information is contained in the article. Using the favorites pattern, the user can mark the items of interest and have easier access to those articles. Additionally, the combination of using filters and search allow the user to refine the content that is displayed on the screen by category (nutrition, physical activity, obesity) or by keywords.

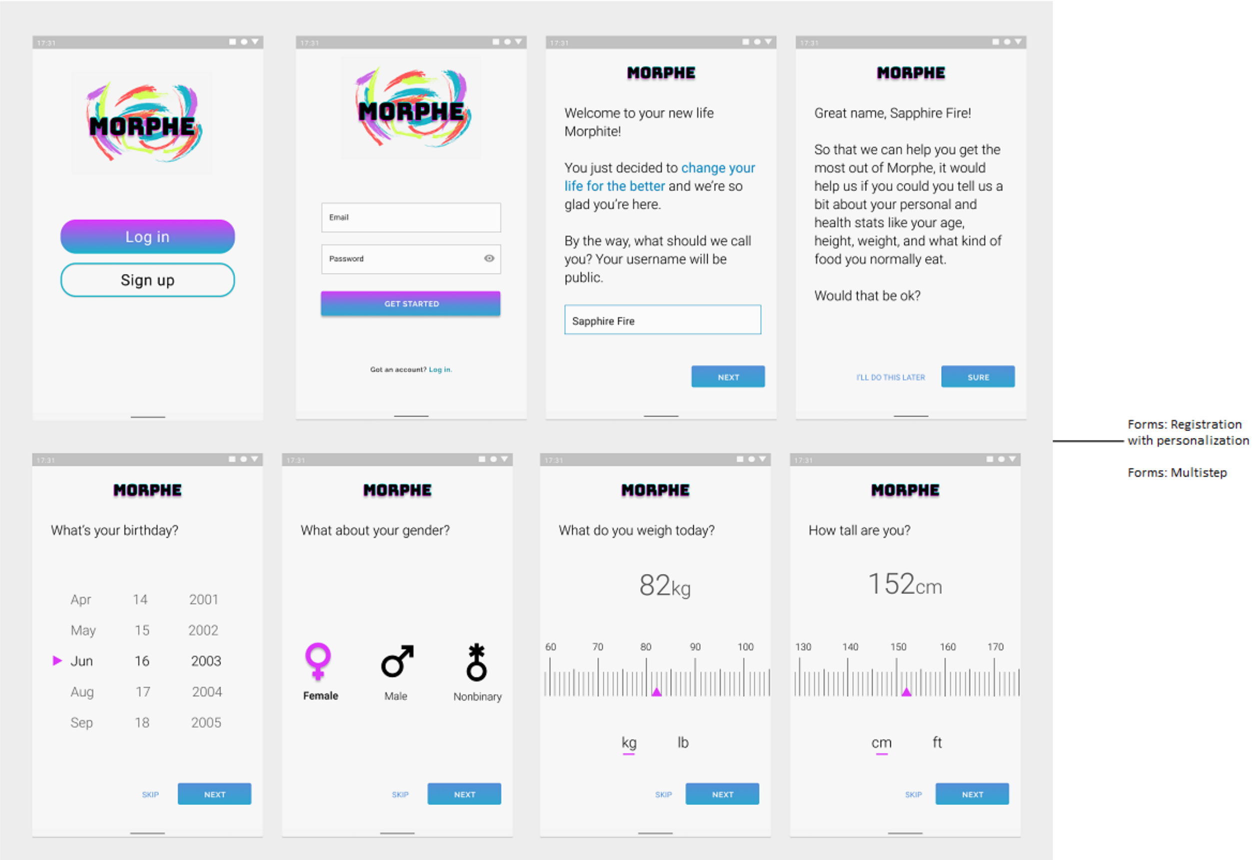
Figure 4. Morphe app educational content.



Use Case 2: User Data

Forms are used to collect data from the user. These data are primarily used to personalize users' experiences as assisting with goal setting and providing a baseline for behavioral monitoring. The Morphe app uses a multistep form to gather data about the user in the registration process (Figure 5). The multistep patterns allow each option to be accessible by some other means as well. For example, a BMI calculator that displays a weight range overview dependent on weight and height input can be used to help a user define a goal weight or to see their progress toward a goal weight each time they log in their weight.

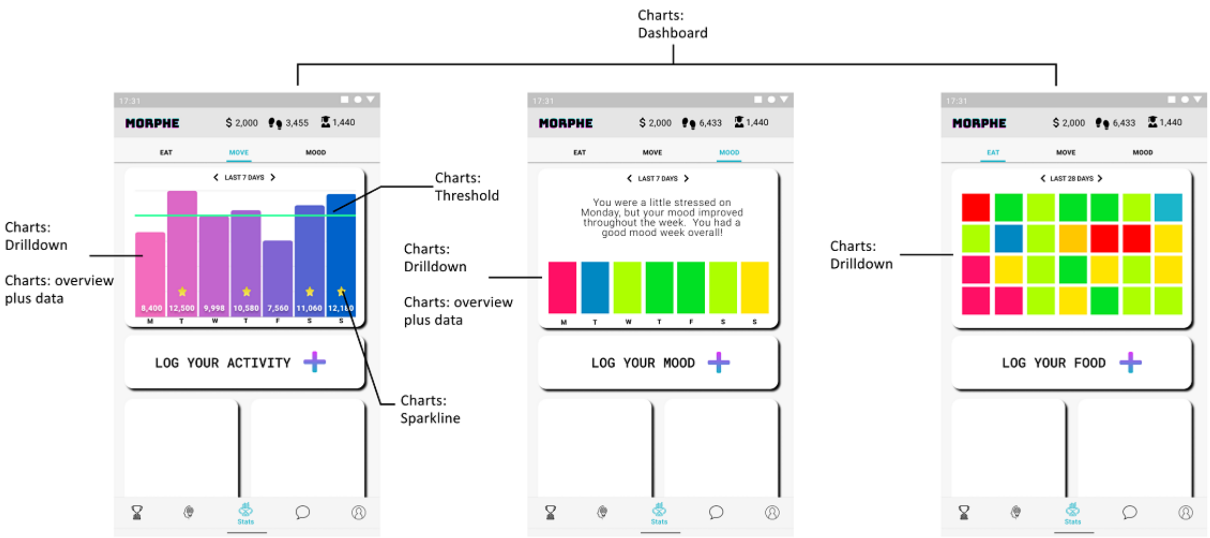
Figure 5. Morphe app registration process.



Use Case 3: Use of Charts

Charts give the user a way of interacting with their data, including goal and performance history and progress. In the context of behavior change, charts can be used to present a review of the behavioral and outcome goals, represent data from periods in which the user achieved their goals, as well as provide feedback on the user's performance. In the given example (Figure 6), the statistics functionality provides an overview of the user's effort and progress concerning the set goals. The stats page includes submenus for diet and nutrition (eat submenu item), exercise (move submenu item), and mental health and well-being (mood submenu item). In order to visualize the progress and achievements, a few types of charts are used: dashboard, drilldown, threshold, sparklines, and overview plus data.

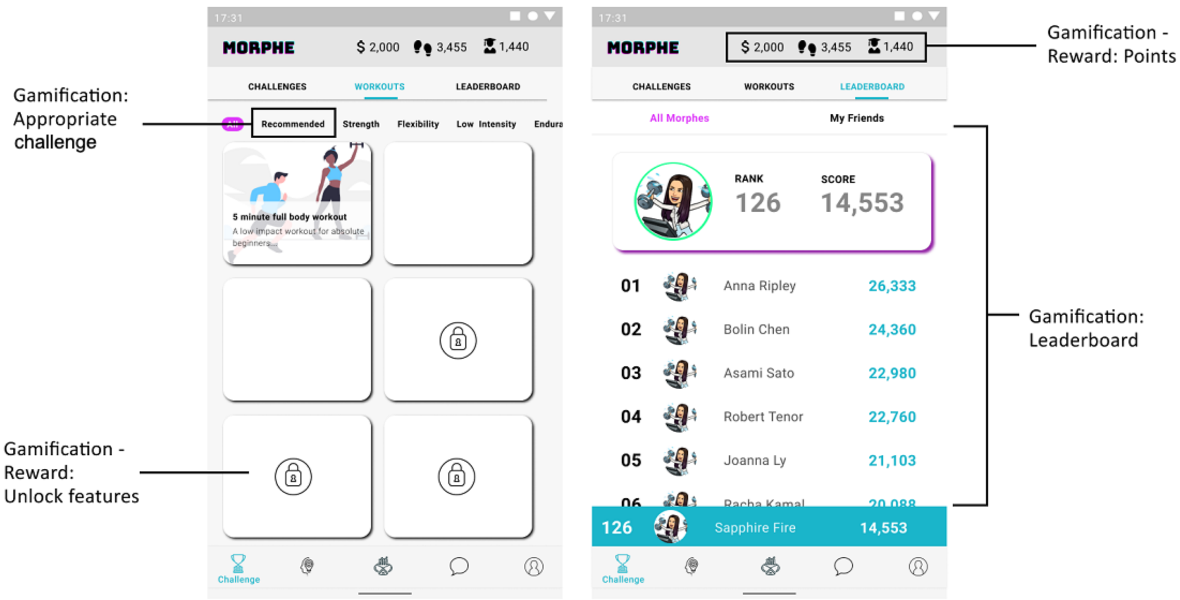
Figure 6. Morphe app statistics.



Use Case 4: Gamification

To engage users in behavior change and achieve their goals, several gamification patterns can be used. The challenge section in the Morphe app (Figure 7) provides access to proposed workouts, challenges, and users' leaderboards. The workouts submenu allows the user to access workouts and exercises across several categories. Additionally, the user can choose the challenge that is most suitable for achieving a certain goal. As the user progresses throughout the workouts, different features are unlocked as a reward for the effort. Another type of reward is points. The user can be awarded a number of points by winning a physical activity challenge, performing, or making progress toward a goal behavior. In the case of the Morphe app, the points can be used to modify a user's avatar or "purchase" entry to other challenges. The leaderboard submenu allows the users to view their achievements relative to other users and their friends. This pattern enables the user to compare their achievements with other users, thus identifying their role model and working toward reaching better results.

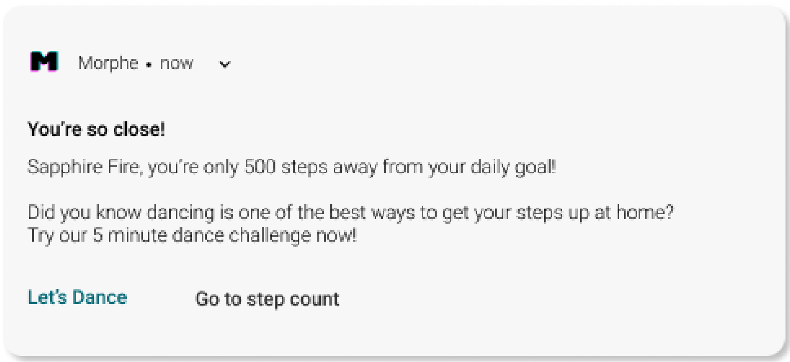
Figure 7. Morphe app challenge.



Use Case 5: Notifications

There are 2 types of notifications: Kairos and Triggers. Kairos are app nudges that utilize personalization and customization patterns to intervene at specific times when the user will be open to receiving advice or performing the goal behavior. For example, the Morphe app uses Kairos (Figure 8) to let the user know when he/she is close to achieving a particular goal and as such, is more likely to feel motivated to give effort toward it.

Figure 8. Morphe app notifications.



Discussion

Behavior change requires highly motivated users; therefore, mHealth apps need to provide features and functionalities that support users' intrinsic and extrinsic motivation. The apps can increase users' intrinsic motivation if activities that are interesting, challenging, and that have aesthetic appeal are introduced, while extrinsic motivation can be underpinned if options perceived as valuable, meaningful, and important by the users are presented [23]. Additionally, the expectation for mHealth interventions is to be able to achieve effectiveness in line with the traditional delivery of behavioral interventions. Therefore, the apps developed to support physical activities and healthy eating habits should be engaging enough to motivate the continuation of use.

The proposed design guidelines aim to provide new design considerations by incorporating and supporting BCTs through the use of the recommended UIDPs in the development of mHealth apps. The benefits of UIDPs are to make task completion quicker and easier by reducing cognitive load, thus helping users to achieve behavior change and higher engagement with the apps. The self-regulatory BCTs, including self-monitoring of behavior, goal setting of behavior, and providing feedback on the performance, can be used as feedback processes that are very important in self-management and behavior control [24]. These BCTs have been consistently coupled with positive changes in physical activity [25], and interventions have been found to be more effective when individuals utilize these techniques [26-28]. Moreover, using different patterns from the content, charts, forms, personalization, or gamification categories can increase the successful intervention engagement; therefore, people will engage with mHealth apps in the long term.

Research shows that reward-seeking behavior is more prominent among young people because this age group receives less significant positive responses from rewards, driving them to pursue reinforcers that increase dopamine-related circuitry [29]. As Bryan et al [30] indicate, rewards are experienced in the context of other available rewards, and young people may be particularly sensitive to these changing contexts. Additionally, Davidow et al [31] note that "adolescents are notorious for engaging in reward-seeking behaviors," and much research in the behavioral health field suggests that the most successful rewards for motivating young people are tied to achieving goals that are immediate, simple, and socioculturally reinforced. This represents an important opportunity to support the users' extrinsic motivation by diversifying the rewards in the apps by using gamification patterns such as the use of a points system, introducing levels, or offering opportunities to unlock new features. These patterns can provide experiences of autonomy, competence, and relatedness by adding fun and excitement to the activities [32]. Although the implementation of the "Training to use prompts" behavior technique and utilizing the "Onboarding: tutorials" design patterns represent a clear opportunity in future app development, training and navigation menus are important in the early stages of app interaction and adoption. However, the provided instructions need to keep the text to a minimum and written to a sixth-grade level [33]; otherwise, it can be difficult to read, presenting a further barrier to engagement.

The literature regarding the social support for BCTs presents opposite findings. Although some research suggests that young people find peer interactions more rewarding [30], others find that young people find social posting not desirable since they do not want to bother their friends or share achievements that they considered to be uninteresting [34]. Another concern is sharing personal or sensitive data with others [35]. In this context, during the development phase of the app, designs that implement social patterns with relative anonymity that will support the young people's engagement need to be considered. Stress management BCT has been identified as a necessary component for health behavior change, especially for young people with obesity who might have greater levels of stress [36]. UIDPs included in the content group can be useful in the implementation of this behavior technique by providing generic information about mindfulness, stress, and anxiety management.

BCT categories such as "prompt self-talk" or "prompt identification as a role model" can increase user engagement by focusing on intrinsic and social motivation elements that are recorded as having bolstered engagement [37]. However, effective implementation of prompts needs consideration in timing, frequency, and tailoring [38] that can be designed using Kairos, Triggers, and Personalization in combination with Gamification or Social Patterns. Further, personalized prompts for a given situation are proven to be more effective in behavior change [39]. Problem-solving solutions for barriers to physical activities can motivate individuals to increase their activity [40]. Therefore, features that could assist users with solving barriers to physical activity, such as well-timed notifications for inclement weather or recommendations for suitable in-door exercises can be beneficial. The proposed design guidelines link the BCTs with UIDPs, which can maximize the impact and increase the adoption and continuation of use of mHealth apps for obese young people. However, the design process of such apps is complex and requires the involvement of relevant stakeholders: public health and clinical experts for content creation, app developers for designing the apps' features and functionality, and young people as prospective users of the app.

A few limitations in this research need to be noted. The time frame of the app analysis should be considered since there is a possibility that some of the features could be revealed to the user after using the app over a longer period of time. The applicability of the proposed BCT-UI design guidelines was presented by designing the prototype app Morphe. However, the prototype was created without insight from the young people community. Furthermore, the guidelines need to be validated with the real users of an mHealth app targeting young people with obesity.

In conclusion, the analysis of 17 mHealth apps has shown that the utilization of UIDPs in implementing BCTs is limited. Taking into account the importance of BCT and UIDP in improving the efficacy of the mHealth apps, in this paper, we proposed BCT-UI design guidelines. The aim of these guidelines is to support the development of mHealth apps that are easy to use and effective for long-term adoption by young people. Additionally, 5 use cases of the Morphe app targeting overweight and obese young people were presented to showcase the usability of the design guidelines. Future research should involve the development and deployment of the Morphe app and validation of its usability and effectiveness in obesity and overweight management within the young people community. However, since the proposed guidelines are generalized, exploring its utilization in the design of mHealth apps for the management of other health conditions as well as various age groups can be valuable.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding constraints of the Coventry, Aberdeen, and London-Refined (CALO-RE) taxonomy of behavior change techniques [10].
[PDF File (Adobe PDF File), 777 KB - mhealth_v11i1e37716_app1.pdf]

Multimedia Appendix 2
User interface design pattern definitions [19,21].
[PDF File (Adobe PDF File), 535 KB - mhealth_v11i1e37716_app2.pdf]

Multimedia Appendix 3
Behavior change techniques implemented in the reviewed apps.
[PDF File (Adobe PDF File), 229 KB - mhealth_v11i1e37716_app3.pdf]

Multimedia Appendix 4
Proposed usage of the user interface design patterns in the context of the behavior change techniques.
[PDF File (Adobe PDF File), 591 KB - mhealth_v11i1e37716_app4.pdf]

References

1. Obesity and overweight. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2022-11-10]
2. Finer N. Medical consequences of obesity. Medicine 2015 Feb;43(2):88-93. [doi: 10.1016/j.mpmed.2014.11.003]
3. Sahoo K, Sahoo B, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192 [FREE Full text] [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
4. Griffiths LJ, Parsons TJ, Hill AJ. Self-esteem and quality of life in obese children and adolescents: a systematic review. Int J Pediatr Obes 2010 Aug;5(4):282-304. [doi: 10.3109/17477160903473697] [Medline: 20210677]
5. Vardanjani AE, Reisi M, Javadzade H, Pour ZG, Tavassoli E. The Effect of nutrition education on knowledge, attitude, and performance about junk food consumption among students of female primary schools. J Educ Health Promot 2015;4:53 [FREE Full text] [doi: 10.4103/2277-9531.162349] [Medline: 26430680]
6. Scully M, Morley B, Niven P, Crawford D, Pratt IS, Wakefield M. Factors associated with high consumption of soft drinks among Australian secondary-school students. Public Health Nutr 2017 Feb 27;20(13):2340-2348 [FREE Full text] [doi: 10.1017/s1368980017000118]
7. Kwon S, Janz KF, Letuchy EM, Burns TL, Levy SM. Developmental Trajectories of Physical Activity, Sports, and Television Viewing During Childhood to Young Adulthood: Iowa Bone Development Study. JAMA Pediatr 2015 Jul;169(7):666-672 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.0327] [Medline: 25984811]
8. Adolescent health in the south-east Asia region. World Health Organization. URL: https://www.who.int/southeastasia/health-topics/adolescent-health [accessed 2022-11-10]
9. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
10. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
11. Children and parents: media use and attitudes report 2020/2021. Ofcom. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf [accessed 2022-11-10]
12. Romão T. Challenges in Designing Smarter Mobile User Experiences. Front. ICT 2016 May 06;3:7. [doi: 10.3389/fict.2016.00007]
13. Yang Q, Zimmerman J, Steinfeld A, Tomasic A. Planning adaptive mobile experiences when wireframing. 2016 Presented at: Proceedings of the 2016 ACM Conference on Designing Interactive Systems; June; Brisbane, QLD, Australia. [doi: 10.1145/2901790.2901858]
14. Nikolaou CK, Tay Z, Leu J, Rebello SA, Te Morenga L, Van Dam RM, et al. Young People's Attitudes and Motivations Toward Social Media and Mobile Apps for Weight Control: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Oct 10;7(10):e11205 [FREE Full text] [doi: 10.2196/11205] [Medline: 31603431]

15. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
16. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
17. Rosales-Morales VY, Sánchez-Morales LN, Alor-Hernández G, Sánchez-Cervantes JL, Cruz-Ramírez N, García-Alcaraz JL. Identification of UIDPs for developing medical apps. 2018 Presented at: International Conference on Software Process Improvement; October 19-21; Guadalajara, Mexico p. 175-185. [doi: 10.1007/978-3-030-01171-0_16]
18. Chan A, Kow R, Cheng JK. Adolescents' perceptions on smartphone applications (apps) for health management. Journal of Mobile Technology in Medicine 2017 Aug;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]
19. Neil T. Mobile Design Pattern Gallery: UI Patterns for Smartphone Apps. Sebastopol, California: O'Reilly Media, Inc; May 2014.
20. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile Health Interventions for Improving Health Outcomes in Youth: A Meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
21. Toxboe A. User interface design pattern library. UI Patterns. URL: http://ui-patterns.com/ [accessed 2022-11-10]
22. Lewis C. Irresistible Apps: Motivational Design Patterns for Apps, Games, and Web-based Communities. New York, NY, USA: Apress; 2014.
23. Ryan R, Deci E. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary Educational Psychology 2020 Apr;61:101860 [FREE Full text] [doi: 10.1016/j.cedpsych.2020.101860]
24. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
25. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
26. Cradock KA, ÓLaighin G, Finucane FM, Gainforth HL, Quinlan LR, Ginis KAM. Behaviour change techniques targeting both diet and physical activity in type 2 diabetes: A systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Feb 08;14(1):18 [FREE Full text] [doi: 10.1186/s12966-016-0436-0] [Medline: 28178985]
27. King AC, Whitt-Glover MC, Marquez DX, Buman MP, Napolitano MA, Jakicic J, et al. Physical Activity Promotion: Highlights from the 2018 Physical Activity Guidelines Advisory Committee Systematic Review. Med Sci Sports Exerc 2019 Jun;51(6):1340-1353. [doi: 10.1249/MSS.0000000000001945] [Medline: 31095090]
28. Hankonen N, Sutton S, Prevost AT, Simmons RK, Griffin SJ, Kinmonth AL, et al. Which behavior change techniques are associated with changes in physical activity, diet and body mass index in people with recently diagnosed diabetes? Ann Behav Med 2015 Feb;49(1):7-17 [FREE Full text] [doi: 10.1007/s12160-014-9624-9] [Medline: 24806469]
29. Galvan A. Adolescent development of the reward system. Front Hum Neurosci 2010;4:6 [FREE Full text] [doi: 10.3389/neuro.09.006.2010] [Medline: 20179786]
30. Bryan CJ, Yeager DS, Hinojosa CP, Chabot A, Bergen H, Kawamura M, et al. Harnessing adolescent values to motivate healthier eating. Proc Natl Acad Sci U S A 2016 Sep 27;113(39):10830-10835 [FREE Full text] [doi: 10.1073/pnas.1604586113] [Medline: 27621440]
31. Davidow JY, Foerde K, Galván A, Shohamy D. An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron 2016 Oct 05;92(1):93-99 [FREE Full text] [doi: 10.1016/j.neuron.2016.08.031] [Medline: 27710793]
32. Nurmi J, Knittle K, Ginchev T, Khattak F, Helf C, Zwickl P, et al. Engaging Users in the Behavior Change Process With Digitalized Motivational Interviewing and Gamification: Development and Feasibility Testing of the Precious App. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884 [FREE Full text] [doi: 10.2196/12884] [Medline: 32003750]
33. Joice A, Harley A. UX design for teenagers (ages 13-17). Nielsen Norman Group. URL: https://www.nngroup.com/reports/teenagers-on-the-web/ [accessed 2022-11-10]
34. Munson SA, Lauterbach D, Newman MW, Resnick P. Happier together: integrating a wellness application into a social network site. 2010 Presented at: International Conference on Persuasive Technology; October; Copenhagen, Denmark p. 27-39. [doi: 10.1007/978-3-642-13226-1_5]
35. Gibson K, Trnka S. Young people's priorities for support on social media: “It takes trust to talk about these issues”. Computers in Human Behavior 2020 Jan;102:238-247 [FREE Full text] [doi: 10.1016/j.chb.2019.08.030]
36. Ruiz LD, Zuelch ML, Dimitratos SM, Scherr RE. Adolescent Obesity: Diet Quality, Psychosocial Health, and Cardiometabolic Risk Factors. Nutrients 2019 Dec 23;12(1):43 [FREE Full text] [doi: 10.3390/nu12010043] [Medline: 31877943]
37. Antezana G, Venning A, Blake V, Smith D, Winsall M, Orlowski S, et al. An evaluation of behaviour change techniques in health and lifestyle mobile applications. Health Informatics J 2020 Mar;26(1):104-113 [FREE Full text] [doi: 10.1177/1460458218813726] [Medline: 30501364]

38. Caraban A, Karapanos E, Gonçalves D, Campos P. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300733]
39. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
40. Sullivan AN, Lachman ME. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]

Abbreviations

BCT: behavior change technique
BCT-UI: behavior change technique–user interface
CALO-RE: Coventry, Aberdeen, and London-Refined
mHealth: mobile health
UIDP: user interface design pattern

Edited by L Buis; submitted 03.03.22; peer-reviewed by R Raeside, K Mooses, K Steinbeck; comments to author 12.04.22; revised version received 15.08.22; accepted 22.11.22; published 19.01.23.

Please cite as:
Vlahu-Gjorgievska E, Burazor A, Win KT, Trajkovik V
mHealth Apps Targeting Obesity and Overweight in Young People: App Review and Analysis
JMIR Mhealth Uhealth 2023;11:e37716
URL: https://mhealth.jmir.org/2023/1/e37716
doi:10.2196/37716
PMID:36656624

Original Paper

Loss-Framed Adaptive Microcontingency Management for Preventing Prolonged Sedentariness: Development and Feasibility Study

Woohyeok Choi[1], PhD; Uichin Lee[2], PhD

[1]Information & Electronics Research Institute, Korea Advanced Institute of Science & Technology, Daejeon, Republic of Korea

[2]School of Computing, Korea Advanced Institute of Science & Technology, Daejeon, Republic of Korea

Corresponding Author:
Uichin Lee, PhD
School of Computing
Korea Advanced Institute of Science & Technology
291 Daehak-ro, Yuseong-gu
Daejeon, 34141
Republic of Korea
Phone: 82 42 350 3544
Email: uclee@kaist.edu

Abstract

Background: A growing body of evidence shows that financial incentives can effectively reinforce individuals' positive behavior change and improve compliance with health intervention programs. A critical factor in the design of incentive-based interventions is to set a proper incentive magnitude. However, it is highly challenging to determine such magnitudes as the effects of incentive magnitude depend on personal attitudes and contexts.

Objective: This study aimed to illustrate loss-framed adaptive microcontingency management (L-AMCM) and the lessons learned from a feasibility study. L-AMCM discourages an individual's adverse health behaviors by deducting particular expenses from a regularly assigned budget, where expenses are adaptively estimated based on the individual's previous responses to varying expenses and contexts.

Methods: We developed a mobile health intervention app for preventing prolonged sedentary lifestyles. This app delivered a behavioral mission (ie, suggesting taking an active break for a while) with an incentive bid when 50 minutes of uninterrupted sedentary behavior happened. Participants were assigned to either the fixed (ie, deducting the monotonous expense for each mission failure) or adaptive (ie, deducting varying expenses estimated by the L-AMCM for each mission failure) incentive group. The intervention lasted 3 weeks.

Results: We recruited 41 participants (n=15, 37% women; fixed incentive group: n=20, 49% of participants; adaptive incentive group: n=21, 51% of participants) whose mean age was 24.0 (SD 3.8; range 19-34) years. Mission success rates did not show statistically significant differences by group (P=.54; fixed incentive group mean 0.66, SD 0.24; adaptive incentive group mean 0.61, SD 0.22). The follow-up analysis of the adaptive incentive group revealed that the influence of incentive magnitudes on mission success was not statistically significant (P=.18; odds ratio 0.98, 95% CI 0.95-1.01). On the basis of the qualitative interviews, such results were possibly because the participants had sufficient intrinsic motivation and less sensitivity to incentive magnitudes.

Conclusions: Although our L-AMCM did not significantly affect users' mission success rate, this study configures a pioneering work toward adaptively estimating incentives by considering user behaviors and contexts through leveraging mobile sensing and machine learning. We hope that this study inspires researchers to develop incentive-based interventions.

(JMIR Mhealth Uhealth 2023;11:e41660) doi:10.2196/41660

KEYWORDS

contingency management; incentive; sedentary behavior; sedentariness; behavior change; health promotion; financial incentives; health intervention; user compliance; incentive-based intervention; mobile phone

Introduction

Background

Intrinsic motivation refers to an inherent motive to perform a target behavior [1], whereas extrinsic motivation is a specific type of motivation for obtaining a certain outcome that can be separated from the behavior [2]. Although it is clear that intrinsic motivation is essential for behavior change, numerous studies have presented evidence that extrinsic motivation can also greatly contribute to behavior change. A representative behavioral therapy that uses extrinsic motivation is contingency management, which provides external rewards (typically financial incentives) contingent on the occurrence of behaviors of interest for reinforcing positive behavior change [3]. Such a therapeutic approach has already shown effectiveness on behavior change in various fields, including physical activity promotion and dietary tracking [4-9], prevention of drug abuse [10-17], smoking cessation [18-22], productivity and academic performance [23-25], and driving behavior [26,27]. In addition to its application in academic fields, companies also use financial incentives as core motivators for behavior changes, such as health insurance discount programs for healthy behaviors [28] and a commitment contract that allows the company to send money from a user's account to a particular person or organization (eg, charities) if one fails to reach a self-defined goal [29].

However, the design of incentives in these contingency management interventions has several issues that hinder the achievement of the goal of these interventions, which is to promote successful behavior change. For example, socioeconomic status probably contributes to incentive effectiveness [30,31]. In addition, the assumption of a trade-off between ability and motivation (namely, users with low and high ability require high and low motivation for behavior change, respectively) may imply that the magnitude of an (extrinsic) motivator should differ by context and one's physical and cognitive capabilities for eliciting behavior change [32]. Another aspect of incentive design that should be considered is the delay between behavior occurrence and incentive delivery, with a shorter delay having shown greater effectiveness in eliciting behavior change [18,33]. Other potential contributors to the effectiveness of contingency management include incentive framing (eg, providing incentives for positive behaviors vs deducting expenses for negative behaviors) [23,26,34,35], incentive magnitude adjustments throughout the intervention [20,34,36,37], and incentive magnitude certainty (eg, fixed vs lottery incentives) [4,12,38].

Although previous studies have shown the effectiveness of the incentive designs of their proposed contingency management interventions, they had several limitations. For example, positive behavior was not immediately rewarded, and only behavioral outcomes from long-term behavior adherence were rewarded at the end of an intervention (eg, lump-sum provision) [7,21]. In addition, incentive magnitudes were often fixed [23,39] or randomly sampled from a predefined range of incentives [4,38] (ie, they did not change by context at the individual level). Moreover, although several studies have proposed an escalating reinforcer where incentive magnitudes increase at each positive behavior occurrence [20,34,36,37], such a design requires intervention practitioners to configure a detailed plan manually (eg, the amount of incentive increment) based on their domain knowledge.

Objectives

This study proposes a novel incentive-based mobile intervention named loss-framed adaptive microcontingency management (L-AMCM), which immediately discourages users' microbehaviors that cause adverse health effects by providing a personally and contextually tailored incentive. In more detail, this approach delivers a prompt recommending a positive behavior change when the user is susceptible to health risks. Each prompt presents a particular expense framed as a loss (ie, a loss-framed incentive), in which, if the user does not change their behavior in response to the prompt, that expense is deducted from an individual budget that the intervention regularly allocates. In addition, the deducted amount presented in each prompt is dynamically adjusted based on individuals' responses to prompts over varying incentives and contexts. To this end, the L-AMCM continuously monitors users' behavior changes in response to prompts presenting varied contexts and expenses. It iteratively learns an individual's behavioral model, which describes the likelihood of a behavior change in a given context and at a given expense. On the basis of the learned model, the L-AMCM estimates how much each prompt needs to bid to elicit positive behavior, at least to some extent.

To evaluate the feasibility of the L-AMCM, we applied it to a mobile health intervention app that delivers active break missions (ie, standing up and moving around for a while) with an estimated incentive via individuals' smartphones to discourage prolonged sedentary behavior (ie, 50-minute uninterrupted sitting sessions). This study illustrates the lessons learned regarding its application via a 3-week field study with 41 participants. We hope that this study will provide new research directions for incentive-based mobile interventions.

Methods

Design of the L-AMCM Intervention

Motivating Scenario

Herein, we illustrate an exemplar intervention scenario with microincentives in the domain of prolonged sedentary behavior interventions, similar to those in previous research [40,41]. When users uninterruptedly sit down for a long time, a given health intervention app triggers a prompt containing a behavioral suggestion for breaking the sedentary period (eg, standing up and moving around for a while) and bids a certain amount of monetary incentive that will be withdrawn from an individually assigned budget if users do not adhere to the suggestion. Users then examine whether the compensation is sufficient to make them adhere to the behavior suggestions in a given context.

For example, if they receive the prompt late at night, a period in which they may be feeling somewhat tired, they may choose not to adhere to the suggestion if the incentive is low; they would rather continue engaging in sedentary behavior. However, if the intervention bids a larger incentive, they may consider accepting the behavioral suggestion. In addition, if the prompt is coincidentally delivered immediately or closely after they have spent some time working very hard at the office and may be feeling the need to refresh, they may be more willing to take an active break even with a lower incentive. However, it is clear that the tendency to accept behavioral suggestions with incentives will differ by user. For example, users who already know the health risks of prolonged sitting sessions may be willing to try to comply with more active break suggestions even with lower incentives.

A core assumption of the presented scenario is that *users are more likely to accept the behavioral suggestion as the incentive grows*, which stems from the evidence of various studies showing that a larger incentive magnitude corresponds to a larger effect on health behavior change [7,10,18,31,33]. Another assumption is that *the incentive magnitude necessary for eliciting behavior change may differ by user and context*, which is grounded in the Fogg Behavior Model, a practical framework illustrating the underlying factors relevant to behavior change [32]. In this model, a particular behavior happens through the interplay of an individual's inherent *motivation* toward the behavior; an individual's *ability* to perform the behavior; and an external *prompt* that elicits behavior change by reminding the behavior, reinforcing motivation, or simplifying the behavior. In the presented scenario, the amount of incentive plays a role in sparking positive behavior change, and the change in incentive magnitude across users and contexts is based on several important aspects of the Fogg Behavior Model, as shown in Textbox 1.

Textbox 1. Key aspects of the Fogg Behavior Model considered in the proposed incentive mechanism.

- Fogg Behavior Model key aspects
  - *Motivation and ability have a trade-off relationship* (eg, lower ability requires more motivation for behavior change); thus, the amount of incentive necessary for the positive behavior might need to change across different levels of motivation and ability. For example, for people with enough adherence motivation, the amount of incentive required for the behavior change might be smaller compared with less motivated people. In addition, people with less ability might need to be compensated with more incentives for behavior change.
  - *Motivation and ability differ among individuals*; thus, the amount of incentive required for the behavior change would be different by individual.
  - *Ability differs by context*; thus, the amount of incentive required for positive behavior change might vary by context.

Hypothetical User Behavior on Incentives and Contexts

On the basis of these assumptions, we hypothesized an equation for a user's behavior occurrence likelihood, $y \in [0,1]$, with a given incentive magnitude $r \in R$ and context $c \in C$ as follows: $f: r, c \rightarrow y$ such that $\forall r, r' \in R$ and $c, c' \in C f(r, c) \leq f(r', c')$ if $r \leq r'$ and $c = c'$.

Various functions satisfying the aforementioned equation can be used to model the hypothetical user behavior we propose. This study considers a logistic regression (LR) model as it naturally maps an input into a probability output and is easily implemented and interpreted [42]. Then, assuming a vector of one-hot encoded discrete contexts, $C=\{c_1, c_2,..., c_{n-1}, c_n\}$, and the corresponding coefficients (ie, the effect of context on behavior occurrence likelihood), $B=\{\beta_1, \beta_2,..., \beta_{n-1}, \beta_n\}$, the hypothesized user behavior can be modeled as follows: , where β_0 indicates the effect of incentive magnitude on behavior occurrence likelihood and $\in$ is an intercept term.

In addition, β_0 is assumed to be >0. A user's behavior model with a nonpositive β_0 implies that the incentive does not influence or even deteriorates behavior occurrence likelihood, and such a situation contradicts the assumption that larger incentives are more likely to elicit behaviors.

Throughout the intervention, the hypothetical user behavior is rebuilt after each incentive bidding and behavior occurrence observation; namely, the rejection or acceptance of a particular incentive magnitude results in gradual changes in the estimated incentive for the next bid (Figure 1). Accordingly, incentive magnitudes are dynamically adjusted by the proposed incentive estimation to find the appropriate incentive magnitude that can elicit behavior occurrence with a probability, $\bar{y}$.

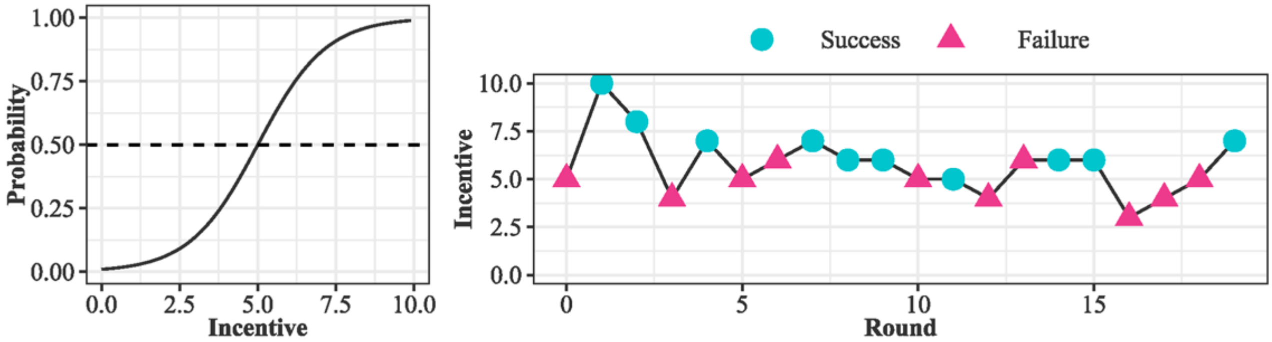
In addition, we only consider recent behaviors for building the hypothetical user behavior (and estimating an appropriate incentive) as user behavior occurrence likelihood related to incentives may change over time. For example, initially, users may choose to adhere to a behavior suggestion because they know that they will receive monetary compensation for the suggestion, not because the behavior may improve health outcomes; however, as they adhere to the behavior because of the knowledge of subsequent compensation, they may eventually perceive the usefulness of performing the target behavior, thereafter potentially becoming intrinsically motivated to conduct the suggested behavior with little or no incentive [1]. Considering this, it may be that more recent response behaviors are more important than older ones in modeling behavior occurrence. A detailed algorithm for the proposed incentive strategy is presented in Multimedia Appendix 1.

Estimation of Incentive Magnitude

The hypothesized user behavior can also be rearranged for estimating the incentive magnitude necessary to elicit a target behavior with a given probability, $\bar{y}$, as follows: .

Importantly, incentive providers can choose the $\bar{y}$ depending on their policies. For example, if greater costs are not of concern, a large $\bar{y}$ will make users highly likely to comply with the behavior suggestions, whereas $\bar{y}$ close to a half probability will make adherence to behavior suggestions highly uncertain.

Figure 1. A toy example that illustrates how the proposed strategy works. The left panel shows a user's behavior probability across different incentives, and the right panel describes a trace for finding a particular incentive magnitude that is necessary for eliciting positive behavior change. y ¯ is set at 0.5.



Loss-Framed Incentive

A major characteristic of the proposed incentive mechanism is to bid higher incentives as the target behavior becomes less likely. For example, once a user rejects a behavioral suggestion for a given incentive magnitude and context, our mechanism would assume that such a magnitude is insufficient to elicit behavior change. Therefore, the subsequent behavioral suggestion triggered in an identical context will bid a greater magnitude. Otherwise, the user is offered the same or a smaller incentive at the next behavioral suggestion. Such a characteristic may yield gaming behavior if the user is rewarded for succeeding in behavioral missions (ie, a gain-framed incentive). For example, the user may deliberately reject the current bid suggested by the intervention prompt and maintain an unhealthy state to earn higher incentives at successive bids.

To discourage such behavior, we used a loss-framed incentive (ie, a deposit contract) that deducts estimated amounts during mission failures from budgets paid in advance. Combined with the loss-framed incentive, our incentive mechanism gradually increases the amount deducted if the user consecutively rejects the bids, whereas if the user is more likely to accept the bids and comply with behavioral missions, the amount deducted for mission failures decreases. In such a mechanism, the optimal strategy for obtaining as many incentives as possible is to maintain a healthy behavior (eg, regularly interrupting prolonged sedentariness) to keep behavioral missions (which are designed to deduct incentives from the budget) from being triggered and comply with behavioral missions regardless of incentive amounts if missions are triggered. Not only does this strategy keep budgets without deduction, but it also decreases the amount deducted for mission failures because of unavoidable reasons.

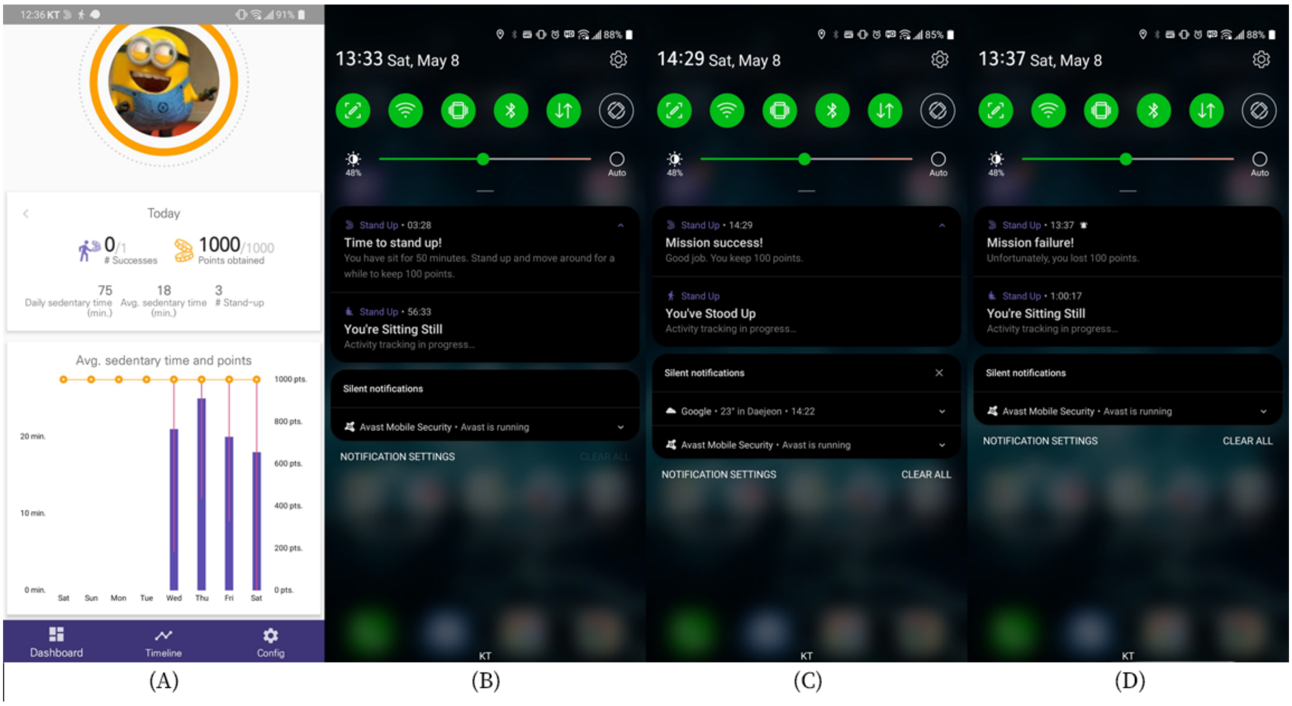
In addition to the prevention of gaming behavior, another reason for using the loss-framed incentive is that the loss-framed incentive is more likely to elicit behavior change than the gain-framed incentive because of people's tendency to place a greater emphasis on losses than gains, as stated by the prospect theory [43]. In practice, previous studies have demonstrated a better effect of the loss-framed incentive on health outcomes than the gain-framed incentive in a variety of intervention domains, including mitigating smartphone overuse [23], promoting physical activity [35], and improving driving behavior [26].

Implementation of Mobile Health Intervention

Overview

To explore how users respond to the proposed incentive strategy, we implemented a research app prototype named *StandUp*. This prototype comprises 4 major components: sedentary behavior tracking, context sensing, incentive estimation, and active break mission delivery. The StandUp user interface is presented in Figure 2.

Figure 2. Overview of the StandUp user interface. From the left, (A) dashboard summarizing mission results and compensation, (B) mission trigger notification, (C) mission success notification, and (D) mission failure notification.



Sedentary Behavior Tracking

Our prototype app monitors step counts through the user's smartphone to detect sedentary behavior. The prototype assumes that the user is stationary when <10 steps are recorded in 1 minute. Moving or taking an active break is defined as >45 steps being recorded within a minute. In contrast, we considered that 10 to 45 steps within a minute were transitions between sedentariness and movement (or vice versa); hence, we refrained from exactly determining whether the user is sedentary in these cases. Specifically, 10 and 45 steps correspond to approximately 6.6 m to 7.9 m and 29.7 m to 35.6 m of movement, respectively [44,45]. The rationale behind the hard-coded threshold for step counts (ie, 10 and 45 steps) was derived from an internal pilot test wherein these step numbers corresponded to walking for 30 to 60 seconds, respectively. StandUp schedules an intervention prompt to appear after 50 minutes when users become stationary. In addition, the scheduled prompt is canceled if a substantial movement change (ie, at least 10 steps within a minute) is detected.

Context Sensing

Context sensing is used for tailoring incentives to different contexts. As location substantially contributes to users' decisions to comply with interventions [40], our prototype considers location as the key context variable. Once any stationary event occurs, StandUp retrieves the latitude and longitude of the current location from the smartphone's GPS sensor.

Active Break Mission Delivery

If there is no mobility state change for 50 minutes after an intervention prompt has been scheduled, a user receives a mission that suggests taking an active break in the form of a smartphone notification. Each mission lasts 10 minutes and informs about a specific expense deducted from a budget upon failing that mission, where the budget is individually assigned at the start of every day. The 10-minute threshold for adherence to the mission is based on the finding that people see incoming smartphone notifications within 10 minutes on average from notification arrival even when the ringer mode is set to silent [46].

After the notification appears on the smartphone, StandUp begins to check via sedentary behavior tracking whether a user takes an active break within 10 minutes. If a given mission expires without behavior change (ie, no mobility is detected within 10 minutes of mission delivery), StandUp reminds the user of the amount lost via a notification and deducts the amount from the user's budget. Otherwise, a message of mission success is displayed on the notification. In the case of mission failures, StandUp reschedules the next active break mission to be delivered after 50 minutes. After each mission is completed, StandUp records the mission result (ie, success vs failure), amount of suggested expenses, and GPS coordinates of the current location. These data are stored in the user's smartphone's internal storage and used for incentive estimation in subsequent missions. In addition, they are later uploaded to a server via the Wi-Fi network.

Incentive Estimation

StandUp supports either a fixed or adaptive incentive strategy. StandUp with a fixed incentive strategy presents a predefined expense without any estimation. Regarding the adaptive incentive strategy, StandUp obtains the mission results (ie, success or failure), expense bids, and GPS coordinates of the locations where missions were initiated within the most recent 7 days. The continuous GPS coordinates should be transformed into discrete factors for modeling user behaviors in response to varying expenses and locations. We used a geohash for this purpose, which maps all locations on Earth onto rectangular grids and represents each rectangle as a short alphanumeric string. Our implementation maps GPS coordinates within 150-m by 150-m square grids to a single 7-character geohash string (ie, a 7-bit geohash) so that continuous GPS coordinates are discretized. Geohashed representations of locations are then factored with one-hot encoding. Consequently, mission results, expense bids, and factored locations were used for user behavior modeling and incentive estimation (Multimedia Appendix 1). In addition, the current implementation sets the probability of expected behavior occurrence (ie, $\bar{y}$) to 0.5. Such a parameter may allow the adaptive incentive strategy to actively explore the smaller incentive magnitude that is potentially optimal for eliciting behavior change.

Study Design

For 3 weeks, we conducted a single-blind, between-group study with 2 groups: fixed incentive and adaptive incentive. All participants received US $1.50, which is presented as 100 points in StandUp, as a daily budget each morning during the intervention period. We used this specific value (US $1.50) as it is the median value of the daily incentives used in previous studies on incentive interventions for improving physical activity [31]. The fixed incentive group lost US $0.30 whenever participants failed a given active break mission. In the adaptive incentive group, participants lost an amount of incentive estimated by the proposed incentive strategy, in which the incentive ranged from US $0.30 to $3 with a US $0.30 increment (namely, US $0.30, US $0.60,..., US $2.70, and US $3) and the closest to the estimated one within that range was bid. For example, if the estimated incentive was US $1.40, the real incentive presented to users was US $1.50. If the daily budget was exhausted, participants did not receive any incentives on that day. The field trial was conducted between April 2020 and May 2020.

Recruitment and Procedures

We recruited participants from our web-based campus community and Facebook. The inclusion criteria were as follows: having a sedentary occupation, spending >6 hours sitting on weekdays, and possessing a smartphone with an Android version 7.00 or higher. The participants were randomly assigned to the fixed and adaptive groups so that there was no significant difference between the groups regarding demographics such as age (*P*=.72; $t_{38.238}$=0.363) and gender (*P*>.99; N=41, χ^2_1<0.0). Before participating in the field study, they received information on the health risks of prolonged sedentary behavior and how to use StandUp. In addition, we briefly instructed participants in the adaptive group on how incentive amounts were estimated (eg, as they become less likely to adhere to behavioral missions, a larger deducted amount is presented). However, we did not explain the detailed algorithm underlying our incentive mechanism (eg, mathematical equations describing user behaviors in response to incentives and contexts) as we believed it might be difficult for the general population to comprehend.

The first week was the baseline period, with StandUp just displaying the minutes that participants spent in a sedentary state on its dashboard and not delivering any active break missions. This period was intended to minimize the novelty effect of our app on any user behavior. After the baseline period, through SMS text messages, we asked participants to activate the mission delivery option for the second and third weeks. Participants were allowed to choose the start time of the missions from 9 AM to 11:59 AM depending on their preferences. The mission prompts were delivered over 9 hours from the chosen start time (eg, 9 AM-6 PM to 11:59 AM-8:59 PM) every day during the intervention period. Thus, at most, 10 missions were delivered to participants per day if they remained sedentary during the mission activation period.

After the field study, we compensated participants with US $24 for study participation and extra payments for the results of their missions (US $21 extra at maximum). In addition, exit interviews lasting 30 minutes were conducted with each participant to investigate user experiences with StandUp and potential factors relevant to the effectiveness of different incentive strategies.

Exclusion Criteria

To clean the data, we first excluded missions collected at the first date of the intervention period as participants manually activated the active break mission delivery option on the first day of the intervention period (the eighth day of the entire field study) and the missions collected on that date possibly contained noise. In addition, we found that StandUp did not operate for a few days for several participants, resulting in a large loss of mission results. Therefore, we excluded all missions from participants whose data did not show any missions triggered for 2 consecutive days.

Measurements and Data Analysis

The major outcome for evaluating the effectiveness of the proposed incentive strategy was the success rate of active break missions. It was defined as the ratio of the number of successful missions to the number of missions triggered. On the basis of the results of the Shapiro-Wilk normality test, both groups' success rate was normally distributed (for the fixed incentive group: *P*=.08, and *W*=0.905; for the adaptive incentive group: *P*=.51 and *W*=0.953). Therefore, we compared the means of the success rates of the fixed and adaptive incentive groups using the Welch 2-tailed *t* test, which is known to have better control over type-1 errors than the Student *t* test [47].

In addition, we performed follow-up analyses of the adaptive incentive group to investigate in depth the effects of various factors on the mission success rate. First, we conducted a generalized linear mixed model (GLMM) analysis, hypothesizing that the mission success rate may be affected by days passed since the intervention onset, expense bids, and location. A reason for including the days passed since the intervention onset in the GLMM analysis is that repeated provision of intervention prompts during intervention periods would decrease responsiveness to those prompts because of the habituation effect [48]. Other 2 factors, deducted amounts and location, were examined to corroborate a hypothesis regarding our incentive mechanism, namely, that the occurrence of the target behavior would vary by context and incentive.

Before building the GLMM, we preprocessed the location data. First, we converted the GPS coordinates of locations where missions were triggered into 7-bit geohash strings, as our incentive mechanism did. As our participants resided elsewhere, geohashed locations would also be different and, thus, could not be used in the GLMM as a factor. Therefore, we relabeled geohashed locations considering the number of behavioral missions triggered (ie, the number of times prolonged sedentariness happened) at each location. For example, the top-*k* location refers to the geohashed location where missions were the *k*th most frequently triggered out of all geohashed locations. We analyzed the top 5 locations where 90% of the missions were triggered. Consequently, our GLMM included the following fixed effects—expense bids, days passed since the intervention onset, and the top 5 locations where missions were triggered—and the following random intercepts—the participants and top 5 locations within participants (ie, nested random effects). A detailed formula for the GLMM is presented in Multimedia Appendix 2.

Another follow-up analysis was conducted to examine the distribution of coefficients in the hypothetical user behavior model (ie, β_0), which was updated for each behavioral suggestion, to investigate whether our incentive estimation worked as intended.

Ethics Approval

This study was approved by the institutional review board of the Korea Advanced Institute of Science and Technology (KH2019-114), and we obtained written consent from all participants.

Results

Population Characteristics

A total of 41 participants initially took part in the 3-week field trial. The mean age was 24.0 (SD 3.8; range 19-34) years, and there were 37% (15/41) female participants. Most participants were graduate (17/41, 41%) and undergraduate (20/41, 49%) students. The other 10% (4/41) of participants were an office clerk, a graphic designer, a private academy instructor, and a researcher. In addition, of the 41 participants, 20 (49%) and 21 (51%) were assigned to the fixed and adaptive incentive groups, respectively. In total, 2387 missions (n=1021, 42.77% failures) were recorded during the field trial. From the data cleaning, of the 2387 missions, we excluded 179 (7.5%) collected on the first day of the intervention period and 399 (16.72%) from 7 participants for whom StandUp did not operate well. The following analyses were conducted with the remaining 1809 missions (n=684, 37.81% failures) from 34 participants (n=13, 38% female participants; n=17, 50% of participants in each group) whose mean age was 24.2 (SD 4.0; range 19-34) years.

Comparison of Mission Success Rate

Participants in the fixed and adaptive incentive groups received 900 (n=347, 38.6% failures) and 909 (n=337, 37.1% failures) active break missions, respectively. Although the fixed incentive group (mean 0.66, SD 0.23) showed a larger success rate than the adaptive incentive group (mean 0.61, SD 0.22), the Welch *t* test showed that the difference between the groups was not statistically significant (P=.54; $t_{31.85}$=0.62).

Follow-up Analysis of the Adaptive Incentive Group

GLMM Analysis

As noted previously, most missions in the adaptive group were triggered at the top 5 locations (831/909, 91.4%), in which the top-1 to top-5 locations occupied 60.3% (548/909), 15.4% (140/909), 7.7% (70/909), 4.8% (44/909), and 3.2% (29/909) of the missions, respectively. As shown in Table 1, the GLMM analysis revealed that expense bids did not show statistical significance on mission success (P=.18; odds ratio [OR] 0.98, 95% CI 0.95-1.01). Meanwhile, the location where missions were fourth most frequently triggered (P=.04; OR 0.49, 95% CI 0.25-0.96) and the days passed since the intervention onset (P=.03; OR 0.95, 95% CI 0.91-0.99) had a statistically significant influence on mission success.

Table 1. Results of the generalized linear mixed model analysis for behavior occurrence likelihood in the adaptive incentive group. Marginal and conditional R^2 are 0.030 and 0.312, respectively.

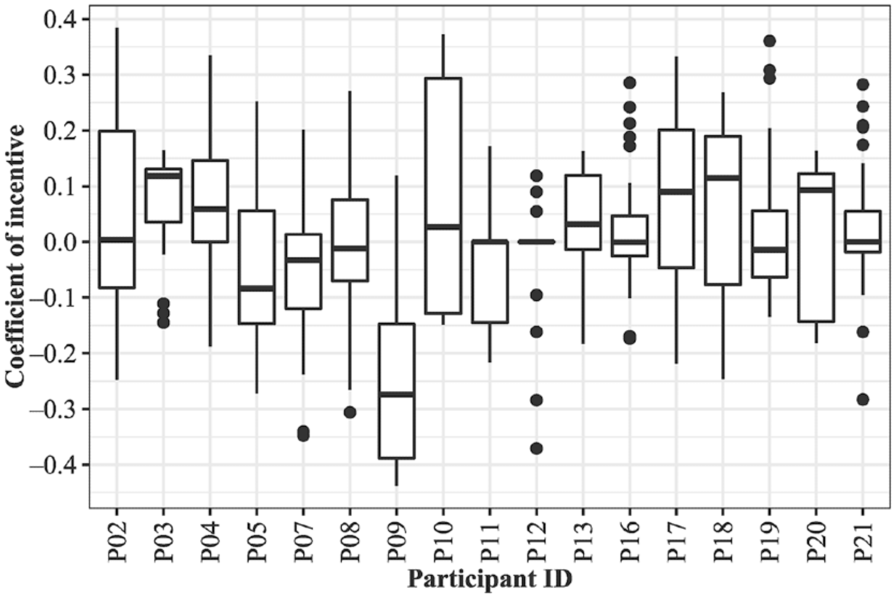
Fixed effects	β (SE)	*z* score	OR[a] (95% CI)	*P* value
Intercept	1.32 (0.46)	2.90	3.75 (1.53-9.18)	.004
Expense bids	−0.02 (0.02)	−1.33	0.98 (0.95-1.01)	.18
Top-*k* location				
Top 1	0.24 (0.22)	1.07	1.26 (0.82-1.96)	.29
Top 2	−0.18 (0.25)	−0.73	0.83 (0.51-1.37)	.47
Top 3	0.02 (0.30)	0.06	1.02 (0.57-1.82)	.96
Top 4	−0.71 (0.34)	−2.08	0.49 (0.25-0.96)	.04
Top 5	0.71 (0.42)	1.66	2.03 (0.88-4.65)	.10
Days since intervention onset	−0.05 (0.02)	−2.23	0.95 (0.91-0.99)	.03

[a]OR: odds ratio.

Coefficients Corresponding to Incentive Magnitudes

We further investigated how the proposed incentive strategy estimated the effect of the incentive magnitude for every expense bidding. As shown in Figure 3, we found that β_0 often became negative where the mean of the coefficient across participants was 0.00 (SD 0.08; 95% CI −0.04 to 0.04). In other words, our incentive strategy often estimated that bidding larger expenses rather inhibited participants' behavior occurrence likelihood.

Figure 3. Distribution of coefficients corresponding to the incentive magnitude across participants.



Corroborating Statistical Analyses via Interviews

Although our statistical analysis did not show a clear relationship between mission success rate and incentive magnitudes, we discovered 2 major behavioral patterns related to incentives in the qualitative interview analysis. One pattern was that participants randomly accepted behavioral suggestions regardless of the expenses offered. Some participants described being intrinsically motivated toward engaging in the active break as they were aware of the risk of sedentary lifestyles or had already felt that their sitting time was too long. These participants typically tried to accomplish active break missions without checking how much expense was bid, as participant 4 noted:

> *I'm having lower back pain when I sit down and keep myself focused on studying for an hour or two. While using this app (StandUp), I stood up every 50 minutes and felt that my back pain was greatly alleviated. The main reason I followed the active break suggestions was for health benefits, not the money.*

Meanwhile, a participant (participant 8) reported choosing to adhere to the mission after 50 minutes of being in a sedentary state regularly to improve his productivity, in a method similar to the Pomodoro technique:

> *Perhaps the original purpose of this app (StandUp) is to prevent some cardiovascular diseases by increasing physical activity. However, I used this app for a different reason; I used to be less efficient when focusing on one thing. Once I engaged in an active break mission, I organized my thoughts for a while as I walked. So, I felt that my productivity improved.*

Another pattern was that participants adhered to behavioral suggestions only when substantial expenses were presented. Subsequently, participants tended to be less sensitive to minor changes in expenses. Several participants had different criteria for the minimum expense that made them consider mission acceptance. Therefore, these participants tended to reject missions when incentives smaller than their criteria were offered, as participant 5 noted:

> *I tried to accept missions when this app (StandUp) will take back at least 0.15 USD. Such an amount is like my psychological Maginot line.*

Discussion

Principal Findings

Financial incentives have been widely regarded as effective behavior reinforcers in diverse health and behavior change domains [31,33,49-54]. However, to the best of our knowledge, most previous studies on incentive-based health interventions often assumed that users' responses to incentives were homogeneous; thus, compensation for positive behavior changes was fixed even in different individuals and contexts. In addition, these studies required intervention providers to manually configure incentive strategies based on their domain knowledge. Meanwhile, this study argues against such one-size-fits-all incentive strategies and explores the feasibility of a novel incentive strategy, L-AMCM. It personally and contextually tailors the incentive magnitude to users, which is then immediately suggested to them to reinforce behavior changes. We hypothesized that we could computationally learn about an individual's preference for incentives by referring to one's previous responses to incentives in varying contexts.

We developed a simple mobile health intervention targeted at discouraging prolonged sedentary behaviors by delivering information for users about prolonged sedentariness (ie, 50-minute sitting sessions) and by suggesting, via app notifications, that they take active breaks with a loss-framed, low-cost financial incentive. We assumed that people would be more likely to adhere to the behavior suggestions as the deducted amount grew; this led to an LR-based context-aware incentive adaptation where the deducted amount dynamically changed depending on adherence to the behavioral suggestion across different incentives and contexts. Unfortunately, our 3-week between-group field trial with 41 participants showed that the proposed incentive strategy failed to promote more adherence to health behaviors than a fixed incentive strategy. Furthermore, the follow-up analyses partially confirmed that location influenced behavior occurrence likelihood. However, it was found that behavior occurrence likelihood did not always increase as incentive magnitude increased, possibly because of enough intrinsic motivation and less sensitivity to incentive magnitude.

Lessons Learned

From our findings, we learned lessons that may improve incentive magnitude tailoring for contingency management interventions. As a previous study pointed out that the effect of incentive magnitude on decision-making is small [55], one lesson is that acceptance of behavioral suggestions may not be proportional to incentive magnitude; namely, our results were not concordant with our initial expectations. At least in the sedentary behavior intervention we designed, there may be a case in which users may accept or reject behavior suggestions without regard to incentive magnitude. Although somewhat counterintuitive and radical, we may design different incentive strategies where a user's behavior occurrence likelihood and incentive magnitude are independent of each other instead of there being a linear relationship between them.

Another lesson is that incentive magnitudes should change in a coarse-grained manner. In this study, incentive magnitudes were set to range from US $0.30 to $3 with a US $0.30 increment; nonetheless, participants reported in the interviews that they were less sensitive to fine-grained changes in incentive magnitudes and that they instead had a rough threshold for considering adherence to behavioral suggestions. Hence, substantial changes in incentive magnitude may be more appropriate for eliciting differential behavioral patterns.

Limitations and Future Work

Our sedentary behavior tracking used step counts obtained from an individual's smartphone; thus, it has constraints such as requiring participants to always carry their smartphones and move around at least 30 m to detect active break sessions. Unfortunately, these constraints made it impossible to capture behavior change when participants did not carry their smartphones and to differentiate a standing activity that can interrupt sedentariness (eg, standing and stretching or working at a standing desk) from sedentary behavior. Detecting sedentary activity by identifying an individual's posture (eg, lying, sitting, or upright position) with wearable sensors (eg, a thigh-attached accelerometer) could be an alternative for tracking sedentary behavior with better precision and granularity [56].

Another limitation of this study was that health-related outcomes were not measured. Given that previous studies have demonstrated the health benefits of contingency management [52] and prompt-based interventions [57], we assumed that our prompt-based contingency management intervention probably had a positive impact on health outcomes in our experimental design phases. Under such an assumption, the primary objective of this study was to compare different incentive mechanisms in terms of the occurrence of the desired behavior and not to confirm the general health effects of the proposed intervention. Nonetheless, it would be beneficial to precisely measure health-related outcomes such as time spent in sedentary or physical activity [57,58] to establish not only the general health benefits of our intervention but also to rigorously compare the effects of various incentive mechanisms.

For ease of implementation, the incentive mechanism presented in this study used only geohashed locations as contextual factors. Unfortunately, it is challenging to interpret geohashed locations intuitively; thus, our GLMM analysis only partially confirmed the impact of location on behavior occurrence and did not provide a comprehensive interpretation of these locations. It would be beneficial to assign semantic meaning to geohashed locations (eg, home, workplace, or eatery) to clearly understand which attributes of locations influence behavior occurrence. For example, a future study may ask participants to name their locations semantically after delivering intervention prompts via ecological momentary assessment [40,56]. In addition, there would be other contextual factors (eg, ongoing tasks and social settings [40]) and intrinsic attributes (eg, self-efficacy and perceived enjoyment [59] and affective responses [60] toward the target behavior) that may influence an individual's response to incentives. Future work may try modeling user behavior with several variables as their results will probably improve our knowledge of appropriate incentive magnitude estimation for contingency management interventions.

Although this study does not reveal the benefits of the L-AMCM in terms of target behavior occurrences over a short intervention period, a long-term and follow-up investigation might disclose intriguing effects on user behaviors, supposing that the L-AMCM works as intended. The mitigation of habituation is one of the potential effects we expect. As with the fixed incentive mechanism, the repeated provision of monotonous incentives (ie, providing stimuli with the same intensity) may diminish the perceived value of incentives over time [61]. In contrast, the unique nature of the L-AMCM to offer varying incentives based on users' responsiveness to incentives (ie, providing stimuli with varying intensities) may keep users from becoming accustomed to intervention prompts to some extent.

Furthermore, it would be an interesting research direction to design multicomponent interventions, including incentive adaptation, by considering challenges specific to sedentary behavior. Previous studies have found, for example, that people tend to identify sedentary behavior as a behavior entailing sitting rather than sedentary behavior itself [62]. Hence, sedentary behavior may be habitual and not purposeful [63], and the time spent in sedentariness was found to be underestimated [64]. This lower awareness of sedentary behavior may make people less aware of its adverse health effects and the health benefits of breaking up long periods of sedentary behavior, possibly leading to decreased motivation toward interventions for preventing sedentary behavior. As an example of how to make people aware of the health risks associated with prolonged sedentariness, the intervention might provide information on behavioral consequences [65]. For example, our mission prompt can be designed to convey specific health outcomes that may result from accepting or rejecting behavioral missions (eg, “Taking an active break now can reduce the risk of a cardiovascular disease by XX%”).

This feasibility study considered prolonged sedentary behavior, which can be easily detected with an off-the-shelf smartphone [40,57], to avoid technical challenges irrelevant to the L-AMCM. However, we believe that the L-AMCM might be used in a wide range of intervention domains that satisfy certain criteria. The first (but not mandatory) criterion is that unhealthy behavior needs to happen somewhat frequently so that responses to incentives are collected to some extent within a short period and a behavior model is quickly learned. The other criterion is that health behavior or outcome changes should be monitored following the bidding of incentives to track responses to particular incentives. For example, smoking cessation would be a suitable intervention domain for applying the L-AMCM; smoking episodes occur frequently and can be automatically detected despite some technical challenges (eg, requiring multiple sensor units such as a wrist-worn sensor for detecting wrist-to-mouth movements and a chest-worn sensor for examining inhalation and exhalation [66]).

Conclusions

This study aimed to devise a novel incentive strategy that adjusts incentive magnitude depending on individuals’ behaviors in different contexts and incentives, as well as explore the feasibility of such a strategy via a field trial. To this end, we first developed the LR-based incentive estimation with the expectation that behavior occurrence likelihood would vary by incentive magnitude and context and increase as the incentive grows. However, the 3-week field study showed that users’ actual behaviors were nonconcordant with our expectations. Thus, the proposed incentive strategy showed no statistically significant differences from the fixed incentive strategy. Interestingly, the follow-up analyses revealed that users might be less sensitive to minor changes in incentive magnitudes. Although the proposed incentive strategy failed to show its effectiveness clearly, we believe that this study was the first step toward incentive adaptation for mobile health interventions and hope that it inspires various other researchers to develop and test adaptive incentive strategies.

Acknowledgments

This work was supported by a National Research Foundation of Korea grant funded by the Korean government (Ministry of Science and ICT; 2022R1C1C2003865) and the Bio & Medical Technology Development Program of the National Research Foundation funded by the Korean government (Ministry of Science and ICT; 2021M3A9E4080780).

Conflicts of Interest

None declared.

Multimedia Appendix 1
An algorithm of a logistic regression–based incentive estimation.
[PNG File , 74 KB - mhealth_v11i1e41660_app1.png]

Multimedia Appendix 2
A generalized linear mixed model formula.
[PNG File , 26 KB - mhealth_v11i1e41660_app2.png]

References

1. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
2. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. J Pers 1994 Mar;62(1):119-142. [doi: 10.1111/j.1467-6494.1994.tb00797.x] [Medline: 8169757]
3. Petry NM. Contingency management: what it is and why psychiatrists should want to use it. Psychiatrist 2011 May;35(5):161-163 [FREE Full text] [doi: 10.1192/pb.bp.110.031831] [Medline: 22558006]
4. Achananuparp P, Lim E, Abhishek V, Yun T. Proceedings of the 2018 International Conference on Digital Health. 2018 Apr Presented at: DH'18: International Digital Health Conference; Apr 23 - 26, 2018; Lyon France. [doi: 10.1145/3194658.3194662]
5. Agapie E, Avrahami D, Marlow J. Staying the course: system-driven lapse management for supporting behavior change. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI'16:

CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA. [doi: 10.1145/2858036.2858142]
6. Ha Y, Karyda M, Lucero A. Exploring virtual rewards in real life: a gimmick or a motivational tool for promoting physical activity? In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. 2020 Jul Presented at: DIS '20: Designing Interactive Systems Conference 2020; Jul 6 - 10, 2020; Eindhoven Netherlands. [doi: 10.1145/3357236.3395477]
7. Finkelstein EA, Linnan LA, Tate DF, Birken BE. A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. J Occup Environ Med 2007 Sep;49(9):981-989. [doi: 10.1097/JOM.0b013e31813c6dcb] [Medline: 17848854]
8. Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G. Financial incentive-based approaches for weight loss: a randomized trial. JAMA 2008 Dec 10;300(22):2631-2637 [FREE Full text] [doi: 10.1001/jama.2008.804] [Medline: 19066383]
9. Lim BY, Kay J, Liu W. How does a nation walk?: interpreting large-scale step count activity with weekly streak patterns. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-46. [doi: 10.1145/3328928]
10. Petry NM, Tedford J, Austin M, Nich C, Carroll KM, Rounsaville BJ. Prize reinforcement contingency management for treating cocaine users: how low can we go, and with whom? Addiction 2004 Mar;99(3):349-360 [FREE Full text] [doi: 10.1111/j.1360-0443.2003.00642.x] [Medline: 14982548]
11. Petry NM, Peirce JM, Stitzer ML, Blaine J, Roll JM, Cohen A, et al. Effect of prize-based incentives on outcomes in stimulant abusers in outpatient psychosocial treatment programs: a national drug abuse treatment clinical trials network study. Arch Gen Psychiatry 2005 Oct;62(10):1148-1156. [doi: 10.1001/archpsyc.62.10.1148] [Medline: 16203960]
12. Petry NM, Alessi SM, Hanson T, Sierra S. Randomized trial of contingent prizes versus vouchers in cocaine-using methadone patients. J Consult Clin Psychol 2007 Dec;75(6):983-991. [doi: 10.1037/0022-006X.75.6.983] [Medline: 18085914]
13. Budney AJ, Moore BA, Rocha HL, Higgins ST. Clinical trial of abstinence-based vouchers and cognitive-behavioral therapy for cannabis dependence. J Consult Clin Psychol 2006 Apr;74(2):307-316. [doi: 10.1037/0022-006X.4.2.307] [Medline: 16649875]
14. Roll JM, Petry NM, Stitzer ML, Brecht ML, Peirce JM, McCann MJ, et al. Contingency management for the treatment of methamphetamine use disorders. Am J Psychiatry 2006 Nov;163(11):1993-1999. [doi: 10.1176/ajp.2006.163.11.1993] [Medline: 17074952]
15. Roll JM, Shoptaw S. Contingency management: schedule effects. Psychiatry Res 2006 Sep 30;144(1):91-93. [doi: 10.1016/j.psychres.2005.12.003] [Medline: 16905197]
16. Peirce JM, Petry NM, Stitzer ML, Blaine J, Kellogg S, Satterfield F, et al. Effects of lower-cost incentives on stimulant abstinence in methadone maintenance treatment: a National Drug Abuse Treatment Clinical Trials Network study. Arch Gen Psychiatry 2006 Feb;63(2):201-208. [doi: 10.1001/archpsyc.63.2.201] [Medline: 16461864]
17. Roll JM, Huber A, Sodano R, Chudzynski JE, Moynier E, Shoptaw S. A comparison of five reinforcement schedules for use in contingency management-based treatment of methamphetamine abuse. Psychol Rec 2017 May 27;56(1):67-81. [doi: 10.1007/bf03395538]
18. Packer RR, Howell DN, McPherson S, Roll JM. Investigating reinforcer magnitude and reinforcer delay: a contingency management analog study. Exp Clin Psychopharmacol 2012 Aug;20(4):287-292. [doi: 10.1037/a0027802] [Medline: 22686494]
19. Romanowich P, Lamb RJ. The effect of framing incentives as either losses or gains with contingency management for smoking cessation. Addict Behav 2013 Apr;38(4):2084-2088 [FREE Full text] [doi: 10.1016/j.addbeh.2013.01.007] [Medline: 23403276]
20. Romanowich P, Lamb RJ. The effects of fixed versus escalating reinforcement schedules on smoking abstinence. J Appl Behav Anal 2015;48(1):25-37 [FREE Full text] [doi: 10.1002/jaba.185] [Medline: 25640764]
21. Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA, et al. A randomized, controlled trial of financial incentives for smoking cessation. N Engl J Med 2009 Feb 12;360(7):699-709 [FREE Full text] [doi: 10.1056/NEJMsa0806819] [Medline: 19213683]
22. Roll JM, Higgins ST. A within-subject comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. Drug Alcohol Depend 2000 Feb 01;58(1-2):103-109. [doi: 10.1016/s0376-8716(99)00073-3] [Medline: 10669060]
23. Park J, Lee H, Park S, Chung K, Lee U. Goldentime: exploring system-driven timeboxing and micro-financial incentives for self-regulated phone use. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 May Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445489]
24. Fryer R, Levitt S, List J, Sadoff S. Enhancing the efficacy of teacher incentives through loss aversion: a field experiment. Am Econ J Econ Policy 2022;14(4):269-299. [doi: 10.1257/pol.20190287]
25. Pope L, Harvey-Berino J. Burn and earn: a randomized controlled trial incentivizing exercise during fall semester for college first-year students. Prev Med 2013 Mar;56(3-4):197-201. [doi: 10.1016/j.ypmed.2012.12.020] [Medline: 23295170]

26. Mortimer D, Wijnands JS, Harris A, Tapp A, Stevenson M. The effect of 'smart' financial incentives on driving behaviour of novice drivers. Accid Anal Prev 2018 Oct;119:68-79 [FREE Full text] [doi: 10.1016/j.aap.2018.06.014] [Medline: 30005270]
27. Cohen YS, Shmueli E. Money drives: can monetary incentives based on real-time monitoring improve driving behavior? Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Dec;1(4):1-22. [doi: 10.1145/3161417]
28. Can a fitness tracker save you money on health insurance? Yahoo Finance Home Page. URL: https://tinyurl.com/47x44ekz [accessed 2022-08-05]
29. Lee H, Kim A, Hong H, Lee U. Sticky goals: understanding goal commitments for behavioral changes in the wild. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 May Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445295]
30. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JP, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis. Prev Med 2015 Jun;75:75-85 [FREE Full text] [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]
31. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2020 Nov;54(21):1259-1268. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
32. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Apr Presented at: Persuasive 2009: Persuasive 2009; 4th International Conference on Persuasive Technology; Apr 26 - 29, 2009; Claremont California USA. [doi: 10.1145/1541948.1541999]
33. Lussier JP, Heil SH, Mongeon JA, Badger GJ, Higgins ST. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. Addiction 2006 Feb;101(2):192-203. [doi: 10.1111/j.1360-0443.2006.01311.x] [Medline: 16445548]
34. Burns RJ, Rothman AJ. Comparing types of financial incentives to promote walking: an experimental test. Appl Psychol Health Well Being 2018 Jul;10(2):193-214. [doi: 10.1111/aphw.12126] [Medline: 29671957]
35. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Heuer J, et al. Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. Ann Intern Med 2016 Mar 15;164(6):385-394 [FREE Full text] [doi: 10.7326/M15-1635] [Medline: 26881417]
36. Mahoney A. A comparison of escalating versus fixed reinforcement schedules on undergraduate quiz taking. J Behav Educ 2017 Apr 25;26(3):264-276. [doi: 10.1007/s10864-017-9268-5]
37. Bachireddy C, Joung A, John LK, Gino F, Tuckfield B, Foschini L, et al. Effect of different financial incentive structures on promoting physical activity among adults: a randomized clinical trial. JAMA Netw Open 2019 Aug 02;2(8):e199863 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9863] [Medline: 31441936]
38. Musthag M, Raij A, Ganesan D, Kumar S, Shiffman S. Exploring micro-incentive strategies for participant compensation in high-burden studies. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Sep Presented at: Ubicomp '11: The 2011 ACM Conference on Ubiquitous Computing; Sep 17 - 21, 2011; Beijing, China. [doi: 10.1145/2030112.2030170]
39. Shaw A, Horton J, Chen D. Designing incentives for inexpert human raters. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work. 2011 Mar Presented at: CSCW '11: Computer Supported Cooperative Work; Mar 19 - 23, 2011; Hangzhou, China. [doi: 10.1145/1958824.1958865]
40. Choi W, Park S, Kim D, Lim Y, Lee U. Multi-stage receptivity model for mobile just-in-time health intervention. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-26. [doi: 10.1145/3328910]
41. Luo Y, Lee B, Wohn D, Rebar A, Conroy D. Time for break: understanding information workers' sedentary behavior through a break prompting system. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Apr Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal, QC, Canada. [doi: 10.1016/j.cophys.2017.07.005]
42. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5-6):352-359 [FREE Full text] [doi: 10.1016/s1532-0464(03)00034-0] [Medline: 12968784]
43. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263-292. [doi: 10.2307/1914185]
44. Murray MP, Drought AB, Kory RC. Walking patterns of normal men. J Bone Joint Surg 1964;46(2):335-360. [doi: 10.2106/00004623-196446020-00009]
45. Murray MP, Kory RC, Sepic SB. Walking patterns of normal women. Arch Phys Med Rehabil 1970 Nov;51(11):637-650. [Medline: 5501933]
46. Mehrotra A, Pejovic V, Vermeulen J, Hendley R, Musolesi M. My phone and me: understanding people's receptivity to mobile notifications. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose, California, USA. [doi: 10.1145/2858036.2858566]
47. Delacre M, Lakens D, Leys C. Why psychologists should by default use Welch's t-test instead of student's t-test. Int Rev Social Psychol 2017;30(1):92-101. [doi: 10.5334/irsp.82]

48. McSweeney FK, Swindell S. General-process theories of motivation revisited: the role of habituation. Psychol Bull 1999 Jul;125(4):437-457. [doi: 10.1037/0033-2909.125.4.437]
49. McPherson SM, Burduli E, Smith CL, Herron J, Oluwoye O, Hirchak K, et al. A review of contingency management for the treatment of substance-use disorders: adaptation for underserved populations, use of experimental technologies, and personalized optimization strategies. Subst Abuse Rehabil 2018;9:43-57 [FREE Full text] [doi: 10.2147/SAR.S138439] [Medline: 30147392]
50. Wall J, Mhurchu CN, Blakely T, Rodgers A, Wilton J. Effectiveness of monetary incentives in modifying dietary behavior:a review of randomized, controlled trials. Nutr Rev 2006 Dec;64(12):518-531. [doi: 10.1111/j.1753-4887.2006.tb00185.x] [Medline: 17274494]
51. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
52. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
53. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 07;19(1):1059 [FREE Full text] [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
54. Barte JC, Wendel-Vos GC. A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. Behav Med 2017;43(2):79-90. [doi: 10.1080/08964289.2015.1074880] [Medline: 26431076]
55. Pulford B, Colman A, Loomes G. Incentive magnitude effects in experimental games: bigger is not necessarily better. Games 2018 Jan 18;9(1):4. [doi: 10.3390/g9010004]
56. Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M. Accuracy of sedentary behavior-triggered ecological momentary assessment for collecting contextual information: development and feasibility study. JMIR Mhealth Uhealth 2020 Sep 15;8(9):e17852 [FREE Full text] [doi: 10.2196/17852] [Medline: 32930668]
57. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
58. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
59. Lewis BA, Williams DM, Frayeh A, Marcus BH. Self-efficacy versus perceived enjoyment as predictors of physical activity behaviour. Psychol Health 2016;31(4):456-469 [FREE Full text] [doi: 10.1080/08870446.2015.1111372] [Medline: 26541890]
60. Rhodes RE, Kates A. Can the affective response to exercise predict future motives and physical activity behavior? A systematic review of published evidence. Ann Behav Med 2015 Oct;49(5):715-731. [doi: 10.1007/s12160-015-9704-5] [Medline: 25921307]
61. Burns RJ, Donovan AS, Ackermann RT, Finch EA, Rothman AJ, Jeffery RW. A theoretically grounded systematic review of material incentives for weight loss: implications for interventions. Ann Behav Med 2012 Dec;44(3):375-388. [doi: 10.1007/s12160-012-9403-4] [Medline: 22907712]
62. Gardner B, Flint S, Rebar AL, Dewitt S, Quail SK, Whall H, et al. Is sitting invisible? Exploring how people mentally represent sitting. Int J Behav Nutr Phys Act 2019 Oct 12;16(1):85 [FREE Full text] [doi: 10.1186/s12966-019-0851-0] [Medline: 31606040]
63. Conroy DE, Maher JP, Elavsky S, Hyde AL, Doerksen SE. Sedentary behavior as a daily process regulated by habits and intentions. Health Psychol 2013 Nov;32(11):1149-1157 [FREE Full text] [doi: 10.1037/a0031629] [Medline: 23477579]
64. Barone Gibbs B, King WC, Davis KK, Rickman AD, Rogers RJ, Wahed A, et al. Objective vs. self-report sedentary behavior in overweight and obese young adults. J Phys Act Health 2015 Dec;12(12):1551-1557 [FREE Full text] [doi: 10.1123/jpah.2014-0278] [Medline: 25710325]
65. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
66. Battalio SL, Conroy DE, Dempsey W, Liao P, Menictas M, Murphy S, et al. Sense2Stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. Contemp Clin Trials 2021 Oct;109:106534 [FREE Full text] [doi: 10.1016/j.cct.2021.106534] [Medline: 34375749]

Abbreviations

GLMM: generalized linear mixed model
L-AMCM: loss-framed adaptive microcontingency management

LR: logistic regression
OR: odds ratio

Edited by L Buis; submitted 07.08.22; peer-reviewed by EK Choe, M Giurgiu; comments to author 06.10.22; revised version received 30.11.22; accepted 21.12.22; published 27.01.23.

Please cite as:
Choi W, Lee U
Loss-Framed Adaptive Microcontingency Management for Preventing Prolonged Sedentariness: Development and Feasibility Study
JMIR Mhealth Uhealth 2023;11:e41660
URL: https://mhealth.jmir.org/2023/1/e41660
doi:10.2196/41660
PMID:36705949

Original Paper

Promoting Hand Hygiene During the COVID-19 Pandemic: Parallel Randomized Trial for the Optimization of the Soapp App

Dario Baretta[1], PhD; Melanie Alexandra Amrein[1], PhD; Carole Bäder[1], MSc; Gian Giacomo Ruschetti[1], MSc; Carole Rüttimann[1], BSc; Maria Del Rio Carral[2], PhD; Carlo Fabian[3], MA; Jennifer Inauen[1], PhD

[1]Institute of Psychology, University of Bern, Bern, Switzerland

[2]Institute of Psychology, University of Lausanne, Lausanne, Switzerland

[3]Institute for Social Work and Health, FHNW School of Social Work, Olten, Switzerland

Corresponding Author:
Dario Baretta, PhD
Institute of Psychology
University of Bern
Fabrikstrasse 8
Bern, 3012
Switzerland
Phone: 41 31 684 58 96
Email: dario.baretta@unibe.ch

Abstract

Background: Hand hygiene is an effective behavior for preventing the spread of the respiratory disease COVID-19 and was included in public health guidelines worldwide. Behavior change interventions addressing hand hygiene have the potential to support the adherence to public health recommendations and, thereby, prevent the spread of COVID-19. However, randomized trials are largely absent during a pandemic; therefore, there is little knowledge about the most effective strategies to promote hand hygiene during an ongoing pandemic. This study addresses this gap by presenting the results of the optimization phase of a Multiphase Optimization Strategy of *Soapp*, a smartphone app for promoting hand hygiene in the context of the COVID-19 pandemic.

Objective: This study aimed to identify the most effective combination and sequence of 3 theory- and evidence-based intervention modules (habit, motivation, and social norms) for promoting hand hygiene. To this end, 9 versions of *Soapp* were developed (conditions), and 2 optimization criteria were defined: the condition with the largest increase in hand hygiene at follow-up and condition with the highest engagement, usability, and satisfaction based on quantitative and qualitative analyses.

Methods: This study was a parallel randomized trial with 9 intervention conditions defined by the combination of 2 intervention modules and their sequence. The trial was conducted from March to August 2021 with interested participants from the Swiss general population (N=232; randomized). Randomization was performed using Qualtrics (Qualtrics International Inc), and blinding was ensured. The duration of the intervention was 34 days. The primary outcome was self-reported hand hygiene at follow-up, which was assessed using an electronic diary. The secondary outcomes were user engagement, usability, and satisfaction assessed at follow-up. Nine participants were further invited to participate in semistructured exit interviews. A set of ANOVAs was performed to test the main hypotheses, whereas a thematic analysis was performed to analyze the qualitative data.

Results: The results showed a significant increase in hand hygiene over time across all conditions. There was no interaction effect between time and intervention condition. Similarly, no between-group differences in engagement, usability, and satisfaction emerged. Seven themes (eg, “variety and timeliness of the task load” and “social interaction”) were found in the thematic analysis.

Conclusions: The effectiveness of *Soapp* in promoting hand hygiene laid the foundation for the next evaluation phase of the app. More generally, the study supported the value of digital interventions in pandemic contexts. The findings showed no differential effect of intervention conditions involving different combinations and sequences of the habit, motivation, and social norms modules on hand hygiene, engagement, usability, and satisfaction. In the absence of quantitative differences, we relied on the results from the thematic analysis to select the best version of *Soapp* for the evaluation phase.

Trial Registration: ClinicalTrials.gov NCT04830761; https://clinicaltrials.gov/ct2/show/NCT04830761

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-055971

(JMIR Mhealth Uhealth 2023;11:e43241) doi:10.2196/43241

KEYWORDS

COVID-19; hand hygiene; behavior change intervention; Multiphase Optimization Strategy; MOST; smartphone apps; motivation; habit; social norm; mobile phone

Introduction

Background

Hand hygiene is an effective behavior for decreasing the transmission of respiratory illnesses [1,2], including COVID-19 [3]. Therefore, recommendations to perform correct hand hygiene at key times have been included in public health guidelines worldwide to counter the spread of COVID-19 [4]. To facilitate the adoption of public health guidelines, the development and evaluation of effective behavior change interventions was identified as a priority of the COVID-19 research agenda, in particular, owing to the fact that limited or no contextualized evidence was available on the effectiveness of behavior change interventions during pandemics [5]. Although evidence synthesis reports became available during the COVID-19 pandemic (July to December 2020), showing a medium, positive effect of hand hygiene interventions developed to counter the spread of various respiratory viruses (eg, influenza virus, respiratory syncytial virus, and adenovirus) [6], their validity and relevance to the COVID-19 pandemic can be questioned. For example, the reviewed studies included interventions targeted at diverse respiratory infections that did not cause pandemics (eg, influenza, flu, and cold) or lead to the spread of a pandemic of the same magnitude as that of the COVID-19 pandemic (eg, pandemic influenza A H1N1).

The need for research on effective behavior change interventions for promoting hand hygiene during a pandemic was further confirmed by the fluctuation in hand hygiene over the course of the COVID-19 pandemic. At first, results indicated high adherence among the public. During the first wave of the pandemic (ie, between March and May 2020), studies suggested that (1) hand hygiene was one of the most adopted protective behaviors against the spread of COVID-19 [7], and (2) the frequency and correctness of hand hygiene behavior in key situations (ie, after coughing, sneezing, blowing one's nose and upon reaching home or workplace) improved compared with the period before the pandemic outbreak [8,9]. However, research including longer periods of the pandemic showed a decrease in hand hygiene over time. For example, a study conducted from May 2020 to August 2021 showed that almost one-third of the adults from the general population did not comply with hand hygiene recommendations, and some of them had no intention to change their behavior [10]. In addition, there is evidence of significant associations between hand hygiene and indicators of the pandemic trajectory (eg, the increase in recent cases of COVID-19 morbidity is associated with an increase in the frequency of self-reported hand hygiene) [11]. Taken together, the literature suggests that hand hygiene is not consistently performed throughout a pandemic and is prone to variations over time. Therefore, fostering sustained hand hygiene through effective behavior change interventions represented a public health priority to counter the spread of COVID-19 and future pandemics.

During an ongoing pandemic in which social contact should be limited, digital interventions have the advantage that no personal contact is required for their use; moreover, they can be personalized and potentially be integrated into the daily lives of an unlimited number of people. Interventions based on smartphone apps can deliver behavior change techniques [12] in real life that could lead to substantial population-level impact and long-term health behavior change [13]. However, recent reviews have pointed out that there is limited knowledge about how to effectively promote hand hygiene using digital interventions in the general population [6,14].

To address these research gaps, we devised a Multiphase Optimization Strategy (MOST) [15] to develop and test Soapp, an effective smartphone-based behavior change intervention for promoting hand hygiene during the ongoing COVID-19 pandemic [16]. In the preparation phase, we developed 3 intervention modules—tackling habit, motivation, and social norms—based on behavior change theory and empirical evidence [16].

Aim of This Study

This study focused on the optimization phase of Soapp, which aimed to identify the most effective combination and sequence of the developed intervention modules to be included in the subsequent evaluation phase. As described in the study protocol [16], during the optimization phase, we compared 9 different combinations of the 3 developed modules (ie, habit, motivation, and social norms). Overall, 2 optimization criteria were defined to select the best intervention version for the subsequent evaluation phase. The optimization criteria were as follows: (1) the condition with the largest increase in hand hygiene at key times at follow-up (T3) and (2) the condition with the highest engagement, usability, and satisfaction. Regarding the first criterion, we tested the following preregistered hypotheses [16]:

Hypothesis (H) 1; H1: The intervention groups show a significant increase in correct hand hygiene at key times after 4 weeks (T3) of intervention compared with baseline (T1).

H2: The intervention groups significantly differ in the effects of the intervention on correct hand hygiene behavior at key times (T1-T3).

In case of significant between-group differences in hand hygiene at key times, post hoc tests were performed to determine the most effective condition. In addition, we investigated the unique contribution of each module by testing the following hypotheses that were not preregistered:

H3: The intervention groups with the *habit* module show a significant increase in correct hand hygiene behavior at key times (T1-T3) compared with the groups without the *habit* module.

H4: The intervention groups with the *motivation* module show a significant increase in correct hand hygiene behavior at key times (T1-T3) compared with the groups without the *motivation* module.

H5: The intervention groups with the *social* module show a significant increase in correct hand hygiene behavior at key times (T1-T3) compared with the groups without the *social* module.

The second optimization criterion leveraged a combination of quantitative and qualitative methods to explore the participants' engagement and satisfaction with the app as well as its usability. This criterion was tested using the following hypotheses:

The intervention groups significantly differ in the engagement with (H6), usability of (H7), and satisfaction (H8) with the intervention (T3).

In addition, semistructured interviews were conducted to explore which aspects and features of Soapp were perceived as more usable and more important for supporting engagement and satisfactory experiences after 34 days of using the app.

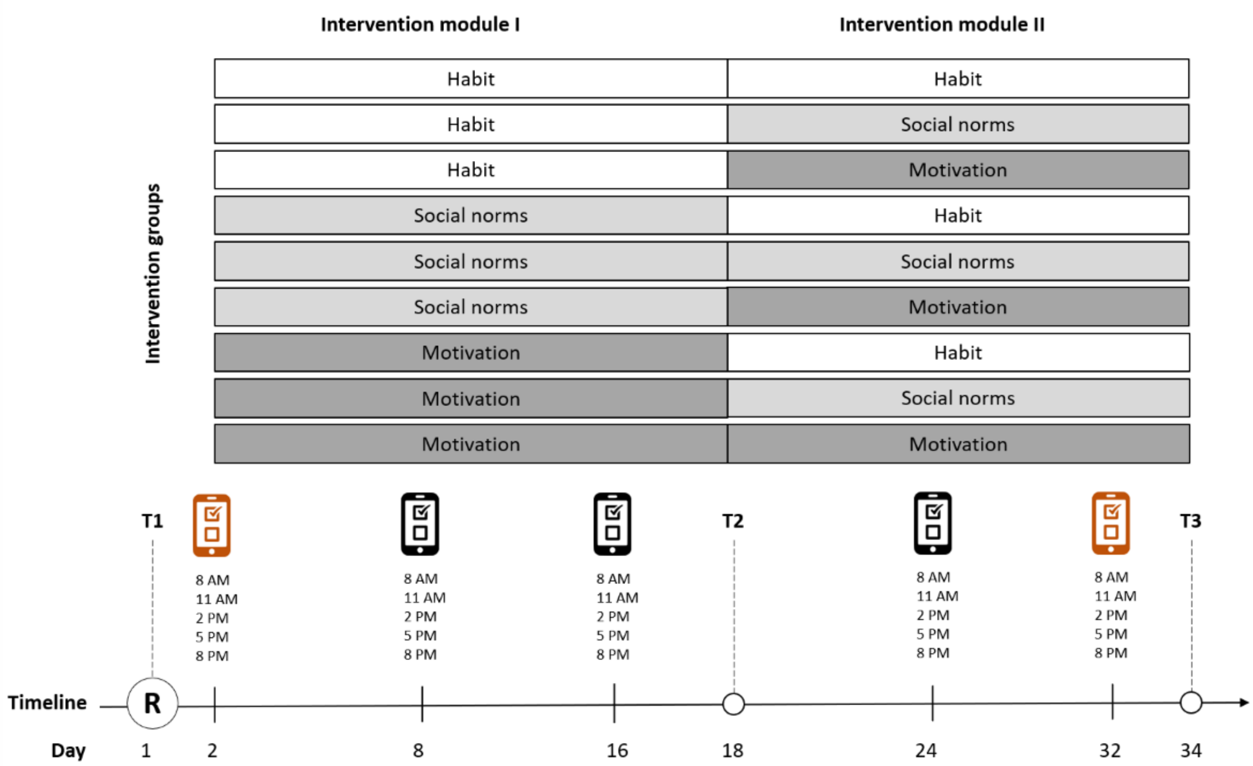
As secondary outcomes, we had preregistered a series of hypotheses regarding the psychological mechanisms and health impact of the intervention that did not inform the optimization decision. We have reported these in Multimedia Appendix 1 for completion.

Methods

Study Design

The study design for the optimization phase was a double-blind parallel randomized trial. The participants were randomized to 1 of 9 intervention groups in a 1:1:1:1:1:1:1:1:1 ratio and completed 2 consecutive intervention modules, as shown in Figure 1. The total duration of the optimization study (ie, recruitment and data collection) was set to 6 months (start: March 26, 2021) or until a total of 465 participants were enrolled, whichever came first. The duration of the optimization trial for each participant (ie, time between T1 and T3) was 34 days. At the end of the study, as a part of the second optimization criterion, qualitative interviews were conducted with a subsample to collect in-depth information about the engagement with, usability of, and satisfaction with the intervention.

Figure 1. Intervention optimization. Red diaries represent baseline (T1) and follow-up (T3) assessments for the primary outcome (hand hygiene). R: randomization.



Ethics Approval

This trial is registered at ClinicalTrials.gov (NCT04830761), and the reporting is in line with the CONSORT (Consolidated Standards of Reporting Trials) guidelines [17] (Multimedia Appendix 2). The trial received ethics approval from the Swiss Ethics Committee of the Canton of Bern (ID: 2021-00164).

Participants

The target population for the Soapp app was German-speaking adults from the Swiss population who were interested in using an app to improve hand hygiene behavior. The inclusion criteria were as follows: (1) being aged at least 18 years, (2) owning a smartphone with mobile access to the internet, (3) being proficient in the German language, and (4) having signed an electronic informed consent form to participate in the study. As presented in the study protocol [16], the initial target sample size for the optimization phase was 387 participants. The sample size was calculated to perform a repeated-measures ANOVA with a within (time: T1-T3)-between (intervention group) interaction. The sample size was determined using an a priori power analysis with G*Power (Heinrich-Heine-Universität Düsseldorf) [18] (β=.80; α=.05; F_8=0.1). Assuming a 20% attrition during the course of the intervention, the target sample size for the study was raised to 465 participants. However, owing to both trial and project timelines, we stopped recruiting after 5 months for a total study duration of 6 months.

A subsample (n=9) participated in qualitative interviews. The recruitment was based on the participants' willingness to participate in semistructured interviews, as assessed at the end of the last survey (T3). The aim of the qualitative interview was to recruit an even number of participants per intervention module according to hand hygiene adherence: low adherence, medium adherence, and high adherence to hand hygiene. The participants in the ≤33rd percentile were assigned to the low adherence group (3/9, 33%), participants in between the 34th and 66th percentiles were assigned to the medium adherence group (3/9, 33%), and participants in the ≥67th percentile were assigned to the high adherence group (3/9, 33%). The sample size (n=9) was smaller than that reported in the study protocol (n=15) because the recruitment was stopped when theoretical saturation was achieved (ie, no new themes emerged) [19].

Outcomes

Primary Outcome

The primary outcome of the study (ie, the first optimization criterion), the frequency of correct hand hygiene at key times at T3, was assessed via ecological momentary assessment with the support of an electronic diary embedded in Soapp. On diary days (days 2, 8, 16, 24, and 32), the participants were prompted 5 times per day to indicate whether each of the 13 key times to perform hand hygiene defined by the Swiss Federal Office of Public Health had occurred (eg, upon arriving home and after using the toilet; Multimedia Appendix 1). For each situation that occurred, the participants were asked how often they correctly washed or disinfected their hands in that specific situation. The 5-point response scale ranged from never (1) to always (5). The main outcome was operationalized as the mean reported frequency of correct hand hygiene across all the indicated key times and ranged from 1 to 5. To test the H1 and H2, the assessment points considered for hand hygiene behavior were the first diary filled out on day 2 (T1) and the last diary filled out on day 32 (T3).

Secondary Outcomes

Engagement, usability, and satisfaction (ie, the second optimization criterion) were measured at T3. User engagement was assessed using the digital behavior change interventions (DBCI) Engagement Scale [20], a 7-point Likert scale ranging from *not at all* (1) to *extremely* (7; Cronbach α=.78). Intervention usability was assessed using the System Usability Scale [21], a 6-point Likert scale ranging from *I do not agree at all* (1) to *I agree completely* (6; Cronbach α=.80). Satisfaction was assessed using the Fragebogen zur Messung der Patientenzufriedenheit (ZUF)-8 [22], a 4-point Likert scale ranging from 0 to 3 (Cronbach α=.89).

Other variables assessed during the study but not relevant to the current report are described in the clinical trial registration, and the corresponding results are presented in Multimedia Appendix 1.

Procedure

The participants were recruited via social media (eg, Facebook [Meta Platforms, Inc] and Instagram [Meta Platforms, Inc]), mailing lists, and leaflets with the help of a market research company and with the aim of recruiting a diverse range of people from the German-speaking adult Swiss general population in terms of gender, age, and socioeconomic status. Interested people who clicked on the campaign link were led to a landing page with the study information. Those who chose to continue were redirected to the study page on REDCap (Research Electronic Data Capture; Vanderbilt University) [23] where they could read and watch a video of the study information, fill out an eligibility and consent survey, and sign the e-consent form. After providing electronic informed consent to participate in the study, the participants received a registration code via email and were guided to download the Soapp app from iTunes (Apple Inc) or Google Play Store (Google LLC) and register on it. The day after the registration, the participants received the T1 questionnaire and were then randomized into one of the intervention groups. Randomization was implemented in Qualtrics (Qualtrics International Inc), which preserved the allocation concealment. In addition, the researchers involved in the study were blinded to the intervention assignment, as the participant identifier was pseudoanonymized before randomization. The day after the randomization, the participants filled out the first-hand hygiene diary. The diary included five 1-minute questionnaires per day to avoid retrospective bias in reporting hand hygiene [24].

The optimization trial lasted 34 days and included two 2-week intervention modules (Figure 1). During the first module, the participants filled out the hand hygiene diary on days 2, 8, and 16. After the first module, the participants received a second questionnaire (T2) and a second intervention module, which followed the same structure as the first. After completing the T2, the participants were offered a small gift (ie, a bar of hand soap and a thank you card) to prevent attrition, which was sent to their homes. During the second intervention module, they filled out the hand hygiene diaries on days 24 and 32. At the end of the second module, the participants received the final questionnaire (T3). The participants were given the chance to win 1 of 3 iPhones (Apple Inc) 12s after both the optimization and evaluation phases of the study were completed. The questionnaires and diaries were integrated into Qualtrics services using Soapp's application programming interface, and the participants' data were stored on Qualtrics.

The participants who were given the option and volunteered to participate in the qualitative study were interviewed via telephone by a study team member (CB). This 30-minute interview was recorded and included questions about the usability of the app and the overall experience with the intervention modules in terms of satisfaction and engagement (Multimedia Appendix 3).

Intervention

In the optimization phase, each arm of the parallel randomized trial was characterized by a unique combination and sequence of 2 of the 3 intervention modules: motivation, habit, and social norms (Figure 1). The modules were defined as the outcome of the preparation phase in which a theory- and evidence-based approach was followed. The resulting content is synthesized in Table 1 and detailed in supplemental material 1 of the protocol paper [16].

Table 1. Contents of the modules.

Module, TDF[a] domain, and behavioral predictor	Behavior change technique
Basic	
Goals	
Intention	• 1.1 Goal setting (behavior)
Skills	
Skills	• 4.1 Instruction on how to perform behavior
Knowledge	
Knowledge	• 5.1 Information about health consequences
Environmental context and resources	
Resources and material resources (availability and management)	• 1.4 Action planning
Motivation	
Goals	
Intention	• 1.1 Goal setting (behavior)
Beliefs about consequences	
Risk perception	• 5.1 Information about health consequences
Attitude	• 5.2 Salience of consequences
Outcome expectancies	• 9.2 Pros and cons
Intention	• 5.2 Salience of consequences
Beliefs about capabilities	
Self-efficacy	• 1.2 Problem solving • 15.1 Verbal persuasion about capabilities • 15.3 Focus on past success
Reinforcement	
Intention	• 10.9 Self-reward
Habit	
Knowledge	
Knowledge	• 4.2 Information about antecedents
Memory, attention, and decision processes	
Action control	• 2.3 Self-monitoring of behavior
Goals	
Action planning	• 1.4 Action planning • 7.1 Prompts and cues
Skills and goals	
Habit strength	• 8.1 Behavioral practice and rehearsal • 8.3 Habit formation
Behavioral regulation	
Habit strength	• 7.1 Prompts and cues (physical cue)
Social norms	

Module, TDF[a] domain, and behavioral predictor	Behavior change technique
Social influences	
Descriptive norm	• 2.1 Monitoring of behavior by others without feedback • 2.2 Feedback on behavior • 6.2 Social comparison • 10.4 Social reward • 10.5 Social incentive
Injunctive norm	• 5.1 Information about health consequences • 6.3 Information about others' approval • 9.1 Credible source • 10.5 Social incentive • 12.1 Restructuring the physical environment

[a]TDF: Theoretical Domain Framework.

The modules were delivered to the participants via their personal smartphones through the study app Soapp. They were comparable in terms of user time and extent of content, and each module took 2 weeks to be completed. In addition, each intervention condition included a basic module that provided information on hand hygiene to all the participants. During the configuration process, the Soapp app underwent various iterative testing cycles to refine the content of each module and improve usability. The Soapp app contained all the information required to use it, and there was no direct contact with the study team.

Data Analysis

Handling of Missing Data

Missing data were handled according to the intention-to-treat (ITT) principle [25]. The ITT analysis included all randomized participants. It ignores noncompliance, protocol deviations from the intervention modules, and anything that occurs after randomization. ITT analysis avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of noncompliers by accepting that noncompliance and protocol deviations that are likely to occur in practice. In line with previous research [26], missing data for hand hygiene behavior were replaced using the last observation carried forward approach.

Hypothesis Testing

To test the hypotheses related to the first optimization criterion, a repeated-measures ANOVA with a within-between interaction was used. The *within* effect was represented by the difference in hand hygiene between T1 and T3 (H1), whereas the *within-between* interaction was represented by the change in correct hand hygiene behavior between T1 and T3 across all 9 intervention groups (H2). If the groups differed significantly, post hoc tests were performed to identify the most effective intervention group. To test hypotheses H3, H4, and H5, 3 dummy variables were created: *habit exposure*, *motivation exposure*, and *social exposure.* These variables indicated whether a participant was exposed to the corresponding module during the intervention. Then, 3 repeated-measures ANOVAs, 1 for each dummy variable, with a within-between interaction were performed. Each ANOVA tested the interaction between time and the exposure to a specific module (Table 2). Finally, for the second optimization criterion, three 1-way ANOVAs were performed to test differences across conditions at T3 in terms of engagement (H6), usability (H7), and satisfaction (H8). For all the hypotheses, a set of sensitivity analyses with robust and nonparametric (ie, Kruskal-Wallis test) ANOVAs was performed to account for potential unequal sample sizes and nonnormal distribution of the data.

Table 2. Summary of hypothesis tests.

H[a]	Preregistered	Dependent variable	Within factor (time)	Between factor
First optimization criterion based on the primary outcome				
H1	Yes	Hand hygiene	T1-T3	N/A[b]
H2	Yes	Hand hygiene	T1-T3	Intervention groups
H3	No	Hand hygiene	T1-T3	Habit exposure
H4	No	Hand hygiene	T1-T3	Motivation exposure
H5	No	Hand hygiene	T1-T3	Social exposure
Second optimization criterion based on the secondary outcomes				
H6	No	Engagement	N/A	Intervention groups
H7	No	Usability	N/A	Intervention groups
H8	No	Satisfaction	N/A	Intervention groups

[a]H: hypothesis.

[b]N/A: not applicable.

Analytical Software

The packages *ez* and *WRS2* from the statistical software R (version 4.1.2; R Foundation for Statistical Computing) were used to perform parametric and robust ANOVAs, respectively. The data and R code used for the main analyses are available on the Open Science Framework repository platform [27].

Qualitative Analysis

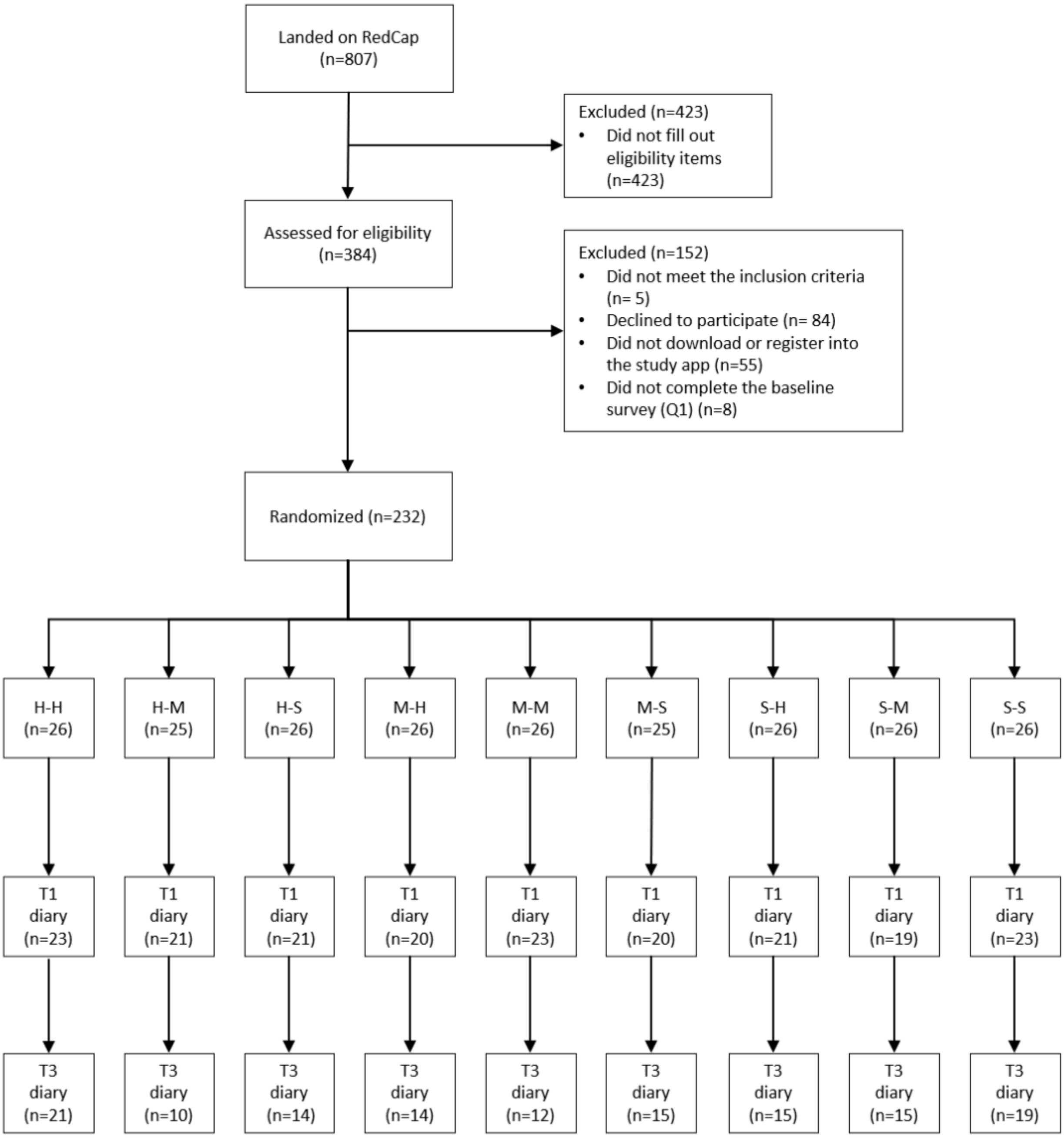
Postintervention user engagement, usability, and satisfaction were explored using semistructured interviews. The interviews were transcribed verbatim, and the transcripts were analyzed using thematic analysis [28]. Thematic analysis is characterized by six phases: (1) familiarizing oneself with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. Data and repeated patterns that were considered pertinent to the aims of the study were coded by a coauthor (CR). New inductive codes were labeled as they were identified during the coding process, and the results of the coding process were iteratively discussed by 2 coauthors (CR and JI). The next stage involved searching for themes; CR reviewed the codes one by one, organizing the findings to combine different codes that focus on similar aspects. The ordered data were reviewed and revised in discussion among 3 coauthors (CR, JI, and DB) and were subsequently organized into themes. The resolution of disagreements and agreement on the final themes was achieved through discussion among CR, JI, and DB. After defining and naming the themes, examples of relevant transcripts were selected to illustrate them. The data were analyzed in their original language to preserve their original meanings. Illustrative quotes were translated by CR. Conclusions were drawn on the possible improvement of Soapp to optimize the effectiveness and usability of the intervention for the evaluation phase.

Results

Overview

The recruitment for the optimization trial began on March 27, 2021, and ended on July 28, 2021. T3 data were collected between April 29 and August 25, 2021. Because of both trial and project timelines, we stopped the trial 6 months after the start of the study, with the recruitment lasting for 5 months. Overall, 232 participants were recruited and randomized into 1 of the 9 intervention conditions. Among these 232 participants, 14 (6%) participants did not fill out any of the 5 hand hygiene diaries, whereas 27 (11.6%) participants did not complete the first diary at T1. Another (1/232, 0.4%) participant completed the first diary but did not encounter any of the key situations to perform hand hygiene during that day. Therefore, these (42/232, 18.1%) participants were excluded from the analysis because the main outcome (ie, hand hygiene) at T1 was missing. Of the 232 participants who were randomized, 190 (81.9%) filled out the hand hygiene diary at T1, and 118 (62.1%; 50.9% of the randomized participants) of them filled out the hand hygiene diary at T3. Figure 2 shows the participants' flow through randomization, T1 diary assessment, and T3 diary assessment for each intervention group. For the secondary analysis, we included only those participants who filled out the T3 panel assessment because the dependent variables were assessed only at T3.

Figure 2. Participant recruitment flow. The intervention groups are specified as follows: H-H: habit-habit; H-M: habit-motivation; H-S: habit-social; M-H: motivation-habit; M-M: motivation-motivation; M-S: motivation-social; S-H: social-habit; S-M: social-motivation; and S-S: social-social. T1: baseline: T3: follow-up.



T1 Characteristics

Sociodemographic and hand hygiene behaviors at T1 are reported in Multimedia Appendix 4. The figures refer to the 190 participants who filled out the first diary at T1. The mean age of the participants was 39.9 (SD 15.9) years. Of the 190 participants, 139 (73.2%) were women, 125 (65.8%) had high school qualifications, 101 (53.2%) were employed, and 49 (25.8%) were living alone. Descriptive statistics for hand hygiene behavior (mean 4.01, SD 0.82; median=4.17; skewness=−1.24) suggested that hand hygiene behavior was already high at T1, with a moderate left-tailed distribution.

Dropout analysis was performed to investigate T1 differences between the participants who completed the study and those who dropped out at any point during the intervention. We analyzed all the 232 randomized participants, and those who did not complete the last panel assessment at T3 were categorized as dropouts (n=83, 36%). The results suggested no T1 differences between dropouts and retainers with respect to age ($F_{1,230}$=2.17; P=.14), sex (χ^2_1=0.4; P=.55), hand hygiene ($F_{1,229}$=0.24; P=.63), or intention to increase hand hygiene behavior ($F_{1,230}$=0.72; P=.40).

First Optimization Criterion: Change in Hand Hygiene Behavior

The main effects of time and the interaction between time and the intervention groups are reported in Table 3. The main effect of time ($F_{1,181}$=10.95; P=.001) was statistically significant (H1) whereas the interaction between the intervention groups and time was not (H2). The results pertaining to H3, H4, and H5 suggested no effect of the exposure to a specific module during the course of the intervention. Sensitivity analysis using a robust approach confirmed the same results (Multimedia Appendix 4). In addition, as a part of a further sensitivity analysis, the main hypotheses were tested without applying any missing value imputation algorithm. The results are available in Multimedia Appendix 4 and confirm the time effect and the null findings for the interaction effect.

Table 3. Main effects and interactions among modules on hand hygiene behavior at key times (N=232).

H[a], outcome, and factor	Participants, n (%)	Parametric ANOVA		
		F test (*df*)	*P* value	Partial eta–squared[b] (95% CI)
H1 and H2	190 (81.9)			
Hand hygiene				
Group[c]		0.33 (8)	.95	0.01 (0.00-1.00)
Time (T1-T3)		*10.95* [d] *(1)*	*.001*	*0.06 (0.01-1.00)*
Time × group		1.19 (8)	.31	0.05 (0.00-1.00)
H3	190 (81.9)			
Hand hygiene				
Habit		1.25 (1)	.27	0.01 (0.00-1.00)
Time (T1-T3)		*10.87 (1)*	*.001*	*0.05 (0.01-1.00)*
Time × habit		1.07 (1)	.30	0.01 (0.00-1.00)
H4	190 (81.9)			
Hand hygiene				
Motivation		0.00 (1)	.99	0.00 (0.00-1.00)
Time (T1-T3)		*10.86 (1)*	*.001*	*0.05 (0.01-1.00)*
Time × motivation		0.94 (1)	.33	0.00 (0.00-1.00)
H5	190 (81.9)			
Hand hygiene				
Social		0.75 (1)	.39	0.00 (0.00-1.00)
Time (T1-T3)		*10.83 (1)*	*.001*	*0.05 (0.01-1.00)*
Time × social		0.41 (1)	.52	0.00 (0.00-1.00)
H6	148 (63.8)			
Engagement				
Group (T3)		*2.19 (8)*	*.03*	*0.11 (0.01-1.00)*
H7	148 (63.8)			
Usability				
Group (T3)		*2.46 (8)*	*.02*	*0.12 (0.01-1.00)*
H8	148 (63.8)			
Satisfaction				
Group (T3)		1.46 (8)	.18	0.11 (0.00-1.00)

[a]H: hypothesis.

[b]Partial eta–squared corresponds to the proportion of variance explained by a variable that is not explained by other variables.

[c]Group: intervention group.

[d]Italicized values indicate significance.

Second Optimization Criterion: Participant Engagement, Usability, and Satisfaction

Quantitative Analysis

The effects of the intervention group on engagement, usability, and satisfaction are shown in Table 3. The results of the parametric ANOVA suggested that the self-reported measures of engagement ($F_{8,139}$=2.19; *P*=.03) and usability ($F_{8,139}$=2.46; *P*=.02) differed across the 9 intervention groups. Nonparametric ANOVA with Kruskal-Wallis test showed significant differences across the intervention groups only for usability (χ^2_8=16.1; *P*=.04). However, both parametric and nonparametric post hoc comparisons with Bonferroni adjustment indicated no mean score differences in engagement and usability between any pair of intervention groups.

Qualitative Analysis

Overview

Across 9 interviews, 7 themes emerged in relation to the research question (refer to Multimedia Appendix 5 for a summary of the themes and for additional extracts illustrating each theme). The themes were named "user experience and app functionality," "importance of guidance," "variety and timeliness of the task load," "reasons for participation," "change in awareness of hand hygiene and its implications," "social interaction," and "personal relevance." In addition, the following 2 subthemes were identified as a part of the "social interaction" theme: "personal communication and connectedness" and "social comparison."

User Experience and App Functionality

The first theme that emerged concerned the user experience with the general aesthetics and functionalities of the app. Overall, satisfaction with the intuitive and simple handling of the app was high. The participants considered the usability to be pleasant. Regarding the app aesthetics, some participants were very satisfied with the simplicity of the layout; however, the majority would have preferred more visual structures:

> *What I liked in particular? Actually, how things were presented. Just the simplicity—all in all it was very simple.* [P7, habit-habit, moderate Adherence]

Another point on which most participants agreed was that certain features of the app showed technical flaws, which negatively affected their motivation:

> *So, when this annoying technical problem occurred—if you were to draw a curve now, it [my motivation] went up quite steadily at the beginning, and then slowly decreased due to this technical problem, and then when it was resolved it [my motivation] got back up again.* [P5, motivation-social, high adherence]

Importance of Guidance

Throughout the interviews, the participants regularly highlighted the importance of receiving guidance within the app. Specifically, they mentioned the importance of clarity and meaning regarding the tasks that the app asked them to do:

> *I also thought it was nice that you kind of knew in the morning "ah today is a day with a big survey," so that you could already plan "okay, today there are maybe a little bit more push messages coming in and I have to pay a little bit more attention."* [P2, habit-habit, high adherence]

The importance of guidance was also manifested as the need for a better overview of the participants' journey during the study. For instance, some participants would have liked more background information about the study to better understand the timeline or the reasons behind receiving certain tasks:

> *And otherwise maybe somehow a little bit more background information about what—why am I being asked these questions, so that I can see even more behind this algorithm and behind this concept and then it would become clearer to me why the same questions keep coming. So, a little bit, so even more background knowledge.* [P3, social-habit, low adherence]

Although guidance was acknowledged as important, too much direction was also perceived as overwhelming, for example, very frequent push notifications:

> *Was that now at 10 o'clock, at 12 o'clock or at 2 o'clock, I do not remember any more in which intervals the push messages came in. At the end, I no longer knew at what point I had I received the last push notification—there, I lost overview.* [P2, habit-habit, high adherence]

Variety and Timeliness of the Task Load

Variety in daily engagement with the content of the app emerged as a central topic in the interviews. A few participants were satisfied with the degree of variety in the task load and the timing of the content offered by the intervention. However, most participants wished for substantially more variety in the task load and timing, particularly toward the end of the intervention:

> *Sometimes, it was just quiet, nothing happened. But later, once again it came "today something is happening," yes, I liked that.* [P2, habit-habit, high adherence]

> *Towards the end, when there were fewer and fewer exercises, I found it almost a bit boring.* [P7, habit-habit, moderate adherence]

Reasons for Participation

In most interviews, the participants mentioned their initial reasons for participating in the study. One of the most frequently reported motives was curiosity and an interest in learning something new:

> *I thought, "yeah, sure, I can wash my hands. But do I know everything when they do a study? I could still learn something at the end, I'm not omniscient." And that's actually what mainly motivated me, this openness, I'm curious to see what else there is to learn.* [P1, habit-habit, moderate adherence]

The participants also elaborated on why they kept using the app. One of the cited motives was the perceived obligation to complete the study:

> *Well, it [my motivation] certainly did not increase, it was more a matter of persevering—in the sense of whoever says A must also say B. It was said that you could drop out at any time, but still.* [P9, motivation-motivation, low adherence]

Change in Awareness of Hand Hygiene and Its Implications

A further theme was represented by the increase in participants' hand hygiene awareness owing to the use of the app. The change in awareness seemed to have been generated by the fact that participants paid more attention to the self-monitoring of the target behavior:

> *That was simply my observation of my reaction then—you observe yourself during these four weeks incredibly—I do not know if you have also heard this from other people, but you start watching yourself.* [P5, motivation-social, high adherence]

The change in awareness generated a positive loop that led to an increase in the frequency of hand hygiene behavior together with a shift in perception of the issue of hand hygiene and its implications:

> *I certainly washed my hands more than I had before. And therefore, I have the feeling that I have certainly benefitted from it [the intervention].* [P4, habit-motivation, high adherence]

Social Interaction

The theme of social interaction came up several times during most interviews. Two subthemes define this main theme according to the different social aspects that came to light during the interviews: personal communication and connectedness and social comparison.

Subtheme 1: Personal Communication and Connectedness

Some participants particularly appreciated that the app communicated with them in a personal manner. This led to a feeling of authenticity; therefore, these participants no longer had the impression of interacting with a machine when using the app:

> *You can say that there is someone behind it. I never felt alone, it was not a one-way kind of communication. I always knew that behind these tasks was indeed a computer, but I still felt connected in a way.* [P2, habit-habit, high adherence]

By contrast, other participants would have preferred an even more human-centered mode of delivery of the app content, for example, receiving direct motivational support from other humans:

> *Maybe, despite everything, a video or something like that—or actually, as is often the case nowadays: a small video with other participants who motivate you. Because reading statistics and news is something else than when someone speaks directly to you.* [P8, social-habit, low adherence]

Some participants described having developed a feeling of connectedness with other app users over time. This led to a sense of community, which made them feel supported:

> *And then I think I had to answer this question three times. And at the end, I think that was at the final question, I thought "yes, I think it is cool that they are taking part, I do not know them, but I think it is cool that they are taking part, and I feel connected to them."* [P2, habit-habit, high adherence]

Subtheme 2: Social Comparison

The participants who were exposed to the social module shared different opinions regarding the opportunity to compare their behavior with that of other participants, which was a feature of the social module only. Indeed, although some participants expressed avoidance of social comparison and fatigue with the competition it created for them, others were pleased about the comparison with other users:

> *For me, personally, it was too much with the community and otherwise, because others cannot motivate me. Whether someone somehow achieved 100% or 50%, that is actually relatively indifferent to me. And it does not encourage me to become more or less active or whatever.* [P8, social-habit, low adherence]

Wish for Personalization

An issue raised by almost every participant was the lack of personal relevance that the list of key moments for hand hygiene entailed. Being regularly asked about key situations that never occurred for them (eg, not having children or not wearing contact lenses) led to a decrease in motivation to fill out the hand hygiene diaries:

> *Things are asked again and again, which do not concern you at all. This leads to a decrease in motivation. Now, I have to spend five minutes filling out the form again, even though it does not apply.* [P3, social-habit, low adherence]

The desire to personalize the app also came up in relation to other intervention content, such as the number of push notifications:

> *But maybe in the beginning you should be able to specify "I would rather have a little more [push notifications] or a little less." But what I have received, however, has been right for me.* [P2, habit-habit, high adherence]

Discussion

Principal Findings

As part of a MOST to develop and test a smartphone-based hand hygiene intervention during the COVID-19 pandemic, this intervention optimization parallel randomized trial aimed to identify the best combination of intervention modules to be included in the subsequent evaluation phase of Soapp. The results from the main analyses confirmed that the participants who participated in the study increased the frequency of correct hand hygiene at key times over time (H1). However, the intervention groups did not differ in their effects on correct hand hygiene at key times (H2). Similarly, the exposure to specific modules was not associated with increased hand hygiene over time (H3, H4, and H5). Taken together, the findings related to the first optimization criterion suggested a promising increase in hand hygiene during the intervention period but did not provide scientific evidence to support the preference of one version of Soapp or a specific module over the others. Similarly, the quantitative results from the second optimization criterion (H6, H7, and H8) did not show any differences in engagement, usability, and satisfaction among the 9 intervention modules at T3.

By contrast, the qualitative results revealed what characteristics and features of Soapp the participants perceived as supportive or, conversely, detrimental in terms of engagement, usability, and satisfaction. The finding that the aesthetics and design of the app are important for participants to better enjoy their interaction with Soapp is in line with a previous study on health-related behavior change [29]. The participants expressed the desire for an app that is simple to use, intuitive, and not cognitively demanding and that allows a smooth use of its functionalities. Such fundamental characteristics are deemed to guarantee satisfactory and engaging user experiences with Soapp. A second relevant aspect raised by the interviewed participants is the desire to receive clear guidance about the tasks that the app proposes and the rationale behind them. The participants also appreciated when they received (1) information regarding the behavior change intervention they committed to and (2) suggestions (eg, tips and problem-solving strategies) on how to adhere to correct hand hygiene. However, to prevent declining engagement, the delivery of guiding content should be balanced and not overwhelming (eg, push notifications). These findings are in line with those found in a previous systematic review and empirical research on engagement with digital behavior change interventions [30-32]. A further topic is variety in the daily interactions with the app and the proposed tasks and activities. A task load that varies daily (ie, days with more tasks and days with fewer tasks) seems to be important for sustaining engagement with Soapp. In addition, the regular provision of content over the course of the intervention was considered an important aspect of the app that might require some improvements. This aspect is of particular relevance, as the receipt of an optimal dose of engagement may increase the effectiveness of digital interventions [30]. Another theme that emerged during the interviews concerned the reasons that led the participants to join and remain engaged with the study. Although curiosity and interest to learn new things were important in triggering initial engagement, perceived obligation was a reason to maintain engagement over time. This result provides further support to participants' demand for a better distributed workload and content over the course of the use of Soapp. On the content side, as in a previous study about adults' perspectives on health behavior change apps [33], the participants appreciated those features that foster an increased sense of awareness around the target health-related behavior (ie, hand hygiene) and the resulting benefits. These results are in line with recent research conducted during the COVID-19 pandemic suggesting that self-monitoring is positively associated with hand washing [34]. Interestingly, the participants reported that the awareness formed mostly because of filling out the hand hygiene diaries that were included in the study as assessment tools and not as behavior change techniques (ie, self-monitoring). This aspect underlines how assessment tools and intervention strategies were not distinguished from one another by the participants but were perceived as part of the same user experience. A further theme that was at the center of the participants' comments regarded social interaction. Consistent with previous findings [30-32,35], features supporting a sense of relatedness owing to both a human-centered communication style (ie, tone of voice) and a feeling of connectedness were considered necessary to create social commitment and, ultimately, for engagement and satisfactory interactions with Soapp. Such a sense of relatedness was generated by the general user experience provided by the app (eg, communication style) and was not related to the features delivered by the social norm module. By contrast, in line with previous findings regarding health-related digital interventions [29,31-33,35], a dual perspective emerged in relation to the features that purposefully provided opportunities for social comparison and were part of the social norms module. Indeed, although some participants expressed avoidance of social comparison because they considered their behavior change journey as a personal dimension of their life, others were pleased about the comparison with other users. Therefore, social comparison features can be seen as a 2-edged sword for engagement, as the preference for such features is expected to vary across individuals. Eventually, the participants believed that receiving more personally relevant content would strengthen their engagement with Soapp. Such comments were partially generated by the participants' experiences in filling out hand hygiene diaries that refer to key times that are not relevant to them.

Implications for the Evaluation Phase of Soapp

Owing to the null findings of the first optimization criterion, we were not able to identify the best intervention group based on the quantitative analysis of the primary outcome. Similarly, no between-group differences emerged in relation to the second optimization criterion (ie, engagement, usability, and satisfaction). Therefore, we relied on the results of the thematic analysis to derive implications for the evaluation phase of Soapp. The resulting intervention design decisions based on this optimization study are summarized in Table 4.

Table 4. Intervention design recommendations for the evaluation phase of Soapp.

Recommendation	Rationale
The social module is excluded from the next evaluation phase.	Habit and motivation modules seem best suited to leverage some of the themes that emerged during the thematic analysis. For instance, themes such as change in awareness and guidance can be better supported by the app features that characterize these modules (ie, action planning tasks, self-monitoring, opportunity to schedule custom reminder, and video on health implication). In addition, the social module might be detrimental for engagement, as it embeds social comparison features, which were perceived as counterproductive by some users.
A parallel delivery of modules is preferrable over a sequential one.	The specific sequence of intervention modules (ie, habit-motivation vs motivation-habit) was not associated with differences in hand hygiene. Therefore, according to the participants' needs identified as part of the theme "variety and timeliness of the task load," a parallel delivery of the selected intervention modules is preferable.
Define a more even distribution of the intervention content and notifications over the course of the study.	A parallel delivery of the modules would allow to distribute each module's content and tasks over 32 days instead of 16 days, as done during the optimization phase. Therefore, there is more flexibility to define the timeline of the intervention with the aim of balancing the daily task load and, ultimately, guaranteeing a more suited dose of content over the course of the intervention.

Limitations

This study is not without limitations. A main weakness is the sample size achieved for testing the main hypotheses. Indeed, an a posteriori achieved power of 0.44 (n=190, 81.9%; α=.05; partial η^2=0.01) suggests that the probability of detecting a true effect of the intervention groups was lower than the recommended standard (ie, 0.80). Different factors contributed to the collection of data from a limited sample of 190 participants. First, we stopped recruitment 5 months after the start of the study, although the target sample size was not achieved. As specified in the study protocol, the criterion of discontinuing the enrollment of participants after 5 months was based on the constraints of the project timeline [16]. This resulted in a sample of 232 randomized participants. A second reason is the dropouts between randomization and T1 assessment. The T1 assessment was scheduled for the day after randomization; however, some (42/232, 18.1%) participants who had been randomized did not complete it. Therefore, they were excluded from the main analyses.

A further limitation that affected the analysis of the primary outcome was attrition. Of the 190 participants who filled out the T1 diary, 118 (62.1%) completed the diary at T3, leading to 38% and 49% attrition compared with the T1 and randomization figures, respectively. The attrition rate was higher than the estimated rate (ie, 20%). To account for potential differences between dropouts and retainers, we conducted a dropout analysis to investigate whether they differed in regard to key variables such as age, sex, hand hygiene, and intention to increase hand hygiene behavior. Because no significant differences emerged, we considered the participants who completed the study to be representative of our target population (ie, adults interested in using an app to improve hand hygiene behavior). A possible explanation for attrition could be the possibility that the longitudinal study design with 5 diary days and further quasidaily tasks might have generated an interaction fatigue. In addition, the pandemic trajectory during the enrollment period flattened in Switzerland, which may have made hand hygiene less of a priority for potential participants. To overcome this issue and in line with the ITT approach, we used the last observation carried forward method to replace the missing observations in the T3 diary with the latest available diary assessment. However, it should be noted that this method is based on the assumption that behavior is stable and, therefore, might have introduced bias.

Furthermore, the recruited sample was characterized by a high prevalence of women (ie, 139/190, 73.2%). This imbalance was in line with previous research on hand hygiene during the COVID-19 pandemic [11]. A plausible explanation for this gender imbalance might be that during the COVID-19 pandemic, women tended to show higher levels of worry and fear of the pandemic and were keener to adopt protective behaviors such as hand hygiene [11,36].

Finally, the self-reported measurement of hand hygiene may be biased. The use of an electronic diary to measure hand hygiene behavior at key times should have had limited retrospective bias. However, social desirability cannot be disregarded. In addition, thematic analysis indicated that the diary may have worked as an unintentional behavior change technique (ie, self-monitoring).

Conclusions

This study described the optimization phase of Soapp, a smartphone app for promoting hand hygiene in the context of the COVID-19 pandemic. By leveraging digital technologies and MOST, we addressed the call raised by public health experts for developing evidence-based behavior change interventions that are designed and optimized to be effective in a pandemic context [5]. In this regard, we provided support for the feasibility and effectiveness of digital interventions promoting hand hygiene behavior during an ongoing pandemic. This aspect is extremely relevant because digital interventions do not require personal contact and can be integrated into the daily lives of an unlimited number of people. Furthermore, our findings contributed to filling an existing research gap and improving the scientific knowledge on the most effective behavior change strategies to promote hand hygiene during a pandemic. Ultimately, Soapp represents a promising ready-to-go digital tool to be used in cases of future pandemics.

Acknowledgments

The authors thank Melanie Bamert, Dominik Zeilstra, Lisa Maria Summermatter, and gfs.bern for study support and their help with app development. Moreover, they thank the Technology Platform of the Faculty of Human Sciences at the University of Bern, especially Stefan Kodzhabashev, for facilitating app development. This study was funded by Ursula-Wirz Stiftung.

Authors' Contributions

DB was involved in formal analysis and the writing of the original draft. MAA contributed to conceptualization and methodology. MDRC and CF contributed to qualitative methodology. GGR contributed to resource accumulation and data curation. CB contributed to resource accumulation, data curation, and qualitative methodology. CR was involved in formal qualitative analysis and the writing of the qualitative report. JI contributed to funding acquisition, conceptualization, methodology, formal qualitative analysis, and supervision. All the authors were involved in discussion and in writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hand hygiene items and secondary hypothesis testing.
[DOCX File , 267 KB - mhealth_v11i1e43241_app1.docx]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[DOCX File , 41 KB - mhealth_v11i1e43241_app2.docx]

Multimedia Appendix 3
Qualitative interview guide.
[DOCX File , 15 KB - mhealth_v11i1e43241_app3.docx]

Multimedia Appendix 4
Supplement to quantitative results.
[DOCX File , 76 KB - mhealth_v11i1e43241_app4.docx]

Multimedia Appendix 5
Summary and quotes from the thematic analysis.
[DOCX File , 23 KB - mhealth_v11i1e43241_app5.docx]

References

1. Rabie T, Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. Trop Med Int Health 2006 Mar;11(3):258-267 [FREE Full text] [doi: 10.1111/j.1365-3156.2006.01568.x] [Medline: 16553905]
2. Aiello AE, Coulborn RM, Perez V, Larson EL. Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis. Am J Public Health 2008 Aug;98(8):1372-1381. [doi: 10.2105/AJPH.2007.124610] [Medline: 18556606]
3. Beale S, Johnson AM, Zambon M, Flu Watch Group, Hayward AC, Fragaszy EB. Hand hygiene practices and the risk of human coronavirus infections in a UK community cohort. Wellcome Open Res 2021 Jun 22;5:98 [FREE Full text] [doi: 10.12688/wellcomeopenres.15796.2] [Medline: 34250260]
4. Lee JK, Bullen C, Ben Amor Y, Bush SR, Colombo F, Gaviria A, (The Lancet COVID-19 Commission Task Force for Public Health Measures to Suppress the Pandemic). Institutional and behaviour-change interventions to support COVID-19 public health measures: a review by the Lancet Commission Task Force on public health measures to suppress the pandemic. Int Health 2021 Sep 03;13(5):399-409. [doi: 10.1093/inthealth/ihab022] [Medline: 33974687]
5. Michie S, West R. Behavioural, environmental, social, and systems interventions against COVID-19. BMJ 2020 Jul 28;370:m2982. [doi: 10.1136/bmj.m2982] [Medline: 32723732]
6. Perski O, Szinay D, Corker E, Shahab L, West R, Michie S. Interventions to increase personal protective behaviours to limit the spread of respiratory viruses: a rapid evidence review and meta-analysis. Br J Health Psychol 2022 Feb;27(1):215-264. [doi: 10.1111/bjhp.12542] [Medline: 34173697]
7. Ainsworth B, Miller S, Denison-Day J, Stuart B, Groot J, Rice C, et al. Infection control behavior at home during the COVID-19 pandemic: observational study of a web-based behavioral intervention (germ defence). J Med Internet Res 2021 Feb 25;23(2):e22197 [FREE Full text] [doi: 10.2196/22197] [Medline: 33566791]

8. Wise T, Zbozinek TD, Michelini G, Hagan CC, Mobbs D. Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. R Soc Open Sci 2020 Sep 16;7(9):200742 [FREE Full text] [doi: 10.1098/rsos.200742] [Medline: 33047037]
9. Moussaoui LS, Ofosu ND, Desrichard O. Social psychological correlates of protective behaviours in the COVID-19 outbreak: evidence and recommendations from a nationally representative sample. Appl Psychol Health Well Being 2020 Dec;12(4):1183-1204. [doi: 10.1111/aphw.12235] [Medline: 32981212]
10. Lippke S, Keller FM, Derksen C, Kötting L, Dahmen A. Hygiene behaviors and SARS-CoV-2-preventive behaviors in the face of the COVID-19 pandemic: self-reported compliance and associations with fear, SARS-CoV-2 risk, and mental health in a general population vs. a psychosomatic patients sample in Germany. Hygiene 2022 Jan 06;2(1):28-43 [FREE Full text] [doi: 10.3390/hygiene2010003]
11. Szczuka Z, Abraham C, Baban A, Brooks S, Cipolletta S, Danso E, et al. The trajectory of COVID-19 pandemic and handwashing adherence: findings from 14 countries. BMC Public Health 2021 Oct 05;21(1):1791 [FREE Full text] [doi: 10.1186/s12889-021-11822-5] [Medline: 34610808]
12. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
13. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
14. White S, Thorseth AH, Dreibelbis R, Curtis V. The determinants of handwashing behaviour in domestic settings: an integrative systematic review. Int J Hyg Environ Health 2020 Jun;227:113512 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113512] [Medline: 32220763]
15. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clin Trials 2014 Aug;11(4):426-434. [doi: 10.1177/1740774514536795] [Medline: 24902922]
16. Amrein MA, Ruschetti GG, Baeder C, Bamert M, Inauen J. Mobile intervention to promote correct hand hygiene at key times to prevent COVID-19 in the Swiss adult general population: study protocol of a multiphase optimisation strategy. BMJ Open 2022 Mar 28;12(3):e055971 [FREE Full text] [doi: 10.1136/bmjopen-2021-055971] [Medline: 35351716]
17. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
18. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
19. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY, USA: Routledge; 1999.
20. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
21. Peres SC, Pham T, Phillips R. Validation of the system usability scale (SUS): SUS in the wild. Proc Hum Factors Ergon Soc Annu Meet 2013 Sep 30;57(1):192-196. [doi: 10.1177/1541931213571043]
22. Schmidt J, Lamprecht F, Wittmann WW. Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. Psychother Psychosom Med Psychol 1989;39(7):248-255 [FREE Full text]
23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. Contzen N, De Pasquale S, Mosler HJ. Over-reporting in handwashing self-reports: potential explanatory factors and alternative measurements. PLoS One 2015 Aug 24;10(8):e0136445 [FREE Full text] [doi: 10.1371/journal.pone.0136445] [Medline: 26301781]
25. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
26. Garnett C, Oldham M, Angus C, Beard E, Burton R, Field M, et al. Evaluating the effectiveness of the smartphone app, Drink Less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6-month follow-up: protocol for a randomised controlled trial. Addiction 2021 Feb;116(2):412-425 [FREE Full text] [doi: 10.1111/add.15287] [Medline: 33067856]
27. Baretta D, Inauen J. SOAPP - Hand Hygiene promotion in the context of the COVID-19 pandemic. Open Science Framework. 2022 Aug 31. URL: https://osf.io/pzhqx/ [accessed 2023-01-17]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

29. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
30. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
31. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: a mixed-methods study. Digit Health 2018 Jul 16;4:2055207618785841 [FREE Full text] [doi: 10.1177/2055207618785841] [Medline: 31463077]
32. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
33. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
34. Keller J, Kwasnicka D, Wilhelm LO, Lorbeer N, Pauly T, Domke A, et al. Hand washing and related cognitions following a brief behavior change intervention during the COVID-19 pandemic: a pre-post analysis. Int J Behav Med 2022 Oct;29(5):575-586. [doi: 10.1007/s12529-021-10042-w] [Medline: 34843096]
35. D'Addario M, Baretta D, Zanatta F, Greco A, Steca P. Engagement features in physical activity smartphone apps: focus group study with sedentary people. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e20460 [FREE Full text] [doi: 10.2196/20460] [Medline: 33196450]
36. Bronfman N, Repetto P, Cordón P, Castañeda J, Cisternas P. Gender differences on psychosocial factors affecting COVID-19 preventive behaviors. Sustainability 2021 May 30;13(11):6148. [doi: 10.3390/su13116148]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DBCI: digital behavior change interventions
H: hypothesis
ITT: intention-to-treat
MOST: Multiphase Optimization Strategy
REDCap: Research Electronic Data Capture
T1: baseline
T3: follow-up
ZUF: Fragebogen zur Messung der Patientenzufriedenheit

Edited by L Buis; submitted 06.10.22; peer-reviewed by I Shubina, H Wang; comments to author 09.12.22; revised version received 20.12.22; accepted 20.12.22; published 03.02.23.

Please cite as:
Baretta D, Amrein MA, Bäder C, Ruschetti GG, Rüttimann C, Del Rio Carral M, Fabian C, Inauen J
Promoting Hand Hygiene During the COVID-19 Pandemic: Parallel Randomized Trial for the Optimization of the Soapp App
JMIR Mhealth Uhealth 2023;11:e43241
URL: https://mhealth.jmir.org/2023/1/e43241
doi:10.2196/43241
PMID:36599056

Original Paper

Associations Between Product Type and Intensity of Tobacco and Cannabis Co-use on the Same Day Among Young Adult Smokers: Smartphone-Based Daily-Diary Study

Nhung Nguyen[1,2], PhD; Johannes Thrul[3,4,5], PhD; Torsten B Neilands[6], PhD; Pamela M Ling[1,2], MD, MPH

[1]Center for Tobacco Control Research and Education, University of California San Francisco, San Francisco, CA, United States
[2]Division of General Internal Medicine, University of California San Francisco, San Francisco, CA, United States
[3]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States
[4]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, United States
[5]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia
[6]Center for AIDS Prevention Studies, Division of Prevention Science, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Nhung Nguyen, PhD
Center for Tobacco Control Research and Education
University of California San Francisco
530 Parnassus Ave
San Francisco, CA, 94143
United States
Phone: 1 415 476 2265
Email: Nhung.Nguyen@ucsf.edu

Abstract

Background: Co-use of tobacco and cannabis is highly prevalent among young US adults. Same-day co-use of tobacco and cannabis (ie, use of both substances on the same day) may increase the extent of use and negative health consequences among young adults. However, much remains unknown about same-day co-use of tobacco and cannabis, in part due to challenges in measuring this complex behavior. Nuanced understanding of tobacco and cannabis co-use in terms of specific products and intensity (ie, quantity of tobacco and cannabis use within a day) is critical to inform prevention and intervention efforts.

Objective: We used a daily-diary data collection method via smartphone to capture occurrence of tobacco and cannabis co-use within a day. We examined (1) whether the same route of administration would facilitate co-use of 2 substances on the same day and (2) whether participants would use more tobacco on a day when they use more cannabis.

Methods: This smartphone-based study collected 2891 daily assessments from 147 cigarette smokers (aged 18-26 years, n=76, 51.7% female) during 30 consecutive days. Daily assessments measured type (ie, cigarette, cigarillo, or e-cigarette) and intensity (ie, number of cigarettes or cigarillos smoked or number of times vaping e-cigarettes per day) of tobacco use and type (ie, combustible, vaporized, or edible) and intensity (ie, number of times used per day) of cannabis use. We estimated multilevel models to examine day-level associations between types of cannabis use and each type of tobacco use, as well as day-level associations between intensities of using cannabis and tobacco. All models controlled for demographic covariates, day-level alcohol use, and time effects (ie, study day and weekend vs weekday).

Results: Same-day co-use was reported in 989 of the total 2891 daily assessments (34.2%). Co-use of cigarettes and combustible cannabis (885 of the 2891 daily assessments; 30.6%) was most commonly reported. Participants had higher odds of using cigarettes (adjusted odds ratio [AOR] 1.92, 95% CI 1.31-2.81) and cigarillos (AOR 244.29, 95% CI 35.51-1680.62) on days when they used combustible cannabis. Notably, participants had higher odds of using e-cigarettes on days when they used vaporized cannabis (AOR 23.21, 95% CI 8.66-62.24). Participants reported a greater intensity of using cigarettes (AOR 1.35, 95% CI 1.23-1.48), cigarillos (AOR 2.04, 95% CI 1.70-2.46), and e-cigarettes (AOR 1.48, 95% CI 1.16-1.88) on days when they used more cannabis.

Conclusions: Types and intensities of tobacco and cannabis use within a day among young adult smokers were positively correlated, including co-use of vaporized products. Prevention and intervention efforts should address co-use and pay attention to all forms of use and timeframes of co-use (eg, within a day or at the same time), including co-use of e-cigarettes and vaporized cannabis, to reduce negative health outcomes.

(JMIR Mhealth Uhealth 2023;11:e40736) doi:10.2196/40736

KEYWORDS

tobacco; cannabis; substance co-use; young adults; intensive longitudinal data; EMA; mHealth; smartphone-based data collection; data collection; smartphone data; substance use

Introduction

Co-use of tobacco and cannabis is highly prevalent among young US adults. National data indicates that 21% of the general population of young adults has used both tobacco and cannabis in the past 30 days [1]. The use of cannabis is associated with persistent cigarette smoking and may pose a barrier to successful tobacco cessation [2,3]. At the person level, combined use of tobacco and cannabis can increase the risk for addiction and negative health outcomes (eg, mental health and respiratory problems) among people who co-use both products compared to those who use only a single substance [4,5]. This public health impact from co-use underscores the need to prevent this behavior during young adulthood. However, much remains unknown about co-use of tobacco and cannabis at both the personal level (ie, comparing people who co-use to those using a single substance) and at the event level (eg, comparing co-use to single-substance use within a day), in part due to challenges in measuring this complex behavior [6,7].

At the event level, the inherent complexity of co-use behavior includes a variety of products and timeframes, which adds extra burden to assessment and intervention [6]. People can use both substances in any combination of forms across the wide array of tobacco and cannabis products available on the marketplace [8]. While co-use is commonly defined in survey research as the use of both tobacco and cannabis within a month or year, it can also occur in a shorter timeframe (eg, within the same occasion or day). Studies indicate that the extent to which individuals use tobacco and cannabis closely in time is associated with more cigarettes smoked per day, greater nicotine dependence, and worse physical and mental functioning [9-11]. In addition, exposure to toxicants may vary by route of coadministration (eg, smoking vs vaping), posing differential health impacts [5]. Smoking both tobacco and cannabis is a well-known route, including the use of "blunts" (cannabis rolled in a cigar leaf for smoking), "spliffs" (combining cannabis and loose-leaf tobacco in a joint), or "chasing" (smoking cigarettes after smoking cannabis). A newer route of co-use is with vaporized products, in which liquid- or leaf-vaporizing devices are used to deliver both nicotine and tetrahydrocannabinol (THC—the main psychoactive component in cannabis), sometimes with the same device, on the same occasion, or in quick succession [12]. As such, understanding co-use among young adults at the event level, taking into account specific products and timeframes, is critical to inform prevention and intervention efforts [7].

Existing evidence at the event level, however, has predominantly focused on co-use in general (eg, any tobacco and cannabis) or has been limited to only combustible products (eg, blunts). In addition, co-use is mostly measured as any use of both tobacco and cannabis in the past 30 days, and little is known about intensity of co-use (defined in this study as quantity of tobacco and cannabis use within a day). Studies are lacking that address co-use via newer products and in shorter timeframes, yet these patterns of use may result in greater substance use and associated health impacts. Cross-sectional surveys and retrospective behavioral measures used in prior research have asked few questions about the nuances of co-use [6,7]. Newer data collection methods (eg, daily-diary assessments and ecological momentary assessments) have the potential to capture occurrence of tobacco and cannabis co-use within a day or moment and generate a richer picture of the behavior [7]. Using this approach, a few studies have indicated that cannabis use increased the odds of cigarette use on the same day [13] or within 4-hour windows [14,15] and that same-day co-use was more prevalent among young sexual-minority adults than their heterosexual peers [10]. These studies, however, have not examined the intensity of co-use within a day as well as co-use of noncombustible products (eg, e-cigarettes and vaporized cannabis). A better understanding of co-use at the event level, including whether and how types and intensity of cannabis use would drive tobacco use within a day, may be beneficial in developing interventions targeting young adults with problematic use of both substances.

To address the aforementioned gaps in knowledge of co-use of tobacco and cannabis at the day level, we analyzed smartphone-based daily assessment data collected among 147 young adult cigarette smokers during 2016 and 2017. We examined day-level associations between types (ie, combustible, vaporized, and edible) and intensity (ie, number of times) of cannabis use and types of tobacco product use (ie, cigarettes, cigarillos, and e-cigarettes). Based on the aforementioned research [10,13-15], we hypothesized that (1) participants would use more tobacco on a day when they use more cannabis; and (2) the same route of administration would facilitate co-use of 2 substances (eg, participants would smoke/vape tobacco on a day when they smoke/vape cannabis, respectively).

Methods

Study Design

This study analyzed daily assessments from a smartphone-based study conducted in California during 2016 and 2017. The study procedure was described in detail elsewhere [10]. Initially, participants completed a baseline survey on their demographics and substance use history. They were then trained on use of the study app to collect data every day for a 30-day period. Each day, participants were prompted between 10 and 11 AM to complete a daily assessment reporting their use of tobacco and cannabis on the entire previous day, including substance use occurrences late at night. To increase participant study compliance and retention, incentives were contingent on level of data-collection completion.

Ethics Approval

Electronic informed consent was obtained from all participants. The study was approved by the University of California San Francisco Institutional Review Board (15-18033).

Study Participants

Participants were recruited through social media and online advertisements (eg, Facebook and Craigslist). To conduct a nested qualitative substudy, participants were also recruited via the websites of sexual-minority youth organizations, and we oversampled women identifying as a sexual minority. Eligible participants were aged 18 to 26 years, had smoked at least 100 cigarettes in their lifetime, and currently smoked at least one cigarette per day at least 3 days per week. Since the parent study focused on cigarette smoking, cannabis use was not part of the inclusion criteria. Of 184 participants who completed the baseline assessments, 147 who completed at least one daily assessment were included in the analytic sample. There was no statistical difference between the analytic sample (n=147) and those who were excluded from the analysis (n=37) in baseline characteristics (ie, age, sex, educational attainment, race, ethnicity, and past-30-day use of tobacco and cannabis). Baseline characteristics of the study sample are presented in Table 1. The sample had a mean age of 22.7 (SD 2.4) years, 51.7% (76/147) of participants were female, 40.8% (60/147) of participants were non-Hispanic White, and 76.9% (113/147) of participants were currently in college or had a college degree or higher. At baseline, a majority of participants reported past-30-day use of cannabis (96/147, 65.3%) and alcohol (136/147, 92.5%). We included 51 participants who did not report past 30-day use of cannabis at baseline in our sample, since these participants could report co-use during the daily-diary period (and n=7 did), allowing for comparison between co-use and single-substance use within a day.

Table 1. Sample characteristics.

Characteristics	Total (n=147)
Age (years), mean (SD)	22.7 (2.4)
Sex at birth, n (%)	
Male	71 (48.3)
Female	76 (51.7)
Race, n (%)	
Non-Hispanic White	60 (40.8)
Non-Hispanic Asian	30 (20.4)
Hispanic	31 (21.1)
Other/multiracial	17 (11.6)
Education, n (%)	
Less than college	33 (22.5)
College or higher	113 (76.9)
Past 30-day substance use at baseline, n (%)	
Tobacco use	
Cigarettes	145 (98.6)
e-Cigarettes	47 (32)
Cannabis use	96 (65.3)
Alcohol use	136 (92.5)

Measures

Outcome Variables (Type and Intensity of Tobacco Use)

For each day, participants reported whether they used cigarettes, cigarillos, or e-cigarettes. These binary variables (yes/no) indicated types of tobacco used in each daily assessment. We examined cigarillos rather than other types of cigars since cigarillos are the most common cigar type used by young adults [16]. Regarding intensity of tobacco use, participants were asked, “Yesterday: How many [cigarettes, cigarillos] used?” and “How many times [disposable e-cigarettes, rechargeable e-cigarettes, tanks, or pod-mods] used?” We assessed tobacco products and different types of e-cigarettes separately. These types of devices included 4 generations of e-cigarettes available on the marketplace at the time of the study (eg, the first generation refers to disposable e-cigarettes, the second generation refers to rechargeable e-cigarettes, the third generation refers to tank devices, and the fourth generation refers to pod-mods). A total intensity of e-cigarette use for each day was calculated by summing the intensities of using all 4 types of e-cigarettes. To not overburden participants, response options provided categories of increasing intensity of use of each product (ie, 0, 1, 2-5, 6-10, 11-15, 16-20, 21-30, and ≥31 cigarettes, cigarillos, or times vaping e-cigarettes per day).

Independent Variables (Type and Intensity of Cannabis Use)

For each day, participants were asked, “How many times did you use marijuana or hash?” Answer options ranged continuously from 0 to 7 or more times. Those who reported any cannabis use were then asked, “How did you use marijuana or hash?” with answer options including smoking, vaping, and edibles. While we asked about intensity of use of cannabis in general, we did not ask about intensity of use of each cannabis product separately, to avoid overburdening participants. As such, depending on a participant’s interpretation, a smoking occasion of combustible cannabis, a hit of a cannabis vaporizer, or consuming 1 edible may have been considered as a single occasion of cannabis use in our study.

Covariates

Demographic characteristics were collected at baseline. Age was calculated based on self-reported date of birth. Sex assigned at birth was measured as female or male. Race/ethnicity was categorized into 4 groups: non-Hispanic White, non-Hispanic Asian, Hispanic, and other/multiracial. Educational attainment was dichotomized as “less than college” and “college or higher,” since having a college education is associated with tobacco and cannabis use among young adults [17]. Participants also reported alcohol use (yes/no) in each daily assessment. A dummy variable was created to indicate the study day of each daily assessment, ranging from day 1 to day 30. As use of substances may be different between weekends and weekdays [18], another dummy variable was created to indicate weekend or weekday.

Statistical Analyses

Statistical analyses were performed using Stata (version 15; Stata Corp). Descriptive statistics of sample characteristics at baseline and substance use in daily assessments were summarized. First, to examine associations of type of tobacco and cannabis co-use on the same day, we fitted multilevel logistic regression models examining associations of use of cannabis products (combustible, vaporized, and edible) with each of the binary outcomes (ie, any use of cigarettes, cigarillos, or e-cigarettes on a given day). Second, to examine associations between intensities of tobacco and cannabis use on the same day, we fitted multilevel mixed-effects ordered logistic regression models examining intensity of cannabis use (ie, number of times using cannabis on a given day) with each of the ordinal outcomes (ie, numbers of cigarettes or cigarillos smoked and number of times using e-cigarettes on a given day) [19]. The models also included random intercepts for participants to control for variation in tobacco use intensity attributable to individual participants.

The variable of intensity of cannabis use was decomposed into 2 elements: personal mean (ie, average intensity of cannabis use for each participant, indicating comparisons between participants, in other words, between-person effects), and deviation (ie, the difference between intensity in a particular daily observation and the personal mean, indicating comparisons across study days within a certain participant, that is, within-person effects) [20]. For each ordinal outcome, the proportional odds assumption was checked by fitting a generalized multinomial logit model and comparing its likelihood ratio to that of the ordinal model [19]; this assumption was satisfied for all the models. All models controlled for demographic covariates, day-level alcohol use [21], and time effects (ie, study day and weekend vs weekday). All tests were 2-tailed with a significance level of $\alpha<.05$. The analyses were not preregistered and thus the results should be considered exploratory.

Results

Daily Assessments of Tobacco and Cannabis Use

During the 30-day study period, 147 participants completed an average of 19.7 (SD 9.6) daily assessments with a completion rate of 65.6% (2891 completed assessments of 4410 prompted assessments). Table 2 describes reports of tobacco and cannabis use among the total of 2891 daily assessments. Co-use was reported in 989 daily assessments (34.2%), while use of tobacco without cannabis was reported in 1501 daily assessments (51.9%). The most common intensity of cigarette use reported in the daily assessments was smoking 2 to 5 cigarettes per day. Not using at all was reported the most in the daily assessments for cigarillos, e-cigarettes, and cannabis. On the days when participants used these products, the common intensities of use were smoking 2 to 5 cigarillos per day, vaping e-cigarettes 2 to 5 times per day and using cannabis once a day.

Table 3 presents same-day co-use in terms of combinations of specific products. The most commonly used tobacco product was cigarettes (2407 of 2891 assessments, 83.3%), while combustible cannabis was the most common type of cannabis use (1040 of 2891 assessments, 36%). The 3 most common product combinations on the same day were cigarettes and combustible cannabis (885 of 2891 assessments, 30.6%), cigarillos and combustible cannabis (197 of 2891 assessments, 6.8%), and cigarettes and vaporized cannabis (147 of 2891 assessments, 5.1%).

Table 2. Daily assessments of substance use among young adult smokers.

Substance use assessments	Assessments (n=2891), n (%)
Daily assessments	
Use of tobacco only	1501 (51.9)
Use of cannabis only	145 (5)
Use of both substances	989 (34.2)
No use	251 (8.7)
Missing data	5 (0.2)
Number of cigarettes smoked in a day	
0	483 (16.7)
1	303 (10.5)
2-5	1250 (43.2)
6-10	644 (22.3)
11-15	145 (5)
16-20	61 (2.1)
21-30	2 (0.1)
≥31	3 (0.1)
Number of cigarillos smoked in a day	
0	2682 (92.8)
1	85 (2.9)
2-5	110 (3.8)
6-10	11 (0.4)
11-15	3 (0.1)
Number of times vaping e-cigarettes in a day	
0	2664 (92.2)
1	17 (0.6)
2-5	113 (3.9)
6-10	52 (1.8)
11-15	22 (0.8)
16-20	15 (0.5)
21-30	7 (0.2)
≥31	1 (<0.1)
Number of times using cannabis in a day	
0	1754 (60.7)
1	351 (12.1)
2	255 (8.8)
3	231 (8)
4	131 (4.5)
5	77 (2.7)
6	17 (0.6)
≥7	75 (2.6)
Daily assessments with alcohol use	1032 (35.7)
Daily assessments on weekend	804 (27.8)
Daily assessments on weekday	2087 (72.2)

Table 3. Same-day co-use of specific tobacco and cannabis products among young adult smokers (n=2891 assessments). Proportions were calculated as frequency of a given product combination out of the total daily assessments.

Types	Any cannabis (n=1136, 39.3%), n (%)	Combustible cannabis (n=1040, 36%), n (%)	Vaporized cannabis (n=190, 6.6%), n (%)	Edible cannabis (n=36, 1.3%), n (%)
Any tobacco (n=2490, 86.1%)	989 (34.2)	915 (31.7)	151 (5.2)	28 (1)
Cigarette (n=2407, 83.3%)	956 (33.1)	885 (30.6)	147 (5.1)	27 (0.9)
e-Cigarette (n=240, 8.3%)	101 (3.5)	79 (2.7)	39 (1.4)	4 (0.1)
Cigarillo (n=209, 7.2%)	197 (6.8)	197 (6.8)	8 (0.3)	5 (0.2)

Associations Between Type of Cannabis and Tobacco Products Used on the Same Day

Results from the mixed-effects models are shown in Table 4. Participants had higher odds of reporting using cigarettes (adjusted odds ratio [AOR] 1.92, 95% CI 1.31-2.81) and cigarillos (AOR 244.29, 95% CI 35.51-1680.62) on days when they used combustible cannabis. Notably, participants had higher odds of using e-cigarettes on days when they used vaporized cannabis (AOR 23.21, 95% CI 8.66-62.24). It should be noted that the CIs for cigarillos and e-cigarettes were quite wide due to the small number of daily assessments with use of these products. In addition, participants had higher odds of smoking cigarettes on days with alcohol use (AOR 2.73, 95% CI 1.99-3.76). The study day was negatively associated with the odds of smoking cigarettes (AOR 0.96, 95% CI 0.95-0.98) and cigarillos (AOR 0.96, 95% CI 0.92-0.99). Older participants (vs younger peers) had higher odds of reporting cigarette smoking (AOR 1.31, 95% CI 1.10-1.57), while Hispanic participants (vs non-Hispanic White peers) had higher odds of reporting cigarillo smoking (AOR 22.93, 95% CI 3.30-159.49); however, this estimate was very wide due to a small number of cigarillo-use reports.

Table 4. Day-level associations between tobacco use product types (outcomes) and cannabis use product types (independent variables) among young adult cigarette smokers (n=2891 assessments), controlled for time-varying covariates (day-level) and demographic covariates (participant-level). The outcomes were binary variables (ie, any use of a tobacco product on a given day). All variables were included in a mixed-effects logistic regression model for each outcome.

Independent variables	Model 1: cigarette smoking AOR[a] (95% CI)	*P* value	Model 2: cigarillo smoking AOR (95% CI)	*P* value	Model 3: e-cigarette vaping AOR (95% CI)	*P* value
Type of cannabis use						
Combustible cannabis	1.92 (1.31-2.81)	.001	244.29 (35.51-1680.62)	<.001	1.84 (0.93-3.67)	.08
Vaporized cannabis	1.58 (0.83-3.03)	.17	0.83 (0.30-2.25)	.71	23.21 (8.66-62.24)	<.001
Edible cannabis	0.58 (0.21-1.63)	.30	1.18 (0.28-5.02)	.82	4.90 (0.90-26.58)	.07
Time-varying covariates						
Alcohol use	2.73 (1.99-3.76)	<.001	0.91 (0.51-1.63)	.76	1.63 (0.90-2.93)	.10
Weekend vs weekday	0.85 (0.65-1.12)	.26	1.02 (0.60-1.72)	.96	0.79 (0.48-1.31)	.36
Study day	0.96 (0.95-0.98)	<.001	0.96 (0.92-0.99)	.01	1.00 (0.97-1.03)	.84
Demographic covariates						
Age	1.31 (1.10-1.57)	.003	1.00 (0.75-1.33)	.98	0.92 (0.69-1.22)	.55
Female vs male	1.06 (0.48-2.34)	.88	2.32 (0.57-9.44)	.24	0.84 (0.24-2.98)	.78
Education (college or higher vs less)	0.57 (0.20-1.62)	.29	0.70 (0.11-4.52)	.71	0.49 (0.10-2.52)	.40
Race (reference non-Hispanic White)						
Non-Hispanic Asian	1.47 (0.51-4.20)	.48	6.84 (0.90-52.23)	.06	2.01 (0.38-10.64)	.41
Hispanic	2.85 (0.88-9.25)	.08	22.93 (3.30-159.49)	.002	0.82 (0.14-4.94)	.83
Other/multiracial	1.19 (0.41-3.43)	.75	4.29 (0.65-28.21)	.13	0.77 (0.13-4.56)	.78

[a]AOR: adjusted odds ratio.

Associations Between Intensity of Cannabis and Tobacco Use on the Same Day

Results from the multilevel ordinal models are shown in Table 5. Participants had higher odds of reporting a greater intensity of using cigarettes (AOR 1.35, 95% CI 1.23-1.48), cigarillos (AOR 2.04, 95% CI 1.70-2.46), and e-cigarettes (AOR 1.48, 95% CI 1.16-1.88) on days when they used more cannabis. In addition, alcohol use on a given day was positively associated with intensity of cigarette use (AOR 1.41, 95% CI 1.35-1.49).

Participants with higher average intensity of cannabis use had higher average intensity of cigarillo use (AOR 3.72, 95% CI 2.41-5.73). The study day was negatively associated with intensity of smoking cigarettes (AOR 0.97, 95% CI 0.96-0.98) and cigarillos (AOR 0.94, 95% CI 0.91-0.97), meaning that intensity of smoking cigarettes and cigarillos decreased slightly over the study period. Those with education attainment of college or higher reported lower intensity of cigarette smoking (AOR 0.30, 95% CI 0.09-0.97), while older participants reported higher intensity of cigarette smoking (AOR 1.52, 95% CI 1.22-1.89). Hispanic participants (vs non-Hispanic White peers) had higher odds of reporting higher intensity of cigarillo smoking (AOR 11.46, 95% CI 1.82-72.36); however, this estimate was very wide due to a small number of cigarillo-use reports.

Table 5. Day-level associations between tobacco use intensity for different products (as outcomes) and intensity of cannabis use (as independent variables) among young adult cigarette smokers, controlling for time-varying (day-level) and demographic (participant-level) covariates (n=2891 assessments). The outcomes were categorical variables (ie, 0, 1, 2-5, 6-10, 11-15, 16-20, 21-30, and ≥31 cigarettes, cigarillos, or times vaping e-cigarettes in a given day). All variables were included in a multilevel mixed-effects ordered logistic regression model for each outcome.

Independent variables	Model 1: cigarette smoking intensity AOR[a] (95% CI)	*P* value	Model 2: cigarillo smoking intensity AOR (95% CI)	*P* value	Model 3: e-cigarette vaping intensity AOR (95% CI)	*P* value
Intensity of cannabis use						
Intensity of cannabis use in a given day	1.35 (1.23-1.48)	<.001	2.04 (1.70-2.46)	<.001	1.48 (1.16-1.88)	.001
Personal mean of cannabis use intensity	1.05 (0.78-1.42)	.75	3.72 (2.41-5.73)	<.001	1.16 (0.75-1.78)	.51
Time-varying covariates						
Intensity of alcohol use in a given day	1.41 (1.35-1.49)	<.001	1.02 (0.90-1.15)	.77	1.04 (0.93-1.16)	.48
Personal mean of alcohol use intensity	0.95 (0.63-1.43)	.80	0.86 (0.44-1.66)	.65	1.08 (0.59-1.99)	.80
Weekend vs weekday	0.95 (0.80-1.12)	.52	1.05 (0.67-1.65)	.84	0.86 (0.57-1.30)	.48
Study day	0.97 (0.96-0.98)	<.001	0.94 (0.91-0.97)	<.001	1.01 (0.99-1.04)	.25
Demographic covariates						
Age	1.52 (1.22-1.89)	<.001	1.00 (0.72-1.37)	.98	0.82 (0.60-1.14)	.24
Female vs male	0.45 (0.18-1.13)	.09	3.89 (0.95-15.91)	.06	1.06 (0.28-4.04)	.94
Education (college or higher vs less)	0.30 (0.09-0.97)	.045	0.82 (0.15-4.57)	.82	0.67 (0.12-3.81)	.66
Race (reference: non-Hispanic White)						
Non-Hispanic Asian	1.05 (0.30-3.65)	.94	6.39 (0.90-45.55)	.06	1.35 (0.23-7.82)	.74
Hispanic	1.58 (0.44-5.61)	.48	11.46 (1.82-72.36)	.01	0.31 (0.04-2.35)	.26
Other/multiracial	0.72 (0.19-2.73)	.63	2.96 (0.37-23.54)	.31	0.60 (0.09-4.07)	.60

[a]AOR: adjusted odds ratio.

Discussion

Principal Results

This study is one of very few examining young adult co-use of tobacco and cannabis within shorter timeframes (ie, a day) than the typical survey measure of past-30-day use, and it is among the first to examine same-day co-use of tobacco and cannabis products that are not smoked (ie, e-cigarettes and vaporized cannabis), including day-level intensity of co-use. The main findings were as hypothesized and showed that the more cannabis participants reported using on a given day, the greater the intensity of tobacco product use (cigarettes, cigarillos, and e-cigarettes). Notably, participants reported smoking cigarettes or cigarillos on the days they smoked cannabis, and vaping e-cigarettes on the days they vaped cannabis, indicating the same routes of administration may play a role in facilitating same-day co-use.

Comparison With Prior Work

Since traditional measures are insufficient to fully capture and monitor co-use of tobacco and cannabis, recent research called for more accurate measures of this behavior [6] and highlighted the potential of digital health applications for collecting fine-grained data and specifying co-use patterns [7]. As a methodological example, our study used a daily-diary design and smartphone-based data collection to generate intensive longitudinal data on co-use patterns on a daily basis over 30 consecutive days, providing a nuanced understanding of the extent of co-use within a day. Another strength of this study was an examination of use of a variety of tobacco and cannabis products, including co-use of vaporized products (ie, e-cigarettes and vaporized cannabis), for which more evidence is needed. In addition to our smartphone-based daily assessment method, future research should also consider using other mobile-data collection methods (eg, ecological momentary assessments and mobile sensors) that may more comprehensively assess co-use of tobacco and cannabis [6,7]. Furthermore, while there are only a handful of studies, including this study, that have directly examined co-use as the focal outcome, many prior studies indirectly addressed co-use by adjusting for use of both tobacco and cannabis in the same analytic models. Systematic review or meta-analysis of both direct and indirect evidence may be warranted to provide comprehensive insights on co-use and its effects.

The positive associations between use of the same types of tobacco and cannabis products on the same day indicate that there may be behavioral cues from shared routes of administration that may facilitate co-use of tobacco and cannabis (eg, smoking or vaping one substance triggers smoking or vaping the other) [22]. Indeed, a combination of cigarettes and combustible cannabis was the most common same-day co-use pattern in our sample of young adult cigarette smokers. It was also the most common pattern of past-30-day and past-year co-use found in other samples of young adults [8,11]. In addition to well-documented co-use patterns via smoking (eg, cigarettes/cigarillos and combustible cannabis), we also found that participants reported vaping e-cigarettes more on the days when they vaped cannabis. This finding, coupled with the high prevalence of vaping among young populations, indicates that more attention to emerging co-use patterns via vaping is needed [12,23]. Previous research has reported vaping-related harms among covapers, such as lung impairments [24] and increased odds of having COVID-19 symptoms and diagnoses [25]. Further investigation of tobacco and cannabis covaping among young adults and its health consequences is warranted. Moreover, our participants also reported using other product combinations across the spectrum of tobacco and cannabis products, underlining the heterogeneity of co-use patterns. Further exploration of unique reasons and contexts for different patterns of co-use would help to identify targets for tailored prevention and treatment strategies.

While one might expect a potential drug substitution effect, in which people use cannabis as a substitute for tobacco [26], our finding of positive associations between intensity of tobacco and cannabis use on the same day suggests the substitution effect did not occur in our sample of young adult smokers. Instead, as explained by the theory of synergistic effects, individuals may use 2 substances at the same time or use 1 substance under the effect of the other to amplify positive effects or counteract negative effects between nicotinic and endocannabinoid systems [22]. Relatedly, shared contexts (eg, being with friends and socializing) may also facilitate intensity of same-day co-use of tobacco and cannabis [27,28]. In addition, our participants with higher average intensity of cannabis use also reported higher average intensity of cigarillo use. This finding could be due to our participants using cigarillos for blunt smoking. Although we did not directly ask about blunt use in the daily assessments, previous studies indicated that young adults perceived cigarillos were frequently used for blunts [29,30]. We also found that participants smoked more cigarettes on the days when they drank alcohol. This finding could be explained by well-known rewarding effects when cigarettes and alcohol are used together [13,31-33].

In addition, several subgroups in our sample demonstrated greater average intensity of tobacco use. Participants who were older and had less than a college education reported greater intensity of cigarette smoking, whereas Hispanic participants reported a greater intensity of cigarillo use. These findings are consistent with previous research documenting high prevalence of tobacco use in these subgroups [17,34,35]. Interestingly, participants' use of cigarettes and cigarillos decreased over the study period. This may be due to a Hawthorne effect or other impacts of research participation, as the process of reporting on their own behaviors may induce reflection and influence participants' behaviors [36]. To our knowledge, this reactivity effect was rarely observed in previous research. Future studies using experience-sampling methods (eg, ecological momentary assessments or daily diaries) should explore reactivity effects and potential impacts on behavioral outcomes.

Study Implications

Collectively, our study has implications for efforts to support smoking cessation among young adults. As co-use of tobacco and cannabis was common, and this may increase harm and addiction, smoking cessation programs may need to address co-use of multiple tobacco products or tobacco and cannabis to improve efficacy with this age group. Most available interventions to reduce tobacco use in young people may not address engaging in co-use [7,37]. A recent study found that when young people reduced their tobacco use, their cannabis use also decreased, suggesting the potential benefits of dual cessation treatment for co-users [38]. In addition, treatment strategies should be expanded to include co-use of nonsmoking products to meet cessation needs of covapers [7,39]. As such, tailored interventions that adapt supports to individuals' co-use patterns may be more effective for reducing the use of both substances. Moreover, tailored interventions may be needed to reach those with high rates of co-use, such as those without college education or those who identify as Hispanic.

Limitations

Several limitations ought to be considered. The data were collected during 2016 and 2017. Since then, there have been rapid changes in public policy related to both tobacco and cannabis, in patterns of use (eg, increasing use of vaporization devices), in cannabis legalization, and in product availability in the marketplace. As such, more recent data are needed to replicate our findings. The convenience sampling procedure via online recruitment in California and the oversampling of young sexual minority adults limit our study's generalizability to other young adult samples or geographic regions. While co-use of tobacco and cannabis is common among smokers, our sample included a minority of young adults who did not report past-30-day cannabis use at baseline; further research should examine co-use of tobacco and cannabis among young adults who report recent use of both substances. Although using categories for intensity of use of tobacco products provided a general measure of increasing intensity, this may result in limitations to interpretation of actual effects for each product, given that the increase of using e-cigarettes from, for example, one time per day to 2 to 5 times per day may be different from the increase of smoking from one to 2 to 5 cigarillos per day. Data on cigar use were not collected, and simultaneous use of tobacco and cannabis and their overlapping effects were not directly assessed in our study. Moreover, data on timing or ordering in use of tobacco and cannabis were not collected; thus, we could not identify temporal relationships in use of these substances. Likewise, we did not collect data on cannabis concentrations and specific intensity of use by type of cannabis. The use of concentrated cannabis could impact the same-day co-use of tobacco and cannabis in meaningful ways depending on complementing versus supplementing behaviors. In addition, our participants were not trained in defining intensity of cannabis use and the meaning of "times of cannabis use" may vary depending on personal definitions of use sessions and types of cannabis use. Future research should consider collecting these data and developing more accurate measures of daily use of tobacco and cannabis in order to provide a better understanding of co-use. Missing data due to participants' compliance with daily assessments may impact the study's internal validity; however, our compliance rate is within the range of previous studies using the same data collection methods [10,13-15] and the models in our analysis are generally robust to missing data under the missing-at-random assumption [40,41].

Conclusion

By using smartphone-based daily assessments, this study identified a substantial correlation of product types and intensities of tobacco and cannabis co-use at the day level, with young adults reporting more tobacco use on days when they used more cannabis, including same-day co-use of e-cigarettes and vaporized cannabis. Future research and interventions should address co-use in all forms, especially co-use via new products and in short timeframes, to better prevent and reduce use of both tobacco and cannabis and related health impacts among young people.

Acknowledgments

This research was supported by the California Tobacco-Related Disease Research Program (grants T31FT1564 and T32KT5071 to NN and 25FT0009 to JT), by the Food and Drug Administration Center for Tobacco Products and the National Heart, Lung, and Blood Institute (grant U54 HL147127 to PML), and by the National Cancer Institute (grants U01 CA154240 to PML and R01 CA246590 to JT). NN is also supported by the University of California San Francisco Clinical and Translational Science Institute (UL1 TR001872-06). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. We would like to thank Dr Donald Hedeker at the University of Chicago for his advice on our data analysis.

Authors' Contributions

NN, JT, TBN, and PML were involved in writing the manuscript. NN conceptualized the study, obtained funding, drafted the initial manuscript, analyzed and interpreted the data, and contributed to all subsequent drafts of the manuscript. JT and TBN analyzed and interpreted the data and reviewed and revised the manuscript. PML supervised and reviewed and revised the manuscript. All authors have read and approved the final manuscript for submission.

Conflicts of Interest

None declared.

References

1. Cohn AM, Abudayyeh H, Perreras L, Peters EN. Patterns and correlates of the co-use of marijuana with any tobacco and individual tobacco products in young adults from Wave 2 of the PATH Study. Addict Behav 2019 May;92:122-127. [doi: 10.1016/j.addbeh.2018.12.025] [Medline: 30623805]
2. Strong DR, Myers MG, Pulvers K, Noble M, Brikmanis K, Doran N. Marijuana use among US tobacco users: Findings from wave 1 of the population assessment of tobacco health (PATH) study. Drug Alcohol Depend 2018 May 01;186:16-22. [doi: 10.1016/j.drugalcdep.2017.12.044] [Medline: 29529455]
3. Weinberger A, Delnevo C, Wyka K, Gbedemah M, Lee J, Copeland J, et al. Cannabis use is associated with increased risk of cigarette smoking initiation, persistence, and relapse among adults in the United States. Nicotine Tob Res 2020 Jul 16;22(8):1404-1408 [FREE Full text] [doi: 10.1093/ntr/ntz085] [Medline: 31112595]
4. Hindocha C, Brose LS, Walsh H, Cheeseman H. Cannabis use and co-use in tobacco smokers and non-smokers: prevalence and associations with mental health in a cross-sectional, nationally representative sample of adults in Great Britain, 2020. Addiction 2021 Aug;116(8):2209-2219. [doi: 10.1111/add.15381] [Medline: 33345423]
5. Meier E, Hatsukami DK. A review of the additive health risk of cannabis and tobacco co-use. Drug Alcohol Depend 2016 Sep 01;166:6-12. [doi: 10.1016/j.drugalcdep.2016.07.013] [Medline: 27476751]
6. Hindocha C, McClure EA. Unknown population-level harms of cannabis and tobacco co-use: if you don't measure it, you can't manage it. Addiction 2021 Jul;116(7):1622-1630 [FREE Full text] [doi: 10.1111/add.15290] [Medline: 33047862]
7. Nguyen N, Nguyen C, Thrul J. Digital health for assessment and intervention targeting tobacco and cannabis co-use. Curr Addict Rep 2020 Sep;7(3):268-279 [FREE Full text] [doi: 10.1007/s40429-020-00317-9] [Medline: 33643768]
8. Nguyen N, Barrington-Trimis JL, Urman R, Cho J, McConnell R, Leventhal AM, et al. Past 30-day co-use of tobacco and marijuana products among adolescents and young adults in California. Addict Behav 2019 Nov;98:106053 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106053] [Medline: 31357072]

9. Akbar SA, Tomko RL, Salazar CA, Squeglia LM, McClure EA. Tobacco and cannabis co-use and interrelatedness among adults. Addict Behav 2019 Mar;90:354-361 [FREE Full text] [doi: 10.1016/j.addbeh.2018.11.036] [Medline: 30522075]
10. Nguyen N, McQuoid J, Neilands T, Dermody S, Holmes L, Ling P, et al. Same-day use of cigarettes, alcohol, and cannabis among sexual minority and heterosexual young adult smokers. Psychol Addict Behav 2021 Mar;35(2):215-223 [FREE Full text] [doi: 10.1037/adb0000678] [Medline: 32804517]
11. Tucker JS, Pedersen ER, Seelam R, Dunbar MS, Shih RA, D'Amico EJ. Types of cannabis and tobacco/nicotine co-use and associated outcomes in young adulthood. Psychol Addict Behav 2019 Jun;33(4):401-411 [FREE Full text] [doi: 10.1037/adb0000464] [Medline: 30985164]
12. Reboussin BA, Wagoner KG, Ross JC, Suerken CK, Sutfin EL. Tobacco and marijuana co-use in a cohort of young adults: Patterns, correlates and reasons for co-use. Drug Alcohol Depend 2021 Oct 01;227:109000 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109000] [Medline: 34507062]
13. Roche D, Bujarski S, Green R, Hartwell E, Leventhal A, Ray L. Alcohol, tobacco, and marijuana consumption is associated with increased odds of same-day substance co- and tri-use. Drug Alcohol Depend 2019 Jul 01;200:40-49 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.02.035] [Medline: 31085377]
14. Berg CJ, Haardörfer R, Payne JB, Getachew B, Vu M, Guttentag A, et al. Ecological momentary assessment of various tobacco product use among young adults. Addict Behav 2019 May;92:38-46 [FREE Full text] [doi: 10.1016/j.addbeh.2018.12.014] [Medline: 30579116]
15. Wilhelm J, Abudayyeh H, Perreras L, Taylor R, Peters EN, Vandrey R, et al. Measuring the temporal association between cannabis and tobacco use among co-using young adults using ecological momentary assessment. Addict Behav 2020 May;104:106250 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106250] [Medline: 31918167]
16. Corey C, King B, Coleman B, Delnevo C, Husten C, Ambrose B, Centers for Disease Control and Prevention. Little filtered cigar, cigarillo, and premium cigar smoking among adults--United States, 2012-2013. MMWR Morb Mortal Wkly Rep 2014 Aug 01;63(30):650-654 [FREE Full text] [Medline: 25078654]
17. Odani S, Soura B, Tynan M, Lavinghouze R, King B, Agaku I. Tobacco and marijuana use among US college and noncollege young adults, 2002-2016. Pediatrics 2019 Dec;144(6):e20191372. [doi: 10.1542/peds.2019-1372] [Medline: 31712275]
18. Camenga DR, Haeny AM, Krishnan-Sarin S, O'Malley SS, Bold KW. Pilot ecological momentary assessment study of subjective and contextual factors surrounding e-cigarette and combustible tobacco product use among. Int J Environ Res Public Health 2021 Oct 20;18(21):11005 [FREE Full text] [doi: 10.3390/ijerph182111005] [Medline: 34769525]
19. Hedeker D. Methods for multilevel ordinal data in prevention research. Prev Sci 2015 Oct;16(7):997-1006 [FREE Full text] [doi: 10.1007/s11121-014-0495-x] [Medline: 24939751]
20. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Stat Med 2012 Nov 30;31(27):3328-3336 [FREE Full text] [doi: 10.1002/sim.5338] [Medline: 22419604]
21. Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychol Methods 2007 Jun;12(2):121-138. [doi: 10.1037/1082-989X.12.2.121] [Medline: 17563168]
22. Rabin RA, George TP. A review of co-morbid tobacco and cannabis use disorders: possible mechanisms to explain high rates of co-use. Am J Addict 2015 Mar;24(2):105-116. [doi: 10.1111/ajad.12186] [Medline: 25662704]
23. Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction 2015 Nov;110(11):1699-1704 [FREE Full text] [doi: 10.1111/add.13036] [Medline: 26264448]
24. Braymiller JL, Barrington-Trimis JL, Leventhal AM, Islam T, Kechter A, Krueger EA, et al. Assessment of nicotine and cannabis vaping and respiratory symptoms in young adults. JAMA Netw Open 2020 Dec 01;3(12):e2030189. [doi: 10.1001/jamanetworkopen.2020.30189] [Medline: 33351085]
25. Merianos AL, Russell AM, Mahabee-Gittens EM, Barry AE, Yang M, Lin H. Concurrent use of e-cigarettes and cannabis and associated COVID-19 symptoms, testing, and diagnosis among student e-cigarette users at four U.S. Universities. Addict Behav 2022 Mar;126:107170 [FREE Full text] [doi: 10.1016/j.addbeh.2021.107170] [Medline: 34776303]
26. McClure EA, Tomko RL, Salazar CA, Akbar SA, Squeglia LM, Herrmann E, et al. Tobacco and cannabis co-use: Drug substitution, quit interest, and cessation preferences. Exp Clin Psychopharmacol 2019 Jun;27(3):265-275 [FREE Full text] [doi: 10.1037/pha0000244] [Medline: 30556733]
27. Lemyre A, Poliakova N, Bélanger RE. The relationship between tobacco and cannabis use: a review. Subst Use Misuse 2019;54(1):130-145. [doi: 10.1080/10826084.2018.1512623] [Medline: 30362881]
28. Berg CJ, Payne J, Henriksen L, Cavazos-Rehg P, Getachew B, Schauer GL, et al. Reasons for marijuana and tobacco co-use among young adults: a mixed methods scale development study. Subst Use Misuse 2018 Feb 23;53(3):357-369 [FREE Full text] [doi: 10.1080/10826084.2017.1327978] [Medline: 28792283]
29. Giovenco DP, Miller Lo EJ, Lewis MJ, Delnevo CD. "They're pretty much made for blunts": product features that facilitate marijuana use among young adult cigarillo users in the United States. Nicotine Tob Res 2017 Nov 01;19(11):1359-1364 [FREE Full text] [doi: 10.1093/ntr/ntw182] [Medline: 27613943]
30. Antognoli E, Cavallo D, Trapl E, Step M, Koopman Gonzalez S, Perez R, et al. Understanding nicotine dependence and addiction among young adults who smoke cigarillos: a qualitative study. Nicotine Tob Res 2018 Feb 07;20(3):377-382 [FREE Full text] [doi: 10.1093/ntr/ntx008] [Medline: 28184877]

31. Thrul J, Gubner NR, Nguyen N, Nguyen C, Goodell EA, Holmes LM, et al. Perceived reward from using cigarettes with alcohol or cannabis and concurrent use: A smartphone-based daily diary study. Addict Behav 2021 Mar;114:106747 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106747] [Medline: 33307406]
32. Gubner NR, Thrul J, Kelly OA, Ramo DE. Young adults report increased pleasure from smoking cigarettes when drinking alcohol but not when using marijuana. Addict Res Theory 2018;26(1):71-76 [FREE Full text] [doi: 10.1080/16066359.2017.1311877] [Medline: 29371859]
33. Thrul J, Gubner NR, Tice CL, Lisha NE, Ling PM. Young adults report increased pleasure from using e-cigarettes and smoking tobacco cigarettes when drinking alcohol. Addict Behav 2019 Jun;93:135-140 [FREE Full text] [doi: 10.1016/j.addbeh.2019.01.011] [Medline: 30710807]
34. Drope J, Liber AC, Cahn Z, Stoklosa M, Kennedy R, Douglas CE, et al. Who's still smoking? Disparities in adult cigarette smoking prevalence in the United States. CA Cancer J Clin 2018 Mar;68(2):106-115 [FREE Full text] [doi: 10.3322/caac.21444] [Medline: 29384589]
35. Chen-Sankey JC, Mead-Morse EL, Le D, Rose SW, Quisenberry AJ, Delnevo CD, et al. Cigar-smoking patterns by race/ethnicity and cigar type: a nationally representative survey among U.S. adults. Am J Prev Med 2021 Jan;60(1):87-94 [FREE Full text] [doi: 10.1016/j.amepre.2020.07.005] [Medline: 33341182]
36. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
37. McClure EA, Rabin RA, Lee DC, Hindocha C. Treatment implications associated with cannabis and tobacco co-use. Curr Addict Rep 2020 Dec;7(4):533-544 [FREE Full text] [doi: 10.1007/s40429-020-00334-8] [Medline: 33777645]
38. Nguyen N, Neilands T, Lisha N, Lyu J, Olson S, Ling P. Longitudinal associations between use of tobacco and cannabis among people who smoke cigarettes in real-world smoking cessation treatment. J Addict Med 2022;16(4):413-419 [FREE Full text] [doi: 10.1097/ADM.0000000000000920] [Medline: 34619713]
39. Berg CJ, Krishnan N, Graham AL, Abroms LC. A synthesis of the literature to inform vaping cessation interventions for young adults. Addict Behav 2021 Aug;119:106898 [FREE Full text] [doi: 10.1016/j.addbeh.2021.106898] [Medline: 33894483]
40. Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Stat Med 1998;17(5-7):781-796. [doi: 10.1002/(sici)1097-0258(19980315/15)17:5/7<781::aid-sim821>3.0.co;2-o] [Medline: 9549823]
41. Glantz S, Slinker B, Neilands T. Primer of Applied Regression and Analysis of Variance. 3rd Ed. Columbus, OH: McGraw Hill Education; 2016.

Abbreviations

AOR: adjusted odds ratio

Edited by L Buis; submitted 04.07.22; peer-reviewed by T Chung, A Sokolovsky, R Gunn; comments to author 29.11.22; revised version received 10.01.23; accepted 07.02.23; published 20.02.23.

Please cite as:
Nguyen N, Thrul J, Neilands TB, Ling PM
Associations Between Product Type and Intensity of Tobacco and Cannabis Co-use on the Same Day Among Young Adult Smokers: Smartphone-Based Daily-Diary Study
JMIR Mhealth Uhealth 2023;11:e40736
URL: https://mhealth.jmir.org/2023/1/e40736
doi:10.2196/40736
PMID:36806440

Original Paper

Evaluating the Effects of the Supportive Parenting App on Infant Developmental Outcomes: Longitudinal Study

Shefaly Shorey[1], PhD; Yap Seng Chong[2], MBBS; Luming Shi[3], MBBS; Jing Shi Chua[1], BSocSci; Thilagamangai[4]; Jancy Mathews[5], MN; Siew Hoon Lim[6], PhD; Ruochen Du[2], MPH; Yiong Huak Chan[2], PhD; Thiam Chye Tan[7], MMed; Cornelia Chee[8], MMed; Evelyn Law[8], MD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore
[2]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[3]Singapore Clinical Research Institute, Singapore, Singapore
[4]Division of Nursing, KK Women's and Children's Hospital, Singapore, Singapore
[5]National University Polyclinics, Singapore, Singapore
[6]Singapore General Hospital, Singapore, Singapore
[7]Mount Elizabeth Novena Specialist Centre, Singapore, Singapore
[8]National University Hospital, Singapore, Singapore

Corresponding Author:
Shefaly Shorey, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Clinical Research Centre, Level 2, Block MD11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294
Email: nurssh@nus.edu.sg

Abstract

Background: Previous studies have investigated the various effects of parenting on infant developmental outcomes. In particular, parental stress and social support have been found to significantly affect the growth of the newborn. Although many parents today use mobile apps to obtain more support in parenting and perinatal care, few studies have examined how these apps could affect infant development.

Objective: This study aimed to examine the effectiveness of the Supportive Parenting App (SPA) in improving infant developmental outcomes during the perinatal period.

Methods: This study adopted a 2-group parallel prospective longitudinal design and recruited 200 infants and their parents (N=400 mothers and fathers). The parents were recruited at 24 weeks of gestation for a randomized controlled trial conducted from February 2020 to July 2022. They were randomly allocated to either the intervention or control group. The infant outcome measures included cognition, language, motor skills, and social-emotional development. Data were collected from the infants when they were aged 2, 4, 6, 9, and 12 months. Linear and modified Poisson regressions were used to analyze the data to examine between- and within-group changes.

Results: At 9 and 12 months post partum, the infants in the intervention group were found to have better communication and language skills than those in the control group. An analysis of motor development revealed that a larger proportion of the infants in the control group fell under the at-risk category, where they scored approximately 2 SDs below the normative scores. The control group infants scored higher on the problem solving domain at 6 months post partum. However, at 12 months postpartum, the infants in the intervention group performed better on cognitive tasks than those in the control group. Despite not being statistically significant, the intervention group infants were found to have consistently scored better on the social components of the questionnaires than the control group infants.

Conclusions: Overall, the infants whose parents had received the SPA intervention tended to fare better in most developmental outcome measures than those whose parents had received standard care only. The findings of this study suggest that the SPA intervention exerted positive effects on the communication, cognition, motor, and socioemotional development of the infants.

Further research is needed to improve the content and support provided by the intervention to maximize the benefits gained by infants and their parents.

Trial Registration: ClinicalTrials.gov NCT04706442; https://clinicaltrials.gov/ct2/show/NCT04706442

(JMIR Mhealth Uhealth 2023;11:e43885) doi:10.2196/43885

KEYWORDS

infant development; parenting; mobile health technology; social support; psychoeducation; peer support; mobile phone

Introduction

Background

The effects of parenting on infant development are a widely investigated topic. Studies have found that parenting knowledge, parental stress, and parental perceived support have significant impacts on the growth of an infant [1-3]. For example, parents' knowledge of and participation in their child's play contribute significantly to the development of the child, as it helps boost executive functioning, encourage prosocial behavior, and enhance creativity [4]. Greater parenting stress and low levels of perceived social support have also been found to be associated with depression among mothers, which is correlated with developmental delays in infants [1].

The lack of social support has been linked to various parental outcomes, such as postnatal depression [5], parental stress [6], and anxiety [7]. Social support is often regarded as a protective factor for parents, especially during the perinatal period. Receiving support from others often helps parents feel less overwhelmed, aiding them with the transition to parenthood or helping them cope with having to care for multiple children. Some studies [8-10] conducted in Singapore found that parents desire more informational and familial support. Specifically, some parents perceived having a lack of knowledge of infant care and wanted to have access to reliable information sources [8]. With the rapid advancements in technology over the last 2 decades, parents today tend to look for information from web-based sources or seek support from web-based parenting communities, as these sources are extremely convenient and accessible [10-12]. They are mostly aware that information found on the web can be fabricated or exaggerated [12]; therefore, they tend to prefer gathering information directly from health care professionals such as obstetricians and neonatologists [8]. This is not always possible, especially during the postnatal period, as parents no longer have regular appointments with their obstetricians or gynecologists. As a result, parents try to obtain accurate information on the web by visiting only reputable sites that disseminate information provided by health care professionals or going to less commercially based websites [13]. Even then, these sites might not be consistent in the information they provide, especially when the information is not contextualized and does not incorporate cultural norms, and this may cause confusion among parents regarding what the right childcare practices are [14]. This justifies the need to create evidence-based programs tailored to the present generation of tech-savvy parents to improve their well-being and aid them in developing competent childcare skills [15].

According to Milgrom et al [5], it is recommended to implement programs to improve parental well-being during the perinatal period, as social support plays a large role in mediating the relationship between postnatal depression and child development during this time. During the perinatal period, low levels of social support such as insufficient partner support [5], lack of reliable information sources [16], and caring for a newborn without aid from others [17] often induce much stress and negative moods in parents. To fulfill the support needs of parents during the perinatal period, a mobile health (mHealth) app known as the Supportive Parenting App (SPA) was developed. The mHealth SPA was developed as a one-stop resource center because past studies have found technology-based interventions to be effective in offering parenting education and support [18,19]. Such remote interventions were found to be helpful specifically for parents who were facing childcare issues but were not always able to seek immediate advice from health care professionals [18,19].

The SPA is a theory- and evidence-based psychoeducational app developed using different theoretical frameworks, such as Singh et al's [20] mHealth user engagement pyramid, Bandura's [21] social cognitive theory, and Bowlby's [22] attachment theory. Through SPA, parents were able to obtain information on childcare and parenting-related topics to aid them in their parenting journey. In addition, unique to SPA, a peer support feature was included to provide parents with emotional support from trained peer volunteers. Although various parenting interventions have been developed and evaluated to improve parental outcomes, a recent review by Adina et al [23] found that few studies have explored how these interventions can indirectly affect infant or child developmental outcomes. This is unexpected, as the improvement of child developmental outcomes is often cited as a reason for developing these parenting programs [1,15]. Therefore, although these interventions are often directed at parents, it is important to examine how the development of infants may be affected as a consequence.

Aims and Hypotheses

This study aimed to examine whether the SPA intervention had any indirect effects on the developmental outcomes of the app users' infants from birth to 12 months of age. The direct effects on parenting outcomes have been reported separately [24]. It was hypothesized that infants in the intervention group would exhibit better language, motor, cognitive, and social skills than their counterparts in the control group.

Methods

Study Design

A 2-group parallel prospective longitudinal design was adopted for this study, which was conducted from February 2020 to July 2022. Expecting parents were recruited from 2 public health care institutions in Singapore. The study was part of a randomized controlled trial (RCT) investigating the effectiveness of SPA in improving perinatal parental outcomes such as postnatal depression and anxiety [24]. Along with their parents, the infants in this study were randomly allocated to either the SPA intervention group or standard care control group.

Eligibility Criteria

Parents were considered eligible for the study if they met the following criteria: (1) both parents were aged ≥21 years; (2) both parents were able to read and speak English; (3) the pregnancy was at low risk with >24 weeks of gestation (age of viability in Singapore); and (4) both parents owned a smartphone with internet access. Parents were excluded from the study if they had high-risk pregnancies (eg, pregnancy-induced hypertension, preeclampsia, and placenta previa major). Infants who were born via a complicated assisted delivery where the mother required prolonged hospitalization and admitted to the neonatal intensive care unit and infants with congenital issues were excluded from the study to minimize confounding influences on the outcome variables.

Sample Size Calculation

As this study was part of an RCT investigating the effectiveness of the SPA intervention on parental outcomes, the parents enrolled in the RCT and their infants were recruited for this study. Considering the medium-sized effect of SPA, a Cohen *d* of 0.5 (90% power and 0.05 significance), and an attrition rate of 20% (based on another study) [25], 200 couples were recruited for the main RCT. Two couples had twins; therefore, 202 infants were recruited for this study.

Intervention

The control group parents received the standard perinatal care offered by the hospitals they were recruited from, which consisted of antenatal checkups, optional antenatal classes, care during their stay in the ward, and a postnatal review scheduled 6 weeks post partum. Perinatal care was provided to the parents by obstetricians, nurses, neonatologists, and lactation consultants. The intervention group parents received the standard perinatal care as well, but they were also granted access to the mHealth intervention SPA upon recruitment into the study. In addition, they were matched with trained peer volunteers, who were experienced mothers trained by the research team to provide peer support for the parents in the RCT.

SPA included a variety of pregnancy-, childbirth-, postpartum-, and infant care–related information. This included articles, audio files, and videos about birth preparation, bonding and attachment across the perinatal period, breastfeeding, baby care–related tasks (from bathing to safe sleep habits), and involvement of both fathers and mothers in baby care tasks. The information was curated by the health care professionals involved in the study so that parents could conveniently access reliable and accurate information. Expert advice, discussion forums, and frequently asked questions were also features of the mobile app that aimed to resolve any pregnancy- or childcare-related queries that the parents might have. The parents were encouraged to interact with the peer volunteer with whom they were matched if they needed emotional or informational support from experienced mothers who had previously had and recovered from postnatal depression. Detailed features of the SPA mobile app and peer volunteer intervention can be found in the published development study [26]. The SPA intervention was made available to the intervention group parents from the point of recruitment until 6 months post partum.

Procedure

Couples were recruited by a research assistant during their scheduled antenatal checkups at 2 tertiary hospitals in Singapore. They were provided with an explanation of the study, and interested couples were screened for eligibility before giving them an informed consent form where they could indicate their willingness to participate in the study. Subsequently, the couples were randomly allocated to the intervention or control group. The estimated due date of the couples was recorded, and the couples were then contacted shortly after their due date to gather information regarding their childbirth (eg, gender of the baby and whether they attended prenatal classes). The parents also entered this information into SPA so that the app could send them information that is specific and relevant to the infant's age and respective postpartum time points.

The parents were contacted via SMS text messages to complete the follow-up questionnaires at 1, 2, 4, 6, 9, and 12 months post partum. Mothers tended to be the ones who completed the infant-related questionnaires. A house visit was also scheduled at 6 and 12 months, during which a trained research assistant visited the participants' homes to assess the infant using the Bayley-4.

Outcome Measures

Conducting research with very young children involves various challenges regarding the accuracy of the data collected, as infants are not verbal, and thus it is difficult to obtain information directly from them. Therefore, the following instruments were used to measure the constructs examined in this study to provide an accurate representation of the infants' developmental progress.

Ages and Stages Questionnaire—Third Edition

The parent-reported Ages and Stages Questionnaire (ASQ) was used to measure the development of the infants across 5 domains: personal-social, gross motor, fine motor, problem solving, and communication [27]. There are 21 sets of ASQ, each catering to a different developmental time point; these can be used to assess infants or children aged 2 to 66 months. Existing literature has found that Cronbach α for the ASQ ranges from .49 to .87, depending on the domain and time point [28]. In this study, the ASQ sets were administered at 2, 4, 6, 9, and 12 months. Each set of ASQ consisted of 30 items, and the parents were asked to select "yes" (10 points), "sometimes" (5 points), or "not yet" (0 points) to indicate whether their child had demonstrated the milestone described in that particular item. Cutoff scores, which were 2 SDs below the normative mean, were provided to indicate whether the infant's development required further monitoring and assessment (Table 1). Cronbach α for the ASQ sets administered in this study are also presented in Table 1.

Table 1. Cronbach α and cutoff scores for the Ages and Stages Questionnaire.

	2 months	4 months	6 months	9 months	12 months
Cronbach α	.746	.883	.864	.877	.910
Cutoff scores[a]					
Communication	22.77	34.60	29.65	13.97	15.64
Gross motor	41.84	38.41	22.25	17.82	21.49
Fine motor	30.16	29.62	25.14	31.32	34.50
Problem solving	24.62	34.98	27.72	28.72	27.32
Personal-social	33.71	33.16	25.34	18.91	21.73

[a]Scores below the cutoff indicate that the child could be at risk of neurodevelopmental conditions, and further assessment with a professional might be needed.

Bayley Scales of Infant and Toddler Development—Fourth Edition

The Bayley-4 consists of 5 scales: cognitive, language, motor, social-emotional, and adaptive behavior (ADBE) [29]. The assessment was administered by a research assistant involved in the study and scored based on the research assistant's observation of the infant's performance in various tasks. The Bayley-4 assessment was conducted at 6 and 12 months post partum, and the number of items administered varied depending on the infant's performance. The Bayley-4 items were scored on a 3-point Likert scale ranging from 0 to 2. Points were added to form a total raw score, which could be converted into scaled or standard scores. Cutoff scores were also provided, where standard scores <85 were marked as "at-risk" to indicate possible developmental delay [30]. Prior studies [31,32] that administered Bayley-III assessments have reported Cronbach α ranging from .88 to .96 across all scales. Cronbach α of the newest Bayley-4 assessment has not been reported in the existing literature; in this study, the average Cronbach α of the Bayley-4 was found to be .61.

Brief Infant Toddler Social Emotional Assessment

The 42-item parent-reported Brief Infant Toddler Social Emotional Assessment (BITSEA) was divided into 2 scales: the problem scale (31 items) and the competence scale (11 items) [33]. The BITSEA items were scored on a 3-point Likert scale ranging from 0 to 2. The scores for the items in each scale were added to obtain the respective total score. Higher scores on the problem scale indicated a higher frequency and range of behavioral and emotional problems, whereas higher scores on the competence scale indicated a higher level of social competence. For the competence scale, the cutoff score was 11, whereas for the problem scale, the cutoff score was 13 for girls and 12 for boys. The BITSEA was administered only during the 12-month follow-up, as it is meant to be administered to infants from 12 months onward. The Cronbach α for the BITSEA was .71, similar to that in a previous study [34], where the Cronbach α for the problem and competence scales were .82 and .72, respectively.

Data Analysis

Data analyses were conducted using SPSS (version 27.0; IBM Corp) [35], and statistical significance was set at $P<.05$. Descriptive statistics are presented as mean (SD) for continuous variables and as n (%) for categorical variables. Linear regression was used to examine the association between continuous outcome scores and the intervention adjusted for baseline measures and other covariates. Each participant's score on the 3 instruments was subsequently categorized based on the cutoff scores determined by the respective developers of each instrument. Modified Poisson regression was used to analyze the association between binary outcome scores and the intervention adjusted for baseline measurements and other covariates based on the cutoff scores for each instrument. This was done to compare the proportion of infants in each group who fell into the at-risk category for each domain. Established correlations between the main outcomes and covariates based on previous studies [36,37] were used to determine which covariates were needed to be statistically corrected for.

Ethics Approval

Before the commencement of the study, ethics approval was obtained from the National Health Group Domain Specific Review Board (NHG DSRB:2019/00875). The parents of the infants involved in this study were provided with information on the study and its procedures before they provided their written consent. It was communicated to the parents that participation was voluntary and that they had the right to withdraw anytime without incurring any consequences.

Results

Overview

In total, 200 couples and their infants were recruited for this study. However, owing to the attrition rate of 28.5% that was reported in the main RCT [24], only 79% (158/200) of infants were included in the analysis (the remaining 42/200, 21% parent-infant dyads dropped out of the study). The demographic characteristics of the participants are presented in Table 2. The mean age of the parents was 31.4 (SD 4.93) years, and Malay

(115/316, 36.4%) and Chinese (125/316, 39.6%) were the most common ethnicities. Most (87/316, 55.1%) of the infants were male. Most (257/316, 81.3%) parents did not attend any prenatal courses.

Table 2. Demographic characteristics of the parents and their infants.

Demographic characteristics	Intervention group	Control group
Age of parents, mean (SD)		
Mothers	29.9 (4.2)	30.5 (4.2)
Fathers	32.1 (4.9)	33.3 (5.4)
Parent's ethnicity (intervention: n=83; control: n=75), n (%)		
Mothers	83 (100)	75 (100)
Chinese	31 (37.3)	28 (37.3)
Malay	32 (38.6)	27 (36)
Indian	14 (16.9)	9 (12)
Others	6 (7.2)	11 (14.7)
Fathers	83 (100)	75 (100)
Chinese	35 (42.2)	31 (41.3)
Malay	29 (34.9)	27 (36)
Indian	14 (16.9)	10 (13.3)
Others	5 (6)	7 (9.3)
Sex of baby (intervention: n=86; control: n=72), n (%)[a]		
Male	44 (51.2)	43 (59.7)
Female	42 (48.8)	29 (40.3)
The educational level of parents (intervention: n=83; control: n=75), n (%)		
Mothers	83 (100)	75 (100)
Primary school	0 (0)	0 (0)
Secondary school	11 (13.3)	4 (5.3)
ITE[b], polytechnic, or junior college	25 (30.1)	35 (46.7)
University	47 (56.6)	36 (48)
Fathers	83 (100)	75 (100)
Primary school	0 (0)	2 (2.7)
Secondary school	17 (20.5)	7 (9.3)
ITE, polytechnic, or junior college	27 (32.5)	34 (45.3)
University	39 (47)	32 (42.7)
Monthly household income (SGD $; intervention: n=164; control: n=147), n (%)[c]		
<1000 (<US $761.88)	14 (8.5)	5 (3.4)
1000-3000 (US $761.88-$2285.64)	34 (20.7)	40 (27.2)
3000-5000 (US $2285.64-$3809.39)	50 (30.5)	47 (32.0)
>5000 (>US $3809.39)	66 (40.2)	55 (37.4)
Attended prenatal courses (intervention: n=166; control: n=150), n (%)		
Yes	33 (19.9)	26 (17.3)
No	133 (80.1)	124 (82.7)

[a]Only 158 infants (86 in the intervention group and 72 in the control group) were included in the analysis because 42 parent-infant dyads dropped out of the study by 6 months post partum.

[b]ITE: Institute of Technical Education.

[c]Not all parents provided this information.

Communication

The mean and SD scores for the ASQ and Bayley-4 are presented in Tables 3 and 4, respectively, along with the proportion of infants who scored below the cutoff scores (labeled as "at-risk"). Results from the generalized linear regression model indicated that the infants in the intervention group scored significantly higher on the communication domain of the ASQ at 6 (effect size=3.31, 95% CI 0.10-6.53; *P*=.04) and 9 (effect size=6.14, 95% CI 0.90-11.38; *P*=.02) months post partum. However, this difference was not significant after the Bonferroni adjustment (at 4 months: *P*=.22; at 6 months: *P*=.11). The Poisson regression results showed that the intervention group infants were less likely to fall under the at-risk category, but there were only a few at-risk cases (Table 3); therefore, the estimation might not be reliable. Results from the linear (effect size=9.304, 95% CI 5.58-13.13; *P*<.001) regression model of the 12-month Bayley-4 assessment also showed that the infants from the intervention group tended to perform better than those from the control group on the language scale.

Table 3. Ages and Stages Questionnaire scores based on domains.

	2 months (n=158)		4 months (n=146)		6 months (n=143)		9 months (n=146)		12 months (n=140)	
	Intervention (n=87)	Control (n=71)	Intervention (n=83)	Control (n=63)	Intervention (n=76)	Control (n=67)	Intervention (n=81)	Control (n=65)	Intervention (n=78)	Control (n=62)
Communication										
Values, mean (SD)	48.72 (12.16)	48.68 (12.97)	53.35 (7.83)	53.11 (8.69)	50.47 (7.54)	47.31 (9.52)	43.38 (14.82)	37.71 (13.56)	48.03 (11.53)	46.11 (13.09)
At risk[a], n (%)	4 (4.7)	4 (5.6)	1 (2)	1 (2.7)	0 (0)	1 (1.9)	2 (2.9)	2 (4.2)	1 (1.5)	2 (4.4)
Gross motor										
Values, mean (SD)	52.38 (11.50)	49.65 (10.76)	54.02 (8.78)	50.41 (10.30)	43.26 (11.82)	41.11 (14.90)	42.57 (16.10)	41.35 (17.59)	48.64 (13.60)	48.11 (12.58)
At risk, n (%)	11 (12.8)	15 (20.8)	4 (7.8)	5 (13.5)	4 (6.1)	5 (9.6)	6 (8.8)	5 (10.4)	2 (3)	2 (4.4)
Fine motor										
Values, mean (SD)	47.56 (10.28)	45.49 (10.08)	46.47 (14.36)	45.00 (13.54)	45.85 (12.52)	42.12 (15.91)	46.91 (13.74)	45.00 (12.68)	46.97 (13.21)	49.00 (11.01)
At risk, n (%)	10 (11.6)	8 (11.1)	6 (11.8)	3 (8.1)	5 (7.6)	11 (21.2)	12 (17.6)	6 (12.5)	10 (15.2)	4 (8.9)
Problem solving										
Values, mean (SD)	43.08 (15.93)	43.75 (12.58)	48.63 (12.77)	48.47 (11.52)	45.77 (12.75)	60.10 (138.72)	42.65 (14.36)	41.28 (12.75)	40.23 (15.60)	40.22 (15.11)
At risk, n (%)	11 (12.8)	4 (5.6)	8 (15.7)	5 (13.5)	7 (10.6)	9 (17.3)	9 (13.2)	9 (18.8)	14 (21.2)	12 (26.7)
Personal-social										
Values, mean (SD)	48.60 (10.62)	47.57 (10.45)	49.12 (10.43)	48.38 (11.49)	45.62 (11.84)	43.65 (12.88)	37.57 (12.77)	34.26 (12.51)	38.38 (15.74)	37.33 (15.90)
At risk, n (%)	7 (8.1)	8 (11.1)	4 (7.8)	4 (10.8)	5 (7.6)	7 (13.5)	3 (4.4)	4 (8.3)	12 (18.2)	9 (20)

[a]The at-risk group refers to infants who scored below the cutoff scores stated in Table 1.

Table 4. Mean and SD of the Bayley-4 standard scores based on domains.

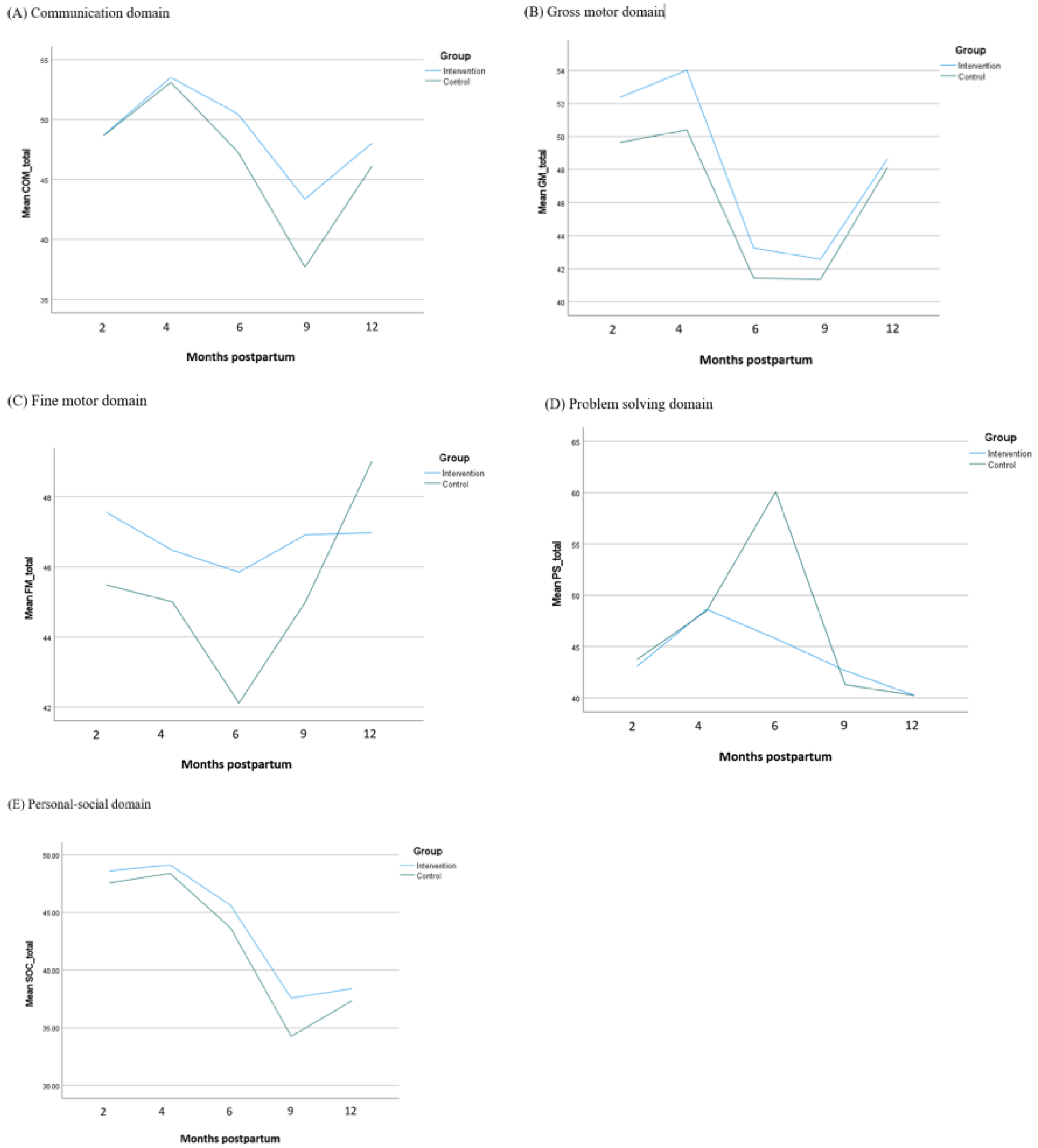
	6 months (n=109)		12 months (n=105)	
	Intervention (n=59)	Control (n=50)	Intervention (n=55)	Control (n=50)
Cognitive				
Values, mean (SD)	100.76 (9.64)	95.12 (15.59)	98.09 (10.34)	89.52 (10.41)
At risk[a], n (%)	3 (5.1)	6 (12)	3 (5.5)	10 (23.8)
Motor				
Values, mean (SD)	—[b]	—	100.18 (10.36)	98.64 (10.76)
At risk, n (%)	—	—	5 (9.1)	8 (19)
Language				
Values, mean (SD)	—	—	95.42 (7.87)	93.02 (6.91)
At risk, n (%)	—	—	4 (7.3)	7 (16.7)
Social-emotional				
Values, mean (SD)	—	—	97.09 (20.68)	100.00 (18.35)
At risk, n (%)	—	—	15 (27.3)	7 (16.7)
Adaptive behavior				
Values, mean (SD)	99.88 (9.20)	96.96 (9.43)	98.18 (8.71)	96.10 (8.57)
At risk, n (%)	3 (5.1)	4 (8)	3 (5.5)	4 (9.5)

[a]The at-risk group refers to infants with standard scores <85.

[b]Data not available.

Figure 1A shows the changes in the ASQ scores in the communication domain from 2 to 12 months post partum in both groups. The infants from both groups showed similar trends: there was an initial increase in communication scores at 2 and 4 months post partum, and, subsequently, there was a steep decrease in communication scores from the 4- to 9-month time points before they increased again at 12 months post partum. From 4 months onward, the infants in the intervention group scored higher in the communication domain than those in the control group. The largest difference was observed at 9 months post partum. Overall, the infants from the intervention group demonstrated better communication skills than those from the control group.

Figure 1. Trend graphs for the changes in the Ages and Stages Questionnaire scores over time: (A) communication domain, assessing speech and language; (B) gross motor domain, assessing ability to produce large movements; (C) fine motor domain, assessing smaller movements; (D) problem solving domain, assessing cognitive and intellectual skills; (E) personal-social domain, assessing emotional and social skills.



Motor Skills

The infants from the control group were significantly more likely to score below the cutoff score of the ASQ gross motor domain at 2 months post partum (risk ratio [RR]=0.417, 95% CI 0.20-0.85; *P*=.02). Although the infants from the intervention group tended to score higher than those from the control group in the gross motor domain (Table 3), this difference was not significant (*P*=.71).

The intervention group infants were found to have better fine motor skills than the control group infants at 6 months post partum, based on the results of the logistic regression model analysis. The infants from the control group were more likely to score below the cutoff score on the ASQ fine motor domain than those from the intervention group (RR=0.25, 95% CI 0.08-0.76; *P*=.02). In addition, the infants from the intervention group had significantly higher scores on the ASQ fine motor domain than those from the control group (effect size=6.02, 95% CI 1.03-11.02; *P*=.02). However, there were no significant differences between both groups in the Bayley-4 motor scale scores at 12 months post partum (effect size=−4.91, 95% CI −12.96 to 3.14; *P*=.23).

Similar to the trend graph of the ASQ communication domain, the ASQ gross motor graph showed that the intervention group scored better than the control group on the gross motor items (Figure 1B). The gross motor scores of both groups slightly increased during the first 2 time points before sharply decreasing at 6 months post partum. There was a steady increase in the gross motor scores of both groups from 9 to 12 months post partum, accompanied by a decreasing difference in gross motor scores. Figure 1C shows the trend graph of the ASQ fine motor scores. The fine motor scores of both groups of infants reduced from 2 to 6 months post partum before gradually increasing again. At 12 months post partum, the infants from the control group had higher scores than those from the intervention group, but this difference was not significant (effect size=−0.98, 95% CI −5.87 to 3.91; *P*=.69).

Cognition

At 6 months post partum, the infants from the control group had a higher chance of being in the at-risk group for the problem solving domain of the ASQ than those from the intervention group (RR=0.34, 95% CI 0.12-0.91; *P*=.03). However, the linear regression model did not find any significant differences in problem solving scores between the 2 groups (effect size=−16.88, 95% CI −52.34 to 18.59; *P*=.35). The infants from the intervention group fared significantly better than their control group counterparts on the cognition scale of the Bayley-4 assessment (effect size=9.30, 95% CI 5.48-13.12; *P*<.001) at 12 months post partum.

According to Figure 1D, the infants from the control group generally scored higher in the problem solving domain than those from the intervention group. There was a sharp increase in the problem solving scores of the control group infants at the 6-month time point. Following this increase, an equally sharp decrease in problem solving scores was found at 9 months post partum, where the control group scores fell below those of the intervention group.

Social-Emotional Skills

No significant group differences were found in the scores related to the social-emotional skills of the infants across all time points (at 2 months: *P*=.28; at 4 months: *P*=.61; at 6 months: *P*=.17; at 9 months: *P*=.06; at 12 months: *P*=.57). This was true for the personal-social domain of the ASQ, the social-emotional scale of the Bayley-4 assessment, and the competence scale of the BITSEA.

Figure 1E shows the changes in the ASQ personal-social scores of the infants from both groups. Scores on the personal-social domain were relatively high during the 2- and 4-month time points but decreased from 4 to 9 months post partum. The personal-social scores increased subsequently at 12 months post partum. Overall, the intervention group appeared to perform better in terms of social-emotional skills than the control group.

Behavioral Outcomes

The ADBE scale of the Bayley-4 and the problem scale of the BITSEA assessed the behavioral outcomes of the infants. The BITSEA problem scale covered externalizing behaviors, dysregulating behaviors, and maladaptive behaviors. The infants from the control group were found to have significantly higher scores on the BITSEA problem scale than those from the intervention group (effect size=−5.87, 95% CI −10.44 to −1.70; *P*=.006). As such, the control group infants tended to exhibit more problem behaviors, as described in the BITSEA. By contrast, the ADBE scale mainly focused on the infants' ability to engage in functional developmental tasks that are critical to their survival. This included feeding oneself and communicating basic needs. No significant group differences in the ADBE scores were found at both 6 months (effect size=2.05, 95% CI −2.74 to 6.84; *P*=.40) and 12 months (effect size=0.954, 95% CI −2.53 to 4.44; *P*=.59) post partum.

Analysis of Covariates

Parents attending prenatal courses was found to significantly influence whether their infants' ASQ scores on the communication (RR=0.13, 95% CI 0.02-0.75; *P*=.01), gross motor (RR=0.43, 95% CI 0.21-0.90; *P*=.03), and personal-social (RR=0.28, 95% CI 0.10-0.82; *P*=.02) domains fell below the cutoff at 2 months post partum. The infants whose parents had attended prenatal courses tended to score better in these domains.

The education level of the parents was also a predictor of the infants' motor skills. The infants of parents with secondary educational qualifications had a higher chance of being in the at-risk group at 6 months post partum than those of parents who graduated from universities (RR=13.41, 95% CI 2.27-79.09; *P*=.004). The generalized linear regression model for fine motor skills at 6 months post partum also showed that the infants of parents who received up to secondary-level education scored significantly lower than those of parents who graduated from universities (effect size=−13.55, 95% CI −24.95 to −2.15; *P*=.02). Thus, the results suggest that the infants of parents with higher educational levels tended to have developed better motor skills at 6 months post partum.

In this study, monthly household income was found to significantly affect the cognition and motor skills of the infants. The infants from households with a monthly income between SGD $3000 (US $2285.64) and SGD $5000 (US $3809.39) were more likely to belong to the at-risk group of the cognition domain of the Bayley-4 assessment than those from households earning >SGD $5000 (US $3809.39) monthly (RR=14.79, 95% CI 4.96-44.15; *P*<.001). Those with higher monthly household income also scored higher on the motor skills domain of the Bayley-4 assessment at 12 months post partum. Those with household income >SGD $5000 (US $3809.39) per month scored significantly higher in the motor skills domain than those with a household income of SGD $3000 (US $2285.64) to SGD $5000 (US $3809.39) per month (effect size=−10.65, 95% CI −15.42 to −5.89; *P*<.001) and those with a household income <SGD $1000 (US $761.88) per month (effect size=−14.97, 95% CI −22.97 to −5.17; *P*=.002). The generalized linear regression model also found similar results with the gross motor scale of the ASQ (effect size=−9.91, 95% CI −15.76 to −4.06; *P*<.001). Nonetheless, the modified Poisson regression model did not find any significant differences in the number of infants at risk for delayed motor skill development based on their income groups (*P*=.68).

Discussion

Principal Findings

This study examined the effects of the SPA intervention on infants' developmental outcomes during the first 12 months of life. The infants in the intervention group mostly scored better in domains assessing communication, cognition, and social-emotional development. More infants from the control group fell under the at-risk category for motor skills than those from the intervention group. Findings from the main RCT reported a high attrition rate of 28.5% [24]. The outbreak of the COVID-19 pandemic in early February 2020 may have affected the parents' and infants' participation in the study, especially for the home visits. Hence, this led to a smaller-than-expected sample size, which likely affected the statistical power of the study. In general, although the infants from the intervention group tended to exhibit better developmental outcomes than those from the control group, these differences were modest. Therefore, the results of this study are not completely in line with the hypothesis.

Communication

From 2 to 12 months post partum, the infants from the intervention group were found to exhibit better communication skills than their control group counterparts. According to Bortfeld and Gabouer [38], early infant communication lays the groundwork for language. Previous studies [38,39] have emphasized that communication development often begins in the womb, where the fetus receives auditory inputs that enable them to start learning how to distinguish sounds. The SPA knowledge base included content encouraging the intervention group parents to communicate with their newborn and suggested ways to enhance parent-child interactions, even during pregnancy. For example, expecting parents can respond to the kicks made by the fetus during late pregnancy to communicate with them. With newborns, parents can play soothing music or read stories aloud to them to facilitate better communication and language development. According to the SPA traffic data, many parents in the intervention group accessed these materials in the mobile app and thus communicated with their infants more effectively, boosting their communication development. The decrease in communication scores from 4 to 9 months in both groups was unsurprising, as normative scores on the ASQ vary across ages and domains. For communication, the normative scores were approximately 42.5 at 4 months, 34 at 6 months, and only 30 at 9 months. Therefore, the reduction in mean communication scores during this period does not indicate that the infants' communication abilities did not progress during this period. They mostly performed well above the normative scores.

On the basis of the trend graphs, the difference in the communication scores between the 2 groups increased over time. Topping et al [40] explained that parent interaction during the infant's preliteracy development was important in enhancing the child's future language abilities. Parenting intervention programs have been found to have a positive impact on children's language development. It is possible that the intervention group parents acted upon the things that they had read about, such as child play and infant milestones, which enabled them to interact more with their newborns. This, in turn, would have enhanced their child's future language development. Therefore, the implementation of the SPA intervention might have led to the widening difference in communication scores between the 2 groups from 6 months onward. Future research could investigate in greater detail the relationship between the materials that parents read and what they put into practice.

Motor Skills

Similar to the communication trend graphs, the fluctuations in gross and fine motor skills indicated by the ASQ scores can be attributed to the differences in normative scores. Overall, the motor skills of the intervention group infants developed to a greater extent than those of the control group infants. This might have been because the parents in the intervention group had read the guide on how they could engage in play with their children on SPA. The guide included some toy recommendations that can help improve the motor skills of infants by allowing them to practice movements such as grasping and head turning. Semistructured interviews with some of the parents [10] revealed that the parents enjoyed having access to SPA, as it included various localized information that applied to them, which was unique because other parenting apps were more general. The parents could also anticipate and encourage the growth of their infants, as they read about developmental milestones. Consequently, the infants from the intervention group were exposed to more opportunities to practice and develop their motor skills. Descriptions of developmental milestones were also provided to educate parents on the motor skills that their children should achieve at each stage of development. However, the results suggest that the positive effects of the SPA intervention on the infants' motor skills did not persist beyond 6 months post partum. As the SPA intervention only lasted up to 6 months post partum, information related to infant motor skill development beyond 6 months was rather scarce. The achievement of various motor-related milestones often drastically changes the subsequent behaviors of infants; thus, the toys or games used to engage them during earlier months might be less relevant in facilitating more advanced motor skill development [2]. Further research is needed to determine whether more age-appropriate information on motor skill development would help improve the motor skills of infants aged >6 months.

Cognition

Results from the problem solving domain of the ASQ revealed that the control group infants generally fared better than the intervention group infants during the first 6 months of their lives. However, the results of the 12-month Bayley-4 assessment revealed that the intervention group infants did better on the cognition scale than the control group infants. This finding is contrary to the hypothesis of this study, which proposed that the intervention group infants would perform better on cognition tasks than the control group infants. A reason for this might be that SPA did not include much content on enhancing the cognition and problem-solving skills of infants. Most of the parenting information available was related to childcare tasks such as feeding and swaddling or was related to parent-child communication and motor skills. As reported in a published qualitative paper [10] regarding the perspectives of the parents in the study, the control group parents described how they took the initiative to explore web-based resources and installed other parenting mobile apps to obtain more parenting-related information than what the standard care offered [10]. Some mobile apps used by the control group parents might have included information on how to build on the cognitive development of the infants during the first few months post partum. Therefore, more research is needed to obtain greater insight into the parenting resources used by Singaporean parents and how they influence infant development. More importantly, health care providers developing educational programs such as SPA should consider including content focusing on the holistic development of infants and children.

Social-Emotional Skills

Although there were no statistically significant group differences in the personal-social domain of the ASQ and social-emotional scale of the Bayley-4 assessment, the intervention group generally demonstrated greater social-emotional development. This lack of significance could be attributed to the fact that SPA did not include much information regarding the development of social-emotional skills in infants. However, the encouragement provided to parents to engage in age-appropriate parent-child play and increase parent-child interactions might have contributed to the slightly higher social-emotional and personal-social scores [38]. As mentioned, the COVID-19 pandemic hit Singapore in early 2020, immediately after the study began. During this period, various restrictions were set in place to minimize social interactions to prevent the spread of the virus [41]. This included the closure of infant care and prohibition of social gatherings. Hence, there were fewer opportunities for infants to engage in social situations with other infants and foster peer relationships. Existing literature has found that playing with peers is an important activity that allows for better development of prosocial behaviors and the formation of relationships with others [42]. The lack of such interactions owing to the pandemic might have further undermined the significance of the group difference in social-emotional development.

Behavioral Outcomes

The infants from the intervention group were found to engage in more ADBEs than those from the control group. On the basis of the parents' responses to the BITSEA, the control group infants exhibited more problem behaviors as well. Maternal responsivity and sensitivity to infant distress are important factors in predicting ADBE in infants. Higher maternal responsivity is associated with greater emotional regulation and fewer behavioral issues [43]. Lorber et al [44] also found that well-known predictors of externalizing behavior include daily parenting hassles, authoritarian parenting, and poor parent-child bonding. The parents who received the SPA intervention were more educated on how to interact and bond with their newborns, which possibly led to improved parent-child bonding. They were also provided with support from the peer volunteers, which helped reassure them and provide them with an avenue to discuss parenting-related worries [10]. Ultimately, the parents in the intervention group received more informational, appraisal, and emotional support than those in the control group [10,26]. This helped them better adjust to the newborn care tasks that they had to take on after childbirth. Therefore, this study suggests that the SPA intervention was effective in facilitating the development of ADBEs in infants.

Prenatal Courses, Education, and Household Income

This study found that prenatal classes had a positive impact on communication, gross motor, and personal-social scores at 2 months post partum. Prior research [45] has found that prenatal education courses can help reduce anxiety during pregnancy, which can help improve prenatal bonding. This is especially true for first-time parents, who experience greater fear of childbirth and parenting self-efficacy [46]. Improved maternal well-being would then facilitate better infant outcomes, such as fewer maladaptive tendencies and greater levels of social competence [47]. However, most of the parents in this study did not attend prenatal classes. During interviews, the parents revealed that they were unable to attend these classes, as they were canceled because of the COVID-19 pandemic [10]. This left many new parents unprepared for the transition to parenthood, causing them to feel stressed and clueless. The implementation of social distancing restrictions also amplified this problem, as parents were unable to seek instrumental support from their family members or nannies [10]. Therefore, it is crucial for maternal care institutions to prepare and provide parents with sufficient support, especially through remote means at times when the availability of support is affected. This study was unable to examine how the development of children of new parents and that of children of experienced parents differed, as information regarding whether the parents were new or experienced was not collected in the main RCT. Therefore, it is crucial for future research to gather this information to further examine how the intervention impacts the differing needs of these parents.

The infants from families with higher household incomes were found to have more developed cognition. This is supported by previous literature, where it was found that children from families with low socioeconomic status (SES) had lower cognitive flexibility [48]. It was explained that the consequences of living with low SES are less favorable for children's development. This includes greater exposure to stress, which affects children's performance on cognitive tasks, and reduces maternal sensitivity and verbal stimulation [48]. Unsafe living conditions and stressors associated with low SES may also result in more negative or authoritarian parenting, which can affect cognitive outcomes.

Both higher parental education and higher monthly household income were significantly associated with stronger motor skills. Freitas et al [49] evaluated the relationship between SES and the availability of resources to promote motor development in infants. This study found that SES is a crucial factor influencing the availability of motor affordances at home. Educational level was also found to significantly affect the provision of toys to infants [49]. Parental education often affects the SES of the family, as higher education levels tend to open up job opportunities that offer higher income. Having more income would then lead to the purchase of more play materials or even the ownership of a larger home that provides more physical space for motor development [49,50]. As a result, infants from low-SES families and infants of parents with lower education levels tend to develop motor skills more slowly. Hence, researchers, health care providers, and policy makers may focus their efforts on developing interventions focusing on family factors that contribute to infant developmental outcomes across SES.

Strengths and Limitations

Given that social support has been proven to improve parental well-being and, in turn, promote infant growth in various areas, the SPA intervention was developed. The intervention aimed to meet the support needs of Singaporean parents during the perinatal period, thus helping them adjust to parenting roles and infant-care tasks. This study found that technology-based parenting interventions such as SPA can lead to benefits beyond enhancing parental well-being. The findings of this study are crucial for the future development of not only mobile apps for parents in Singapore but also those for parents in other countries. Providing parenting education and emotional support can indirectly improve infant developmental outcomes. However, it is important to recognize and consider the cultural beliefs, practices, and support needs of parents from other countries. This would allow app developers to provide a knowledge base and appraisal support that would be respectful of and helpful in meeting their individualized needs.

Another strength of this study is that it used questionnaires that were completed by both parents and trained personnel. Parent-completed measures are advantageous in that parents spend the most time with their infants and are the most knowledgeable about them. However, existing literature [51] has also pointed out that there are biases associated with parent-reported questionnaires. Generally, parents are not trained in evaluating the development of infants, which can make them susceptible to overestimating or underestimating their child's abilities and thus render their responses less reliable. By contrast, although the Bayley-4 was administered by a trained research assistant and can, therefore, provide a more objective assessment of the infant's growth, infants tend to behave differently with unfamiliar individuals [51]. As such, Miller et al [51] expressed that a fuller picture of the infant's development could be obtained if both types of assessments were used.

This study has some limitations. One of its limitations is its high attrition rate. Because of the COVID-19 pandemic, parents were more cautious of physical interactions, as they did not want themselves, their infants, or other family members to contract the virus. Therefore, many parents declined home visits for the Bayley-4 assessment. This may have affected the accuracy of the findings in representing the sample recruited for this study. Moreover, the longitudinal nature of the study might have also contributed to the high attrition rate. Parents tend to become busier post partum owing to the need for them to adjust to parenting responsibilities; in the case of this study, the need to take extra precautions to prevent contracting the COVID-19 virus added to these responsibilities. Therefore, it is paramount to devise strategies to keep parents interested in and willing to participate in the study. For example, research team members can frequently contact parents to build stronger rapport and remind them to access SPA if they have parenting-related concerns. Although the research team originally intended to do so, some technical issues led to the absence of chat notifications, affecting the communication between the team and parents. Furthermore, many of the research team members were also frontline health care workers; therefore, they were unable to meet often and resolve these issues in a timely manner.

Another limitation of this study is the lack of information regarding whether the parents were experienced or new. This is an important limitation, as the struggles and support needs that new and experienced parents encounter may differ widely. Hence, future studies should take note to collect such information from parents to provide deeper perspectives regarding the effectiveness of parenting interventions in new and experienced mothers and fathers. Subsequent research may also investigate parental sensitivity and responsivity to provide further insight into how they may affect infant behavioral outcomes.

Conclusions

This study examined the effects of the SPA intervention on infant developmental outcomes. The results showed that the infants from the intervention group generally developed better in terms of communication, motor skills, cognition, and social-emotional skills than those from the control group. The peer support and informational support that the SPA intervention offered to the intervention group parents were thus helpful in indirectly influencing the development of infants. More research is needed to obtain an in-depth understanding of what functions of the intervention influenced the infant outcomes and what information the current generation of parents hopes to see in parenting mobile apps. This would facilitate the creation of more effective mHealth app–based support for parents. In the future, interventions targeting infant growth and development should be created to measure the direct effects of educational interventions. In addition, future parenting interventions should focus on providing more support to families with lower SES to help promote the development of infants and support parents from these families.

Acknowledgments

The authors would like to thank the Ministry of Health for funding this project through the Health Services Research Grant (HSRGHP18may-0001). They also extend their thanks to all the study participants for sharing their insights despite being new parents during the pandemic.

Conflicts of Interest

None declared.

References

1. Huang CY, Costeines J, Ayala C, Kaufman JS. Parenting stress, social support, and depression for ethnic minority adolescent mothers: impact on child development. J Child Fam Stud 2014 Feb;23(2):255-262 [FREE Full text] [doi: 10.1007/s10826-013-9807-1] [Medline: 24653641]
2. Bornstein MH. Parenting infants. In: Bornstein MH, editor. Handbook of Parenting: Volume I: Children and Parenting. 3rd edition. Oxfordshire, UK: Routledge; 2019:3-55.
3. Razurel C, Kaiser B, Antonietti JP, Epiney M, Sellenet C. Relationship between perceived perinatal stress and depressive symptoms, anxiety, and parental self-efficacy in primiparous mothers and the role of social support. Women Health 2017 Feb;57(2):154-172 [FREE Full text] [doi: 10.1080/03630242.2016.1157125] [Medline: 26909523]
4. Marjanovič-Umek L, Fekonja-Peklaj U. The roles of child gender and parental knowledge of child development in parent-child interactive play. Sex Roles 2017 Jan 21;77(7-8):496-509 [FREE Full text] [doi: 10.1007/s11199-016-0734-7]
5. Milgrom J, Hirshler Y, Reece J, Holt C, Gemmill AW. Social support-a protective factor for depressed perinatal women? Int J Environ Res Public Health 2019 Apr 21;16(8):1426 [FREE Full text] [doi: 10.3390/ijerph16081426] [Medline: 31010090]
6. Racine N, Plamondon A, Hentges R, Tough S, Madigan S. Dynamic and bidirectional associations between maternal stress, anxiety, and social support: the critical role of partner and family support. J Affect Disord 2019 Jun 01;252:19-24 [FREE Full text] [doi: 10.1016/j.jad.2019.03.083] [Medline: 30954841]
7. Philpott LF, Savage E, FitzGerald S, Leahy-Warren P. Anxiety in fathers in the perinatal period: a systematic review. Midwifery 2019 Sep;76:54-101 [FREE Full text] [doi: 10.1016/j.midw.2019.05.013] [Medline: 31176080]
8. Ong SF, Chan WC, Shorey S, Chong YS, Klainin-Yobas P, He HG. Postnatal experiences and support needs of first-time mothers in Singapore: a descriptive qualitative study. Midwifery 2014 Jun;30(6):772-778 [FREE Full text] [doi: 10.1016/j.midw.2013.09.004] [Medline: 24161493]
9. Shorey S, Dennis CL, Bridge S, Chong YS, Holroyd E, He HG. First-time fathers' postnatal experiences and support needs: a descriptive qualitative study. J Adv Nurs 2017 Dec;73(12):2987-2996 [FREE Full text] [doi: 10.1111/jan.13349] [Medline: 28557020]
10. Shorey S, Tan TC, Thilagamangai, Mathews J, Lim SH, Shi L, et al. Perinatal support for couples during COVID-19: a descriptive qualitative study. J Clin Nurs (forthcoming) 2022 Jun 30 [FREE Full text] [doi: 10.1111/jocn.16432] [Medline: 35773956]
11. Plantin L, Daneback K. Parenthood, information and support on the internet. A literature review of research on parents and professionals online. BMC Fam Pract 2009 May 18;10:34 [FREE Full text] [doi: 10.1186/1471-2296-10-34] [Medline: 19450251]
12. Suárez A, Rodrigo MJ, Muñetón Ayala MA. Parental activities seeking online parenting support: is there a digital skill divide? Rev Cercet Interv Soc 2016 Sep;54:36-54 [FREE Full text] [doi: 10.33788/rcis]
13. Dworkin J, Connell J, Doty J. A literature review of parents’ online behavior. Cyberpsychology 2013 Jul 01;7(2):2 [FREE Full text] [doi: 10.5817/cp2013-2-2]
14. Walsh A, Edwards H, Fraser J. Influences on parents' fever management: beliefs, experiences and information sources. J Clin Nurs 2007 Dec;16(12):2331-2340 [FREE Full text] [doi: 10.1111/j.1365-2702.2006.01890.x] [Medline: 17419783]
15. Gutierrez-Galve L, Stein A, Hanington L, Heron J, Ramchandani P. Paternal depression in the postnatal period and child development: mediators and moderators. Pediatrics 2015 Feb;135(2):e339-e347 [FREE Full text] [doi: 10.1542/peds.2014-2411] [Medline: 25560437]
16. Wee KY, Skouteris H, Pier C, Richardson B, Milgrom J. Correlates of ante- and postnatal depression in fathers: a systematic review. J Affect Disord 2011 May;130(3):358-377. [doi: 10.1016/j.jad.2010.06.019] [Medline: 20599275]
17. Alzahrani AD. Risk factors for postnatal depression among primipara mothers. Span J Psychol 2019 Jul 12;22:E35 [FREE Full text] [doi: 10.1017/sjp.2019.33] [Medline: 31298175]
18. Shorey S, Ng ED. Evaluation of mothers' perceptions of a technology-based supportive educational parenting program (part 2): qualitative study. J Med Internet Res 2019 Feb 13;21(2):e11065 [FREE Full text] [doi: 10.2196/11065] [Medline: 30758295]
19. Hannan J. APN telephone follow up to low-income first time mothers. J Clin Nurs 2013 Jan;22(1-2):262-270 [FREE Full text] [doi: 10.1111/j.1365-2702.2011.04065.x] [Medline: 22845337]
20. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11 [FREE Full text] [Medline: 26934758]
21. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. 2nd edition. Hoboken, NJ, USA: Prentice Hall; 1986.
22. Bowlby J. Attachment and Loss. Vol. I. Attachment. New York, NY, USA: Basic Books; 1969.

23. Adina J, Morawska A, Mitchell AE, McBryde M. Effect of parenting interventions on perinatal depression and implications for infant developmental outcomes: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2022 Jun;25(2):316-338 [FREE Full text] [doi: 10.1007/s10567-021-00371-3] [Medline: 34580804]
24. Shorey S, Law E, Thilagamangai, Mathews J, Lim SH, Shi L, et al. Evaluating the effectiveness of the supportive parenting app on parental outcomes: randomized controlled trial. J Med Internet Res 2023 Jan 16;25:e41859 [FREE Full text] [doi: 10.2196/41859] [Medline: 36645699]
25. Shorey S, Ng YP, Ng ED, Siew AL, Mörelius E, Yoong J, et al. Effectiveness of a technology-based supportive educational parenting program on parental outcomes (part 1): randomized controlled trial. J Med Internet Res 2019 Feb 13;21(2):e10816 [FREE Full text] [doi: 10.2196/10816] [Medline: 30758289]
26. Shorey S, Tan TC, Thilagamangai, Mathews J, Yu CY, Lim SH, et al. Development of a supportive parenting app to improve parent and infant outcomes in the perinatal period: development study. J Med Internet Res 2021 Dec 24;23(12):e27033 [FREE Full text] [doi: 10.2196/27033] [Medline: 36260376]
27. Singh A, Yeh CJ, Boone Blanchard S. Ages and Stages Questionnaire: a global screening scale. Bol Med Hosp Infant Mex 2017 Jan;74(1):5-12 [FREE Full text] [doi: 10.1016/j.bmhimx.2016.07.008] [Medline: 29364814]
28. Squires J, Bricker D, Potter L. Revision of a parent-completed development screening tool: ages and stages questionnaires. J Pediatr Psychol 1997 Jun;22(3):313-328 [FREE Full text] [doi: 10.1093/jpepsy/22.3.313] [Medline: 9212550]
29. Bayley N, Aylward GP. Bayley Scales of Infant and Toddler Development, Technical Manual. 4th edition. New York, NY, USA: Pearson; 2019.
30. Aylward GP. Bayley 4 Clinical Use and Interpretation. Cambridge, MA, USA: Academic Press; 2020.
31. McHenry MS, Oyungu E, Yang Z, Hines AC, Ombitsa AR, Vreeman RC, et al. Cultural adaptation of the Bayley scales of infant and toddler development, 3rd edition for use in Kenyan children aged 18-36 months: a psychometric study. Res Dev Disabil 2021 Mar;110:103837 [FREE Full text] [doi: 10.1016/j.ridd.2020.103837] [Medline: 33453695]
32. Nicolson S, Carron SP, Paul C. Supporting early infant relationships and reducing maternal distress with the newborn behavioral observations: a randomized controlled effectiveness trial. Infant Ment Health J 2022 May;43(3):455-473 [FREE Full text] [doi: 10.1002/imhj.21987] [Medline: 35531961]
33. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The brief infant-toddler social and emotional assessment: screening for social-emotional problems and delays in competence. J Pediatr Psychol 2004 Mar;29(2):143-155 [FREE Full text] [doi: 10.1093/jpepsy/jsh017] [Medline: 15096535]
34. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, Tasdemir GN, et al. The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behav Dev 2009 Jun;32(3):291-297 [FREE Full text] [doi: 10.1016/j.infbeh.2009.03.003] [Medline: 19411111]
35. IBM SPSS Statistics for Windows, Version 27.0. IBM Corp. 2020. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27 [accessed 2022-08-01]
36. Shorey S, Chan SW, Chong YS, He HG. A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. J Adv Nurs 2015 Jun;71(6):1260-1273 [FREE Full text] [doi: 10.1111/jan.12590] [Medline: 25496615]
37. Shorey S, Ying L, Yobas P. Parenting outcomes and predictors of parenting satisfaction in the early postpartum period. West J Nurs Res 2020 Jan;43(1):13-24 [FREE Full text] [doi: 10.1177/0193945920914593] [Medline: 32389069]
38. Bortfeld H, Gabouer A. Infant communication. In: Vangelisti AL, editor. The Routledge Handbook of Family Communication. 3rd edition. Oxfordshire, UK: Routledge; 2021:221-232.
39. Leman P, Bremner A, Parke RD, Gauvain M. Developmental Psychology. 2nd edition. New York, NY, USA: McGraw Hill; 2019.
40. Topping K, Dekhinet R, Zeedyk S. Parent–infant interaction and children's language development. Educ Psychol 2013 Jul;33(4):391-426 [FREE Full text] [doi: 10.1080/01443410.2012.744159]
41. Pazos RV, Adeline S, Singson TC, Linn TM, Veira XG. A year upended: how governments (still) fight to keep Covid-19 in check - and what may work best. The Straits Times. 2020. URL: https://sigmaawards.org/a-year-upended-how-governments-still-fight-to-keep-covid-19-in-check-and-what-may-work-best/ [accessed 2022-09-11]
42. Davis B, Degotardi S. Educators' understandings of, and support for, infant peer relationships in early childhood settings. J Early Child Res 2014 Aug 01;13(1):64-78 [FREE Full text] [doi: 10.1177/1476718x14538600]
43. Leerkes EM, Weaver JM, O'Brien M. Differentiating maternal sensitivity to infant distress and non-distress. Parent Sci Pract 2012 Jan 01;12(2-3):175-184 [FREE Full text] [doi: 10.1080/15295192.2012.683353] [Medline: 22798728]
44. Lorber MF, Del Vecchio T, Slep AM. The emergence and evolution of infant externalizing behavior. Dev Psychopathol 2015 Aug;27(3):663-680 [FREE Full text] [doi: 10.1017/S0954579414000923] [Medline: 25212988]
45. Bellieni CV, Ceccarelli D, Rossi F, Buonocore G, Maffei M, Perrone S, et al. Is prenatal bonding enhanced by prenatal education courses? Minerva Ginecol 2007 Apr;59(2):125-129 [FREE Full text] [Medline: 17505454]
46. Shakarami A, Mirghafourvand M, Abdolalipour S, Jafarabadi MA, Iravani M. Comparison of fear, anxiety and self-efficacy of childbirth among primiparous and multiparous women. BMC Pregnancy Childbirth 2021 Sep 21;21(1):642 [FREE Full text] [doi: 10.1186/s12884-021-04114-8] [Medline: 34548055]

47. Jones D, Feinberg M, Hostetler M, Roettger M, Paul IM, Ehrenthal DB. Family and child outcomes 2 years after a transition to parenthood intervention. Fam Relat 2018 Apr;67(2):270-286 [FREE Full text] [doi: 10.1111/fare.12309] [Medline: 30140113]
48. Clearfield MW, Niman LC. SES affects infant cognitive flexibility. Infant Behav Dev 2012 Feb;35(1):29-35 [FREE Full text] [doi: 10.1016/j.infbeh.2011.09.007] [Medline: 22018718]
49. Freitas TC, Gabbard C, Caçola P, Montebelo MI, Santos DC. Family socioeconomic status and the provision of motor affordances in the home. Braz J Phys Ther 2013 Jul;17(4):319-327 [FREE Full text] [doi: 10.1590/S1413-35552013005000096] [Medline: 24072221]
50. Pace A, Luo R, Hirsh-Pasek K, Golinkoff RM. Identifying pathways between socioeconomic status and language development. Annu Rev Linguist 2017 Jan 14;3(1):285-308 [FREE Full text] [doi: 10.1146/annurev-linguistics-011516-034226]
51. Miller LE, Perkins KA, Dai YG, Fein DA. Comparison of parent report and direct assessment of child skills in toddlers. Res Autism Spectr Disord 2017 Sep;41-42:57-65 [FREE Full text] [doi: 10.1016/j.rasd.2017.08.002] [Medline: 28919924]

Abbreviations

ADBE: adaptive behavior
ASQ: Ages and Stages Questionnaire
BITSEA: Brief Infant Toddler Social Emotional Assessment
mHealth: mobile health
RCT: randomized controlled trial
RR: risk ratio
SES: socioeconomic status
SPA: Supportive Parenting App

Edited by G Eysenbach, R Kukafka; submitted 28.10.22; peer-reviewed by R Otte, Y Wang; comments to author 24.11.22; revised version received 04.12.22; accepted 12.01.23; published 22.02.23.

Please cite as:
Shorey S, Chong YS, Shi L, Chua JS, Thilagamangai, Mathews J, Lim SH, Du R, Chan YH, Tan TC, Chee C, Law E
Evaluating the Effects of the Supportive Parenting App on Infant Developmental Outcomes: Longitudinal Study
JMIR Mhealth Uhealth 2023;11:e43885
URL: https://mhealth.jmir.org/2023/1/e43885
doi:10.2196/43885
PMID:36811952

Original Paper

Effectiveness of a Mobile App to Increase Risk Perception of Tobacco, Alcohol, and Marijuana Use in Mexican High School Students: Quantitative Study

Patricia María del Carmen Fuentes A[1*], PhD; Alberto Jiménez Tapia[1*], MSc; Eunice M Ruiz-Cortés[1*], BSc; Fernando Bolaños-Ceballos[2*], PhD; Julio César Flores Castro[*], BSc; Rafael Gutiérrez[1*], MSc; Catalina González-Forteza[1*], PhD

[1]Direction of Epidemiological and Psychosocial Research, National Institute of Psychiatry Ramón de la Fuente Muñiz, Mexico City, Mexico

[2]Universidad Autónoma del Estado de Hidalgo, Actopan, Hidalgo, Mexico

[*]all authors contributed equally

Corresponding Author:
Alberto Jiménez Tapia, MSc
Direction of Epidemiological and Psychosocial Research
National Institute of Psychiatry Ramón de la Fuente Muñiz
Calzada Mexico Xochimilco, 101
Mexico City, 14370
Mexico
Phone: 52 5541605154
Email: alberjt1969@gmail.com

Abstract

Background: Young people have the highest rate of drug use worldwide. Recent data from Mexico in this population show that the prevalence of illicit drug use doubled between 2011 and 2016 (2.9%-6.2%), with marijuana being the one with the highest increase (2.4%-5.3%), but also point out that alcohol and tobacco use have remained steady or decreased. Mexican adolescents are at high risk for drug use owing to a low perception of risk and the availability of drugs. Adolescence is an ideal period to reduce or prevent risky behaviors using evidence-based strategies.

Objective: In this study, we aimed to test the short-term effectiveness of a mobile intervention app ("What Happens if you Go Too Far?" ["¿Qué pasa si te pasas?"]) that seeks to increase risk perception of tobacco, alcohol, and marijuana use in a sample of Mexican high school students.

Methods: A nonexperimental evaluation based on pretest-posttest design was used to measure the effectiveness of a preventive intervention using a mobile app, "What Happens If You Go Too Far?" The dimensions analyzed were knowledge of drugs and their effects, life skills, self-esteem, and risk perception. The intervention was conducted on a high school campus with 356 first-year students.

Results: The sample included 359 first-year high school students (mean 15, SD 0.588 years; women: 224/359, 62.4% men: 135/359, 37.6%). The intervention increased the overall risk perception of tobacco (χ^2_4=21.6; *P*<.001) and alcohol use (χ^2_4=15.3; *P*<.001). There was no significant difference in the perception that it is dangerous to smoke 5 cigarettes, and there was a marginal difference in the perception that it is very dangerous to smoke 1 cigarette or to use alcohol or marijuana. We used a generalized estimating equation method to determine the impact of the variables on risk perception. The results showed that knowledge about smoking increased the risk perception of smoking 1 cigarette (odds ratio [OR] 1.1065, 95% CI 1.013-1.120; *P*=.01), and that knowledge about marijuana use (OR 1.109, 95% CI 1.138-1.185; *P*=.002) and self-esteem (OR 1.102, 95% CI 1.007-1.206; *P*=.04) produced significant increases in the risk perception of consuming 5 cigarettes. Resistance to peer pressure and assertiveness also increased the perceived risk of using tobacco and alcohol.

Conclusions: The intervention has the potential to increase the perception of risk toward drug use in high school students by providing knowledge about the effects and psychosocial risks of drug use and by strengthening life skills that are associated with increased risk perception. The use of mobile technologies in intervention processes may broaden the scope of preventive work for adolescents.

(JMIR Mhealth Uhealth 2023;11:e37873) doi:10.2196/37873

KEYWORDS

adolescents; students; risk perception; tobacco; alcohol; marijuana; mobile apps; apps; substance use; prevention

Introduction

Background

Drug use is shaped by the interaction of personal, social, and contextual factors. Lack of knowledge about substance use dynamics (social influence, positive attitudes toward use, peer group use, availability, and inadequate parental supervision) are elements that favor use in adolescents [1,2]. The most recent data in this population not only show that the prevalence of illicit drug use doubled between 2011 and 2016 (2.9%-6.2%), with marijuana being the one with the highest increase (2.4%-5.3%), but also point out that alcohol and tobacco use have remained steady or decreased [3]. The negative consequences of use affect general well-being, academic performance, physical and mental health, family dynamics, and peer relationships, and they increase the likelihood of fatal accidents [4].

Adolescents and young adults have the highest rates of substance use throughout the world [5]. The data show that adolescents are the population group with the highest risk for substance use in Mexico and that they start at increasingly younger ages. Their perception of risk is low, and drug availability is increasing [6,7]. This situation highlights the need for evidence-based early interventions with a public health approach and an ecological perspective.

During adolescence (10-19 years of age), there are important biological, psychological, and social transitions resulting from rapid physical, psychological, sexual, and social development that involves changes in the brain, cognition, and emotions [1,8]. Adolescents are in a vulnerable position that brings them closer to risk behaviors such as drug abuse [9-12], and it has become a public health problem.

Adolescents are exposed to individual, family, and social risks and protective factors that can increase or decrease the probability of drug use. Family risk factors include conflict, lack of parental supervision, and family members who are abusive physically, emotionally, and sexually. Social risk factors include peer pressure to use drugs as a form of socialization, peers who use it, and the availability of drugs. Individual risk factors include inclination toward experimentation, curiosity, rebelliousness, and impulsivity, as well as low self-esteem, lack of emotional regulation, depression, anxiety, behavioral problems, poor school performance, previous experiences with drugs, and low risk perception toward drug use [13-16]. Risk perception may contribute to increasing the odds of drug use in adolescents, and there is evidence that it may act in both ways as a risk or protective factor; that is, when it is low, it increases the risk of drug use and vice versa [13,16].

Health behavior theories such as the Health Belief Model and the Theory of Planned Behavior suggest that risk perception is an important factor in health behavior and that the level of risk perception determines the likelihood of the occurrence of risk behaviors such as adolescent drug use [14,17-19]. Because risk perception is an attitude that represents the evaluation of an object through its favorable and unfavorable attributes [17], levels of risk perception have been considered important determinants of risky behaviors.

There is evidence indicating that the probability of drug use increases as people perceive little or no risk of associated harm, that is, a higher perceived risk can be considered as a protective factor against drug use [13]. A longitudinal study on risk perception toward tobacco, alcohol, and cannabis use with a 10-year follow-up reported that it reduced the probability of consumption in German adolescents aged 14 and 15 years [18]. Other studies in the Latin American population reported a low risk perception associated with drug use [20], which is usually at these levels for tobacco and alcohol and higher for other drugs. However, a reduction in the perception of risk of marijuana use has also been reported between 2000 and 2014, which has been associated with an increase in the use of this drug in the adolescent population [3].

Evidence shows that a lower perception of risk is related to higher rates of use [21] and that protective dynamics can emerge within the perception of negative consequences or health risks [22-24]. Work is necessary in this regard because the perceived risk of regular marijuana use has decreased by 40% between 1995 and 2019, but the potency of the drug and its consumption have increased; therefore, it is important to reduce this gap to lower its impact on public health [25].

Adolescence is an ideal period for interventions designed to reduce or prevent risky behaviors [26]. Life skills, including positive and adaptive social skills, enable adolescents to cope with everyday challenges [27,28]. The life skills approach considers the following: (1) the recognition and evidence of the role of cognitive, interpersonal, and coping skills in psychosocial development; (2) the effect of skills on young people's ability to protect their health, adopt positive behaviors, and foster healthy relationships; (3) the application of skills in managing education, violence, and human rights; (4) the reinforcement of protective factors such as self-awareness, self-confidence, and self-esteem; and (5) the mastery and application of skills in everyday situations to feel self-confident, self-efficient, and self-worthy [29,30]. The development of life skills is part of the learning, competence, and education that underpin adolescent well-being [31].

Interventions based on life skills training have proven to be effective for the prevention of drug use in adolescents, are easy to adapt and disseminate [29,32-35], and are an effective strategy to promote health care in schools [27,28,36]. However, their design requires providing knowledge and skills to the professionals who perform them, which in turn requires economic, human, and time resources [8].

Schools are effective settings for preventive interventions, given the ease of access to adolescents and because they are spaces designed to foster learning and socialization [32,37]. Studies on prevention and early intervention based on life skills to address drug use in educational settings have shown excellent results [8]. The COVID-19 pandemic has underscored the importance of measuring the effects of these interventions when they are executed using digital technologies [38].

Mobile apps are effective tools for prevention and intervention in health care. There are various examples of the use of this technology to address depression, anxiety, drug use, and suicide [10,11,39]. Some strategies provide only information and education, and others focus on giving advice, strategies, or skills training, but few offer the possibility of self-assessing drug use and providing feedback [11,39]. Additional research is needed to evaluate the effectiveness of apps, and funding is needed to develop apps using evidence-based techniques [40].

Interventions that are designed for the internet or mobile devices using current evidence-based approaches and resources are more likely to be successful in prevention [11,39,41]. The use of these technologies, including life skills training, has the advantages of accessibility, portability, interactivity, feedback, ease of use, and wide reach at low cost; they reach adolescent and young adult populations because of their ubiquity and mobility and because young people know, accept, and integrate them easily into their lifestyle [32,42].

Objective

There are few studies that have evaluated the characteristics and effectiveness of digital mobile interventions. It is necessary to evaluate their effectiveness and applicability in real-world contexts to verify their potential and usefulness [32,39-41]. Consequently, the objective of this study was to test the short-term effectiveness of a mobile intervention app ("What Happens if you Go Too Far?" ["¿Qué pasa si te pasas?"]). This study aimed to increase risk perception of tobacco, alcohol, and marijuana use in a sample of Mexican high school students.

Methods

Ethics Approval

All participants' parents signed an informed consent form, and all participants provided written informed consent. The study was approved by the Research Ethics Committee of the Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (approval number CEI/C/003/2016).

Design and Procedure

Overview

This study used a nonexperimental pretest-posttest design, with assessments at baseline and at the end of the intervention. The intervention was performed in a high school campus, as part of the introductory curriculum for the school year, in a classroom with 45 computers using the Android emulator BlueStacks (BlueStacks) to execute the app "What Happens if you Go Too Far?" ("¿Qué pasa si te pasas?"). The intervention evaluation questionnaire (Brief Life Skills Scale for Adolescents) (González-Forteza et al, unpublished data, August 2022) was administered as a Google Form. The Brief Life Skills Scale for Adolescents was developed using items from several more extensive scales, each including 35 to 65 items that assess the skills separately and have been validated for Mexican adolescents. These skills include planning for the future, assertiveness, expression of emotions, taking responsibility, decision-making, and resistance to peer pressure.

Before its implementation, teachers and psychology interns received training in the management of the intervention, which included modules on the effects and risks of drug use in adolescents and on strengthening 6 life skills and addressing the emotions that occur when these skills are applied.

The intervention was performed during the orientation week for the incoming students. The students were invited to participate in the study voluntarily and anonymously. The school authorities and teachers were informed of the project and granted access to the school. The objectives of the research as well as the risks and benefits of participating in the study were explained to the students; they were informed that their participation was voluntary, their answers would be anonymous, and the results would not affect their activities or evaluation in school.

In total, 10 groups of first-year high school students from the 2019-2020 classes participated. The intervention lasted 10.5 hours: three 90-minute sessions for each group at school, plus 2 hours per day of individual activities at home using the app during the same week the intervention was implemented. In each of the 3 sessions, there was an average of 36 students per group. Participants did not receive any incentives for their involvement in the study.

Session 1

The first session aimed to apply the pretest evaluation, present the mobile app, and explain life skills. The evaluation questionnaire took 20 minutes to complete. After the evaluation, participants were instructed to download the "What Happens If You Go Too Far?" apps on mobile devices (iOS and Android). Downloading the app required students to connect to the internet via a mobile network or Wi-Fi, and once the app was installed on their devices, it could be used offline. The session also included a description of the World Health Organization (WHO)–recommended life skills (decision-making, resistance to peer pressure, problem-solving, future planning, and assertiveness), an explanation of basic emotions, and the link between life skills and the absence or presence of emotions. We also showed the objectives and options included in the main menu of the app and the available resources (comics, video games, quizzes, trivia, agendas, and news). Participants were asked to form teams of 5 to 6 members to share their answers as a group in subsequent sessions on the effects and risks of drug use and the use of life skills in different cases. Finally, we assigned the comic resources on the topics of alcohol consumption and women or men as homework.

Session 2

The second session was designed to (1) identify the effects and risks of alcohol use in women and men; (2) apply decision-making skills, resistance to peer pressure, problem-solving, and goal achievement skills; and (3) identify the consequences and emotions associated with the application of the skills. Participants were instructed to read the comic

"Alcohol and Women" to understand the risks and immediate effects of alcohol use on women, to use the cases of Mony and Lucy to choose 3 options and their consequences, and to identify the decision-making and resistance to peer pressure involved. Subsequently, they were asked to answer the questions on the "What Happens If You Go Too Far?" app activity form, based on their interaction with the comic to identify the immediate effects of alcohol use and recognize or describe different situations that encourage alcohol use, some of the consequences and emotions involved in using drugs, and some strategies to resist peer pressure. The facilitator asked the teams to answer some questions related to the scenarios described before to reinforce the identification of the immediate social and health effects of alcohol use in women, apply their skills to situations that lead to drug use, associate them with emotions that may occur, and think about the consequences of the characters' decisions (Lucy and Mony) in the situations they face. Participants then completed the trivia and quiz on alcohol and women to reinforce their knowledge. The second part of the session comprised the same procedure with the comic "Alcohol and Men," in which problem-solving and goal setting were applied to the cases of the male characters (Beto and Angel). Participants then completed the trivia and quiz on alcohol and men to reinforce their knowledge, and they received orientation on using the comics on tobacco and marijuana to apply negotiation, assertive communication, problem-solving, and resistance to peer pressure.

Session 3

The third session sought to (1) identify the effects and risks of tobacco and marijuana use; (2) apply skills related to negotiation, assertive communication, problem-solving, and resistance to peer pressure; (3) identify consequences and emotions associated with the application of the skills; and (4) administer the posttest evaluation questionnaire. Participants were asked to read the comic about tobacco to review the risks and effects of using it. The characters of the comic (Susy, Alma, and Lizet) were used to model the consequences and identify the negotiations and assertive communication involved. Participants then answered the questions on the "What Happens If You Go Too Far?" app activity form to check their understanding of the steps of negotiation, their ability to identify situations that encourage tobacco use, if they could recognize decision-making options and discuss consequences and emotions involved, and if they could apply the life skills in an everyday situation. The facilitator asked the teams to answer some questions related to the scenarios described before to reinforce the identification of the immediate social and health effects of tobacco and marijuana use, apply their skills to situations that lead to drug use, associate them with emotions that may occur, and think about the consequences of the characters' decisions in the situations they face. Participants then completed the trivia and quiz on tobacco to reinforce their knowledge. The second part of the session followed the same procedure with the comic "Marijuana" to understand the risks and immediate effects of using drugs, and identify the skills involved in problem-solving and resistance to peer pressure involved. This was applied using the cases of the characters (Agus, Diana, Pedro, and Omar). Participants then completed the trivia and quiz on marijuana to reinforce their knowledge. Finally, we administered the posttest evaluation questionnaire (evaluation questionnaire, 20 minutes).

Intervention to Prevent Addiction: the "What Happens If You Go Too Far?" App

The "What Happens If You Go Too Far?" [43] intervention app seeks to strengthen life skills and increase risk perception of drug use. It is based on the WHO Skills for Health model and Bandura's Social Learning Theory.

The intervention acts at cognitive and behavioral levels to increase the risk perception of drug use. At the cognitive level, it facilitates the acquisition of specific knowledge of the effects of drugs on the brain and behavior. It contains evidence-based scientific information translated into textual and visual language to generate interactive comics, trivia, quizzes, and games that facilitate understanding and encourage thinking about substance use. At the behavioral level, it strengthens life skills and improves adolescents' ability to relate with their peers, resist pressure to use substances, solve problems effectively, and make responsible decisions with awareness of consequences. The comics in the app depict and simulate everyday situations related to drug use to apply these skills.

The comics include three elements: (1) information about drugs, their immediate effects, and their risks; (2) situations experienced by young people that are associated with drug use; and (3) skills involved in decision-making, problem-solving, negotiation, assertiveness, and resistance to peer pressure that help to face the challenges of everyday life.

The trivia is a question-and-answer game that facilitates immediate self-assessment of knowledge of the effects of drug use to reinforce the knowledge acquired with the comics. The quiz is a brief, flexible, and self-administered resource based on the WHO alcohol, smoking, and substance involvement screening test, which assesses drug use using 8 questions and indicates low-, moderate-, or high-risk scores [44,45]. It allows participants to identify their individual level of risk and provides them with information to recognize likely consequences and risks.

The video game models situations associated with drug use experienced by the characters Beto and Bety and facilitates the application of skills to making decisions, communicating assertively, negotiating, and resisting peer pressure; it provides automatic reinforcement. The news feature facilitates constant updating of (1) preventive messages or specific events for timely dissemination and (2) links of interest to youth. The agenda feature connects to hotlines that focus on drug use problems.

The Technology Behind the "What Happens If You Go Too Far?" App

The configuration of comics and game resources uses the potential of interactive technology, using stories that model everyday situations associated with drug use in young people, to practice decision-making with various options and their consequences. This serves to reinforce knowledge and encourage thinking about the risks and effects of drug use.

The interactive dynamic favors (1) reinforcing the knowledge of drugs and their effects, (2) thinking about the risks and consequences of consumption, (3) interactively applying decision-making and its automatic and immediate feedback in the situations presented, (4) concluding reinforcement with infographics of the different skills in each case, and (5) giving immediate feedback on the trivia and quiz. The app content was developed to be both didactic and informative.

The app is available for free on the 2 most important platforms in the market (Google Android [Google LLC] and Apple iOS [Apple Inc]). It can be installed on a wide range of mobile devices and has a broad reach among young people. A third option is to use an Android emulator that allows the app to be used offline on PCs for students and schools without internet access.

The app was developed using Unity 3D and state-of-the-art web technologies. It is based on the current standards of responsive web design and user experience to generate an accessible and optimized product that is fast and easy to download. The front-end offers interactive content and resources that can be used individually or in groups (the game, trivia, and quiz). The information is complemented by an agenda that facilitates contact with specialized care centers and services, as well as links to relevant news and web content. The back end includes various web tools for web-based editing and updating of content, as well as data consultation and statistical reports for use analysis by the developers, administrators, and researchers in charge of the project.

Data Analysis

In this study, 2 types of analyses were used to identify differences between pre- and posttest measurements. The first was used to measure variations in the perceived risk of using marijuana, alcohol, and tobacco (smoking 1 cigarette and smoking 5 cigarettes) in both measurements. The original categories of risk perception were "not dangerous," "dangerous," and "very dangerous." There was no statistically significant difference between "not dangerous" and "dangerous"; therefore, the former category was excluded from the analysis, and differences in pre- and posttest measurements of risk perception were made between the categories of "dangerous" and "very dangerous."

Comparisons were performed using McNemar test, with 95% certainty considered statistically significant. This is a nonparametric analysis of the comparison of proportions for 2 related samples whose function is to compare the change in the distribution of proportions between 2 measurements of a dichotomous variable and determine that the difference is not due to chance. In this case, there was no dependent or independent variable, as they were related or paired measurements. The impact of variables on risk perception was calculated using the generalized estimating equation (GEE) method, which extends the generalized linear model to allow for the effect of repeated measurements and other related observations. GEE is a method for modeling longitudinal or pooled data, and is often used with nonnomal data, such as binary or count data. This method uses a set of equations that are solved to obtain parameter estimates. This modeling strategy uses a quasi-likelihood function that assumes only a relationship between μ and Var(Y) rather than a specific distribution for Y. This allows deviation from the usual assumptions, such as overdispersion caused by correlated observations or unobserved explanatory variables. To do this, the quasi-likelihood approach takes the usual formula for variance but multiplies it by a constant that is estimated using the data. GEE is designed for simple clustering or repeated measures; it is not easily adaptable to more complex designs, such as nested or cross-group designs [46].

We took care in the analysis that (1) the mean structure was correctly specified (all relevant variables were included and all irrelevant variables were removed), (2) the observations between clusters were unrelated (there is no higher-level clustering mechanism), (3) the sample size was large enough for asymptotic inference (356 records); (4) the normality of residuals was not assumed with GEE; and (5) the database was restructured to obtain the necessary information. As there is no field called TIME, we used the SAMPLE field, which in our case, expresses the ratio of time between evaluations as a function of TIME. Although it was designed for longitudinal studies, in this case, it was applied to 2 measurements because missing measurements are common in longitudinal designs and are assumed to be caused by chance. Therefore, missing values were imputed using IBM SPSS Modeler (version 18.3; IBM Corp) tool, which has different imputation methods (fixed, random, expression, and algorithm). In this analysis, we used the algorithm method, which replaces a predicted value with a model based on the classification and regression trees algorithm. In each field imputed with this method, there is a separate classification and regression trees model, along with a fill node that replaces empty and null values with the value predicted by the model.

Results

Participants

The sample for the initial evaluation included 359 (90% of the 399 enrolled students) first-year middle-class high school students at the Escuela Superior Actopan, affiliated with the Universidad Autónoma del Estado de Hidalgo. The mean age of the participants was 15 (SD 0.588) years. Of the total sample, 224 (62.4%) of the students were women and 135 (37.6%) were men; 182 (50.7%) participants were in the morning session and 176 (49%) participants were in the afternoon session. Regarding drug use history, 64 (17.8%) participants reported tobacco use sometime in life (women: 36/224, 16.1%; men: 28/135, 20.7%) and 27 (7.5%) in the past 3 months (women: 11/224,4.9%; men: 16/135, 11.9%); 174 (48.5%) students reported alcohol use sometime in life (women: 99/224, 44.2%; men: 75/135, 55.6%) and 126 (35.1%) in the past 3 months (women: 81/224, 36.2%; men: 45/135, 33.3%); 18 (5%) participants reported marijuana use sometime in life (women: 7/224, 3.1%; men: 11/135, 8.1%) and 6 (1.7%) in the past 3 months (women: 1/224, 0.4%; men: 5/135, 3.7%; Multimedia Appendix 1).

The follow-up evaluations were completed by 356 (99.2%) of the 359 study participants. The mean age of the participants was 15 (SD 0.574) years. Of them, 224 (62.9%) participants were women, 132 (37.1%) were men; 224 (62.9%) were in the morning session, and 132 (37.1%) were in the afternoon session.

The inclusion criterion was enrollment in the first semester of high school at Escuela Superior Actopan.

Effectiveness of the App

Of the 356 students, the proportion of those who perceived that it was dangerous to smoke 1 cigarette decreased from 197 (55.3%) to 161 (46%) comparing the pre- and posttest measurements, but the perception that it was very dangerous increased from 85 (23.9%) to 142 (39.9%). Of the 356 students, the proportion perceiving that it was dangerous to smoke 5 cigarettes increased from 62 (17.4%) to 73 (20.5%) comparing the pre- and posttest measurements, and the perception that it was very dangerous decreased from 288 (80.9%) to 271 (76.1%).

The proportion perceiving that it was dangerous to consume alcohol decreased from (193/356, 54.2%) to (131/356, 36.8%) comparing the pre- and posttest measurements, and the perception that it was very dangerous increased from (156/356, 43.8%) to (210/356, 60%).

Similarly, the proportion of the students perceiving that it was dangerous to use marijuana decreased from (141/356, 39.6%) to (111/356, 31.2%) and that it was very dangerous increased from (203/356, 57%) to (235/356, 66%).

The overall risk perception of the use of these drugs showed significant differences only for the use of 1 cigarette (χ^2_4=21.6; *P*<.001) and the use of alcohol (χ^2_4=15.3; *P*<.001); there was a marginal difference for the use of marijuana (χ^2_4=3.6; *P*=.057). There was no significant difference in the risk perception of smoking the 5 cigarettes.

Tables 1-4 show the GEE models in which gender, drug knowledge, life skills, and self-esteem were included as covariates (Multimedia Appendix 2). The analysis showed that knowledge of smoking was the variable that generated a significant increase in risk perception (odds ratio [OR] 80.060, 95% CI 0.009-0.110; *P*=.02) of smoking 1 cigarette (Table 1).

Knowledge of marijuana (OR 0.091, 95% CI 0.023-,0.160; *P*=.009), and self-esteem (OR 0.131, 95% CI 0.046-0.217; *P*=.003), were the variables that produced significant increases in the risk perception of consuming 5 cigarettes (Table 2).

The variables that increased the risk perception of alcohol consumption were knowledge of smoking (OR 0.089, 95% CI 0.043-0.135; *P*<.001), assertiveness (OR 0.102, 95% CI 0.004-0.200; *P*=.04), resistance to peer pressure (OR 0.078, 95% CI 0.023-0.134; *P*=.006), and self-esteem (OR 0.140, 95% CI 0.063-0.216; *P*<.001; Table 3).

For perceived risk of marijuana use, the variables that significantly increased risk perception were knowledge of tobacco (OR 0.074, 95% CI 0.026-0.121; *P*=.002), knowledge of marijuana (OR 0.110, 95% CI 0.047-0.174; *P*=.001), resistance to peer pressure (OR 0.117, 95% CI 0.060-0.174; *P*<.001), and self-esteem (OR 0.165, 95% CI 0.085-0.244; *P*<.001) (Table 4).

Table 1. Generalized estimating equation model for the risk perception of smoking one cigarette.

Variable	Exp(B)	95% Wald CI for exp(B)	*P* value
(Intersection)	−6.555	−8.725 to −4.385	<.001
Gender (men)	0.017	−0.323 to 0.357	.92
Knowledge about tobacco	0.060	0.009 to 0.110	.02
Knowledge about marijuana	0.038	−0.030 to 0.106	.28
Knowledge about alcohol	0.002	−0.030 to 0.034	.90
Planning for the future	−0.015	−0.090 to 0.059	.68
Assertiveness	0.083	−0.023 to 0.189	.12
Expression of emotions	0.069	−0.024 to 0.162	.14
Resistance to peer pressure	0.003	−0.057 to 0.062	.93
Decision-making	0.023	−0.035 to 0.080	.44
Taking responsibility	0.015	−0.115 to 0.145	.82
Self-esteem	0.059	−0.025 to 0.142	.17

Table 2. Generalized estimating equation model for the risk perception of smoking 5 cigarettes.

Variable	Exp(B)	95% Wald CI for exp(B)	*P* value
(Intersection)	−3.265	−5.452 to −1.078	.003
Gender (men)	0.132	−0.255 to 0.518	.50
Knowledge of tobacco	0.023	−0.029 to 0.075	.39
Knowledge of marijuana	0.091	0.023 to 0.160	.009
Knowledge of alcohol	−0.006	−0.042 to 0.031	.77
Planning for the future	−0.062	−0.157 to 0.032	.20
Assertiveness	0.077	−0.039 to 0.193	.20
Expression of emotions	0.007	−0.096 to 0.110	.89
Resistance to peer pressure	0.027	−0.035 to 0.090	.39
Decision-making	−0.024	−0.087 to 0.039	.46
Taking responsibility	−0.028	−0.175 to 0.118	.70
Self-esteem	0.131	0.046 to 0.217	.003

Table 3. Generalized estimating equation model for the risk perception of frequent alcohol use.

Variable	Exp(B)	95% Wald CI for exp(B)	*P* value
(Intersection)	−8.326	−10.404 to −6.248	<.001
Gender (men)	0.087	−0.238 to 0.412	.60
Knowledge of tobacco	0.089	0.043 to 0.135	<.001
Knowledge of marijuana	0.005	−0.056 to 0.066	.87
Knowledge of alcohol	0.010	−0.020 to 0.041	.50
Planning for the future	−0.047	−0.143 to 0.049	.34
Assertiveness	0.102	0.004 to 0.200	.04
Expression of emotions	0.019	−0.068 to 0.106	.66
Resistance to peer pressure	0.078	0.023 to 0.134	.006
Decision-making	0.020	−0.033 to 0.073	.46
Taking responsibility	0.062	−0.060 to 0.183	.32
Self-esteem	0.140	0.063 to 0.216	<.001

Table 4. Generalized estimating equation model for the risk perception of marijuana use.

Variable	Exp(B)	95% Wald CI for exp(B)	*P* value
(Intersection)	−9.428	−11.611 to −7.246	<.001
Gender (men)	0.314	−0.030 to 0.658	.07
Knowledge of tobacco	0.074	0.026 to 0.121	.002
Knowledge of marijuana	0.110	0.047 to 0.174	.001
Knowledge of alcohol	−0.011	−0.044 to 0.022	.51
Planning for the future	−0.040	−0.106 to 0.026	.23
Assertiveness	0.045	−0.059 to 0.149	.40
Expression of emotions	−0.033	−0.126 to 0.061	.49
Resistance to peer pressure	0.117	0.060 to 0.174	<.001
Decision-making	0.049	−0.007 to 0.105	.08
Taking responsibility	−0.025	−0.153 to 0.104	.71
Self-esteem	0.165	0.085 to 0.244	<.001

Discussion

Principal Findings

The objective of this study was to evaluate the effectiveness of a mobile app in increasing the risk perception of tobacco, alcohol, and marijuana use based on life skills and self-esteem training components in a sample of Mexican high school students. The results of the study showed four main findings regarding the intervention: (1) it increased the overall risk perception of tobacco and alcohol use, (2) it increased the perception that it was dangerous to smoke 5 cigarettes, (3) it increased the perception that it was very dangerous to smoke 1 cigarette or to use alcohol or marijuana, and (4) the variables that significantly increased the risk perception of using these drugs were knowledge of tobacco, knowledge of marijuana, resistance to peer pressure, assertiveness, and self-esteem.

Risk perception of the possible consequences of drug use is usually low in adolescents and young adults [47], and evidence has shown that this low perception of risk can increase the likelihood of using drugs [48]. Risk perception is also associated with the type and knowledge of drugs [49]. For example, the recreational use of marijuana in Mexico is still not legal. Although the Supreme Court voted in favor of personal possession and consumption for recreational purposes in 2022, the stigma surrounding the use of this substance is still considerable, which could have an important impact on the perception of risk toward its use.

Interventions that are designed to increase risk perception have the potential to be useful tools in developing strategies to address drug use in this population. Interventions involving life skills training have shown satisfactory effects in reducing drug use and on attitudes and beliefs toward drug use [38]. It is therefore plausible to think that life skills education could also have a positive effect on risk perception and help reduce the likelihood of drug use.

Evidence shows that life skills education helps modify the risk perception of potentially harmful behaviors [50] and drug use [51].

Our results showed that the self-esteem component of the intervention significantly increased the risk perception of tobacco, alcohol, and marijuana use. This finding should be viewed with caution, given that the evidence on the relationship between self-esteem and the perception and display of risky behaviors is inconsistent. Some studies have reported weak to moderate relationships, while others have reported no correlation between these constructs [52]. However, there is also evidence that self-esteem is a protective element against potentially dangerous behaviors [53-57]. Despite the controversy in the data and in the conceptualization of self-esteem, we believe that its inclusion in intervention programs that seek to reduce substance use could be an important way to improve the results because it has been found to be relevant in improving life skills that contribute to reducing exposure to risk factors and improving the ability to protect themselves from situations that adolescents experience daily [58,59].

Limitations

The results of this study should be considered in light of its limitations. It is necessary to explore the reasons why there was no significant difference in the perception of risk between not dangerous and dangerous and to design strategies to achieve change in these categories. Another limitation is that no variables that could moderate the effect of the modification of risk perception were identified; therefore, the impact of these variables is unknown. We are not unaware of the possible influence of threats to the internal validity of this study. For example, it is possible that there were some events between the 2 measurements that could have influenced the results and of which we had no record. In addition, the selection of the participants could have had some effect, that is, because they were new students to a prestigious school, they could have responded orientated by social desirability and by certain knowledge of or familiarity with the measurement instrument. These threats can be minimized in future studies using quasi-experimental or true experimental designs. It is also important to consider the conditions for the implementation of the intervention, for example, the optimal institutional conditions of technological infrastructure, the selection and training of personnel to administer the intervention, and its incorporation into the school program as a support tool in the curriculum. Another limitation is that given the sample selection strategy, the results are not generalizable to the entire adolescent population. We also acknowledge that our intervention is very brief and this may pose a threat for effectiveness, but our results provide evidence of the potential of this intervention to become a cost-effective tool for prevention strategies and for continuing work on the impact of life skills in modifying risk perception and for the evaluation of this type of intervention in longitudinal studies. Another important element to consider is that it would be very enriching to include more detailed measures of risk perception that would allow an exhaustive taxonomy for the analysis, for example, the perception of risk of possible harm and of sanctions or the risk related to stigma and the emotional risk associated with drug use.

Conclusions and Perspectives

The development of the "What Happens If You Go Too Far?" app focused on increasing the risk perception of substance use in adolescents, based on the evidence that this perception may reduce the probability of use [16] and that the perception of negative health consequences can generate protective dynamics [6,22,24,48]. It is important to note that more research using more robust designs and sampling strategies is necessary. Our results show that the app has the potential to increase the risk perception of alcohol, tobacco, and marijuana use in high school students, and by providing evidence-based content on the psychosocial effects and risks of the use of each of these drugs and strengthening of life skills (assertiveness and resistance to peer pressure), which were significantly associated with increased risk perception, it would be plausible to improve the effectiveness of interventions aimed at preventing drug use in adolescents and youth.

The app is an accessible resource that can be included in intervention strategies for prevention, as it strengthens life skills that are useful as self-care strategies for adolescents and favors prevention with different social actors, such as educational and health centers that interact with this population, using basic technological resources. The app may also become popular in a youth population that has accepted and integrated mobile technologies into its lifestyle [8,42], thanks to their accessibility, usability, portability, interactivity, and ease of use. It provides a broad reach at low cost, with the likelihood of expansion among the adolescent population.

The incorporation of mobile technologies as tools to reduce drug use or delay the onset of drug use can be fundamental in reaching larger numbers of people with knowledge of the most commonly used drugs that are considered gateway substances to more serious drugs [8,25,41].

Acknowledgments

Funding for this study was provided by the Fundación Gonzalo Río Arronte. The authors would like to thank the Escuela Superior de Actopan affiliated with the Universidad Autónoma del Estado de Hidalgo for the facilities granted to conduct this research. The authors acknowledge the collaboration of the psychologist Julio César Flores Castro in the statistical analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Substance use history pre-intervention.
[PDF File (Adobe PDF File), 75 KB - mhealth_v11i1e37873_app1.pdf]

Multimedia Appendix 2
Code for the analyses.
[DOC File , 23 KB - mhealth_v11i1e37873_app2.doc]

References

1. World Health Organization. Global Accelerated Action for the Health of Adolescents (AA-HA!) Guidance to Support Country Implementation. Geneve: World Health Organization; 2017.
2. World Drug Report 2021. Booklet 2. Global Overview of Drug Demand and Drug Supply. United Nations Office on Drugs and Crime. 2021. URL: https://www.unodc.org/unodc/en/data-and-analysis/wdr-2021_booklet-2.html [accessed 2021-11-01]
3. Villatoro Velázquez JA, Bustos Gamino MN, Fregoso Ito DA, Fleiz Bautista C, Gutiérrez López MD, Amador Buenabad NG, et al. Contextual factors associated with marijuana use in school population. Salud Mental 2017 Jun 09;40(3):93-101 [FREE Full text] [doi: 10.17711/SM.0185-3325.2017.012]
4. Zhang S, Wu S, Wu Q, Durkin DW, Marsiglia FF. Adolescent drug use initiation and transition into other drugs: a retrospective longitudinal examination across race/ethnicity. Addict Behav 2021 Feb;113:106679 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106679] [Medline: 33032193]
5. Executive summary | Impact of COVID-19 | policy implications. United Nations. 2020. URL: https://wdr.unodc.org/wdr2020/en/exsum.html [accessed 2021-11-01]
6. Encuesta Nacional de Consumo de Drogas, Alcohol y Tabaco, ENCODAT 2016-2017. Comisión Nacional contra las Adicciones. 2017. URL: https://www.gob.mx/salud%7Cconadic/acciones-y-programas/encuesta-nacional-de-consumo-de-drogas-alcohol-y-tabaco-encodat-2016-2017-136758 [accessed 2021-10-01]
7. Encuesta de Consumo de Drogas en Estudiantes 2012 Ciudad de México. Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz. 2013 Jul 5. URL: http://www.uade.inpsiquiatria.edu.mx/pagina_contenidos/investigaciones/estudiantes_df/Estudiantes%20DF_2012/Encuesta%202012/index.html [accessed 2021-11-30]
8. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub MP. A mobile phone-based life skills training program for substance use prevention among adolescents: pre-post study on the acceptance and potential effectiveness of the program, Ready4Life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
9. Chaves M. Jóvenes, territorios y complicidades: Una antropología de la juventud urbana. Palo Alto, CA: Space; 2010.
10. Stapinski LA, Prior K, Newton NC, Deady M, Kelly E, Lees B, et al. Protocol for the inroads study: a randomized controlled trial of an internet-delivered, cognitive behavioral therapy-based early intervention to reduce anxiety and hazardous alcohol use among young people. JMIR Res Protoc 2019 Apr 12;8(4):e12370 [FREE Full text] [doi: 10.2196/12370] [Medline: 30977742]
11. Birrell L, Furneaux-Bate A, Chapman C, Newton NC. A mobile peer intervention for preventing mental health and substance use problems in adolescents: protocol for a randomized controlled trial (the mind your mate study). JMIR Res Protoc 2021 Jul 30;10(7):e26796 [FREE Full text] [doi: 10.2196/26796] [Medline: 34328426]

12. Fuentes P, González FC, Ruiz E, Vega L, Gutiérrez R. Violencia psicológica en relaciones de pareja de jóvenes. In: Fuentes de Iturbe P RL, editor. Violencia y salud mental: Perspectiva desde la Neurociencia, la Clínica y la Salud Pública. México: Grañén Porrúa e Instituto Nacional de Psiquiatría; 2022:393-417.
13. Nawi AM, Ismail R, Ibrahim F, Hassan MR, Manaf MR, Amit N, et al. Risk and protective factors of drug abuse among adolescents: a systematic review. BMC Public Health 2021 Nov 13;21(1):2088 [FREE Full text] [doi: 10.1186/s12889-021-11906-2] [Medline: 34774013]
14. Larsman P, Eklöf M, Törner M. Adolescents' risk perceptions in relation to risk behavior with long-term health consequences; antecedents and outcomes: A literature review. Safety Sci 2012 Nov;50(9):1740-1748. [doi: 10.1016/j.ssci.2012.04.009]
15. Whitesell M, Bachand A, Peel J, Brown M. Familial, social, and individual factors contributing to risk for adolescent substance use. J Addict 2013;2013:579310 [FREE Full text] [doi: 10.1155/2013/579310] [Medline: 24826363]
16. Las drogas, el cerebro y la conducta: bases científicas de la adicción. National Institute on Drug Abuse. 2020 Jun. URL: https://www.drugabuse.gov/sites/default/files/soa_sp.pdf [accessed 2022-09-29]
17. Leung KS, Ben Abdallah A, Copeland J, Cottler LB. Modifiable risk factors of ecstasy use: risk perception, current dependence, perceived control, and depression. Addict Behav 2010 Mar;35(3):201-208 [FREE Full text] [doi: 10.1016/j.addbeh.2009.10.003] [Medline: 19880258]
18. Grevenstein D, Nagy E, Kroeninger-Jungaberle H. Development of risk perception and substance use of tobacco, alcohol and cannabis among adolescents and emerging adults: evidence of directional influences. Subst Use Misuse 2015 Feb 12;50(3):376-386. [doi: 10.3109/10826084.2014.984847] [Medline: 25496046]
19. Li W, Gao L, Chen H, Cao N, Sun B. Prediction of injunctive and descriptive norms for willingness to quit smoking: the mediating role of smoking risk perception. J Substance Use 2017 Nov 07;23(3):274-279. [doi: 10.1080/14659891.2017.1394378]
20. Bejarano J, Ahumada G, Sánchez G, Cadenas N, de Marco M, Hynes M, et al. Perception of risk and drug use: an exploratory analysis of explanatory factors in six Latin American countries. J Int Drug Alcohol Tobacco Res 2011;1(1):9-17.
21. Deligianni E, Daniel OJ, Corkery JM, Schifano F, Lione LA. Impact of the UK Psychoactive Substances Act on awareness, use, experiences and knowledge of potential associated health risks of novel psychoactive substances. Br J Clin Pharmacol 2020 Mar;86(3):505-516 [FREE Full text] [doi: 10.1111/bcp.14123] [Medline: 31495968]
22. Becoña E. Bases psicológicas de la prevención del consumo de drogas. Papeles del Psicólogo 2007;28(1):11-20 [FREE Full text]
23. Uribe I, Verdugo J, Zacarías X. Relación entre percepción de riesgo y consumo de drogas en estudiantes de bachillerato. Psicología y salud 2011;21(1):47-55 [FREE Full text] [doi: 10.25009/pys.v21i1.586]
24. Las Drogas, el Cerebro y el Comportamiento La Ciencia de la Adicción. National Institute on Drug Abuse. 2014. URL: https://tinyurl.com/bdfu5p9v [accessed 2021-09-10]
25. World Drug Report 2021 Executive Summary. Viena, Austria: United Nations Office on Drugs and Crime; 2021.
26. Guidelines on mental health promotive and preventive interventions for adolescents: helping adolescents thrive. Executive summary. World Health Organization. 2021. URL: https://apps.who.int/iris/handle/10665/341140 [accessed 2022-05-01]
27. Skills for Health. Skills-based Health Education Including Life Skills: An Important Component of a Child-friendly/ Health Promoting School. Geneva: World Health Organization; 2003.
28. World Health Organization. Life Skills Education School Handbook: Prevention of Noncommunicable Diseases: Introduction. Geneva: World Health Organization; 2020.
29. Botvin G, Griffin K, Williams C. Preventing daily substance use among high school students using a cognitive-behavioral competence enhancement approach. World J Preventive Med 2015;3(3):48-53 [FREE Full text] [doi: 10.12691/jpm-3-3-1]
30. United Nations Children's Fund. Importancia del desarrollo de habilidades transferibles en América Latina y el Caribe. Fondo de las Naciones Unidas para la Infancia (UNICEF). 2020. URL: https://www.unicef.org/lac/sites/unicef.org.lac/files/2020-07/Importancia-Desarrollo-Habilidades-Transferibles-ALC_0.pdf [accessed 2021-11-01]
31. Ross DA, Hinton R, Melles-Brewer M, Engel D, Zeck W, Fagan L, et al. Adolescent well-being: a definition and conceptual framework. J Adolesc Health 2020 Oct;67(4):472-476 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.06.042] [Medline: 32800426]
32. Haug S, Paz Castro R, Wenger A, Schaub MP. A mobile phone-based life-skills training program for substance use prevention among adolescents: cluster-randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 13;9(7):e26951 [FREE Full text] [doi: 10.2196/26951] [Medline: 34255703]
33. Velasco V, Griffin KW, Antichi M, Celata C. A large-scale initiative to disseminate an evidence-based drug abuse prevention program in Italy: lessons learned for practitioners and researchers. Eval Program Plann 2015 Oct;52:27-38. [doi: 10.1016/j.evalprogplan.2015.03.002] [Medline: 25913236]
34. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
35. Moulier V, Guinet H, Kovacevic Z, Bel-Abbass Z, Benamara Y, Zile N, et al. Effects of a life-skills-based prevention program on self-esteem and risk behaviors in adolescents: a pilot study. BMC Psychol 2019 Dec 17;7(1):82 [FREE Full text] [doi: 10.1186/s40359-019-0358-0] [Medline: 31847888]

36. Martínez LB, Oviedo M, Palos PA, de la Barrera CB, de Oca AM. Habilidades para la vida, consumo de tabaco y alcohol en adolescentes. Revista Española de Drogodependencia 2010;35(1):67-77 [FREE Full text]
37. Krauskopf D, Maddaleno M, Pan American Health Organization. Estado del arte de los programas de prevención de la violencia en ámbitos escolares. Pan American Health Organization. 2006. URL: https://iris.paho.org/handle/10665.2/2732 [accessed 2021-06-10]
38. Tremblay M, Baydala L, Khan M, Currie C, Morley K, Burkholder C, et al. Primary substance use prevention programs for children and youth: a systematic review. Pediatrics 2020 Sep;146(3):e20192747. [doi: 10.1542/peds.2019-2747] [Medline: 32769198]
39. Chapman C, Champion KE, Birrell L, Deen H, Brierley M, Stapinski LA, et al. Smartphone apps about crystal methamphetamine ("Ice"): systematic search in app stores and assessment of composition and quality. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e10442 [FREE Full text] [doi: 10.2196/10442] [Medline: 30463834]
40. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630 [FREE Full text] [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
41. Giroux I, Goulet A, Mercier J, Jacques C, Bouchard S. Online and mobile interventions for problem gambling, alcohol, and drugs: a systematic review. Front Psychol 2017;8:954 [FREE Full text] [doi: 10.3389/fpsyg.2017.00954] [Medline: 28649211]
42. Birrell L, Deen H, Champion KE, Newton NC, Stapinski LA, Kay-Lambkin F, et al. A mobile app to provide evidence-based information about crystal methamphetamine (ice) to the community (cracks in the ice): co-design and beta testing. JMIR Mhealth Uhealth 2018 Dec 20;6(12):e11107 [FREE Full text] [doi: 10.2196/11107] [Medline: 30573443]
43. Qué pasa si te pasas? Google Play. URL: https://play.google.com/store/apps/details?id=mx.ohkasystems.fgra&hl=es_MX&gl=US [accessed 2022-10-10]
44. Humeniuk R, Ali R, Babor TF, Farrell M, Formigoni ML, Jittiwutikarn J, et al. Validation of the Alcohol, Smoking And Substance Involvement Screening Test (ASSIST). Addiction 2008 Jun;103(6):1039-1047. [doi: 10.1111/j.1360-0443.2007.02114.x] [Medline: 18373724]
45. Humeniuk R, Ali R, Babor T, Souza-Formigoni M, de Lacerda RB, Ling W, et al. A randomized controlled trial of a brief intervention for illicit drugs linked to the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in clients recruited from primary health-care settings in four countries. Addiction 2012 May;107(5):957-966. [doi: 10.1111/j.1360-0443.2011.03740.x] [Medline: 22126102]
46. Agresti A. An Introduction to Categorical Data Analysis. Hoboken, New Jersey: Wiley-Interscience; 2007.
47. Arterberry BJ, Treloar HR, Smith AE, Martens MP, Pedersen SL, McCarthy DM. Marijuana use, driving, and related cognitions. Psychol Addict Behav 2013 Sep;27(3):854-860 [FREE Full text] [doi: 10.1037/a0030877] [Medline: 23276319]
48. González-Roz A, Aonso-Diego G, Martínez-Loredo V, Cuesta M, Secades-Villa R. Effects of risk perception and accessibility on cannabis use among young population in Spain: findings from the 2016 national survey (ESTUDES). Subst Use Misuse 2022;57(1):36-46. [doi: 10.1080/10826084.2021.1981387] [Medline: 34678115]
49. Bujalski M, Wieczorek L, Sierosławski J. Risk perception and risk aversion among people who use New Psychoactive Substances. Int J Drug Policy 2021 Nov;97:103326. [doi: 10.1016/j.drugpo.2021.103326] [Medline: 34146791]
50. Boam S, Pulford BD. Experiencing risk: the effect of the experiential life-skills centre 'Warning Zone' on children's risk perception. J Risk Res 2017 Jul 12;22(2):177-190. [doi: 10.1080/13669877.2017.1351481]
51. Kröninger-Jungaberle H, Nagy E, von Heyden M, DuBois F, Ullrich J, Wippermann C, et al. REBOUND: a media-based life skills and risk education programme. Health Educ J 2014 Nov 16;74(6):705-719. [doi: 10.1177/0017896914557097]
52. Arsandaux J, Montagni I, Macalli M, Bouteloup V, Tzourio C, Galéra C. Health risk behaviors and self-esteem among college students: systematic review of quantitative studies. Int J Behav Med 2020 Apr;27(2):142-159. [doi: 10.1007/s12529-020-09857-w] [Medline: 32072455]
53. Banstola RS, Ogino T, Inoue S. Self-esteem, perceived social support, social capital, and risk-behavior among urban high school adolescents in Nepal. SSM Popul Health 2020 Aug;11:100570 [FREE Full text] [doi: 10.1016/j.ssmph.2020.100570] [Medline: 32258358]
54. Holman MS, Williams MN. Suicide risk and protective factors: a network approach. Arch Suicide Res 2022;26(1):137-154. [doi: 10.1080/13811118.2020.1774454] [Medline: 32522102]
55. Deng Y, Li X, Liu L, Chui WH. Suicide attempts and perceived social support among Chinese drug users: the mediating role of self-esteem and depression. Int J Environ Res Public Health 2020 Dec 30;18(1):208 [FREE Full text] [doi: 10.3390/ijerph18010208] [Medline: 33396608]
56. Colmsee IO, Hank P, Bošnjak M. Low self-esteem as a risk factor for eating disorders. Zeitschrift für Psychologie 2021 Jan;229(1):48-69. [doi: 10.1027/2151-2604/a000433]
57. Schick MR, Nalven T, Spillane NS. Drinking to fit in: the effects of drinking motives and self-esteem on alcohol use among female college students. Subst Use Misuse 2022;57(1):76-85 [FREE Full text] [doi: 10.1080/10826084.2021.1990334] [Medline: 34678113]

58. Mohammadzadeh M, Awang H, Ismail S, Kadir Shahar H. Improving emotional health and self-esteem of Malaysian adolescents living in orphanages through Life Skills Education program: a multi-centre randomized control trial. PLoS One 2019;14(12):e0226333 [FREE Full text] [doi: 10.1371/journal.pone.0226333] [Medline: 31877163]
59. Arur A, DeJaeghere J. Decolonizing life skills education for girls in Brahmanical India: a Dalitbahujan perspective. Gender Educ 2019 Apr 03;31(4):490-507. [doi: 10.1080/09540253.2019.1594707]

Abbreviations

GEE: generalized estimating equation
OR: odds ratio
WHO: World Health Organization

Edited by L Buis; submitted 10.03.22; peer-reviewed by G Tzilos Wernette, K Cummins; comments to author 17.06.22; revised version received 11.08.22; accepted 21.11.22; published 09.03.23.

Please cite as:
Fuentes A PMDC, Jiménez Tapia A, Ruiz-Cortés EM, Bolaños-Ceballos F, Flores Castro JC, Gutiérrez R, González-Forteza C
Effectiveness of a Mobile App to Increase Risk Perception of Tobacco, Alcohol, and Marijuana Use in Mexican High School Students: Quantitative Study
JMIR Mhealth Uhealth 2023;11:e37873
URL: https://mhealth.jmir.org/2023/1/e37873
doi:10.2196/37873
PMID:36892918

Original Paper

The Relationship Between How Participants Articulate Their Goals and Accomplishments and Weight Loss Outcomes: Secondary Analysis of a Pilot of a Web-Based Weight Loss Intervention

Danielle E Jake-Schoffman[1,2], PhD; Molly E Waring[3,4], PhD; Joseph DiVito[3,4], BS; Jared M Goetz[3,4], BA; Cindy Pan[3,4], BA, BS; Sherry L Pagoto[3,4], PhD

[1]Department of Health Education and Behavior, College of Health and Human Performance, University of Florida, Gainesville, FL, United States

[2]Center for Integrative Cardiovascular and Metabolic Disease, University of Florida, Gainesville, FL, United States

[3]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

[4]UConn Center for mHealth & Social Media, University of Connecticut, Storrs, CT, United States

Corresponding Author:
Danielle E Jake-Schoffman, PhD
Department of Health Education and Behavior
College of Health and Human Performance
University of Florida
PO Box 118210
Gainesville, FL, 32611
United States
Phone: 1 352 294 1046
Fax: 1 352 392 1909
Email: djakeschoffman@ufl.edu

Abstract

Background: In behavioral weight loss interventions, participants are asked to set weekly goals to support long-term habits that lead to weight loss. Although participants are asked to set and accomplish weekly goals, we do not know how often they do this and whether doing so is associated with weight loss. Web-based weight loss interventions allow for the analysis of participant engagement data, including how participants articulate their goals and accomplishments.

Objective: Using engagement data from a web-based weight loss intervention, we examined whether participants articulating their goals and accomplishments in measurable and repeating terms were associated with greater weight loss.

Methods: Adults with overweight or obesity received a 12-week Facebook-delivered weight loss intervention based on the Diabetes Prevention Program Lifestyle Intervention. Participants replied to conversation threads that queried about their goals and accomplishments. Two independent coders classified participants' posts that articulated goals or accomplishments as measurable or repeating. Crude and age-adjusted linear regression models were used to examine the relationship between the frequency of post type and percent weight loss.

Results: Participants (N=53; n=48, 91% female; n=48, 91% non-Hispanic White) were on average 46.2 (SD 10.5) years old with a mean BMI of 32.4 (SD 4.8) kg/m^2. Over 12 weeks, participants shared a median of 4 (IQR 1-8) posts that reported goals and 10 (IQR 4-24) posts that reported accomplishments. Most participants shared ≥1 post with a goal (n=43, 81%) and ≥1 post with an accomplishment (n=47, 89%). Each post reporting a goal was associated with 0.2% greater weight loss (95% CI −0.3% to 0.0%). Sharing ≥1 post with a repeating goal was associated with an average of 2.2% greater weight loss (95% CI −3.9% to −0.4%). Each post with a repeating goal was associated with an average of 0.5% greater weight loss (95% CI −1.0% to 0.0%). Sharing ≥1 post with measurable and repeating goals was associated with an average of 1.9% greater weight loss (95% CI −3.7% to −0.2%). Sharing each post with an accomplishment was associated with an average of 0.1% greater weight loss (95% CI −0.1% to 0.0%). Every post with an accomplishment that was repeating was associated with an average of 0.2% greater weight loss (95% CI −0.3% to 0.0%). Sharing other types of goals and accomplishments was not associated with weight loss.

Conclusions: In a web-based weight loss intervention, stating goals in repeating or both measurable and repeating terms was associated with greater weight loss, but simply stating them in measurable terms was not. For accomplishments, only those articulated in repeating terms were associated with greater weight loss. Posts about one-time goals and accomplishments represent an opportunity to encourage planning for future behaviors. Future research should examine if stating goals and accomplishments in repeating terms signals habit formation.

(JMIR Mhealth Uhealth 2023;11:e41275) doi:10.2196/41275

KEYWORDS

weight loss; social media; goal setting; web-based program; behavior change; habit formation; diabetes; Facebook; lifestyle

Introduction

Nearly 40% of US adults have obesity [1], putting them at risk for chronic diseases including diabetes [2], heart disease [3], and some cancers [4]. Behavioral interventions are effective at reducing these risks. For example, the Diabetes Prevention Program (DPP) Lifestyle Intervention [5] has been shown to produce significant weight loss and reduce the risk of diabetes [6]. However, behavioral interventions have not been widely disseminated in part due to the high cost [7] and barriers such as scheduling and transportation issues [8,9]. Technology tools have the potential to expand the reach of behavioral interventions through increased accessibility and access.

Behavioral weight loss programs that have traditionally been delivered in-person have now been adapted for web-based delivery [10] through novel platforms, including commercial social media platforms [11]. A recent systematic review of 21 studies of technology-delivered interventions based on the DPP Lifestyle Intervention found promising weight loss efficacy [10]. Engagement in web-based interventions entails participant actions like viewing content, commenting on posts, posting, and reacting to posts (eg, hitting a like button). In general, engagement in web-based weight loss interventions is variable, ranging from <1 to >100 mean posts made per participant [12]. However, greater participant engagement is generally associated with more weight loss [12-15]. Less is known about the type and quality of engagement that has the most impact on weight loss. As a first step in this line of inquiry, the relationship between overall participant post counts and broad categories of posts with weight loss was examined [16] in a Facebook-delivered weight loss intervention. We found that while the overall volume of participant posts was associated with weight loss, not all types of participant posts were predictive of weight loss [16]. Specifically, participant posts that mentioned a goal or an accomplishment were associated with weight loss, but posts reporting problems losing weight or that simply served to support other group members were not [16].

Posting more about goals and accomplishments, which are predictive of weight loss, is not surprising because goal setting is an effective behavioral strategy. Participants in the DPP are coached on setting weekly "SMART" (Specific, Measurable, Action-oriented, Relevant, and Time-bound) goals, which are goals that are specific, measurable, attainable, relevant, and time-bound [17] and following up on them effectively [5,18]. SMART goals may more effectively lead to habit formation than goals without these parameters. A goal is considered measurable if it is articulated in a way that can be quantified. For example, an exercise goal that mentions frequency, duration, or intensity would be considered measurable. A goal is considered time-bound if it specifies when the action is going to occur. Ideally, time-bound goals specify a repeating habit as opposed to a one-time action. For example, an exercise goal that mentions a frequency greater than 1 episode is more likely to develop into a habit than an exercise goal that mentions a frequency of 1 episode (eg, "I am going to work out 3 days this week" vs "I am going to work out today") [19]. Habit formation theory suggests that bridging the intention-translation gap, or the disconnect between the desire to routinize new healthy behaviors and actually doing so, might be overcome in part with goal setting that involves careful planning and tracking progress [18]. Thus, the articulation of goals in measurable and repeating ways is likely to be associated with better outcomes. Similarly, the articulation of accomplishments in measurable and repeating ways may also be associated with better outcomes because this would also seem to be evidence of habit formation. Articulation of accomplishments in specific terms may also be related to successful feedback loops in the goal-setting process, whereby participants are effectively receiving information about their progress in relation to a goal and are able to use that information to reflect on whether they also need to recalibrate their goals [20]. However, behavioral interventions do not coach participants on how to articulate their accomplishments per se, so we know little about how accomplishments are articulated and whether how people articulate their accomplishments is associated with better outcomes.

In this study, we examined how participants articulated both their goals and accomplishments in a web-based behavioral weight loss intervention in which all communication occurred in written discussion threads. All participant-reported goals and accomplishments were coded as measurable, repeating, both, or neither. The relationship between the frequency with which participants articulated their goals and accomplishments in these ways and their percent weight loss at 12 weeks was then examined. It was hypothesized that greater sharing goals and accomplishments that were measurable or repeating, particularly those that were *both* measurable and repeating, would be associated with greater percent weight loss than sharing of goals and accomplishments that were neither measurable nor repeating.

Methods

Sample

This study is a secondary analysis of data from a previously described randomized feasibility pilot trial of a Facebook-delivered weight loss intervention [21]. Participants were recruited from the Worcester, MA community and were eligible if they used a smartphone, used Facebook, were interested in losing weight, and had a BMI of 25-45 kg/m^2.

Participants were ineligible if they had type 1 or 2 diabetes, were unable to attend assessment visits, had participated in a previous weight loss study with our team, were in a concurrent weight loss program, were taking certain medications, had bariatric surgery, had plans to move during the study period, had a medical condition that would limit their physical activity or diet, did not have clearance from their primary care provider, were pregnant or breastfeeding, did not have a body weight scale at home, or did not speak English.

Eligible participants (N=56) were randomized to 1 of 2 conditions, both of which received a 12-week Facebook-delivered weight loss intervention. However, in 1 group, 3 participants received financial compensation to post in the group daily to be a role model for engagement and social support. The results of the overall trial have been reported previously [16,21]. The 3 participants who received financial incentives were excluded from the present analyses because the content of their posts could have been influenced by the incentives, yielding an analytic sample of 53 participants. As conditions did not differ in weight loss or frequency of participant posts [21], we combined the conditions for the present analyses. The sample size for the parent study was based on the necessities for examining feasibility and acceptability.

Intervention

All participants received a 12-week Facebook-delivered weight loss intervention based on the DPP Lifestyle Intervention [5]. Two weight loss counselors were responsible for providing counseling to each private, invitation-only Facebook group. Intervention posts were prescheduled to occur twice daily from the counselors' accounts, and then counselors logged in twice a day to generate discussion, field questions, and provide support and problem-solving to participants [21]. The program was delivered asynchronously, meaning no meeting time was required to participate but rather participants could participate whenever convenient, any time of the day or night. Participants were given a calorie goal to help them lose 1-2 pounds per week and an exercise goal of acquiring at least 175 minutes of moderate-intensity physical activity per week. However, participants were also asked to set small weekly goals for behaviors to work on, report their progress, discuss challenges, and problem solve with the counselors. They were encouraged to set SMART goals and were given instructions on what a SMART goal is. The weight loss counselors were trained to provide corrective feedback, and often participants clarified or specified their goals in a follow-up reply.

Measures

Weight Loss

Baseline and follow-up weights were measured in the lab by research staff using a calibrated balance beam scale. For participants missing weight at follow-up (n=2), we assumed no weight change (baseline value carried forward). The percent weight loss was calculated by dividing the pounds lost between baseline and follow-up by the baseline weight and multiplying by 100.

Demographics

At baseline, participants reported demographics on a survey, including age, sex, marital status, and educational attainment.

Facebook Posts

Participants' posts, including original posts and replies to other posts, were extracted from Facebook using Facebook's application programming interface with a program developed for this purpose.

Analytic Plan

Content Analysis

We previously conducted a directed content analysis of all original posts and replies shared by participants [16]. In the original analysis, all posts were classified as 1 of 10 types (eg, goal and accomplishment) to describe the overall intent of the post. In our original content analysis, posts that included a report of a goal or accomplishment could have been coded as another post type. For example, a post in which a participant shared an accomplishment and a challenge or slip was coded as a challenge or slip post. For this analysis, we re-reviewed participant posts and replies that were originally coded in other categories for the purpose of the previous study. This analysis includes all participant posts with content classified as a goal (ie, reported an intention or plan to make a healthy lifestyle change) or an accomplishment (ie, reported that they completed an action toward a goal). A subset of posts was independently reviewed by a second coder (n=113, 35% of goal posts and n=394, 50% of accomplishment posts). Codes were compared, and discrepancies were resolved through team discussion. Interrater reliability was examined by calculating percent agreement and κ statistics.

Goals

Coders classified whether the goals were measurable (ie, specified a particular behavior and how often it will happen; see Table 1). An example of a measurable physical activity goal is "I'm going to walk 3 times this week," and an example of a specific calorie goal is "I'm going to stay under 1,500 calories each day this week." Examples of nonmeasurable goals include "I will try harder to eat healthy this week," "I will increase my willpower," "I will do my best to stay active," or any other language that does not specify a specific behavior and how often it will happen. Agreement for measurable goal codes was 81.5% (κ=0.62, 95% CI 0.49-0.78). The 2 coders also determined if a goal was repeating (eg, "I will walk three times a week") versus a one-time goal (eg, "I will eat a light salad for lunch today"; Table 1). The agreement for repeating codes was 89.7% (κ=0.62, 95% CI 0.49-0.75).

Table 1. Participant posts sharing goals and accomplishments in a Facebook-delivered lifestyle intervention by whether posts were phrased in measurable or repeating terms.

Type of post	All posts, n (%)	Participants with any posts, n (%)	Number of posts per participant, median (IQR; range)	Example post
Goal posts				
All goal posts	322 (100.0)	43 (81.1)	4 (1-8; 0-24)	N/A[a]
Measurable and repeating	64 (19.9)	28 (52.8)	1 (0-2; 0-8)	"My first exercise goal is to walk for at least fifteen minutes a day this week."
Measurable, not repeating	162 (50.3)	41 (77.4)	2 (1-5; 0-14)	"I'm putting together a shopping list now so that I can get meals prepped this weekend for my lunches and dinners after work."
Not measurable, repeating	16 (5.0)	11 (20.8)	0 (0-0; 0-3)	"...stay away from junk food and hitting the gym more"
Not measurable, not repeating	80 (24.8)	34 (64.2)	1 (0-2; 0-9)	"I simply want to feel better mentally and physically. I hope to learn more about nutrition."
Accomplishment posts				
All accomplishment posts	789 (100.0)	47 (88.7)	10 (4-24; 0-62)	N/A
Measurable and repeating	146 (18.5)	39 (73.6)	2 (0-4; 0-11)	"...I did everything right this week. I walked a lot every day. I [swam] laps one day and took that zumba class. I drank water like it was my job and stayed within my calorie range every day except yesterday."
Measurable, not repeating	287 (36.4)	43 (81.1)	3 (1-6; 0-36)	"I did a mile without stopping. Took about 20 minutes. It's a start."
Not measurable, repeating	146 (18.5)	36 (67.9)	2 (0-4; 0-11)	"I have been following my plan with healthy eating and walking. Slow and steady wins the race!"
Not measurable, not repeating	210 (26.6)	39 (73.6)	3 (0-6; 0-14)	"I'm like a little kid and being told no makes the treat all the more desirable. So that is why I will do a bite or two as a cheat and then walk away!"

[a]N/A: not applicable.

Accomplishments

Coders classified accomplishments as measurable or repeating. Measurable accomplishments specified a particular behavior and how often it occurred. An example of a measurable accomplishment is reporting a daily step total (eg, "Got 10K steps today"), a healthy food selection (eg, "Chose the grilled chicken sandwich for lunch"), or eliminating an unhealthy habit (eg, "Cut out my after dinner snack today"; see Table 1). Agreement for measurable codes was 82.4% (κ=0.65; 95% CI 0.58-0.72). An example of a nonmeasurable accomplishment is "I choose to not get down on myself about bad days" or "I have been going out for walks." The 2 coders also determined if an accomplishment was repeating (eg, "I was under my calorie goal on five days this week") or a one-time accomplishment (eg, "I was under my calorie goal today"). Agreement for repeating codes was 86.3% (κ=0.70; 95% CI 0.63-0.76).

Analytic Plan

Because the number of each type of goal and accomplishment per participant was not normally distributed, we reported medians, IQRs, and ranges. The relationship between frequency of each post type and percent weight loss with crude and age-adjusted linear regression models was assessed. Because the total number of posts sharing goals or accomplishments of different types was small for some post types (eg, repeating but not measurable goals), the association between posting one or more of each type versus none and percent weight loss was first examined. The association between the number of posts of each type and percent weight loss was then examined. Together, these analyses describe if sharing just 1 goal or accomplishment post of each type (ie, measurable and repeating) is associated with weight loss, as well as specifically how much weight loss is associated with each post of a specific type. Similar to other studies examining the relationship between engagement in a web-based intervention and weight loss [13], we planned to also adjust for gender and race or ethnicity; however, we did not have an adequate distribution in our sample to do so, and thus our adjusted regression model includes only age. All regression models were checked to ensure they met the assumptions of linear modeling. Scatter plots of any posts and number of posts and percent weight loss (ie, a linear relationship) were reviewed. Diagnostic plots were examined to assess the distribution of residuals (ie, normally distributed and homogeneity of variance), and Shapiro-Wilk tests were used to additionally assess whether residuals were normally distributed. We also assessed for the presence of outlying or potentially influential observations. Results from these analyses indicated no clear violations of the model assumptions and adequate model fit. Analyses were conducted using SAS 9.4 (SAS Institute, Inc).

Ethics Approval

The University of Massachusetts Medical School Institutional Review Board approved this study (H00001484). As this was a pilot trial, with data collection initiated before 2017, it did not meet the Applicable Clinical Trial final rule for ClinicalTrials.gov and was not preregistered.

Results

Demographics and General Result

Participants (N=53) were predominantly female (n=48, 91%), non-Hispanic White (n=48, 91%), married (n=35, 66%), and college educated (n=30, 57%). On average, participants were 46.2 (SD 10.5) years old with a baseline BMI of 32.4 (SD 4.8) kg/m^2. As previously reported, during the 12-week program, participants lost an average of 2.6% ±3.5% (range −12.5% to 5.4%) of their body weight, with 26% (n=14) losing 5% or more of their baseline body weight [19]. Overall, participants made 2918 posts to the Facebook group, with a median of 37 (IQR 16-76; range 0-262) posts per participant. Higher post frequency in general was significantly associated with greater percent weight loss (*r*=−0.38; *P*=.005) [16].

Goal Posts

Overview

During the 12-week intervention, participants posted 322 posts that included a goal, of which 19.9% (n=64) were measurable and repeating, 50.3% (n=162) were measurable but not repeating, 5% (n=16) were not measurable but were repeating, and 24.8% (n=80) were not measurable or repeating. The majority (n=43, 81%) of participants posted at least 1 goal, and participants shared a median of 4 (IQR 1-8; range 0-24) posts that included goals (Table 1). In adjusted models, the frequency of sharing goal posts was associated with weight loss, and each goal post shared was associated with an average of 0.2% greater weight loss (95% CI −0.3% to 0.0%; Table 2).

Table 2. Distribution of participant posts sharing measurable or repeating goals and percent weight loss in a Facebook-delivered lifestyle intervention.

Type of goal post	All goal posts, n (%)	Shared 1+ goal posts[a]			Number of goal posts[b]		
		Participants with any goal posts, n (%)	Crude β (95% CI)	Adjusted[c] β (95% CI)	Number of goal posts per participant, median (IQR; range)	Crude β (95% CI)	Adjusted[c] β (95% CI)
All goal posts	322 (100.0)	43 (81.1)	−1.5 (−4.0 to 0.9)	−0.7 (−3.0 to 1.7)	4 (1-8; 0-24)	*−0.2 (−0.4 to −0.1)*[d]	*−0.2 (−0.3 to 0.0)*
Measurable	226 (70.2)	42 (79.3)	−2.1 (−4.4 to 0.2)	−1.3 (−3.5 to 0.9)	3 (1-7; 0-17)	*−0.3 (−0.5 to −0.1)*	−0.2 (−0.4 to 0.0)
Not measurable	96 (29.8)	34 (64.2)	−1.6 (−3.6 to 0.3)	−1.1 (−2.9 to 0.7)	1 (0-2; 0-10)	*−0.5 (−0.9 to −0.1)*	−0.4 (−0.7 to 0.0)
Repeating	80 (24.8)	30 (56.6)	*−2.8 (−4.5 to −1.0)*	*−2.2 (−3.9 to −0.4)*	1 (0-2; 0-8)	*−0.7 (−1.2 to −0.3)*	*−0.5 (−1.0 to 0.0)*
Not repeating	242 (75.2)	43 (81.1)	−1.5 (−4.0 to 0.9)	−0.7 (−3.0 to 1.7)	3 (1-7; 0-20)	*−0.2 (−0.4 to −0.1)*	−0.2 (−0.4 to 0.0)
Measurable and repeating	64 (19.9)	28 (52.8)	*−2.6 (−4.4 to −0.8)*	*−1.9 (−3.7 to −0.2)*	1 (0-2; 0-8)	*−0.7 (−1.2 to −0.1)*	−0.4 (−1.0 to 0.2)
Measurable, not repeating	162 (50.3)	41 (77.4)	−1.7 (−4.0 to 0.5)	−0.8 (−3.0 to 1.4)	2 (1-5; 0-14)	*−0.4 (−0.7 to −0.1)*	−0.2 (−0.5 to 0.0)
Not measurable, repeating	16 (5.0)	11 (20.8)	*−3.1 (−5.3 to −0.9)*	*−2.5 (−4.6 to −0.5)*	0 (0-0; 0-3)	—[e]	—
Not measurable, not repeating	80 (24.8)	34 (64.2)	−1.6 (−3.6 to 0.3)	−1.1 (−2.9 to 0.7)	1 (0-2; 0-9)	*−0.5 (−1.0 to 0.0)*	−0.3 (−0.7 to 0.2)

[a]β represents the difference in mean percent weight loss for participants who shared 1+ goal post versus those who did not share any goal posts.

[b]β represents the difference in mean percent weight loss for each goal post shared.

[c]Adjusted for age.

[d]Italicized entries indicate statistical significance (*P*<.05).

[e]Due to the small sample size, this type of post did not meet the assumptions for the regression model.

Measurable Goals

Most participants (n=42, 79%) shared at least 1 post that included a measurable goal. In adjusted models, sharing posts that included measurable goals was not associated with greater weight loss (Table 2).

Repeating Goals

More than half of the participants (n=30, 57%) shared at least 1 post that included a repeating goal. In adjusted models, sharing at least 1 post that included a repeating goal was associated with an average of 2.2% greater weight loss (95% CI −3.9% to −0.4%), and each post that mentioned a repeating goal was associated with an average of 0.5% greater weight loss (95% CI −1.0% to 0.0%).

Measurable and Repeating Goals

More than half of the participants (n=28, 53%) shared at least 1 post that included a goal that was both measurable and repeating (Table 1). In adjusted models, sharing at least 1 post that included a measurable and repeating goal was associated with an average of 1.9% greater weight loss (95% CI −3.7% to −0.2%; Table 2); sharing each post that mentioned a goal that was measurable and repeating was not associated with greater weight loss.

Accomplishment Posts

Overview

During the 12-week intervention, participants posted 789 posts that included an accomplishment, of which 18.5% (n=146) were measurable and repeating, 36.4% (n=287) were measurable but not repeating, 18.5% (n=146) were not measurable but were repeating, and 26.6% (n=210) were neither measurable nor repeating (Table 1). The majority of participants (n=47, 88%) posted at least 1 accomplishment, and participants shared a median of 10 (IQR 4-24; range 0-62) posts that included accomplishments (Table 1). In adjusted models, the frequency of sharing accomplishment posts was associated with weight loss, where each accomplishment post was associated with an average of 0.1% greater weight loss (95% CI −0.1% to 0.0%; Table 3).

Table 3. Distribution of participant posts sharing measurable or repeating accomplishments and percent weight loss in a Facebook-delivered lifestyle intervention.

Type of accomplishment post	All accomplishment posts, n (%)	Shared 1+ accomplishment posts[a]			Number of accomplishment posts[b]		
		Participants with any accomplishment posts, n (%)	Crude β (95% CI)	Adjusted[c] β (95% CI)	Number of accomplishment posts per participant, median (IQR; range)	Crude β (95% CI)	Adjusted[c] β (95% CI)
All accomplishment posts	789 (100.0)	47 (88.7)	0.0 (−3.0 to 3.0)	0.2 (−2.6 to 2.9)	10 (4-24; 0-62)	*−0.1 (−0.2 to 0.0)*[d]	*−0.1 (−0.1 to 0.0)*
Measurable	433 (54.9)	46 (86.8)	0.1 (−2.7 to 2.9)	0.1 (−2.5 to 2.7)	6 (2-9; 0-49)	*−0.1 (−0.2 to 0.0)*	−0.1 (−0.2 to 0.0)
Not measurable	356 (45.1)	42 (79.3)	−1.7 (−4.1 to 0.6)	−1.2 (−3.3 to 1.0)	4 (2-9; 0-25)	*−0.2 (−0.4 to −0.1)*	*−0.2 (−0.3 to 0.0)*
Repeating	292 (37.0)	41 (77.4)	−1.8 (−4.1 to 0.4)	−1.9 (−3.9 to 0.2)	4 (1-9; 0-23)	*−0.3 (−0.4 to −0.1)*	*−0.2 (−0.3 to 0.0)*
Not repeating	497 (63.0)	46 (86.8)	−1.3 (−4.1 to 1.6)	−0.6 (−3.2 to 2.0)	6 (3-13; 0-42)	*−0.1 (−0.2 to 0.0)*	−0.1 (−0.2 to 0.0)
Measurable and repeating	146 (18.5)	39 (73.6)	−1.4 (−3.5 to 0.8)	−1.7 (−3.6 to 0.3)	2 (0-4; 0-11)	*−0.3 (−0.6 to 0.0)*	−0.2 (−0.5 to 0.1)
Measurable, not repeating	287 (36.4)	43 (81.1)	−1.7 (−4.1 to 0.7)	−0.9 (−3.2 to 1.4)	3 (1-6; 0-36)	−0.1 (−0.3 to 0.0)	−0.1 (−0.2 to 0.0)
Not measurable, repeating	146 (18.5)	36 (67.9)	*−2.0 (−4.0 to 0.0)*	−1.7 (−3.6 to 0.1)	2 (0-4; 0-11)	*−0.5 (−0.8 to −0.3)*	*−0.4 (−0.7 to −0.2)*
Not measurable, not repeating	210 (26.6)	39 (73.6)	−1.3 (−3.5 to 0.8)	−0.9 (−2.9 to 1.1)	3 (0-6; 0-14)	*−0.3 (−0.5 to 0.0)*	−0.2 (−0.4 to 0.1)

[a]β represents the difference in mean percent weight loss for participants who shared 1+ goal post versus those who did not share any goal posts.

[b]β represents the difference in mean percent weight loss for each goal post shared.

[c]Adjusted for age.

[d]Italicized entries indicate statistical significance (P<.05).

Measurable Accomplishments

Most participants (n=46, 87%) shared at least 1 post that included a measurable accomplishment. In adjusted models, sharing posts that included measurable goals was not associated with greater weight loss (Table 3).

Repeating Accomplishments

Most participants (n=41, 77%) shared at least 1 post that included a repeating accomplishment. While sharing at least 1 post that included a repeating accomplishment was not associated with greater weight loss overall, sharing each post that included an accomplishment that was repeating was associated with an average of 0.2% greater weight loss (95% CI −0.3% to 0.0%; Table 3).

Measurable and Repeating Accomplishments

Most participants (n=39, 74%) shared at least 1 post that mentioned an accomplishment that was both measurable and repeating (Table 1). In adjusted models, sharing such posts was not associated with greater weight loss (Table 3).

Discussion

Principal Findings

Engagement data from web-based weight loss programs that rely on text-based interactions allows us to study how participants discuss their goals and accomplishments. This study provides insights into the specific ways people articulate their goals and accomplishments that may signal the development of habits that promote weight loss. A greater frequency of sharing goals and accomplishments were both independently related to weight loss. In particular, goals that were articulated in either repeating or both measurable and repeating terms were associated with greater weight loss, but for accomplishments, only those articulated in repeating terms were associated with greater weight loss.

In the DPP Lifestyle Intervention, participants are taught to set SMART goals each week [5,18], but variability exists in the extent to which they actually do so. In this study, we found that setting SMART goals and doing so frequently were associated with greater weight loss. Participants who specified parameters about their goals may have been more likely to have had a specific plan in place, which is an important element of goal setting. Goal setting, and specifically setting the context, frequency, duration, or intensity of a goal, has been defined within the Behavior Change Taxonomy as encompassing action planning as well [22]. Indeed, setting specific parameters around a goal is associated with improved outcomes in behavioral interventions. Our findings regarding goals are in line with past research that showed that participants who set physical activity goals that specified the time a goal would be enacted were more likely to follow through with the goal [23]. This may be particularly important for participants with ambitious weight-loss goals. One study found that among people who set large weight loss goals, those with more specific diet and exercise goals lost more weight [24].

These findings suggest several areas for additional research. First, research is needed about how best to support participants in a web-based setting to set SMART goals and how to provide corrective feedback when goals are stated in a way that is neither measurable nor repeating. The weight loss counselors provided corrective feedback; however, participants did not always modify their goal in a reply, though it is possible they may have changed their plan without sharing this in the group. Further, little is known about why participants do not set SMART goals when they are coached to do so. Research is needed to understand reluctance to articulate specific targets in a goal (eg, frequency or duration), as it may be related to low self-efficacy, poor planning skills (eg, executive function deficits), or barriers to accomplishing the goal. As an example of the latter, a participant who feels little control over their schedule may be reluctant to publicly share a measurable and repeating exercise goal because they are not confident that they will be able to accomplish the goal. It may also be that participants set SMART goals but do not share them in the group in a way that articulates the features of SMART. Web-based behavioral interventions (where participants engage in conversation threads) provide an opportunity to flag and intervene upon goals that are not expressed in measurable or repeating terms. Digital tools that assist participants in setting SMART goals also may be useful in web-based behavioral weight-loss programs. Additional content on the merits of SMART goals and research supporting their efficacy may be useful.

In regards to accomplishments, findings revealed that more frequent sharing of accomplishments in general was predictive of weight loss in general, as was sharing of accomplishments expressed in repeating terms (eg, went for a walk every day at lunch last week). This suggests that the expression of repeating accomplishments may signal habit formation [25], an important behavioral milestone associated with long-term weight loss maintenance [26]. Indeed, a growing number of recent studies find that habit, as opposed to intentional choices, contributes more to behaviors affecting energy balance, such as physical activity, diet, and sedentary time [25,27,28]. Further, sharing repeating accomplishments may also be associated with participants more effectively applying feedback when reflecting on their goal progress [20]. This reflection could then lead to a recalibration of goals when necessary [20].

Contrary to our hypothesis, participant posts sharing accomplishments in measurable terms were not predictive of weight loss. In fact, posts sharing accomplishments in *nonmeasurable* terms were predictive of weight loss. We did not coach participants on how to apply the SMART framework to accomplishments in the same way that we did goals. For example, a participant might have accomplished their goal to stay under their daily calorie goal but reported successful execution of this goal more succinctly as “I did well with my eating this week!” Further research should explore how people prefer to articulate their accomplishments in a web-based group setting and the barriers to doing so. Perhaps people who accomplish ambitious goals avoid being overly descriptive of their accomplishments so as not to come off as braggadocios in the group, or it may simply feel like providing all the details of their accomplishment in their post is too time-consuming to type out. People are likely to post their accomplishments to solicit positive reinforcement from the counselor and the group. Thus, great detail about the accomplishment may be felt unnecessary to produce such a response.

In this study, goals and accomplishments were solicited through a variety of posts at different times, so they were not necessarily linked to each other in a way that would show reliably whether a goal was reported later as an accomplishment. On Sunday mornings, participants were asked to report on how they did on their goals each week, and some responded, but others shared at different times or not at all, and this varied from week to week. Future studies could more directly link specific goals to associated accomplishments to examine the extent to which setting a goal and reporting on the outcome are more predictive of weight loss than how goals and accomplishments are articulated.

Limitations and Strengths

This study has several limitations. First, it is possible that how participants articulated their goals and accomplishments was not exactly how they thought of their goals and accomplishments. For example, a participant could have set SMART goals for themselves but articulated them in the group in a way that was neither measurable nor repeating. However, we did find a relationship between how goals and accomplishments were articulated in the group and weight loss that suggests that what is shared in the group may be a proxy for how they were thinking about their goals and accomplishments. Second, our sample was predominantly female and non-Hispanic White, similar to many behavioral weight loss intervention trials [29,30], and the analyses conducted here were planned post hoc and therefore not taken into account in the original sample size estimation. Combined with our modest sample size, this homogeneous sample limited our ability to control for additional variables in our analyses, such as gender and race or ethnicity. Research is needed with larger and more diverse (eg, racially or ethnically, educationally, and by gender) samples to explore if similar traits of goals and accomplishments are associated with weight loss. Finally, the parent study excluded participants who did not have a body weight scale at home or did not speak English, which reduced the generalizability of the results.

This study also has several strengths. A major strength of this study is the use of objective engagement data to define goals and accomplishments. Content analysis of micro-level engagement data from web-based weight loss programs provides the opportunity to dive deeply into participation in ways that have not previously been available to researchers [31]. While participation and engagement have traditionally been conceptualized as session attendance [32-34], web-based programs that allow people to engage via written exchanges offer a transcript of every conversation that occurred during the intervention. These data can be used to study the nuances of how specific types of engagement impact weight loss. While we focus here on using these data to the specifics of setting goals and reporting accomplishments that are related to weight loss, this method has the potential to be adapted to better understand a range of discussions and post types and their relationships with behavioral and clinical outcomes.

Conclusions

Much remains to be explored to fully understand not just how much engagement but what types of engagement are associated with better outcomes in web-based behavioral weight loss programs [31]. A deeper understanding of what types of engagement are associated with greater weight loss can help us refine intervention content in ways that solicit these types of engagement from participants. This type of data may also offer guidance to participants on how to engage in ways that may increase their success in web-based lifestyle interventions.

Acknowledgments

This work was supported by K24HL124366 (to SLP). We would like to thank the research participants and counselors for their support of this research.

Conflicts of Interest

None declared.

References

1. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020(360):1-8 [FREE Full text] [Medline: 32487284]
2. Verma S, Hussain ME. Obesity and diabetes: an update. Diabetes Metab Syndr 2017;11(1):73-79. [doi: 10.1016/j.dsx.2016.06.017] [Medline: 27353549]
3. Khan SS, Ning H, Wilkins JT, Allen N, Carnethon M, Berry JD, et al. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. JAMA Cardiol 2018;3(4):280-287 [FREE Full text] [doi: 10.1001/jamacardio.2018.0022] [Medline: 29490333]
4. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K, International Agency for Research on Cancer Handbook Working Group. Body fatness and cancer—viewpoint of the IARC working group. N Engl J Med 2016;375(8):794-798 [FREE Full text] [doi: 10.1056/NEJMsr1606602] [Medline: 27557308]
5. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
6. National Diabetes Prevention Program: why participate. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/prevention/why-participate.html [accessed 2022-11-28]
7. Pagoto S. The current state of lifestyle intervention implementation research: where do we go next? Transl Behav Med 2011;1(3):401-405 [FREE Full text] [doi: 10.1007/s13142-011-0071-x] [Medline: 24073065]
8. McVay MA, Yancy WS, Scott CN, Wilson-Barlow L, Woolson S, McSherry WC, et al. Patient factors associated with initiation of behavioral weight loss treatment: a prospective observational study in an integrated care setting. Transl Behav Med 2017;7(1):75-83 [FREE Full text] [doi: 10.1007/s13142-016-0430-8] [Medline: 27501799]
9. McVay MA, Yancy WS, Bennett GG, Jung SH, Voils CI. Perceived barriers and facilitators of initiation of behavioral weight loss interventions among adults with obesity: a qualitative study. BMC Public Health 2018;18(1):854 [FREE Full text] [doi: 10.1186/s12889-018-5795-9] [Medline: 29996812]
10. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
11. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social media and obesity in adults: a review of recent research and future directions. Curr Diab Rep 2018;18(6):34. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
12. Pagoto SL, Waring ME, Schneider KL, Oleski JL, Olendzki E, Hayes RB, et al. Twitter-delivered behavioral weight-loss interventions: a pilot series. JMIR Res Protoc 2015;4(4):e123 [FREE Full text] [doi: 10.2196/resprot.4864] [Medline: 26500186]
13. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]

14. West DS, Harvey JR, Krukowski RA, Prewitt TE, Priest J, Ashikaga T. Do individual, online motivational interviewing chat sessions enhance weight loss in a group-based, online weight control program? Obesity (Silver Spring) 2016;24(11):2334-2340 [FREE Full text] [doi: 10.1002/oby.21645] [Medline: 27616628]
15. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
16. Pagoto S, Waring M, Jake-Schoffman D, Goetz J, Michaels Z, Oleski J, et al. What type of engagement predicts success in a Facebook weight loss group? In: Proceedings of the 51st Annual Hawaii International Conference on System Sciences. 2018 Presented at: 51st Annual Hawaii International Conference on System Sciences; January 2-6, 2018; Big Island, HI URL: https://scholarspace.manoa.hawaii.edu/items/c938eb3d-ccd0-445a-975f-2ac7c8f5da10 [doi: 10.24251/hicss.2018.419]
17. Doran GT. There's a SMART way to write management's goals and objectives. J Manag Rev 1981;70(11):35-36.
18. National Diabetes Prevention Program: curricula and handouts. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/prevention/resources/curriculum.html [accessed 2022-12-02]
19. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013;7(suppl 1):S137-S158. [doi: 10.1080/17437199.2011.603640]
20. Swann C, Rosenbaum S, Lawrence A, Vella SA, McEwan D, Ekkekakis P. Updating goal-setting theory in physical activity promotion: a critical conceptual review. Health Psychol Rev 2021;15(1):34-50. [doi: 10.1080/17437199.2019.1706616] [Medline: 31900043]
21. Pagoto S, Waring M, Oleski J, May C, Evans M. The feasibility of incentivizing participation in an online social network weight loss program. In: Proceedings of the 50th Annual Hawaii International Conference on System Sciences. 2017 Presented at: 50th Annual Hawaii International Conference on System Sciences; January 3-7, 2017; Big Island, HI URL: https://scholarspace.manoa.hawaii.edu/items/8da3d452-b44c-4182-b0d6-8af454d98391 [doi: 10.3390/bios8010003]
22. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
23. Fleig L, Gardner B, Keller J, Lippke S, Pomp S, Wiedemann AU. What contributes to action plan enactment? Examining characteristics of physical activity plans. Br J Health Psychol 2017;22(4):940-957. [doi: 10.1111/bjhp.12263] [Medline: 28833976]
24. Dombrowski SU, Endevelt R, Steinberg DM, Benyamini Y. Do more specific plans help you lose weight? Examining the relationship between plan specificity, weight loss goals, and plan content in the context of a weight management programme. Br J Health Psychol 2016;21(4):989-1005. [doi: 10.1111/bjhp.12212] [Medline: 27454908]
25. Gardner B, de Bruijn GJ, Lally P. A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. Ann Behav Med 2011;42(2):174-187. [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]
26. Cleo G, Isenring E, Thomas R, Glasziou P. Could habits hold the key to weight loss maintenance? A narrative review. J Hum Nutr Diet 2017;30(5):655-664. [doi: 10.1111/jhn.12456] [Medline: 28150402]
27. Fürtjes S, King JA, Goeke C, Seidel M, Goschke T, Horstmann A, et al. Automatic and controlled processing: implications for eating behavior. Nutrients 2020;12(4):1097 [FREE Full text] [doi: 10.3390/nu12041097] [Medline: 32326623]
28. Howlett N, Schulz J, Trivedi D, Troop N, Chater A. Determinants of weekly sitting time: construct validation of an initial COM-B model and comparison of its predictive validity with the Theory of Planned Behaviour. Psychol Health 2021;36(1):96-114. [doi: 10.1080/08870446.2020.1763994] [Medline: 32403945]
29. Pagoto SL, Schneider KL, Oleski JL, Luciani JM, Bodenlos JS, Whited MC. Male inclusion in randomized controlled trials of lifestyle weight loss interventions. Obesity (Silver Spring) 2012;20(6):1234-1239 [FREE Full text] [doi: 10.1038/oby.2011.140] [Medline: 21633403]
30. Haughton CF, Silfee VJ, Wang ML, Lopez-Cepero AC, Estabrook DP, Frisard C, et al. Racial/ethnic representation in lifestyle weight loss intervention studies in the United States: a systematic review. Prev Med Rep 2018;9:131-137 [FREE Full text] [doi: 10.1016/j.pmedr.2018.01.012] [Medline: 29616185]
31. Pagoto S, Waring ME. A call for a science of engagement: comment on Rus and Cameron. Ann Behav Med 2016;50(5):690-691. [doi: 10.1007/s12160-016-9839-z] [Medline: 27663577]
32. Skrine Jeffers K, Castellon-Lopez Y, Grotts J, Mangione CM, Moin T, Tseng CH, et al. Diabetes Prevention Program attendance is associated with improved patient activation: results from the Prediabetes Informed Decisions and Education (PRIDE) study. Prev Med Rep 2019;16:100961 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100961] [Medline: 31516814]
33. Centers for Disease Control and Prevention Diabetes Prevention Recognition Program: standards and operating procedures. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/prevention/pdf/dprp-standards.pdf [accessed 2022-12-08]
34. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]

Abbreviations

DPP: Diabetes Prevention Program
SMART: Specific, Measurable, Action-oriented, Relevant, and Time-bound

Edited by L Buis; submitted 21.07.22; peer-reviewed by J Antoun, E Eppes, YJ Oh, M Kapsetaki; comments to author 21.09.22; revised version received 15.12.22; accepted 25.01.23; published 16.03.23.

Please cite as:
Jake-Schoffman DE, Waring ME, DiVito J, Goetz JM, Pan C, Pagoto SL
The Relationship Between How Participants Articulate Their Goals and Accomplishments and Weight Loss Outcomes: Secondary Analysis of a Pilot of a Web-Based Weight Loss Intervention
JMIR Mhealth Uhealth 2023;11:e41275
URL: https://mhealth.jmir.org/2023/1/e41275
doi:10.2196/41275
PMID:36927569

Original Paper

Characterization of Self-reported Improvements in Knowledge and Health Among Users of Flo Period Tracking App: Cross-sectional Survey

Liudmila Zhaunova[1], PhD; Ryan Bamford[2], MSc; Tara Radovic[2,3], MSc; Aidan Wickham[2], PhD; Kimberly Peven[2,4], PhD; Jazz Croft[2], PhD; Anna Klepchukova[2], MD; Sonia Ponzo[2,5], PhD

[1]Flo Health LTU, Vilnius, Lithuania

[2]Flo Health UK Limited, London, United Kingdom

[3]Department of Psychology and Ergonomics, Technische Universitaet Berlin, Berlin, Germany

[4]Maternal, Adolescent, Reproductive & Child Health Centre, London School of Hygiene & Tropical Medicine, London, United Kingdom

[5]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Liudmila Zhaunova, PhD
Flo Health LTU
Saltoniškių g. 2
Vilnius, LT-01109
Lithuania
Phone: 370 60396823
Email: l_zhaunova@flo.health

Abstract

Background: Research shows that poor knowledge and awareness of menstrual and pregnancy health among women are associated with adverse reproductive health and pregnancy outcomes. Menstrual cycle– and pregnancy-tracking mobile apps are promising tools for improving women's awareness of and attitudes toward their reproductive health; however, there is little information about subscribers' perceptions of app functionality and its impact on their knowledge and health.

Objective: This study aimed to explore knowledge and health improvements related to menstrual cycle and pregnancy, as well as improvements in general health among Flo app users. We also investigated what components of the Flo app were associated with the abovementioned improvements and evaluated whether those improvements differed based on education level, country of residence (low- and middle-income vs high-income countries), free or premium subscription to the app, short- or long-term use of the app, and frequency of use.

Methods: Flo subscribers who had been using the app for no less than 30 days, completed a web-based survey. A total of 2212 complete survey responses were collected. The survey included demographic questions and questions about motivations guiding the use of the Flo app and which components of the app improved their knowledge and health, as well as to what extent.

Results: Most study participants reported improvements in menstrual cycle (1292/1452, 88.98%) and pregnancy (698/824, 84.7%) knowledge from Flo app use. Participants with higher levels of education and those from high-income countries reported using the app predominantly for getting pregnant (χ^2_1=4.2, *P*=.04; χ^2_1=52.3, *P*<.001, respectively) and pregnancy tracking (χ^2_1=19.3, *P*<.001; χ^2_1=20.9, *P*=.001, respectively). Participants with less education reported using the app to avoid pregnancy (χ^2_1=4.2; *P*=.04) and to learn more about their body (χ^2_1=10.8; *P*=.001) and sexual health (χ^2_1=6.3; *P*=.01), while participants from low- and middle-income countries intended to mainly learn more about their sexual health (χ^2_1=18.2; *P*<.001). Importantly, the intended use of the app across education levels and country income levels matched areas in which they had gained knowledge and achieved their health goals upon use of the Flo app. Period, fertile days, and ovulation predictions as well as symptom tracking were consistently the top 3 components in the app that helped users with their cycle knowledge and general health. Reading articles or watching videos helped with users' education regarding their pregnancy. Finally, the strongest improvements in knowledge and health were observed in premium, frequent, and long-term users.

Conclusions: This study suggests that menstrual health apps, such as Flo, could present revolutionary tools to promote consumer health education and empowerment on a global scale.

(JMIR Mhealth Uhealth 2023;11:e40427) doi:10.2196/40427

KEYWORDS

health knowledge; menstrual cycle; pregnancy; period-tracking app; digital health; women's health

Introduction

Background

Women's health has long been understudied and underfunded, compromising the amount of health information available to women as well as the quality of health care they receive [1]. Recently, more efforts have been made to include women in clinical trials, and more policies have been introduced to enhance and promote women's health globally [2-4]. However, many cultures still hold different myths and taboos regarding the reproductive health and menstrual cycle, specifically. Studies have shown that there is a large knowledge gap among adolescent girls and adult women, especially in low- and middle-income countries (LMICs), about menstrual cycle health [5,6], fertility [5,7], healthy pregnancy [8,9], and overall women's health and well-being [10,11]. For example, a North Indian study found that only 31% of respondents reported the right definition of the menstrual cycle [6], whereas an extremely high prevalence of inadequate knowledge of symptoms, risk factors, and complications of preeclampsia (88.4%) was found among pregnant women in Ghana [9].

In high-income countries (HICs), despite an increase in growing social and governmental momentum to improve reproductive health education [12-16], there is still a knowledge gap and misconceptions about reproductive and fertility health as well as reluctance to seek medical treatment for some health issues [17,18]. A Chicago-based study reported that one of the main reasons for a delayed fibroid diagnosis was the belief among women that heavy menstrual bleeding was normal [17]. Another study conducted on American women of reproductive age found that almost 60% of all surveyed women did not know when they had higher chances of conception during the menstrual cycle, while more than two-thirds were unaware that pain during menstruation (eg, owing to conditions such as endometriosis) may correlate with a woman's ability to get pregnant [18]. A study conducted in England found that 30% of adolescent girls were unaware if their periods were regular; furthermore, most of them were considerably less aware of endometriosis than of other chronic diseases such as diabetes or epilepsy [19]. Another study found that only 42% of pregnant women in Italy knew of all the main risk factors in pregnancy such as alcohol consumption, smoking, passive smoking, and obesity [20]. Thus, low menstrual cycle and pregnancy health awareness among women and people who menstruate might cause adverse effects on women's and newborn health and well-being and can restrict women from their daily sociocultural activities and impact their quality of life [21,22]. In addition, poor knowledge and misconceptions about reproductive health might delay the time to diagnosis and increase medical costs associated with treatment [17,23].

Given the lack of reproductive health knowledge among women, identifying medically credible, easy-to-access and understand sources of reproductive health–related information is crucial. On the basis of survey data, women's health care providers (HCPs) are one of the main sources of trustworthy reproductive health–related knowledge; however, more than one-third of survey respondents reported visiting their reproductive HCP less than once per year or having never visited one [18]. Furthermore, the use of medical jargon or more dense language by HCPs can confuse patients and result in inability to recall the information that they received during medical consultations [24-27]. Health-related internet websites are another source of education and empowerment for women [18,28,29]. Nevertheless, web-based health sources vary in quality, accuracy, and reliability, and incorrect information or advice from websites may affect health-related behavior and decisions [30].

In the past few years, the rapid growth and acceptance of mobile health apps specifically designed to address women's health needs have made substantial progress in the area of reproductive and pregnancy knowledge improvements [31-33]. Female health apps are often perceived as helpful by their users, as they provide easy-to-access information that helps them feel more knowledgeable and supported [34,35]. In addition to the educational component, apps for women's health often provide their users with tracking functionality for their self-knowledge, monitoring, and recording [35,36]. Such functionality includes menstrual cycle and pregnancy tracking or tracking of cycle-related symptoms, for example, hormone-triggered migraines or mood swings [33,37-39]. Previous studies have shown that health mobile apps also facilitate good habit formation and health promotion by providing a number of functions to map behavior patterns across time, including, but not limited to, sleep, diet, physical activity, and mental health practices etc [40,41]. Despite the prevalence and apparent popularity of women health–tracking apps, there remains a knowledge gap regarding the accuracy and safety of the medical advice provided as well as the effectiveness of digital health products in changing health behaviors. Researchers, clinicians, and patient groups have recently advocated for more rigorous requirements regarding the validation and evaluation of digital health solutions, with scientists highlighting the need for a unified evidence generation framework [42-45]. As such, there is a need for health app developers to conduct and publish research assessing the effectiveness of products at both pre- and postmarket time points, so that end users and clinicians are provided with sufficient evidence to make an informed decision [46,47].

One of the mobile apps dedicated to women's health and well-being is the Flo app (by Flo Health Inc). Flo offers its users artificial intelligence–based period and ovulation predictions and allows them to track their periods, fertile window, ovulation, and symptoms while trying to get pregnant or avoid pregnancy, as well as to track pregnancy symptoms and fetal development. In addition to the tracking functionality, Flo provides its users with evidence-based and expert-reviewed educational content available through the in-app library. Flo app content creation and validation are based on a peer-reviewed practice in which Flo's in-house medical doctors and external medical and science experts review each piece of content before it is published in the Flo application or on the website. Each content unit (article, graphics, courses, etc) within the app has references to peer-reviewed articles, medical guidelines, and links to internationally recognized health advocacy organizations and academic institutions. Moreover, each article has a link to the profile of the medical expert or organization who reviewed it. To check for symptoms against an array of conditions, Flo also has interactive questions available via so-called "*health assistants*" or "*chatbots*" [38]. Finally, Flo has a secure place called "Secret Chats" where women can discuss intimate topics, ask questions anonymously, and obtain support from millions of women worldwide, thus reducing perceived taboos and stigma surrounding topics such as menstrual health and sexual life [48-50].

Objective

Although the Flo app provides a range of functionalities, there is an open question as to whether the Flo users obtain any improvements in knowledge and health by using the app. Hence, the aim of this study was to describe the demographic and app use characteristics of a self-enrolled sample of Flo subscribers and explore the relationship between app use and self-reported improvements in knowledge and health related to the menstrual cycle and pregnancy, as well as improvements in general health in Flo users. We hypothesized that subscribers who use Flo Premium instead of a free version, as well as those who used the app more frequently and for a longer period, would be more likely to report increased health benefits with a greater improvement in their knowledge and understanding across different areas. We also aimed to evaluate the components of the Flo app that were associated with the aforementioned improvements.

Methods

Participants

Users of the Flo app who had the app installed in English for at least 30 days and were >18 years old were eligible to participate in this study. Recruitment took place within the Flo app between December 5, 2021, and January 16, 2022. We collected 5015 partial responses and 2212 complete survey responses. Only responses from the participants who fully completed the survey were used for the analysis (Table 1 provides the sample's demographics).

Table 1. Demographics of study participants (N=2212).

Category	Frequency, n (%)
Country (top 10)	
The United States	758 (34.3)
The United Kingdom	217 (9.8)
Canada	121 (5.5)
India	108 (4.9)
Australia	92 (4.2)
South Africa	86 (3.9)
Nigeria	82 (3.7)
Ghana	63 (2.8)
Philippines	43 (1.9)
Netherlands	32 (1.4)
Age (years)	
18-24	544 (24.6)
25-34	1264 (57.1)
35-44	370 (16.7)
45-54	34 (1.5)
Race	
White, European-American, or Caucasian	1041 (47)
Black or African American	397 (17.9)
Asian or Asian-American	223 (10.1)
Hispanic, Latino, Spanish origin	116 (5.2)
Biracial or multiracial	88 (4)
American Indian or Alaskan	12 (0.5)
Native Hawaiian or Pacific Islander	3 (0.1)
Other	225 (10.2)
Prefer not to say	107 (4.8)
Gender	
Woman	2181 (96.7)
Nonbinary	22 (1)
Genderqueer or gender fluid	19 (0.8)
Questioning or unsure	10 (0.4)
Man	5 (0.2)
Trans man	3 (0.1)
Trans woman	1 (0.1)
Agender	3 (0.1)
I prefer not to say	11 (0.5)
Education	
Doctorate degree	43 (1.9)
Master’s degree	373 (16.9)
Bachelor’s degree	759 (34.3)
Associate’s degree	139 (6.3)
High school graduate or diploma	600 (27.1)

Category	Frequency, n (%)
Some high school, no diploma	115 (5.2)
I prefer not to say	112 (5.1)
Other	71 (3.2)

Materials

Participants completed a demographic questionnaire followed by an investigator-developed quantitative survey (Multimedia Appendix 1 provides the full list of survey questions). The survey included questions addressing the aim, frequency, and duration of the app use, as well as whether and what components of the Flo app (and to what extent) helped users to improve their knowledge and health. Specifically, the survey included questions addressing improvements in knowledge and health across 5 different areas: menstrual health and education, irregular cycle and related conditions, getting pregnant and pregnancy health, general health (mental, sexual, physical health, and health behaviors), and communication with an HCP.

Overall, the survey contained 70 questions. Not every question was available for every participant, as questions were presented only if they were relevant to the participant based on their previous responses. For example, those who reported that they downloaded the app to track their pregnancy were not asked menstrual cycle–related questions. The survey was created using SurveyMonkey software (Momentive Inc).

Procedure

Participants were notified of the possibility of participating in a survey via an in-app notification. Upon clicking on the survey button, the participants were redirected to SurveyMonkey and asked to provide electronic informed consent. Those who consented proceeded to complete the survey on their electronic devices, which took an average of 8 minutes.

Ethical Considerations

The study was reviewed by the Independent Ethical Review Board (WIRB-Copernicus Group Institutional Review Board), which deemed the study exempt (IRB tracking number: 20216374).

Defining Categories

To describe app use patterns among Flo users, the following categories were defined: a short-term user was any participant who had been using the Flo app for <1 year, whereas a long-term user was anyone using the Flo app for ≥1 year. A frequent user was defined as a participant who used the app several times a day to several times a week, whereas an infrequent user was a participant who used the app once a week or less. Premium app users are those who use the paid version of the Flo app, while free app users are those who use the free version of the app. Compared with the free version, the paid app version provides users with full access to all in-app content materials (including text and video materials), all symptom- and cycle-related chats with Flo's health assistant, and all community discussion chats with an additional opportunity to ask for support and advice from a medical professional ("Ask an Expert"). Multimedia Appendix 2 lists the complete app use characteristics.

Users with higher education were defined as those who had either a bachelor's, master's, or doctorate degree, whereas users with less education included those with any education below a completed bachelor's degree. In terms of age, a younger user was anyone between the ages of 18 and 34 years, whereas older users were defined as anyone between the ages of 35 and 54 years (Table 1).

Country income level was defined based on the World Bank Country Income classification [51]. In this study, 62.4% (78/125) of countries were defined as LMICs, whereas the remaining 37.6% (47/125) of countries were classified as HICs. Multimedia Appendices 3 and 4 provide a complete breakdown of HICs and LMICs, respectively.

Statistical Analysis

Data were analyzed using Python Jupyter Notebook (version 6.0.1; Project Jupyter). All variables in the survey were categorical. Throughout the survey, participants had options to answer questions as binary categorical variables (yes or no), nominal categorical variables (eg, symptoms they believe Flo has helped with) and ordinal categorical variables (eg, 5 categories from "Not at all" to "A great deal"). Throughout the survey, depending on the participants' choices, the users were directed to answer different questions. Consequently, the number of participants who answered each question differed. Ordinal categorical variables were recoded as dichotomous variables (not at all—a little vs a moderate amount—a great deal).

Chi-squared tests compared participant responses for the association between user app subscription status (free vs premium), frequency of app use (frequent vs infrequent), and duration of use (short term vs long term). Chi-squared tests were also used to assess whether the country of residence and education level had a relationship with the reasons why they used the app, the areas in which users' knowledge was improved, and to what extent knowledge was improved. To confirm the relationship between demographic factors (age and education) and the reasons for using the app, we fitted univariate logistic regression models.

Results

Overall Knowledge and Health Improvements Upon Flo App Use

Most respondents reported that the Flo app had improved how educated they feel about their overall cycle (1292/1452, 88.98%; Table S1 in Multimedia Appendix 5, question 46) and pregnancy health (698/824, 84.7%; Table S2 in Multimedia Appendix 5, question 34). Users who had tracked their cycle with Flo felt improvements in the following top 3 menstrual health areas: they knew whether their cycle was regular (1085/1292, 84.0%), whether their cycle length was normal (978/1292, 75.7%), and whether it was normal to have certain symptoms during their cycle (755/1292, 58.4%; Table S1 in Multimedia Appendix 5, question 47). Users also reported that they believed Flo had helped them understand how to use their cycle to know when they were most fertile (780/1292, 60.4%) and how their cycle affected their physical (776/1292, 60.1%) and mental health (765/1292, 59.2%; Table S1 in Multimedia Appendix 5, question 51). In addition, 77.3% (1270/1642) of participants reported that Flo had helped them manage their menstrual symptoms, with the top 3 symptoms being "ovulation" (668/1270, 52.6%); "bad mood" (including feeling sad, guilty, irritated, obsessive thoughts, and confused or self-critical; 641/1270, 50.5%); and "cramps" (618/1270, 48.7%; Table S1 in Multimedia Appendix 5; questions 16 and 17). Finally, 65.1% (309/475) of users said that the app had helped them manage their irregular cycles and related conditions (Table S1 in Multimedia Appendix 5, question 21).

Most of the study cohort (627/824, 76.1%) who used the app to track their pregnancy reported getting pregnant while using Flo (Table S2 in Multimedia Appendix 5, question 29). Of those who became pregnant, 73.5% (461/627) of believed that using Flo had helped them become pregnant (Table S2 in Multimedia Appendix 5, question 30). More than 8 in 10 participants (530/627, 84.5%) said that Flo had helped them to prepare for a healthy pregnancy (Table S2 in Multimedia Appendix 5, question 32). Pregnancy-tracking users reported that Flo had improved their knowledge about their body during pregnancy (589/698, 84.4%); their baby's development (577/698, 82.7%); and which symptoms did not require medical attention and which did during pregnancy (505/698, 72.3%; Table S2 in Multimedia Appendix 5, question 36). The study participants also reported that Flo had improved their understanding of how pregnancy affected their physical health (509/698, 72.9%); how to manage their pregnancy symptoms (403/698, 57.7%); and how to optimize their life around their pregnancy (393/698, 56.3%; Table S2 in Multimedia Appendix 5, question 39).

A smaller number of respondents reported improvements in mental (843/2212, 38.1%), sexual (835/2212, 37.7%), and physical health (589/2212, 26.6%); health behaviors (621/2212, 28.1%); and communication with an HCP (587/2212, 26.5%; Table S3 in Multimedia Appendix 5, question 53).

Period predictions, fertile days and ovulation predictions, and symptom tracking were the top 3 components of the app that have helped users with all the abovementioned areas (Tables S1 and S3 in Multimedia Appendix 5) except pregnancy education where "reading articles or watching video sources" in the app was rated the highest (503/698, 72.1%; Table S2 in Multimedia Appendix 5, question 38). Complete responses regarding knowledge and health improvements in cycle, pregnancy, and general health are shown in Tables S1, S2, and S3 in Multimedia Appendix 5, respectively.

Knowledge and Health Improvements in Relation to Education Level

Study participants with less education were more likely to use the Flo app to track irregular cycles and related conditions to learn more about their body and sexual health and to help them not become pregnant (χ^2_1=10.9, *P*<.001; χ^2_1=10.8, *P*=.001; χ^2_1=6.3, *P*=.01; χ^2_1=4.2, *P*=.04, respectively), while higher-educated users were more likely to use the app to help them become pregnant and for pregnancy tracking (χ^2_1=4.2, *P*=.04; χ^2_1=19.3, *P*<.001, respectively; Table 2). Complete statistics on the reasons for using the app in relation to age are shown in Table S1 in Multimedia Appendix 6.

Table 2. Education versus reasons for using the app (N=2212).

Reasons for using the app	HE[a], n (%)	LE[b], n (%)	Chi-square (*df*)	*P* value
Menstrual cycle and symptom tracking	943 (80.3)	851 (82.1)	1.1 (1)	.01
Irregular cycles and related conditions	219 (18.6)	254 (24.5)	10.9 (1)	<.001
Pregnancy tracking	487 (41.4)	335 (32.3)	19.3 (1)	<.001
To help me get pregnant	657 (55.9)	537 (51.5)	4.2 (1)	.04
To help me not get pregnant	117 (10)	133 (12.8)	4.2 (1)	.04
Pregnancy loss	73 (6.2)	78 (7.5)	1.3 (1)	.26
Sexual health	305 (26)	320 (30.9)	6.3 (1)	.01
Get tailored health information	390 (33.2)	343 (33.1)	0.0(1)	.99
To learn more about my body	444 (38)	467 (45)	10.8 (1)	.001

[a]HE: higher education.

[b]LE: lower education.

Univariate logistic regression models confirmed the results of the chi-squared tests. Participants with a higher educational background, for example, master's degree, had significantly higher odds (odds ratio 1.80, 95% CI 1.15-2.86) of using the app for pregnancy tracking and significantly lower odds (odds ratio 0.59, 95% CI 0.39-0.91) of using the app to learn more about their body compared with users with "Some High School, no Diploma." Tables S2 and S3 in Multimedia Appendix 6 show full logistic regression results for education and age, respectively.

Furthermore, we examined whether both lower- and higher-educated users achieved health and knowledge improvements in the intended areas of app use. Users with less education were more likely to report that Flo had helped them improve their understanding of how to manage their menstrual symptoms (χ^2_1=4.7; *P*=.03; Table S1 in Multimedia Appendix 7, question 51) and helped them identify issues related to endometriosis (χ^2_1=8.0; *P*=.004; Table S1 in Multimedia Appendix 7, question 47). In addition, they were more likely to report that Flo had improved their sexual health knowledge (χ^2_1=9.5; *P*=.002), such as sexually transmitted infections (STIs) and how to avoid them (χ^2_1=13.1; *P*<.001); the signs and symptoms of STIs (χ^2_1=19.3; *P*<.001); safe sex (χ^2_1=11.7; *P*<.001); contraception options (χ^2_1=11.1; *P*<.001); the signs and symptoms during sex that were indicative of a health issue (χ^2_1=10.6; *P*<.001); and how to have more pleasure during sex (χ^2_1=6.7; *P*=.009; Table S2 in Multimedia Appendix 7, questions 53, 61, and 62).

Finally, users with less education were more likely to self-report that Flo had helped them improve their skin (χ^2_1=4.8; *P*=.03) or fitness (χ^2_1=5.4; *P*=.02) from a moderate to a great deal (Table S2 in Multimedia Appendix 7, question 57). In addition, they reported that Flo had helped them reduce harmful habits (χ^2_1=12.1; *P*<.001; Table S2 in Multimedia Appendix 7, question 59) and made them more confident about asking their HCPs for resources that they thought they needed (eg, stress relief, birth control, and support: χ^2_1=4.5; *P*=.03; Table S2 in Multimedia Appendix 7, question 64).

In contrast, users with higher education mainly reported improvements in pregnancy health education (Table S3 in Multimedia Appendix 7, questions 34, 36, and 39). In addition, they were more likely to report that they conceived while using Flo (χ^2_1=17.8; *P*<.001; Table S3 in Multimedia Appendix 7, question 29). Complete responses regarding knowledge and health improvements in cycle health, general health, and pregnancy health in relation to education level are shown in Tables S1, S2, and S3, respectively, in Multimedia Appendix 7.

Knowledge and Health Improvements in Relation to Country of Residence (LMICs vs HICs)

The study included participants from both HICs (1513/2212, 68.4%) and LMICs (699/2212, 31.6%), with participants from LMICs having a statistically lower level of education compared with those from HICs (χ^2_1=11.0; *P*<.001). Table 3 shows that participants from HICs were more likely to use the app for menstrual cycle and symptom tracking, for pregnancy tracking, to get help with getting pregnant, and for pregnancy loss (χ^2_1=11.8, *P*<.001; χ^2_1=20.9, *P*<.001; χ^2_1=52.3, *P*<.001; χ^2_1=22.6, *P*<.001, respectively), while participants from LMICs were more likely to use the app to improve their sexual health (χ^2_1=18.2; *P*<.001).

Table 3. Low- and middle-income countries (LMICs) and high-income countries (HICs) versus reasons for using the app (N=2212).

Reasons for using the app	HIC, n (%)	LMIC, n (%)	Chi-square (*df*)	*P* value
Menstrual cycle and symptom tracking	1257 (83.1)	537 (76.9)	11.8 (1)	<.001
Irregular cycles and related conditions	324 (21.4)	149 (21.3)	0.0 (1)	.99
Pregnancy tracking	611 (40.4)	211 (30.2)	20.9 (1)	<.001
To help me get pregnant	894 (59.1)	297 (42.5)	52.3 (1)	<.001
To help me not get pregnant	172 (11.4)	78 (11.2)	0.0(1)	.94
Pregnancy loss	130 (8.6)	21 (3)	22.6 (1)	<.001
Sexual health	385 (25.4)	240 (34.4)	18.2 (1)	<.001
Get tailored health information	502 (33.2)	231 (33)	0.0(1)	.99
To learn more about my body	611 (40.4)	303 (43.4)	1.6 (1)	.20

We further tested whether users from LMICs and HICs achieved their health and knowledge improvement in their intended areas of app use. We found that users from HICs were significantly more likely to report that they became pregnant while using the Flo app (χ^2_1=17.0; *P*<.001; Table S1 in Multimedia Appendix 8, question 29), as well as know more about their baby's development (χ^2_1=4.2; *P*=.04; Table S1 in Multimedia Appendix 8, question 36) and communicate better with their HCPs (χ^2_1=6.2; *P*=.01; Table S2 in Multimedia Appendix 8, question 53) and partners about their cycle (χ^2_1=5.7; *P*=.02; Table S3 in Multimedia Appendix 8, question 47).

Users from LMICs were significantly more likely to self-report that using Flo had helped them improve their understanding of sexual health (χ^2_1=17.7; *P*<.001; Table S2 in Multimedia Appendix 8, question 53)—more specifically, their understanding of STIs and how to avoid them (χ^2_1=8.1; *P*=.004; Table S2 in Multimedia Appendix 8, question 61) and how to have safe sex (χ^2_1=17.6; *P*<.001; Table S2 in Multimedia Appendix 8, question 61), as well as their understanding of their sexuality (χ^2_1=14.7; *P*<.001; Table S2 in Multimedia Appendix 8, question 62) and how to get more pleasure during sex (χ^2_1=5.0; *P*=.03; Table S2 in Multimedia Appendix 8, question 62). In addition to improving their sexual health, users from LMICs also reported improvements in understanding of whether their period flow was normal or not (χ^2_1=6.3; *P*=.01; Table S3 in Multimedia Appendix 8, question 47); how to identify issues related to polycystic ovary syndrome (PCOS; χ^2_1=6.7; *P*=.001; Table S3 in Multimedia Appendix 8, question 47); and how to manage menstrual symptoms (χ^2_1=23.8; *P*<.001; Table S3 in Multimedia Appendix 8, question 51) and postpartum symptoms (χ^2_1=8.9; *P*=.002; Table S1 in Multimedia Appendix 8, question 39). Finally, users from LMICs also reported improvements in physical (χ^2_1=25.1; *P*<.001; Table S2 in Multimedia Appendix 8, question 53) and mental health (χ^2_1=4.7; *P*=.03; Table S2 in Multimedia Appendix 8, question 53), including having more energy (χ^2_1=4.5; *P*=.03; Table S2 in Multimedia Appendix 8, question 57); sleeping better (χ^2_1=6.0; *P*=.01; Table S2 in Multimedia Appendix 8, question 57); having better stress management skills (χ^2_1=5.0; *P*=.03; Table S2 in Multimedia Appendix 8, question 59); and reducing harmful habits (χ^2_1=5.0, *P*=.03; Table S2 in Multimedia Appendix 8, question 59). Complete results regarding knowledge and health improvements in the pregnancy, general health, and cycle health between HICs and LMICs are shown in Tables S1, S2, and S3, respectively, in Multimedia Appendix 8.

Knowledge and Health Improvements in Relation to Paid Versus Free App Use

Similar to users with higher education, Flo Premium users reported using the app mainly for pregnancy-related issues (pregnancy tracking: χ^2_1=109.8, *P*<.001; to help get pregnant: χ^2_1=21.4, *P*<.001) as well as to obtain tailored health information relevant to them (χ^2_1=16.7; *P*<.001; Table 4).

Table 4. Reasons for using the app versus app use characteristics (N=2212).

Reasons for using the app	F[a], n (%)	P[b], n (%)	F vs P		LT[c], n (%)	ST[d], n (%)	LT vs ST		Fr[e], n (%)	I[f], n (%)	Fr vs I	
			Chi-square (*df*)	*P* value			Chi-square (*df*)	*P* value			Chi-square (*df*)	*P* value
Menstrual cycle and symptom tracking	484 (84.9)	674 (80.8)	3.7 (1)	.06	719 (88.4)	439 (74.3)	46.5 (1)	<.001	889 (82.2)	267 (83.3)	0.1 (1)	.73
Irregular cycles and related conditions	132 (23.2)	170 (20.4)	1.4 (1)	.24	194 (23.9)	108 (18.3)	6.0 (1)	.01	217 (20.1)	85 (26.3)	5.4 (1)	.02
Pregnancy tracking	137 (24)	435 (52.2)	109.8 (1)	<.001	338 (41.7)	234 (39.6)	0.5 (1)	.49	481 (44.5)	91 (28.2)	26.8 (1)	<.001
To help me get pregnant	273 (47.9)	505 (60.6)	21.4 (1)	<.001	403 (49.6)	375 (63.5)	26.1 (1)	<.001	643 (59.5)	135 (41.8)	30.8 (1)	<.001
To help me not get pregnant	69 (12.1)	84 (10.1)	1.2 (1)	.27	121 (14.9)	32 (5.4)	30.6 (1)	<.001	114 (10.5)	39 (12.1)	0.5 (1)	.50
Pregnancy loss	23 (4)	75 (9)	12.1 (1)	<.001	61 (7.5)	37 (6.3)	0.6 (1)	.43	85 (7.9)	13 (4)	5.1 (1)	.02
Sexual health	175 (30.7)	234 (28.1)	1.0 (1)	.31	273 (33.6)	136 (23)	18.0 (1)	<.001	315 (29.1)	94 (29.1)	0.0 (1)	.99
Get tailored health information	163 (28.6)	328 (39.3)	16.7 (1)	<.001	310 (38.1)	181 (30.6)	8.1 (1)	.004	389 (36)	102 (31.6)	1.9 (1)	.16
Learn more about my body	233 (40.9)	351 (42.1)	0.2 (1)	.69	368 (45.3)	216 (36.5)	10.3 (1)	.001	456 (42.2)	128 (39.6)	0.6 (1)	.45

[a]P: premium app user.
[b]F: free app user.
[c]LT: long-term app user.
[d]ST: short-term app user.
[e]Fr: frequent user of the app.
[f]I: infrequent user of the app.

We investigated whether Flo Premium users achieved health and knowledge improvements in their intended areas of the app use. First, we found that premium users of Flo were more likely to report knowledge improvements in almost all areas of pregnancy health (Table S1 in Multimedia Appendix 9, questions 29, 32, 34, 36, and 39). In addition, premium users were more likely to report knowledge and health improvements regarding their menstrual health (Table S2 in Multimedia Appendix 9, questions 17, 46, 47, and 51). Finally, Flo Premium users were more likely to report that Flo had helped them improve all aspects of general health (except sexual health), as well as improve communication with HCPs; specifically, Flo Premium users felt more confident in sharing what was going on in their body when communicating with their HCPs (Table S3 in Multimedia Appendix 9, questions 53 and 64).

Complete responses regarding knowledge and health improvements in pregnancy, cycle, and general health in relation to paid versus free app use are shown in Tables S1, S2, and S3, respectively, in Multimedia Appendix 9.

Knowledge and Health Improvements in Relation to Frequency of the App Use

Similar to users with higher education and Flo Premium users, participants who used the app frequently reported using the app mainly for pregnancy-related issues (pregnancy tracking: χ^2_1=26.8; *P*<.001; to get help with becoming pregnant: χ^2_1=30.8; *P*<.001), whereas infrequent app users were more likely to report using the app to track irregular cycles and related conditions (χ^2_1=5.4; *P*=.02; Table 4).

We examined whether frequent users of Flo achieved pregnancy health and knowledge improvements from app use. As shown in Table S1 in Multimedia Appendix 10, frequent Flo app users were more likely to report that Flo app had helped them become pregnant (χ^2_1=4.5; *P*=.03; question 30); prepare for a healthy pregnancy (χ^2_1=4.9; *P*=.03; question 32); and improve their pregnancy health knowledge (χ^2_1=16.2; *P*<.001; question 34). For example, frequent Flo app users reported that Flo had helped them to know more about their baby's development (χ^2_1=6.6; *P*=.01); their body during pregnancy (χ^2_1=6.7; *P*=.01); basic do's and don'ts during pregnancy (χ^2_1=4.5; *P*=.03; Table S1 in Multimedia Appendix 10, question 36); and how pregnancy affects their physical health (χ^2_1=4.4; *P*=.04; Table S1 in Multimedia Appendix 10, question 39).

In addition to improvements in pregnancy health and knowledge, study participants who used the Flo app frequently were more likely to report improvements in menstrual health knowledge (Table S2 in Multimedia Appendix 10) such as what symptoms were normal throughout the cycle (χ^2_1=6.1; *P*=.01; question 47) and how to estimate the most fertile period by using cycle-related knowledge (χ^2_1=32.7; *P*<.001; question 51). Frequent app users were also more likely to say that Flo had helped them to identify issues related to endometriosis (χ^2_1=3.9; *P*=.047) and improved communication with their partners about their cycle (χ^2_1=30.6; *P*<.001; Table S2 in Multimedia Appendix 10, question 47). Finally, frequent Flo app users were more likely to report improvements in communication with a HCPs (χ^2_1=13.1; *P*<.001; Table S3 in Multimedia Appendix 10, question 53), that is, they became more confident in asking questions about their health and body (χ^2_1=6.4; *P*=.01) and they became more confident in understanding their reproductive health (χ^2_1=9.1; *P*=.003; Table S3 in Multimedia Appendix 10, question 64).

Complete responses regarding knowledge and health improvements in pregnancy, cycle, and general health in relation to the frequency of the app use are shown in Tables S1, S2, and S3, respectively, in Multimedia Appendix 10.

Knowledge and Health Improvements in Relation to Long Versus Short App Use

As shown in Table 4, short-term users reported using the app mainly for help with becoming pregnant (χ^2_1=26.1; *P*<.001) as opposed to long-term users who reported using the app to help avoid pregnancy (χ^2_1=30.6; *P*<.001); track menstrual cycle and symptoms (χ^2_1=46.5; *P*<.001); track irregular cycles and related conditions (χ^2_1=6.0; *P*=.01); learn more about their body (χ^2_1=10.3; *P*=.001); learn more about sexual health (χ^2_1=18.0; *P*<.001); and obtain tailored health information relevant to them (χ^2_1=8.1; *P*=.004).

We found that short-term Flo users were more likely to report that Flo had helped them to know how to use their cycle to know when they were most fertile (χ^2_1=8.1; *P*=.008; Table S1 in Multimedia Appendix 11, question 51). At the same time, long-term users were more likely to report improvements across multiple areas of cycle, pregnancy, and general health. Long-term Flo app users were more likely to report that Flo had improved their menstrual health knowledge such as knowing whether their cycle was regular (χ^2_1=10.0; *P*=.002); cycle length was normal (χ^2_1=5.4; *P*=.02); and period flow was normal (χ^2_1=10.3; *P*=.001). In addition, they were more likely to say that Flo had helped them to identify issues related to PCOS (χ^2_1=6.9; *P*=.009) and helped them to reduce premenstrual syndrome symptoms (χ^2_1=14.4; *P*<.001; Table S1 in Multimedia Appendix 11, question 47). In addition, long-term users were more likely to say that Flo had helped them to better understand how their cycle affected their mental (χ^2_1=12.3; *P*<.001) and physical health (χ^2_1=14.7; *P*<.001) and reactions to situations (lack of patience, sadness, etc; χ^2_1=9.3; *P*=.002), as well as how to optimize their life around their cycle (χ^2_1=4.8; *P*=.03; Table S1 in Multimedia Appendix 11, question 51).

In addition, long-term users of Flo reported that had Flo helped them to improve their knowledge of fetal development (χ^2_1=5.0; *P*=.03) and the management of pregnancy symptoms (χ^2_1=10.0; *P*=.002) and postpartum symptoms (χ^2_1=26.0; *P*<.001) as well as the reduction of risk of preterm labor (χ^2_1=6.1; *P*=.01; Table S2 in Multimedia Appendix 11, question 36 and 39). Finally, long-term users were more likely to report improvements in sexual (χ^2_1=11.7; *P*<.001); mental (χ^2_1=12.2; *P*<.001); and physical health (χ^2_1=13.2; *P*<.001) and health behaviors (χ^2_1=10.9; *P*<.001; Table S3 in Multimedia Appendix 11, question 53). Most improvements were reported in sexual health knowledge such as STIs and how to avoid them (χ^2_1=20.9; *P*<.001); the signs and symptoms of STIs (χ^2_1=16.2; *P*<.001);

safe sex (χ^2_1=14.0; *P*<.001); contraception options (χ^2_1=12.3; *P*<.001); and the signs and symptoms during sex that were indicative of a health issue (χ^2_1=8.7; *P*<.003; Table S3 in Multimedia Appendix 11, question 61).

Complete responses regarding knowledge and health improvements in cycle, pregnancy, and general health in relation to short-term versus long-term app use are shown in Tables S1, S2, and S3, respectively, in Multimedia Appendix 11.

Discussion

Summary

The aim of this study was to investigate the effects of the Flo app on health and knowledge improvements, with a specific focus on menstrual cycle and pregnancy health. In addition, we aimed to explore whether such effects differed based on education level, country of residence, and app use characteristics (ie, free or paid subscription, frequent or infrequent app use, and short- or long-term app use). We found that most participants reported menstrual cycle and pregnancy knowledge and health improvements upon using the Flo app. These improvements were mostly associated with the use of period and fertile day predictions, symptom-tracking features, and the consumption of relevant in-app health content (articles and videos). Users with different levels of education, countries of residence, and app use patterns differed in their selected health goals. The strongest improvements in knowledge and health were observed in premium, frequent, and long-term users.

Benefits of Using Menstrual Cycle– and Pregnancy-Tracking Apps

Our study is the first to date examining the effect of the mobile app Flo in improving menstrual and pregnancy knowledge and health as self-reported by its users. Most users who tracked their menstrual cycle with Flo reported that the app had helped them understand whether their menstrual cycle was normal or not; specifically, whether their cycle was regular, whether their cycle length was within the norms, and whether it was typical to have certain symptoms during their cycle. Menstrual cycle is a vital indicator of women's health, and cycle abnormalities might be a sign of an underlying health issue [52]. For example, prolonged cycle length and cycle irregularity, in combination with other symptoms such as excessive body hair and acne, might be signs of PCOS [38], while painful and heavy periods or bleeding between periods might be indicative of endometriosis [53]. By improving the understanding of what a normal menstrual cycle is and what body symptoms and signs are associated with it, individuals are equipped with tools to make informed decisions as to whether symptoms they experience may require medical attention. Studies show that increased self-awareness promotes successful health behavior changes and makes individuals more proactive about seeking appropriate care when needed [31,54]. This is of vital importance given that the diagnosis of certain conditions such as endometriosis can be delayed for 8 to 12 years [53]. Prevention or early detection of reproductive conditions can improve women's quality of life as well as decrease the time to diagnosis and medical costs associated with treatment, thereby reducing the economic burden on health care systems [55,56].

One in 3 participants in our sample reported that the Flo app had helped them to improve communication with an HCP. This is not surprising considering that better self-awareness and understanding of one's own menstrual cycle and body were shown to be important for improving provider-patient communication [31]. According to the UK Women's Health Strategy Survey, 1 in 4 women do not feel comfortable talking to health care professionals about their menstrual cycle, while most women do not feel heard by their physician [57]. Our results suggest that women's health apps could improve the education and self-awareness of one's menstrual health and facilitate patient-physician conversations.

Our survey respondents also reported that the use of the Flo app had helped them manage their menstrual symptoms such as bad mood or cramps. Symptoms that women experience during their menstrual cycle can negatively impact their quality of life, work productivity, and increase health care costs [58]. Dysmenorrhea (a commonly reported cramp-like pain occurring before or during periods) can affect approximately 45% to 95% of women and is associated with productivity loss and absenteeism (ie, absence from work) [59-62]. Furthermore, symptoms of endometriosis such as painful periods or chronic pelvic pain can also negatively impact work performance and can lead to 10.8 hours of lost work per week with an annual total productivity loss of €6298 (US $6838) per woman [55,63], in addition to the profound effects on psychological and social well-being [64,65]. In addition, 65.1% (309/475) of the Flo users reported that the app had helped them manage their irregular cycles and related conditions, with 1 in 3 participants reporting improvements in their mental health. Perceived improvements in cycle symptom management were largely associated with period predictions (935/1276, 73.3%) and reading and watching articles and video sources in the app (787/1276, 61.7%). Given that menstrual cycle symptoms have an impact on several aspects of women's lives, cycle-tracking apps can help with being mentally prepared for upcoming periods (eg, via period predictions) and can provide insights into how to manage cycle symptoms (eg, via educational content).

Most participants who had used Flo to track their pregnancy reported improvements in pregnancy and postnatal health knowledge (eg, knowing how to manage pregnancy symptoms and which symptoms require medical attention) and well-being (eg, body image issues associated with pregnancy). Women's health during pregnancy has profound effects on subsequent generations [66]. Poor maternal health can lead to low neonatal weight and reduced chances of survival, congenital abnormalities, problems with child behavior, poor school performance, and adult health and productivity [67-69]. Furthermore, maternal behaviors such as smoking or alcohol consumption during pregnancy may be significant drivers of infant health issues [70,71]. Therefore, health apps that allow pregnancy tracking are promising tools for improving knowledge about maternal and newborn health, thus decreasing the chances of pregnancy-related complications and child health issues.

Benefits of Using Menstrual Cycle–Tracking Apps for LMICs and HICs

Although participants from LMICs (699/2212, 32.6% of the sample) intended to use the Flo app to improve their sexual health (eg, gaining knowledge of STIs and how to avoid them, safe sex, etc), they also reported significantly higher improvements in multiple areas of their health (menstrual cycle, mental and physical health, and improved healthy behaviors). Compared with HICs, where reproductive and sexual education is mandatory in many schools and menstrual health has started to gain increasing attention from governmental and health care authorities (eg, Women's Health Strategy in England in 2021) [72], LMICs often lack formal menstrual health education, resulting in high levels of stigma and shame surrounding menstruation and sex [73]. Poor period education results in a lack of knowledge about menstrual cycle norms; menstrual hygiene; gynecological conditions and their symptoms; and sexual health (eg, contraception and STIs). Specifically, according to the World Bank classification, LMICs have the highest prevalence of STIs (eg, gonorrhea, trichomoniasis, and syphilis), and the awareness of STIs other than HIV in those countries is very low [74-76]. Therefore, menstrual and sexual health apps, together with systemic governmental and societal education efforts, can mitigate long-term repercussions of poor health literacy, thereby decreasing unfavorable health outcomes such as infertility, abortions, preterm delivery, and perinatal and neonatal morbidities and helping with cutting health care costs associated with STI treatment [77-80].

Flo app users from HICs reported using the app to help them get pregnant and for pregnancy and menstrual cycle tracking. We found that users from HICs were statistically more likely to report that they became pregnant while using the Flo app, while 97.6% (451/462) acknowledged that fertile days and ovulation predictions made by the Flo app helped them become pregnant. Despite an increase in social and governmental momentum to improve menstrual and sexual health literacy in HICs, topics such as fertility and healthy pregnancy have not received the attention they deserve. A survey conducted on a sample of 1000 American participants showed that younger women (aged 18-24 years) demonstrated a low level of knowledge about conception, ovulation, and the effect of age on the length of time to conception and miscarriage risks. Women aged 25 to 40 years were more likely to believe in common myths around fertility and conception. Finally, more than one-quarter of all survey participants were unaware of factors impacting fertility such as STIs or obesity [18]. Therefore, health apps might help to save costs associated with infertility treatments by educating users on factors impacting fertility and providing fertile window– and ovulation-tracking functionalities to optimize conception strategies.

Menstrual cycle– and pregnancy-tracking apps might represent promising tools for improving reproductive health knowledge and health among women globally. This is largely owing to the anonymity, scalability, and accessibility of such digital solutions. According to a 2019 survey by the Kaiser Family Foundation, 1 in 3 women of reproductive age in the United States used a menstrual cycle–tracking app [81]. At the same time, an increasing adoption of mobile health apps is also observed in LMICs [82]. Thus, app developers and governmental and health authorities should collaborate to promote HCPs encouraging the adoption of menstrual cycle–tracking apps by the target group and facilitate the distribution of health apps in LMICs.

Modes (Free vs Paid), Frequency, and Length of App Use

We found that participants who used a paid version of the app, as well as those who used the app more frequently and for a longer period, were more likely to report improvements in knowledge and health compared with free app users and infrequent or short-term users. These results are not surprising, as studies show that health interventions that are used consistently for prolonged periods are more likely to positively impact behavior, thus leading to better health outcomes [83]. For example, a study conducted by Huberty et al [84] found that more frequent use of the meditation app Calm was associated with an increase in the likelihood of noticing changes in mental health, sleep, and stress levels. Another study by Han and Rhee [85] showed that users who frequently monitored their weight, food consumption, and exercise via a weight loss app, Noom, had more efficient weight loss over time. Therefore, health app developers should consider designing user-friendly, easy-to-navigate, and evidence-based technologies to encourage higher and consistent engagement with the app to achieve better health outcomes for the users.

Limitations

Although this study is the first to provide insights into the benefits of using Flo app for women's health and education, it has several limitations. First, most of the study cohort consisted of highly engaged (1650/2212, 74.6% use the Flo app several times a day to several times a week) and loyal (1305/2212, 59% use the Flo app for >1 year) users of the Flo app who may have more favorable opinions about the app and have benefited more from using it. Those who did not find Flo helpful may have elected not to participate.

Another limitation is that this was a cross-sectional, nonexperimental study that does not allow us to infer causality with regard to the impact of Flo app on user knowledge, health, and well-being. Future randomized controlled trials assessing the efficacy of the Flo app with baseline and follow-up measures should explore these relationships prospectively to determine the extent to which changes can be attributed to Flo app use. Furthermore, systematic exploration of the health benefits of available women's health apps is needed to allow clinicians and end users to make informed decisions about effective, safe, and trustworthy solutions. To this end, app developers should conduct and make publicly available pre- and postmarket research assessing the effectiveness and efficacy of their health products. Simultaneously, external organizations should conduct rigorous comparisons of such products to facilitate end users' understanding of benefits and risks. As mentioned in the Introduction section, despite the call for more rigorous evidence generation, research on this topic is scarce, and a more systematic effort is needed across the board.

Finally, the survey questions were specifically designed for this study; thus, they had not been previously validated in scientific research. This study used self-reported statements about perceived health knowledge and health behaviors. Although perceived health knowledge and behaviors do not equate to actual improvements in health literacy and health outcomes, they provide useful preliminary insights into the awareness of one's own health. Future studies should extend the findings of this report by using validated questionnaires to assess changes in menstrual and pregnancy health literacy, as well as knowledge assessment quizzes over time.

Conclusions

This study provides the first insights into the effectiveness of the menstrual cycle and women's health app, Flo, in improving user knowledge and health, and it builds toward a much-needed body of evidence in digital health. Our results highlight the opportunity for menstrual health apps to become useful tools for promoting reproductive health education and empowerment among users globally.

Acknowledgments

The authors thank the Flo users who took part in the study.

Conflicts of Interest

LZ, RB, TR, AW, KP, JC, AK, and SP were employees of Flo Health.

Multimedia Appendix 1
Survey questions.
[PDF File (Adobe PDF File), 150 KB - mhealth_v11i1e40427_app1.pdf]

Multimedia Appendix 2
Flo users' app use characteristics.
[PDF File (Adobe PDF File), 52 KB - mhealth_v11i1e40427_app2.pdf]

Multimedia Appendix 3
List of high-income countries and total number of users.
[PDF File (Adobe PDF File), 47 KB - mhealth_v11i1e40427_app3.pdf]

Multimedia Appendix 4
List of low- and middle-income countries and total number of users.
[PDF File (Adobe PDF File), 48 KB - mhealth_v11i1e40427_app4.pdf]

Multimedia Appendix 5
Cycle, pregnancy, and general health responses.
[PDF File (Adobe PDF File), 76 KB - mhealth_v11i1e40427_app5.pdf]

Multimedia Appendix 6
Reasons for using the app versus age and education.
[PDF File (Adobe PDF File), 82 KB - mhealth_v11i1e40427_app6.pdf]

Multimedia Appendix 7
Education level versus cycle, pregnancy, and general health questions.
[PDF File (Adobe PDF File), 84 KB - mhealth_v11i1e40427_app7.pdf]

Multimedia Appendix 8
High-income and low- and middle-income countries versus cycle, pregnancy, and general health questions.
[PDF File (Adobe PDF File), 76 KB - mhealth_v11i1e40427_app8.pdf]

Multimedia Appendix 9
Premium and free users versus cycle, pregnancy, and general health questions.
[PDF File (Adobe PDF File), 83 KB - mhealth_v11i1e40427_app9.pdf]

Multimedia Appendix 10
Frequent and infrequent app use versus cycle, pregnancy, and general health questions.

[PDF File (Adobe PDF File), 83 KB - mhealth_v11i1e40427_app10.pdf]

Multimedia Appendix 11
Long- and short-term app use versus cycle, pregnancy, and general health questions.
[PDF File (Adobe PDF File), 83 KB - mhealth_v11i1e40427_app11.pdf]

References

1. Mirin AA. Gender disparity in the funding of diseases by the U.S. National Institutes of Health. J Womens Health (Larchmt) 2021 Jul;30(7):956-963. [doi: 10.1089/jwh.2020.8682] [Medline: 33232627]
2. Liu KA, Mager NA. Women's involvement in clinical trials: historical perspective and future implications. Pharm Pract (Granada) 2016 Jan;14(1):708 [FREE Full text] [doi: 10.18549/PharmPract.2016.01.708] [Medline: 27011778]
3. Peters SA, Woodward M, Jha V, Kennedy S, Norton R. Women's health: a new global agenda. BMJ Glob Health 2016 Nov 01;1(3):e000080 [FREE Full text] [doi: 10.1136/bmjgh-2016-000080] [Medline: 28588958]
4. Sommer M, Caruso BA, Torondel B, Warren EC, Yamakoshi B, Haver J, et al. Menstrual hygiene management in schools: midway progress update on the "MHM in Ten" 2014-2024 global agenda. Health Res Policy Syst 2021 Jan 02;19(1):1 [FREE Full text] [doi: 10.1186/s12961-020-00669-8] [Medline: 33388085]
5. Diamond-Smith N, Onyango GO, Wawire S, Ayodo G. Knowledge of menstruation and fertility among adults in rural Western Kenya: gaps and opportunities for support. PLoS One 2020 Mar 03;15(3):e0229871 [FREE Full text] [doi: 10.1371/journal.pone.0229871] [Medline: 32126117]
6. Ahmad A, Garg SG, Gupta S, Alvi R. Knowledge and practices related to menstruation among Lucknow college students in north India: results from cross-sectional survey. medRxiv 2021 Feb 12. [doi: 10.1101/2021.02.10.21251460]
7. Hammarberg K, Zosel R, Comoy C, Robertson S, Holden C, Deeks M, et al. Fertility-related knowledge and information-seeking behaviour among people of reproductive age: a qualitative study. Hum Fertil (Camb) 2017 Jun;20(2):88-95. [doi: 10.1080/14647273.2016.1245447] [Medline: 27778517]
8. Samuel RS, Youness EM, Mohamed HA. Assessment of pregnant women's knowledge and practice about risks of iodine intake deficiency during pregnancy. Minia Sci Nurs J 2021 Dec 30;010(1):2-10 [FREE Full text] [doi: 10.21608/msnj.2021.193794]
9. Fondjo LA, Boamah VE, Fierti A, Gyesi D, Owiredu EW. Knowledge of preeclampsia and its associated factors among pregnant women: a possible link to reduce related adverse outcomes. BMC Pregnancy Childbirth 2019 Dec 02;19(1):456 [FREE Full text] [doi: 10.1186/s12884-019-2623-x] [Medline: 31791264]
10. Hamdanieh M, Ftouni L, Al Jardali B, Ftouni R, Rawas C, Ghotmi M, et al. Assessment of sexual and reproductive health knowledge and awareness among single unmarried women living in Lebanon: a cross-sectional study. Reprod Health 2021 Jan 28;18(1):24 [FREE Full text] [doi: 10.1186/s12978-021-01079-x] [Medline: 33509225]
11. Alshogran OY, Mahmoud FM, Alkhatatbeh MJ. Knowledge and awareness toward menopause and hormone therapy among premenopausal women in Jordan. Climacteric 2021 Apr;24(2):171-178. [doi: 10.1080/13697137.2020.1813099] [Medline: 32914657]
12. Weiss-Wolf J. U.S. policymaking to address menstruation: advancing an equity agenda. In: Bobel C, Winkler IT, Fahs B, Hasson KA, Kissling EA, Roberts TA, editors. The Palgrave Handbook of Critical Menstruation Studies. Singapore: Palgrave Macmillan; Jul 25, 2020:539-549.
13. Committee opinion no. 585: health literacy. Obstet Gynecol 2014 Feb;123(2 PART 1):380-383 [FREE Full text] [doi: 10.1097/01.aog.0000443277.06393.32]
14. Reis M, Ramiro L, Gaspar de Matos M, Diniz JA. The effects of sex education in promoting sexual and reproductive health in Portuguese university students. Procedia Soc Behav Sci 2011;29:477-485 [FREE Full text] [doi: 10.1016/j.sbspro.2011.11.266]
15. Janiak E, Rhodes E, Foster AM. Translating access into utilization: lessons from the design and evaluation of a health insurance web site to promote reproductive health care for young women in Massachusetts. Contraception 2013 Dec;88(6):684-690. [doi: 10.1016/j.contraception.2013.09.004] [Medline: 24120249]
16. Bush D, Brick E, East MC, Johnson N. Endometriosis education in schools: a New Zealand model examining the impact of an education program in schools on early recognition of symptoms suggesting endometriosis. Aust N Z J Obstet Gynaecol 2017 Aug;57(4):452-457. [doi: 10.1111/ajo.12614] [Medline: 28349513]
17. Ghant MS, Sengoba KS, Vogelzang R, Lawson AK, Marsh EE. An altered perception of normal: understanding causes for treatment delay in women with symptomatic uterine fibroids. J Womens Health (Larchmt) 2016 Aug;25(8):846-852 [FREE Full text] [doi: 10.1089/jwh.2015.5531] [Medline: 27195902]
18. Lundsberg LS, Pal L, Gariepy AM, Xu X, Chu MC, Illuzzi JL. Knowledge, attitudes, and practices regarding conception and fertility: a population-based survey among reproductive-age United States women. Fertil Steril 2014 Mar;101(3):767-774 [FREE Full text] [doi: 10.1016/j.fertnstert.2013.12.006] [Medline: 24484995]

19. Randhawa AE, Tufte-Hewett AD, Weckesser AM, Jones GL, Hewett FG. Secondary school girls' experiences of menstruation and awareness of endometriosis: a cross-sectional study. J Pediatr Adolesc Gynecol 2021 Oct;34(5):643-648. [doi: 10.1016/j.jpag.2021.01.021] [Medline: 33548448]
20. Esposito G, Ambrosio R, Napolitano F, Di Giuseppe G. Women's knowledge, attitudes and behavior about maternal risk factors in pregnancy. PLoS One 2015 Dec 29;10(12):e0145873 [FREE Full text] [doi: 10.1371/journal.pone.0145873] [Medline: 26714032]
21. Armour M, Hyman MS, Al-Dabbas M, Parry K, Ferfolja T, Curry C, et al. Menstrual health literacy and management strategies in young women in Australia: a national online survey of young women aged 13-25 Years. J Pediatr Adolesc Gynecol 2021 Apr;34(2):135-143. [doi: 10.1016/j.jpag.2020.11.007] [Medline: 33188935]
22. Armour M, Parry K, Al-Dabbas MA, Curry C, Holmes K, MacMillan F, et al. Self-care strategies and sources of knowledge on menstruation in 12,526 young women with dysmenorrhea: a systematic review and meta-analysis. PLoS One 2019 Jul 24;14(7):e0220103 [FREE Full text] [doi: 10.1371/journal.pone.0220103] [Medline: 31339951]
23. Simpson CN, Lomiguen CM, Chin J. Combating diagnostic delay of endometriosis in adolescents via educational awareness: a systematic review. Cureus 2021 May 20;13(5):e15143 [FREE Full text] [doi: 10.7759/cureus.15143] [Medline: 34164243]
24. Chapple A, Campion P, May C. Clinical terminology: anxiety and confusion amongst families undergoing genetic counseling. Patient Educ Couns 1997 Sep;32(1-2):81-91. [doi: 10.1016/s0738-3991(97)00065-7] [Medline: 9355575]
25. Michie S, McDonald V, Marteau TM. Genetic counselling: information given, recall and satisfaction. Patient Educ Couns 1997 Sep;32(1-2):101-106. [doi: 10.1016/s0738-3991(97)00050-5] [Medline: 9355577]
26. Roter DL, Erby L, Larson S, Ellington L. Oral literacy demand of prenatal genetic counseling dialogue: predictors of learning. Patient Educ Couns 2009 Jun;75(3):392-397. [doi: 10.1016/j.pec.2009.01.005] [Medline: 19250792]
27. Watson M, Lloyd S, Davidson J, Meyer L, Eeles R, Ebbs S, et al. The impact of genetic counselling on risk perception and mental health in women with a family history of breast cancer. Br J Cancer 1999 Feb;79(5-6):868-874 [FREE Full text] [doi: 10.1038/sj.bjc.6690139] [Medline: 10070883]
28. Huberty J, Dinkel D, Beets MW, Coleman J. Describing the use of the internet for health, physical activity, and nutrition information in pregnant women. Matern Child Health J 2013 Oct;17(8):1363-1372. [doi: 10.1007/s10995-012-1160-2] [Medline: 23090284]
29. Horgan Á, Sweeney J. University students' online habits and their use of the Internet for health information. Comput Inform Nurs 2012 Aug;30(8):402-408. [doi: 10.1097/NXN.0b013e3182510703] [Medline: 22525044]
30. Kitchens B, Harle CA, Li S. Quality of health-related online search results. Decis Support Syst 2014 Jan;57:454-462 [FREE Full text] [doi: 10.1016/j.dss.2012.10.050]
31. Levy J, Romo-Avilés N. "A good little tool to get to know yourself a bit better": a qualitative study on users' experiences of app-supported menstrual tracking in Europe. BMC Public Health 2019 Sep 03;19(1):1213 [FREE Full text] [doi: 10.1186/s12889-019-7549-8] [Medline: 31481043]
32. Karasneh RA, Al-Azzam SI, Alzoubi KH, Muflih SM, Hawamdeh SS. Smartphone applications for period tracking: rating and behavioral change among women users. Obstet Gynecol Int 2020 Sep 01;2020:2192387 [FREE Full text] [doi: 10.1155/2020/2192387] [Medline: 32952563]
33. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
34. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
35. Epstein DA, Lee NB, Kang JH, Agapie E, Schroeder J, Pina LR, et al. Examining menstrual tracking to inform the design of personal informatics tools. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May 02;2017:6876-6888 [FREE Full text] [doi: 10.1145/3025453.3025635] [Medline: 28516176]
36. Ford A, De Togni G, Miller L. Hormonal health: period tracking apps, wellness, and self-management in the era of surveillance capitalism. Engag Sci Technol Soc 2021 Oct 05;7(1):48-66 [FREE Full text] [doi: 10.17351/ests2021.655]
37. Gambier-Ross K, McLernon DJ, Morgan HM. A mixed methods exploratory study of women's relationships with and uses of fertility tracking apps. Digit Health 2018 Jul 25;4:2055207618785077 [FREE Full text] [doi: 10.1177/2055207618785077] [Medline: 30225095]
38. Jain T, Negris O, Brown D, Galic I, Salimgaraev R, Zhaunova L. Characterization of polycystic ovary syndrome among Flo app users around the world. Reprod Biol Endocrinol 2021 Mar 03;19(1):36 [FREE Full text] [doi: 10.1186/s12958-021-00719-y] [Medline: 33658043]
39. Grieger JA, Norman RJ. Menstrual cycle length and patterns in a global cohort of women using a mobile phone app: retrospective cohort study. J Med Internet Res 2020 Jun 24;22(6):e17109 [FREE Full text] [doi: 10.2196/17109] [Medline: 32442161]
40. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995 [FREE Full text] [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]

41. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
42. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. It Inf Technol 2019;61(5-6):253-263. [doi: 10.1515/itit-2019-0019]
43. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3:110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
44. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
45. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
46. Scott KM, Gome GA, Richards D, Caldwell PH. How trustworthy are apps for maternal and child health? Health Technol (Berl) 2015 Mar 12;4(4):329-336 [FREE Full text] [doi: 10.1007/s12553-015-0099-x]
47. Wicks P, Chiauzzi E. 'Trust but verify'--five approaches to ensure safe medical apps. BMC Med 2015 Sep 25;13:205 [FREE Full text] [doi: 10.1186/s12916-015-0451-z] [Medline: 26404791]
48. Cook RJ, Dickens BM. Reducing stigma in reproductive health. Int J Gynaecol Obstet 2014 Apr;125(1):89-92. [doi: 10.1016/j.ijgo.2014.01.002] [Medline: 24513258]
49. Johnston-Robledo I, Chrisler JC. The menstrual mark: menstruation as social stigma. Sex Roles 2011 Jul 31;68(2):9-18 [FREE Full text] [Medline: 33347160]
50. Babbar K, Martin J, Ruiz J, Parray AA, Sommer M. Menstrual health is a public health and human rights issue. Lancet Public Health 2022 Jan;7(1):e10-e11 [FREE Full text] [doi: 10.1016/S2468-2667(21)00212-7] [Medline: 34717798]
51. World Bank country and lending groups. World Bank Data Help Desk. 2021. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2022-11-28]
52. Foster C, Al-Zubeidi H. Menstrual irregularities. Pediatr Ann 2018 Jan 01;47(1):e23-e28. [doi: 10.3928/19382359-20171219-01] [Medline: 29323693]
53. Kiesel L, Sourouni M. Diagnosis of endometriosis in the 21st century. Climacteric 2019 Jun;22(3):296-302. [doi: 10.1080/13697137.2019.1578743] [Medline: 30905186]
54. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
55. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. Hum Reprod 2012 May;27(5):1292-1299. [doi: 10.1093/humrep/des073] [Medline: 22422778]
56. Jensen JT, Lefebvre P, Laliberté F, Sarda SP, Law A, Pocoski J, et al. Cost burden and treatment patterns associated with management of heavy menstrual bleeding. J Womens Health (Larchmt) 2012 May;21(5):539-547. [doi: 10.1089/jwh.2011.3147] [Medline: 22360696]
57. Results of the ‘women’s health – let’s talk about it’ survey. Department of Health & Social Care, UK. 2022 Apr 13. URL: https://www.gov.uk/government/consultations/womens-health-strategy-call-for-evidence/outcome/results-of-the-womens-health-lets-talk-about-it-survey [accessed 2022-12-02]
58. Ponzo S, Wickham A, Bamford R, Radovic T, Zhaunova L, Peven K, et al. Menstrual cycle-associated symptoms and workplace productivity in US employees: a cross-sectional survey of users of the Flo mobile phone app. Digit Health 2022 Dec 15;8:20552076221145852 [FREE Full text] [doi: 10.1177/20552076221145852] [Medline: 36544535]
59. Titilayo A, Agunbiade OM, Banjo O, Lawani A. Menstrual discomfort and its influence on daily academic activities and psychosocial relationship among undergraduate female students in Nigeria. Tanzan J Health Res 2009 Oct;11(4):181-188. [doi: 10.4314/thrb.v11i4.50173] [Medline: 20734697]
60. Sharma A, Taneja DK, Sharma P, Saha R. Problems related to menstruation and their effect on daily routine of students of a medical college in Delhi, India. Asia Pac J Public Health 2008;20(3):234-241. [doi: 10.1177/1010539508316939] [Medline: 19124317]
61. Ballagh SA, Heyl A. Communicating with women about menstrual cycle symptoms. J Reprod Med 2008 Nov;53(11):837-846. [Medline: 19097516]
62. Iacovides S, Avidon I, Baker FC. What we know about primary dysmenorrhea today: a critical review. Hum Reprod Update 2015 Nov;21(6):762-778. [doi: 10.1093/humupd/dmv039] [Medline: 26346058]
63. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, World Endometriosis Research Foundation Global Study of Women's Health consortium. Reprint of: impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2019 Oct;112(4 Suppl1):e137-e152 [FREE Full text] [doi: 10.1016/j.fertnstert.2019.08.082] [Medline: 31623725]

64. Unsal A, Ayranci U, Tozun M, Arslan G, Calik E. Prevalence of dysmenorrhea and its effect on quality of life among a group of female university students. Ups J Med Sci 2010 May;115(2):138-145 [FREE Full text] [doi: 10.3109/03009730903457218] [Medline: 20074018]
65. Iacovides S, Avidon I, Bentley A, Baker FC. Reduced quality of life when experiencing menstrual pain in women with primary dysmenorrhea. Acta Obstet Gynecol Scand 2014 Feb;93(2):213-217 [FREE Full text] [doi: 10.1111/aogs.12287] [Medline: 24266425]
66. Zylbersztejn A, Gilbert R, Hjern A, Wijlaars L, Hardelid P. Child mortality in England compared with Sweden: a birth cohort study. Lancet 2018 May 19;391(10134):2008-2018 [FREE Full text] [doi: 10.1016/S0140-6736(18)30670-6] [Medline: 29731173]
67. Kahn RS, Wilson K, Wise PH. Intergenerational health disparities: socioeconomic status, women's health conditions, and child behavior problems. Public Health Rep 2005 Jul;120(4):399-408 [FREE Full text] [doi: 10.1177/003335490512000407] [Medline: 16025720]
68. Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter L, MaternalChild Undernutrition Study Group. Maternal and child undernutrition: consequences for adult health and human capital. Lancet 2008 Jan 26;371(9609):340-357 [FREE Full text] [doi: 10.1016/S0140-6736(07)61692-4] [Medline: 18206223]
69. Ahmad K, Kabir E, Keramat SA, Khanam R. Maternal health and health-related behaviours and their associations with child health: evidence from an Australian birth cohort. PLoS One 2021 Sep 13;16(9):e0257188 [FREE Full text] [doi: 10.1371/journal.pone.0257188] [Medline: 34516561]
70. Burns L, Breen C, Bower C, O' Leary C, Elliott EJ. Counting fetal alcohol spectrum disorder in Australia: the evidence and the challenges. Drug Alcohol Rev 2013 Sep;32(5):461-467. [doi: 10.1111/dar.12047] [Medline: 23617437]
71. Sutin AR, Flynn HA, Terracciano A. Maternal smoking during pregnancy and offspring personality in childhood and adulthood. J Pers 2018 Aug;86(4):652-664 [FREE Full text] [doi: 10.1111/jopy.12342] [Medline: 28833118]
72. Results of the women’s health strategy call for evidence – written responses from organisations and experts. Department of Health & Social Care, UK. 2022 Apr 13. URL: https://tinyurl.com/ysazn7r4 [accessed 2022-11-24]
73. Hennegan J, Shannon AK, Rubli J, Schwab KJ, Melendez-Torres GJ. Women's and girls' experiences of menstruation in low- and middle-income countries: a systematic review and qualitative metasynthesis. PLoS Med 2019 May 16;16(5):e1002803 [FREE Full text] [doi: 10.1371/journal.pmed.1002803] [Medline: 31095568]
74. Mayaud P, Mabey D. Approaches to the control of sexually transmitted infections in developing countries: old problems and modern challenges. Sex Transm Infect 2004 Jun;80(3):174-182 [FREE Full text] [doi: 10.1136/sti.2002.004101] [Medline: 15169997]
75. Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. Bull World Health Organ 2019 Aug 01;97(8):548-62P [FREE Full text] [doi: 10.2471/BLT.18.228486] [Medline: 31384073]
76. Vasudeva M, Nakka R, Stock S, Ghebremichael M. Associations between awareness of sexually transmitted infections (STIs) and prevalence of STIs among sub-Saharan African men and women. Trop Med Infect Dis 2022 Jul 26;7(8):147 [FREE Full text] [doi: 10.3390/tropicalmed7080147] [Medline: 35893654]
77. Wynn A, Bristow CC, Cristillo AD, Murphy SM, van den Broek N, Muzny C, et al. Sexually transmitted infections in pregnancy and reproductive health: proceedings of the star sexually transmitted infection clinical trial group programmatic meeting. Sex Transm Dis 2020 Jan;47(1):5-11 [FREE Full text] [doi: 10.1097/OLQ.0000000000001075] [Medline: 31658242]
78. Adachi K, Xu J, Yeganeh N, Camarca M, Morgado MG, Watts DH, NICHD HPTN 040 Study Team. Combined evaluation of sexually transmitted infections in HIV-infected pregnant women and infant HIV transmission. PLoS One 2018 Jan;13(1):e0189851 [FREE Full text] [doi: 10.1371/journal.pone.0189851] [Medline: 29304083]
79. Korenromp EL, Rowley J, Alonso M, Mello MB, Wijesooriya NS, Mahiané SG, et al. Global burden of maternal and congenital syphilis and associated adverse birth outcomes-estimates for 2016 and progress since 2012. PLoS One 2019 Feb 27;14(2):e0211720 [FREE Full text] [doi: 10.1371/journal.pone.0211720] [Medline: 30811406]
80. Meites E, Gaydos CA, Hobbs MM, Kissinger P, Nyirjesy P, Schwebke JR, et al. A review of evidence-based care of symptomatic trichomoniasis and asymptomatic trichomonas vaginalis infections. Clin Infect Dis 2015 Dec 15;61 Suppl 8(Suppl 8):S837-S848 [FREE Full text] [doi: 10.1093/cid/civ738] [Medline: 26602621]
81. Health Apps and Information Survey. Henry J Kaiser Family Foundation. 2019 Sep. URL: https://files.kff.org/attachment/Topline-Health-Apps-and-Information-Survey-September-2019 [accessed 2022-12-07]
82. Alaiad A, Alsharo M, Alnsour Y. The determinants of m-Health adoption in developing countries: an empirical investigation. Appl Clin Inform 2019 Oct;10(5):820-840 [FREE Full text] [doi: 10.1055/s-0039-1697906] [Medline: 31667819]
83. Kim M, Yang J, Ahn WY, Choi HJ. Machine learning analysis to identify digital behavioral phenotypes for engagement and health outcome efficacy of an mHealth intervention for obesity: randomized controlled trial. J Med Internet Res 2021 Jun 24;23(6):e27218 [FREE Full text] [doi: 10.2196/27218] [Medline: 34184991]
84. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]

85. Han M, Rhee SY. Effect of adherence to smartphone app use on the long-term effectiveness of weight loss in developing and OECD countries: retrospective cohort study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e13496 [FREE Full text] [doi: 10.2196/13496] [Medline: 34255708]

Abbreviations

HCP: health care provider
HIC: high-income country
LMIC: low- and middle-income country
PCOS: polycystic ovary syndrome
STI: sexually transmitted infection

Edited by L Buis; submitted 20.06.22; peer-reviewed by D Reifegerste, S Meherali, L Xu; comments to author 04.11.22; revised version received 16.12.22; accepted 06.03.23; published 26.04.23.

Please cite as:
Zhaunova L, Bamford R, Radovic T, Wickham A, Peven K, Croft J, Klepchukova A, Ponzo S
Characterization of Self-reported Improvements in Knowledge and Health Among Users of Flo Period Tracking App: Cross-sectional Survey
JMIR Mhealth Uhealth 2023;11:e40427
URL: https://mhealth.jmir.org/2023/1/e40427
doi:10.2196/40427
PMID:37099370

Original Paper

Delivering a Postpartum Weight Loss Intervention via Facebook or In-Person Groups: Results From a Randomized Pilot Feasibility Trial

Molly E Waring[1,2], PhD; Sherry L Pagoto[1,2], PhD; Tiffany A Moore Simas[3,4,5,6], MEd, MPH, MD; Loneke T Blackman Carr[2,7], RD, PhD; Madison L Eamiello[1], MA; Brooke A Libby[1], MPH, MS; Lauren R Rudin[1], BS; Grace E Heersping[1], BS

[1]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States

[2]UConn Center for mHealth & Social Media, University of Connecticut, Storrs, CT, United States

[3]Department of Obstetrics & Gynecology, University of Massachusetts Chan Medical School/UMass Memorial Health, Worcester, MA, United States

[4]Department of Pediatrics, University of Massachusetts Chan Medical School/UMass Memorial Health, Worcester, MA, United States

[5]Department of Psychiatry, University of Massachusetts Chan Medical School/UMass Memorial Health, Worcester, MA, United States

[6]Department of Population & Quantitative Health Sciences, University of Massachusetts Chan Medical School/UMass Memorial Health, Worcester, MA, United States

[7]Department of Nutritional Sciences, University of Connecticut, Storrs, CT, United States

Corresponding Author:
Molly E Waring, PhD
Department of Allied Health Sciences
University of Connecticut
358 Mansfield Rd
Unit 1101
Storrs, CT, 06269
United States
Phone: 1 8604861446
Email: molly.waring@uconn.edu

Abstract

Background: Postpartum weight retention contributes to weight gain and obesity. Remotely delivered lifestyle interventions may be able to overcome barriers to attending in-person programs during this life phase.

Objective: This study aimed to conduct a randomized feasibility pilot trial of a 6-month postpartum weight loss intervention delivered via Facebook or in-person groups. Feasibility outcomes were recruitment, sustained participation, contamination, retention, and feasibility of study procedures. Percent weight loss at 6 and 12 months were exploratory outcomes.

Methods: Women with overweight or obesity who were 8 weeks to 12 months post partum were randomized to receive a 6-month behavioral weight loss intervention based on the Diabetes Prevention Program lifestyle intervention via Facebook or in-person groups. Participants completed assessments at baseline, 6 months, and 12 months. Sustained participation was defined by intervention meeting attendance or visible engagement in the Facebook group. We calculated percent weight change for participants who provided weight at each follow-up.

Results: Among individuals not interested in the study, 68.6% (72/105) were not interested in or could not attend in-person meetings and 2.9% (3/105) were not interested in the Facebook condition. Among individuals excluded at screening, 18.5% (36/195) were ineligible owing to reasons related to the in-person condition, 12.3% (24/195) related to the Facebook condition, and 2.6% (5/195) were unwilling to be randomized. Randomized participants (n=62) were a median of 6.1 (IQR 3.1-8.3) months post partum, with a median BMI of 31.7 (IQR 28.2-37.4) kg/m^2. Retention was 92% (57/62) at 6 months and 94% (58/62) at 12 months. The majority (21/30, 70%) of Facebook and 31% (10/32) of in-person participants participated in the last intervention module. Half (13/26, 50%) of Facebook and 58% (15/26) of in-person participants would be likely or very likely to participate again if they had another baby, and 54% (14/26) and 70% (19/27), respectively, would be likely or very likely to recommend the program to a friend. In total, 96% (25/26) of Facebook participants reported that it was convenient or very convenient to log into the Facebook group daily compared with 7% (2/27) of in-person participants who said it was convenient or very convenient to attend group meetings each week. Average weight loss was 3.0% (SD 7.2%) in the Facebook condition and 5.4% (SD 6.8%) in the in-person condition at 6 months, and 2.8% (SD 7.4%) in the Facebook condition and 4.8% (SD 7.6%) in the in-person condition at 12 months.

Conclusions: Barriers to attending in-person meetings hampered recruitment efforts and intervention participation. Although women found the Facebook group convenient and stayed engaged in the group, weight loss appeared lower. Research is needed to further develop care models for postpartum weight loss that balance accessibility with efficacy.

Trial Registration: ClinicalTrials.gov, NCT03700736; https://clinicaltrials.gov/ct2/show/NCT03700736

(JMIR Mhealth Uhealth 2023;11:e41545) doi:10.2196/41545

KEYWORDS

postpartum weight loss; Facebook; social media; pilot study; feasibility; mobile phone

Introduction

Background

Postpartum weight retention contributes to long-term weight gain and obesity among childbearing persons [1-4]. Among women in the multicenter Community Child Health Network study, a third of women with a normal weight BMI prepregnancy had overweight or obesity at 1 year post partum, and 44% of women with overweight prepregnancy transitioned to obesity by 1 year post partum [1]. Among women enrolled in the 2016 Los Angeles Mommy and Baby (LAMB) follow-up study, 35% of women with normal weight BMI prepregnancy had transitioned to overweight or obesity by 2 years after giving birth [5]. Postpartum weight retention varies, and although many women return to their prepregnancy weight by 1 year post partum, a substantial proportion retain substantial amounts of weight [2,6]. In a cohort of women delivering their first child from Pennsylvania, 24% had retained 1-9 pounds (0.5-4 kg) and 24% had retained ≥10 pounds (4.5 kg) at 1 year post partum [6]. In the LAMB cohort, 35% had retained ≥10 pounds at 2 years post partum [5].

Although systematic reviews and meta-analyses have demonstrated the efficacy of lifestyle interventions targeting dietary intake and physical activity for weight loss during the postpartum period [7-10], interventions with numerous in-person sessions are not a good logistical match for the busy lives of many postpartum women [11-14]. Indeed, high attrition from treatment has plagued many postpartum weight loss intervention studies [7,8]. Remotely delivered lifestyle interventions can overcome some of the barriers to attending in-person meetings during the postpartum period (eg, work schedules, childcare, and transportation challenges) [11-14], challenges that have only increased during the COVID-19 pandemic [15]. In addition, remotely delivered lifestyle interventions may be more cost-effective to deliver, especially when accounting for participant costs [16]. Establishing noninferiority of remote versus in-person postpartum weight loss intervention models would advance the science by identifying a potentially more convenient and less costly model of care delivery.

Facebook may be an effective platform for remotely delivering evidence-based weight loss programming to postpartum women. Currently, 70% of US adults aged 18-29 years and 77% of adults aged 30-49 years use Facebook [17], with higher rates of use among mothers (87%) and women aged 18-39 years (84%) [18]. Many mothers turn to Facebook for support and information about parenting issues [19,20], and 80% of parents who use Facebook engage on the platform daily [18]. Using this popular commercial social media platform for intervention delivery allows us to leverage women's daily routines to engage them in behavior change.

Lifestyle interventions that deliver at least some content via Facebook are efficacious for adults generally [21], and pilot studies conducted by our team and others have demonstrated feasibility and acceptability of leveraging Facebook for lifestyle intervention delivery among postpartum women specifically [22-25]. However, our approach is to deliver all the didactic intervention content via Facebook, whereas others have leveraged Facebook along with other treatment modalities (eg, telephone or in-person counseling sessions, text messaging, or an in-person orientation meeting) [23-25]. We previously developed a postpartum weight loss intervention [22] by adapting the Diabetes Prevention Program (DPP) lifestyle intervention [26] to address the needs of postpartum women and for delivery by a trained weight loss counselor via a private Facebook group [27]. In our earlier work, we conducted a 1-arm pilot study of our intervention with 19 postpartum women with overweight or obesity (ie, BMI ≥25 kg/m^2 but <45 kg/m^2) [22]. We were able to retain participants (95% retention) and keep them engaged over the 12-week intervention period [22]. The majority of participants said they would be likely or very likely to participate again if they had another baby, more than 80% would recommend the program to a postpartum friend, and participants lost an average of 4.8% of their baseline weight [22]. Although the results of this 1-arm pilot study are promising, this study did not provide information on the feasibility of recruitment of women able and willing to be randomized to either the Facebook or in-person intervention nor information about sustained participation in the Facebook-delivered intervention beyond 12 weeks. We conducted a randomized pilot trial to answer these feasibility questions (ie, feasibility of recruitment under conditions of randomization, sustained participation through the entire 6-month intervention, and retention at 6- and 12-month assessments) before conducting a large-scale trial to evaluate whether delivery via Facebook groups is noninferior to delivery via in-person group meetings.

Objective

This study aimed to conduct a randomized feasibility pilot trial of a 6-month postpartum weight loss intervention delivered via Facebook versus in-person groups with postpartum women with overweight or obesity. We examined the feasibility of recruitment, sustained participation, contamination, retention, and assessment procedures in both conditions. We also described intervention acceptability. We described percent weight loss at 6 and 12 months in both treatment groups as an exploratory outcome.

Methods

Study Design

We conducted a randomized feasibility pilot trial to compare the delivery of a 6-month postpartum weight loss intervention via Facebook versus in-person groups among women with overweight or obesity. The design of this trial has been described in detail elsewhere [28].

Ethics Approval

The University of Connecticut Institutional Review Board approved this study (protocol #H17-206). The trial was registered on clinicaltrials.gov (NCT03700736).

Recruitment and Eligibility

Participants were recruited in 2 waves starting in August 2018 through October 2019. We recruited women from the Hartford, Connecticut area community by posting recruitment messages on Facebook, Instagram, Twitter, ResearchMatch [29], Craigslist, and University of Connecticut and UConn Health employee email digests, and by posting study flyers in the community. Additional details on recruitment are described elsewhere [28]. Research staff conducted eligibility screenings of interested individuals via phone.

Inclusion criteria were being aged ≥18 years; being at least 8 weeks but <12 months post partum at time of enrollment; having a BMI ≥25 kg/m^2 per measured height and weight at baseline; either owning a scale or being willing to be provided one if needed; being comfortable reading and writing in English; owning an Android or iPhone smartphone; being an active Facebook user defined as accessing Facebook daily and posting or commenting at least weekly over the past 4 weeks; having clearance from their primary care provider or obstetrician or gynecologist; being willing to participate in either treatment condition (Facebook or in-person); being available to attend in-person meetings over the 6-month study period in Hartford, Connecticut; taking <45 minutes to travel to intervention meetings; and being willing and able to provide informed consent.

Women were excluded if they met any of the following criteria: currently pregnant; plan to become pregnant during the study period; current participation in an in-person or web-based clinical weight loss program; diagnosed with type 1 or type 2 diabetes as self-reported or reported by their health care provider; medical conditions or medications affecting weight; incapable of walking a quarter mile unaided without stopping; pain that prevents engagement in exercise; previous bariatric surgery; planned surgery during the study period; plans to move out of the area during the study period; high depressive symptoms or suicidal ideation (a score of ≥12 or positive on question #10 on the Edinburgh Postnatal Depression Scale [EPDS] [30]); positive screen for binge eating disorder [31]; failure to complete any baseline procedures (eg, baseline survey, orientation webinar, or prerandomization survey); or University of Connecticut student or employee supervised or taught by study investigators.

Eligible participants were all biologically female owing to the inclusion criteria of having given birth; we did not ask participants their gender identity. Although not all persons who become pregnant identify as women [32] as recruitment materials included the phase, “we are recruiting women who had a baby in the past year,” it is likely that all participants identified as women, and we refer to participants in this study as “women” or “mothers.”

Assessments

Participants completed assessments at baseline, 6 months (postintervention), and 12 months, and filled out brief weekly surveys during the intervention period, as described in detail elsewhere [28]. Participants were provided gift cards to thank them for completing study assessments at baseline (US $20), 6 months (US $40), and 12 months (US $40). For wave 1, the intervention occurred from February to August 2019, with follow-up assessments in August 2019 and February 2020. For wave 2, the intervention occurred from October 2019 to April 2020, with follow-up assessments in April 2020 and October 2020.

At baseline, participants completed an in-person study visit that included providing informed consent, height and weight measurement, and screenings for elevated depressive symptoms and binge eating disorder. Study staff also provided instructions for downloading and using the MyFitnessPal app and instructions for using the battery settings to report Facebook app use (iPhone users) or a free app to track time on Facebook (Android users). Following this visit, participants completed a 30-minute web-based survey that included demographic and clinical characteristics (including prepregnancy weight for calculation of postpartum weight retention at baseline) and other baseline measures. Next, research staff contacted participants’ primary care provider or obstetrician or gynecologist for medical clearance. After completing their baseline visit and survey, participants completed a 60-minute webinar with other participants to orient them to the scientific process, review study procedures, and discuss the barriers and advantages of each study condition [33]. Following the orientation webinar but before randomization, participants completed a 5-minute web-based survey composed of a randomization agreement, report of app-tracked time on Facebook over the past 7 days, and their Facebook use habits [34]. Weekly during the intervention period, participants in both treatment conditions reported their weight, past 7-day app-tracked time on Facebook, and Facebook use habits [34] via a brief, 5-minute web-based survey.

At the end of the 6-month intervention, participants attended a focus group with other members of their weight loss group to provide qualitative feedback on their experiences in the study. The focus groups started out by asking general questions about participants' experiences in the intervention (eg, "Overall, what do you think of this program?", "What about this program did you find most helpful?", and "How could we improve this weight loss program?"), transitioned to asking questions specific to each treatment modality (eg, "What influenced whether you commented on a post or comment?" in the Facebook condition and "How difficult was it for you to attend the sessions?" in the in-person condition), and finally prompted for any additional feedback (ie, "Do you have any other feedback about this program?"). Participants also completed a 30- to 45-minute web-based survey that included questions about contamination, acceptability, depressive symptoms (EPDS [30]), quality of life (PROMIS-Preference [PROPr] [35,36]), Facebook use habits including time spent on Facebook [17,34], and incident pregnancies. Research staff measured participants' weight at the focus group visit. Participants who could not attend the focus group completed an individual interview and weight measurement at an individual visit. For wave 2, the 6-month focus groups were conducted via video conferencing software owing to the COVID-19 pandemic, and participants self-reported their current weight on the survey.

At 12 months, participants in wave 1 completed an in-person visit to measure weight and completed a 30-minute web-based survey that included measures of depressive symptoms (EPDS [30]), quality of life (PROPr [35,36]), Facebook use habits including time spent on Facebook [17,34], and incident pregnancies. Owing to the COVID-19 pandemic, participants in wave 2 did not attend in-person follow-up visits; follow-up weights were self-reported in the follow-up surveys.

Randomization

Eligible participants who completed all screening and baseline procedures were randomized 1:1 to the Facebook and in-person conditions in randomly permuted blocks of size 4 and 6. Randomization was stratified by months post partum at enrollment (8 weeks to <6 months vs 6-12 months) and type of smartphone (iPhone vs Android). We stratified randomization by months post partum because weight change varies across the postpartum period in the absence of formal intervention [37,38]. We stratified randomization by smartphone type to balance any differences related to methods for measuring time spent on Facebook, as the procedures for collecting these data differed by phone operating system.

Treatment Conditions

Participants in both treatment conditions received a 6-month weight loss intervention based on the DPP lifestyle intervention [26]. As described elsewhere [28], we adapted the intervention content to meet the needs and challenges of the postpartum period [11,39-42]. In the materials for both the in-person and Facebook-delivered interventions (eg, participant handouts and Facebook posts), we included stock images of women with larger bodies with a variety of skin tones, racial or ethnic phenotypes, and family configurations. Weight loss counselors had backgrounds in nutrition and dietetics and completed the National DPP training and training by a licensed clinical psychologist with extensive experience using the DPP in our specific intervention protocols [22,43,44]. The weight loss counselor for wave 1 identified as non-Hispanic White, and the weight loss counselor for wave 2 identified as Hispanic. The intervention goals were 5% to 10% weight loss and increasing physical activity to 150 minutes per week of moderate intensity physical activity. Calorie and physical activity goals were set to facilitate weekly weight loss of 1 to 2 pounds. For women who reported breastfeeding at baseline, initial calorie goals accounted for lactation [45], and calorie goals were adjusted during the intervention, as participants reported changes in breastfeeding. Participants were encouraged to use the free MyFitnessPal app to track their diet, exercise, and weight, and weight loss counselors emailed or messaged participants' feedback on diet and activity records weekly or every 2 weeks (corresponding to the frequency of meetings in the in-person condition). Participants were withdrawn from the intervention if they reported becoming pregnant to the weight loss counselor or study staff. The 2 treatment conditions received the same intervention content; the difference between conditions was the delivery modality: in-person groups versus Facebook groups.

In the in-person condition, the weight loss counselor facilitated 90-minute group discussions, which were held weekly for the first 15 weeks and every other week during weeks 16-25, for a total of 20 meetings. The intervention materials were provided via paper handouts. Participants were reimbursed up to US $5 for parking or bus fare for each intervention meeting attended. Owing to the COVID-19 pandemic, the last 2 meetings of wave 2 were conducted via synchronous videoconferencing software.

In the Facebook condition, the weight loss counselor facilitated discussion about weekly topics via posts and comments in a private ("secret") Facebook group [46]. The counselor posted 2 posts per day during weeks 1-15 and 1 post per day during weeks 16-25, corresponding to the intensity of contact in the in-person condition. We used the Facebook post scheduling tool to schedule daily intervention posts from the weight loss counselor's account. We developed posts covering the intervention content of each module of the DPP lifestyle intervention based on our previous work with postpartum women [22] and adults generally [43,44,47]. Posts provided information and resources related to the topic of the week or asked participants to share their thoughts, experiences, or challenges related to the topic of the week; set goals (Mondays); report their progress toward these goals (Sundays); or report their weekly weight change (Fridays). Additional logistic details about the Facebook group and sample intervention posts are described elsewhere [28].

Participation in both interventions was monitored by the research team. The weight loss counselor recorded attendance at in-person intervention meetings. Research staff reviewed the Facebook group and recorded the date of each participant's latest post or reply (each Monday during weeks 1-15 and every other Monday during weeks 16-25, corresponding to the frequency of in-person intervention meetings). The weight loss counselor emailed participants who did not participate (ie, did not attend in-person meetings or did not post or reply in the Facebook group) in a given week or 2-week period to encourage them to participate during the following week. After 2 consecutive weeks of no participation, the weight loss counselor called the participant, and after 3 consecutive weeks, the research coordinator called the participant. After 4 consecutive weeks without participation, the weight loss counselor sent a final email encouraging participation.

Measures

Primary Outcomes: Feasibility

The feasibility outcomes were recruitment, retention, sustained participation, contamination, and feasibility of the assessment procedures. We also report participant feedback regarding the acceptability of the interventions.

Recruitment

We tracked participants through eligibility screening and study procedures and calculated recruitment rates from the number of individuals contacted, screened, consented, and randomized, overall and by recruitment source. We recorded the reasons for ineligibility and nonparticipation, including unwillingness to be randomized to either the Facebook or in-person condition.

Retention

We calculated retention as the proportion of participants who completed the 6- and 12-month follow-up assessments (ie, provided weight or completed the follow-up survey) in each condition.

Sustained Participation

We assessed sustained participation in the intervention (ie, treatment retention). For the in-person condition, the weight loss counselor recorded attendance at each intervention meeting, and sustained participation was calculated as the last intervention session attended. In the Facebook condition, treatment modules were spread over 1 week (weeks 1-15) or 2 weeks (weeks 16-25) to correspond to the frequency of intervention meetings in the in-person condition. Thus, we assessed participation in each treatment module. We captured engagement data from Facebook using Grytics tools (Grytics, Inc) and then manually abstracted identifiers (eg, participants' Facebook usernames), reactions to posts and comments (including who reacted and what the reaction was), and poll responses (who voted for each option). A second member of the research team reabstracted a random 10% sample of threads (ie, a post plus any associated replies; 99.7% agreement across abstracted data points) to confirm the accuracy of abstraction. We calculated sustained participation as the latest treatment module participated in based on the latest post, reply (comments on posts and comments in replies to comments), poll vote (based on the date of the post that included the poll), or reaction (based on the date of the post or reply reacted to) in the Facebook group. We secondarily calculated whether participants participated in the last (20th) intervention module, and overall participation as the number of intervention modules participated in, and whether they participated in 0 intervention modules, ≥16 (ie, ≥80%), or 20 (ie, 100%). We additionally calculated the total number of posts, replies, polls voted in, and reactions by participants in the Facebook condition.

Contamination

To assess contamination, the 6-month follow-up survey included questions about participation in other weight loss programs (web-based or in person); whether participants sought weight loss support on Facebook or other web-based social networks [48]; and if so, to what extent and reasons they sought this support. One participant in the Facebook group reported during week 13 of the intervention that she had started a 21-week Beachbody program. Although she did not report this in her 6-month survey, we counted this as concurrent use of another weight loss program.

Feasibility of Assessment Procedures

We developed data collection and participant tracking systems and procedures. We assessed the degree of missingness of measures to be included in a large-scale trial to assess intervention efficacy and cost-effectiveness. We created a tracking system in REDCap (Research Electronic Data Capture; Vanderbilt University [49]) for research staff to enter time spent on specific tasks (eg, leading in-person intervention meetings, counseling via the Facebook group, copying participant handouts) [50] that would be needed to implement each intervention in practice (ie, outside the research context) using methods developed by others [51-53]. At baseline, 6 months, and 12 months, the participants completed a quality-of-life measure (PROPr) [35,36].

Acceptability of the Interventions

At 6 months, participants also answered questions regarding intervention acceptability [22]. Participants were asked, "If you had another baby, how likely would you be to participate in this weight loss program again?" and "If you had a friend who recently had a baby, how likely would you be to recommend this program to her?" (response options: "very unlikely," "unlikely," "neutral," "likely," and "very likely"; dichotomized as likely or very likely vs not). Participants were asked how convenient it was for them to log into the private Facebook group daily (Facebook condition) or attend 90-minute group meetings each week (in-person condition; response options: "very convenient," "convenient," "neither convenient nor inconvenient," "inconvenient," or "very inconvenient"; dichotomized as convenient or very convenient vs not). Participants in both groups were asked whether they would find attending weekly in-person group meetings or interacting in a private Facebook group daily more convenient (response options: "Facebook much more convenient," "Facebook more convenient," "Facebook and in-person groups equally convenient," "in-person groups more convenient," and "in-person groups much more convenient"; dichotomized as Facebook much more or more convenient vs not). To help us understand factors influencing intervention participation, after answering acceptability questions, participants were asked: "Thinking about the times when you logged in but did not post or reply to any posts why did you choose not to?" (Facebook condition) or "Thinking about the times when you didn't come to the in-person group meetings, what was the reason?" (in-person condition). Participants in each condition were provided with a list of possible reasons (see tables for response options) and were asked to select all the answers that applied to them.

Exploratory Outcome: Weight Change

At baseline, 6 months, and 12 months, the staff measured weight twice (Tanita C-110 scale), and we calculated the average of the 2 measurements. In wave 1, participants who were unwilling to attend an in-person follow-up visit were asked to self-report their current weight (Facebook: 0/14, 0% at 6 months and 4/14, 29% at 12 months; in-person: 3/15, 20% at 6 months and 5/15, 33% at 12 months). As both follow-up time points for wave 2 occurred during the COVID-19 pandemic (April and October 2020), we pivoted to remote assessments, and participants self-reported their weight at 6 months and 12 months. Thus, all follow-up weights in wave 2 were self-reported. We calculated absolute (lbs) and percent weight change from baseline to 6 months and baseline to 12 months and defined clinically significant weight loss as ≥5% [54,55]. For women who were pregnant at follow-up, we used self-reported prepregnancy weight to calculate weight change. We secondarily calculated weight change assuming no weight change for those missing follow-up weights (baseline observation carried forward).

Power Calculation

The purpose of this pilot trial was to examine feasibility and to identify modifications required before conducting a large randomized controlled trial to assess efficacy for weight loss. As recommended [56,57], we based the sample size on necessities for examining feasibility. We decided a priori that retention of ≥80% would indicate feasibility and that a retention rate in either condition <60% would indicate a lack of feasibility. With the target sample of 72 participants (36 per condition), the lower limit of the 95% CI for the observed retention rate in either treatment condition should not be <60%.

Statistical Analysis

We used REDCap [49] for participant tracking and participant surveys. Data management and analyses were conducted using SAS (version 9.4; SAS Institute, Inc). We described the feasibility outcomes and exploratory outcome of weight loss in both conditions. We compared retention rates with the a priori benchmark for feasibility. After transcribing the focus group and interview recordings, we conducted a thematic analysis of responses [58]. This analysis focused on participant feedback about their assigned intervention modality that might impact acceptability ratings. Specifically, 3 members of the research team reviewed the focus group and interview transcripts as well as notes from focus group facilitators (familiarization) to create an initial list of feedback themes (generation of initial codes), reread transcripts to identify additional passages (searching for themes), and then consolidated feedback into themes (reviewing and defining themes) [58]. The first author then summarized relevant participant feedback identified by the review team (producing the report) [58].

Results

Study Sample

We screened 338 women, of whom 78 (23.1%) were eligible at screening and started baseline assessment procedures (Figure 1). We randomized 62 postpartum women. Two participants were withdrawn owing to pregnancy (1 per condition), and 8 dropped out of treatment (2 in the Facebook condition and 6 in the in-person condition; Figure 1). Overall retention was 92% (57/62) at 6 months and 94% (58/62) at 12 months.

Randomized participants (N=62) were on average aged 32.8 (SD 4.0) years and were a median of 6.1 (IQR 3.1-8.3) months post partum at enrollment (Table 1). In total, 60% (37/62) of participants had obesity at baseline, and median BMI was 31.7 (IQR 28.2-37.4) kg/m^2. Average postpartum weight retention was 13.9 (SD 15.4) pounds. Most (43/62, 69%) of the participants were breastfeeding and 63% (38/62) had ≥2 children. Three-quarters (46/62, 74%) of participants were non-Hispanic White, 85% (53/62) had at least a bachelor’s degree, and 73% (40/62) were employed full-time. Additional characteristics of participants are provided in Table 1.

Figure 1. Participant recruitment and retention. Individuals excluded at eligibility screening and baseline assessment could be excluded for multiple reasons. Only the most common reasons for ineligibility are shown in the figure.

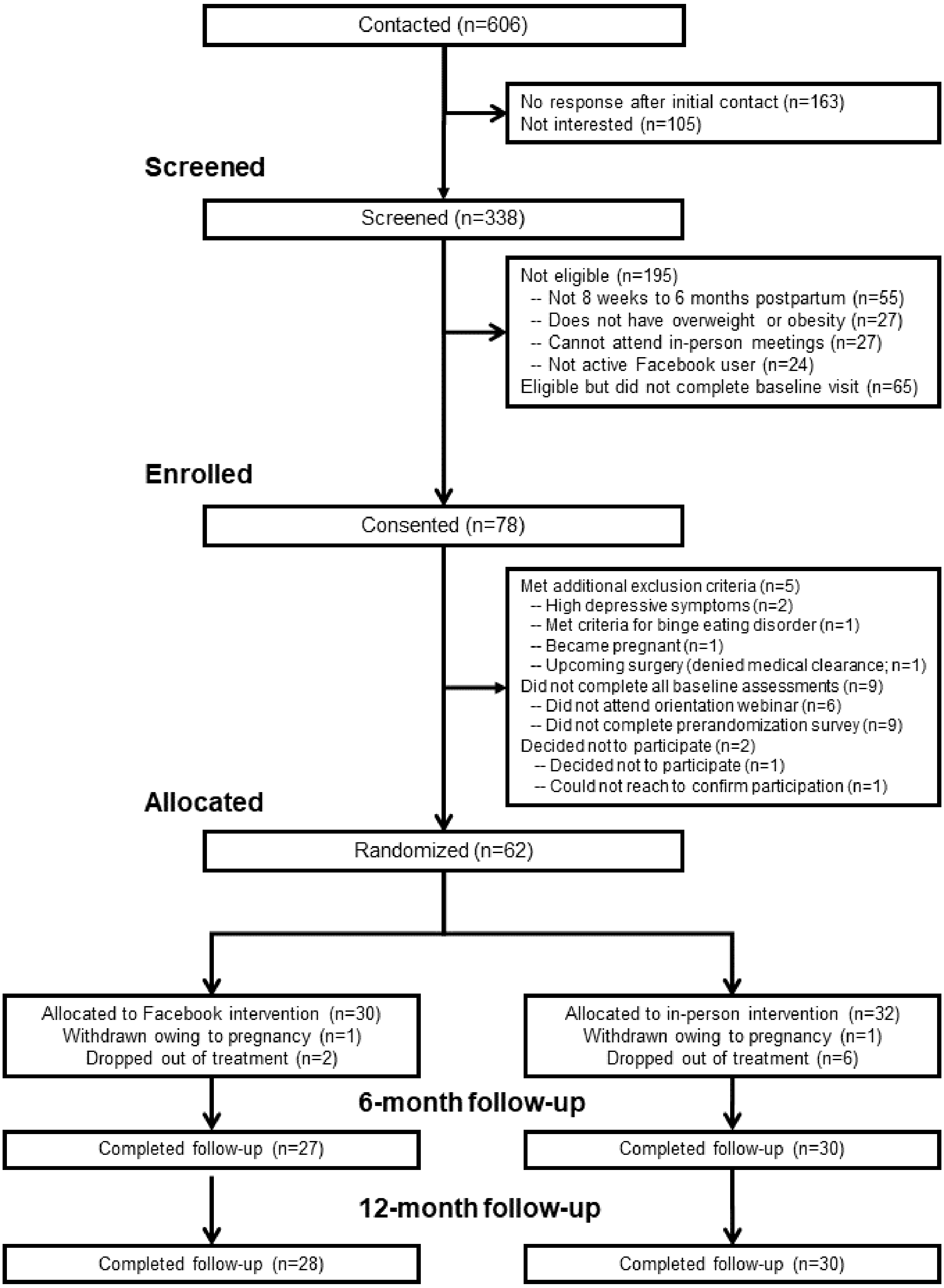


Table 1. Characteristics of postpartum women with overweight or obesity at study enrollment, overall, and by treatment condition.

	All randomized participants (n=62)	Facebook condition (n=30)	In-person condition (n=32)
Smartphone type[a], n (%)			
iPhone	45 (73)	21 (70)	24 (75)
Android	17 (27)	9 (30)	8 (25)
Months post partum[a], median (IQR)	6.1 (3.1-8.3)	5.8 (3.1-8.3)	6.1 (2.9-8.2)
≥8 weeks but <6 months, n (%)	30 (48)	15 (50)	15 (47)
≥6 months but <12 months, n (%)	32 (52)	15 (50)	17 (53)
Singleton gestation, n (%)	59 (95)	28 (93)	31 (97)
Breastfeeding, n (%)	43 (69)	23 (77)	20 (63)
>2 children in her household[b], n (%)	38 (63)	21 (70)	17 (57)
BMI (kg/m^2), median (IQR)	31.7 (28.2-37.4)	32.7 (28.0-36.6)	31.4 (28.3-37.6)
Overweight, n (%)	25 (40)	12 (40)	13 (41)
Obesity, n (%)	37 (60)	18 (60)	19 (59)
Postpartum weight retention (lbs), mean (SD)	13.9 (15.4)	14.6 (13.9)	13.1 (16.9)
Age (years), mean (SD)	32.8 (4.0)	33.3 (3.5)	32.3 (4.4)
Race and ethnicity, n (%)			
Non-Hispanic White	46 (74)	22 (73)	24 (75)
Non-Hispanic Black	3 (5)	1 (3)	2 (6)
Hispanic or Latina	9 (15)	5 (17)	4 (13)
Non-Hispanic Asian	3 (5)	1 (3)	2 (6)
Non-Hispanic multiracial	1 (2)	1 (3)	0 (0)
Education, n (%)			
Less than bachelor’s degree	9 (15)	1 (3)	8 (25)
Bachelor’s degree or graduate courses	20 (32)	13 (43)	7 (22)
Graduate degree	33 (53)	16 (53)	17 (53)
Marital status, n (%)			
Married	55 (89)	28 (93)	27 (84)
Living with partner	5 (8)	2 (7)	3 (9)
Single	2 (3)	0 (0)	2 (6)
Employment status[b], n (%)			
Employed full-time	40 (73)	19 (73)	21 (72)
Employed part-time	8 (15)	4 (15)	4 (14)
Stay-at-home mom (not employed)	7 (13)	3 (12)	4 (14)
Hard to pay for basics, n (%)			
Not at all hard	37 (60)	16 (53)	21 (66)
Somewhat hard	24 (39)	13 (43)	11 (34)
Very hard	1 (2)	1 (3)	0 (0)

[a]Randomization was stratified by smartphone type and months post partum.

[b]n=7 participants missing information on employment (n=4 in the Facebook condition and 3 in the in-person condition), and n=2 participants missing information on number of children in her household (both in the in-person condition).

Feasibility of Recruitment

Among individuals who were not interested in the study and therefore not screened for eligibility, 68.6% (72/105) explicitly reported a lack of interest or barriers related to the in-person condition as the reason (Table 2). In contrast, only 2.9% (3/105) were not interested in participating because they were not interested in the Facebook condition.

Among individuals determined to be ineligible at screening, 33.3% (65/195) were ineligible owing to reasons related to one or both treatment modalities, 18.5% (36/195) owing to reasons related to the in-person condition (ie, not available on the day or time of meetings, >45 minutes of travel to the intervention location, or plans to move out of the area in the next 12 months), 2.6% (5/195) owing to unwillingness to be randomized to either condition (all 5 preferred Facebook), and 12.3% (24/195) owing to their Facebook use habits (eg, no Facebook account, does not browse their feed at least daily, or does not post or reply at least weekly).

Table 2. Reasons[a] potential participants were not interested in participating in the study (N=105).

	Values, n (%)
Any reason related to in-person condition	72 (69)
Not interested in in-person condition	51 (49)
Day or time of meetings does not work	29 (28)
Intervention location too far to travel	19 (18)
Any reason related to Facebook condition	3 (3)
Not interested in Facebook condition	3 (3)
Reasons not explicitly related to either treatment modality	24 (23)
Does not have time to participate	24 (23)
Intervention will not start soon enough	0 (0)

[a]Participants could provide multiple reasons for not being interested in participating in the study.

Retention

Overall, retention was 92% (57/62) at 6 months and 94% (58/62) at 12 months. Retention in the Facebook condition was 90% (27/30) at 6 months and 93% (28/30) at 12 months. Retention in the in-person condition was 94% (30/32) at 6 months and 94% (30/32) at 12 months. Retention in both conditions at both time points exceeded our a priori benchmarks of 80%, and the lower limits for 95% CIs were ≥79% for all conditions at all time points, which was substantially higher than the a priori benchmark of 60%.

Because of the disruptions to study procedures caused by the COVID-19 pandemic, we also explored retention by wave. In wave 1, retention in the Facebook condition was 79% (11/14) at 6 months and 86% (12/14) at 12 months and 93% (14/16) at 6 months and 93% (14/15) at 12 months in the in-person condition. In wave 2, retention in the Facebook condition was 100% (16/16) at 6 months and 100% (16/16) at 12 months and 94% (16/17) at 6 months and 94% (16/17) at 12 months in the in-person condition.

Sustained Participation and Engagement

Over the 6-month intervention, participants in the Facebook condition posted 159 original posts and 3318 replies and contributed 614 poll votes and 1996 reactions. Participants posted a median of 3 (IQR 1-7; range 0-28) original posts and a median of 88 (IQR 50-142; range 11-309) replies and contributed a median of 18.5 (IQR 11-29; range 3-59) poll votes and a median of 47 (IQR 24-94; range 7-208) reactions.

In total, 70% (21/30) of the participants in the Facebook condition posted, replied, voted in a poll, or reacted during the last 2 weeks of the intervention (ie, participated in the last intervention module), compared with 31% (10/32) of participants in the in-person condition who attended the final intervention meeting (Table 3). The latest treatment module participated in (of 20 modules) was median 20 (IQR 19-20) for participants in the Facebook condition and median 18 (IQR 2-20) for participants in the in-person condition (Table 3). In a sensitivity analysis that used a stricter definition of participation for participants in the Facebook condition—posting and replying only—63% (19/30) of participants participated in the last intervention module, and the latest treatment module participated in was median 20 (IQR 17-20). Participants posted or replied in a median of 18 (IQR 13-20) treatment modules; 67% (20/30) of participants engaged in ≥16 modules, and 43% (13/30) engaged in all 20 intervention modules.

In wave 2, the last 2 intervention meetings for the in-person condition were held via videoconference calls owing to the COVID-19 pandemic. In wave 1, a total of 33% (5/15) and 33% (5/15) of participants, respectively, attended the 19th and 20th intervention meeting. In wave 2, a total of 35% (6/17) and 29% (5/17) of participants, respectively, attended these last 2 meetings. In the postintervention focus groups, some participants mentioned liking not needing to travel or arrange childcare, but others mentioned "Zoom fatigue," feeling less connected to other women over video versus in person, or having children in the background was distracting.

When participants in the Facebook condition were asked to select reasons they did not post or reply when they had logged into Facebook, the most common responses selected were not having anything to add to the conversation (22/26, 85%), preferring to lurk rather than actively engaging (13/26, 50%), and not wanting to be the only person posting (10/26, 38%; Table 4). In-person participants were similarly asked why they did not attend their in-person meetings. The most common response, endorsed by 93% (25/27) of participants, was the need to attend to other responsibilities that were more important (Table 4). Other common responses included motivation for weight loss declined (4/27, 15%) and forgetting about the meeting (4/27, 15%; Table 4).

Table 3. Sustained participation[a] in the intervention, by treatment condition.

	Facebook condition (n=30)	In-person condition (n=32)
Latest intervention module participated in, median (IQR)	20 (19-20)	18 (2-20)
Participated in the last intervention module, n (%)	21 (70)	10 (31)
Number of intervention modules participated in, median (IQR)	19 (17-20)	10 (2-14)
Participated in no intervention modules, n (%)	0 (0)	6 (19)
Participated in ≥16 (ie, ≥80%) intervention modules, n (%)	24 (80)	7 (22)
Participated in all 20 intervention modules, n (%)	13 (43)	0 (0)

[a]Participation in a treatment module was attending the intervention meeting for participants in the in-person condition and posting, replying, voting in a poll, or reacting to a post or comment in the Facebook condition.

Table 4. Reasons participants did not post in the Facebook group or attend in-person intervention meetings, by condition.

	Value, n (%)
Reasons participants in the Facebook condition did not post or reply when they logged into the Facebook group (N=26)	
I did not have anything to add to the conversation	22 (85)
I generally prefer to “lurk”—meaning I like to read posts but not say anything	13 (50)
It seemed like nobody in the group was posting so I did not want to be the only one	10 (38)
The topic was not relevant to me	6 (23)
I did not feel comfortable posting my ideas	4 (15)
The topic was not interesting to me	4 (15)
I feared how people would respond to what I would say (eg, I might get judged, ignored, or not supported)	3 (12)
I did not feel like I was a part of the group	1 (4)
I did not understand how to post	0 (0)
I was concerned about privacy in the group	0 (0)
Reasons participants in the in-person condition did not attend the group intervention meetings (N=27)	
I had to attend to other responsibilities that were more important	25 (93)
My motivation to focus on weight loss declined	4 (15)
I forgot we had group that day	4 (15)
I had transportation issues	3 (11)
It seemed like a lot of people weren’t coming so this reduced my motivation to be a part of the group	3 (11)
The topics were not interesting or helpful to me	2 (7)
I did not feel like I was a part of the group	2 (7)
The topic was not relevant to me	1 (4)
I did not feel comfortable in the group	1 (4)
I was concerned about privacy in the group	0 (0)
I feared how people would respond to me (eg, I might get judged, ignored, or not supported)	0 (0)
N/A[a] (I attended all groups)	0 (0)

[a]N/A: not applicable.

Contamination

In total, 15% (4/26) of participants in the Facebook condition and 4% (1/28) of participants in the in-person condition reported that they had used other in-person or web-based weight loss programs during the intervention period. However, when asked for details about these other weight loss programs, only 2 participants reported a structured program (Weight Watchers, a 21-week Beachbody program). The other 3 participants reported activities that would support their weight loss efforts (eg, saw a nutritionist, used a meal plan from dietitian she has been working with for 4 years, and personal training challenge) but do not represent a structured weight loss program. In total, 35% (9/26) of participants in the Facebook condition and 57% (16/28) of participants in the in-person condition sought weight loss information or support on social media during the intervention period. All reports of contamination were to external resources; no participants reported access to the other study intervention.

Feasibility of Assessment Procedures

Regarding patient-reported data needed to evaluate cost-effectiveness, we were able to obtain measured baseline weights on 100% (62/62) of participants in both conditions and measured or self-reported weights for 90% (27/30) and 93% (28/30) of participants in the Facebook condition at 6 and 12 months, respectively, and 94% (30/32) and 94% (30/32) of participants in the in-person condition, respectively. At baseline, 1 participant in each condition (2/62, 3%) did not complete the quality-of-life measure (PROPr) owing to a license agreement issue that delayed inclusion of another quality-of-life measure in the baseline survey. An additional 4 participants missed one of the PROPr items; at baseline, 87% (26/30) and 94% (30/32) of participants in the Facebook and in-person conditions, respectively, completed the quality-of-life measure in full. At 6 and 12 months, respectively, 87% (26/30) and 93% (28/32) of participants in the Facebook condition and 84% (27/30) and 91% (29/32) of participants in the in-person condition completed this measure in full.

Acceptability

Half (13/26, 50%) of participants in the Facebook condition would be likely or very likely to participate again and 54% (14/26) would be likely or very likely to recommend this program to a friend (Table 5). In the in-person condition, 58% (15/26) of participants would participate again and 70% (19/27) would recommend this program to a friend (Table 5). In postintervention focus groups or interviews, participants in the Facebook condition shared that they appreciated the flexibility of being able to engage with the group anytime and from anywhere. However, participants also noted that a downside of this flexibility was a lack of accountability—that it was easy to put off responding to posts or setting goals because there was no set schedule for participation. Participants also noted that they felt it was hard to get to know other mothers and build a sense of community. In contrast, participants in the in-person condition shared that they really got to know other women in their group and felt a strong sense of community. They also felt that meeting in person—and being weighed at intervention meetings—kept them accountable, and having a set meeting time each week helped them prioritize their health. However, participants noted barriers including time to travel to meetings, variable parking availability, and the need to arrange childcare for older children. A few participants also mentioned feeling guilty leaving their children 1 evening per week.

Table 5. Intervention acceptability by treatment condition.

	Facebook condition (n=26), n (%)	In-person condition (n=27), n (%)
If you had another baby, how likely would you be to participate in this weight loss program again? (likely or very likely)[a]	13 (50)	15 (58)
If you had a friend who recently had a baby, how likely would you be to recommend this program to her? (likely or very likely)	14 (54)	19 (70)
How convenient was it for you to log into the private Facebook group each day/attend 90-minute group meetings each week? (convenient or very convenient)	25 (96)	2 (7)
Thinking about attending weekly in-person group meetings versus interacting in a private Facebook group daily, which would you find more convenient? (Facebook more convenient or Facebook much more convenient)[a]	23 (88)	23 (88)

[a]n=1 participant in the in-person condition missing information for this question.

Almost all participants (25/26, 96%) in the Facebook condition reported that it was convenient or very convenient for them to log into the private Facebook group each day (Table 5). In contrast, only 7% (2/27) of participants in the in-person condition said it was convenient or very convenient to attend 90-minute group meetings each week. Most of the participants (23/26, 88%) in each condition agreed that interacting in a private Facebook group daily would be more convenient than attending a weekly in-person group (Table 5). In postintervention focus groups, several participants suggested a hybrid approach. A few participants from the Facebook condition suggested adding video meetings to increase accountability and sense of community, whereas others suggested occasional in-person meetings (eg, to start the group or once a month). Participants from the in-person condition suggested adding a Facebook group for connection and support between in-person meetings.

Weight Change (Exploratory Outcome)

At 6 months, participants in the Facebook condition had lost an average of 3.0% (SD 7.2%) of their baseline weight and participants in the in-person condition had lost an average of

5.4% (SD 6.8%; Table 6). Average percent weight loss at 12 months was 2.8% (SD 7.4%) in the Facebook condition and 4.8% (SD 7.6%) in the in-person condition (Table 6). In a sensitivity analysis using a baseline observation carried forward approach (ie, assuming no weight change) for participants missing follow-up weights (3 at 6 months and 2 at 12 months in the Facebook condition and 2 and 2 in the in-person condition), average percent weight loss at 6 months was 2.7% (SD 6.9%) in the Facebook condition and 5.0% (SD 6.7%) in the in-person condition, with 27% (8/30) and 50% (16/32), respectively, achieving ≥5% weight loss. At 12 months, average percent weight loss was 2.6% (SD 7.2%) in the Facebook condition and 4.5% (SD 7.5%) in the in-person condition, with 33% (10/30) and 47% (15/32), respectively, achieving 5% weight loss.

Table 6. Weight change at 6 and 12 months, by treatment condition.

	Facebook condition[a]	In-person condition[a]
Weight change from baseline (lbs), mean (SD)		
6 months	−4.8 (13.8)	−10.0 (13.0)
12 months	−5.1 (13.8)	−9.2 (15.3)
Weight change from baseline (%), mean (SD)		
6 months	−3.0 (7.2)	−5.4 (6.8)
12 months	−2.8 (7.4)	−4.8 (7.6)
Lost ≥5% of baseline weight, n (%)		
6 months	8 (30)	16 (53)
12 months	10 (36)	15 (50)

[a]At 6 months, weights were available for 27 participants in the Facebook condition and 30 participants in the in-person condition. At 12 months, weights were available for 28 participants in the Facebook condition and 30 participants in the in-person condition.

Discussion

Principal Findings

Feasibility trials provide an opportunity to pilot study procedures and measures in preparation for a large-scale efficacy trial. We assessed the feasibility of recruiting a sample of postpartum women willing and able to participate in a lifestyle intervention delivered either via Facebook or in-person groups. The in-person condition posed challenges to recruitment. Among individuals not interested in the study, 68.6% (72/105) were not interested in or could not attend in-person meetings, and 18.5% (36/195) of screened individuals were excluded because of reasons related to participating in the in-person condition. These findings are not unexpected, as the barriers to attending in-person treatment sessions documented in several previous studies [11-14,41,59-61] motivated the current line of research to develop a Facebook-delivered version of the intervention. The numbers of participants contacted, screened, eligible, and enrolled will inform the timeline for subsequent efficacy testing.

Overall, retention was 92% at 6 months and 94% at 12 months. Although based on the small numbers (approximately 15 participants per condition per wave), it appears that retention may have been higher in the Facebook condition in wave 2 when in-person assessments were not required. In wave 1, retention at 6 and 12 months was 79% and 86% versus 100% and 100%, respectively, in wave 2. For comparison, retention in the in-person condition was 93% and 93% in wave 1 and 94% and 94% in wave 2 at 6 and 12 months, respectively. It may be that women who were used to connecting with the weight loss counselor and their group remotely perceived attending an in-person visit as more burdensome. The only data collected at follow-up that required an in-person visit were weight. Providing participants digital scales in future trials would allow all follow-up assessments to be conducted remotely [62], which may increase completion of follow-up measures by reducing participant burden. Remote assessments would also allow for national recruitment, thus widening the participant pool.

Average weight losses of 3% in the Facebook condition and 5% in the in-person condition are promising, and it appears that women in the in-person condition lost more weight on average than those in the Facebook condition. Future research should explore the mechanisms through which delivering lifestyle interventions in person versus via digital platforms influences weight loss (eg, through increased accountability, stronger connections with the weight loss counselor or group members, and greater impact on participants' motivation to engage in behavior change). However, the weight loss in this trial should be interpreted cautiously for 2 reasons. First, participants in wave 2 self-reported their weight at 6 months and 12 months owing to the COVID-19 pandemic-related disruptions in in-person research assessments. Although weights self-reported as part of a digital lifestyle intervention tend to be accurate [63,64], weights measured with home scales that differ by a few pounds from the study scale can bias estimates of weight change, especially when baseline weights are measured by study staff. Second, the COVID-19 pandemic and particularly the early-pandemic school and childcare shutdowns had immeasurable impact on women's lives and their motivation and ability to make and sustain behavioral changes [65,66]. A study on the impact of the COVID-19 pandemic on current research participants' ability and desire to engage in research found that among those currently enrolled in a group-based behavioral intervention, 52% reported that the pandemic had impacted their ability to adhere to behavioral recommendations a little bit or moderately and 22% reported that the pandemic had impacted their behavior quite a bit or extremely [66]. Indeed, multiple participants in wave 2 reported at their 12-month follow-up assessments that they had regained a substantial amount of weight owing to disruptions and stress related to the pandemic, whereas others said that they lost additional weight owing to changes in their lifestyle (eg, sharp decline in eating out and more exercise). As only a single wave of participants completed the study before the pandemic and a single wave had their experience disrupted by the COVID-19 pandemic, we did not have a sufficiently large sample size to examine the impact of the pandemic on participant experiences or outcomes.

Whether participants stay engaged in treatment influences treatment receipt and, thus, efficacy. Digital health interventions have long been plagued by high dropout rates [67], and systematic reviews of postpartum weight loss interventions have highlighted attrition from treatment as a common challenge [7,8]. In our previous 1-arm pilot of the Facebook-delivered intervention, 63% of participants participated in the last week of the intervention [22]; however, that preliminary pilot study did not provide information on whether participants would stay engaged in treatment for the full 6-month intervention. This feasibility pilot trial examined sustained treatment through the full 6-month intervention period in both the Facebook-delivered and in-person versions of the intervention and found that 70% (21/30) of women in the Facebook condition engaged in the Facebook group during the last intervention module (last 2 weeks of the intervention) and engaged during a median of 19 (IQR 17-20) of the 20 intervention modules. In contrast, only 31% (10/32) of participants in the in-person condition attended their last intervention meeting. Women only attended a median of 10 (IQR 2-14) of the 20 intervention meetings, and 19% (6/32) did not attend a single intervention meeting. To be sure, participating in Facebook requires much less effort than attending a group visit. Indeed, when asked to select from a list of reasons participants did not come to the meetings, 93% (25/27) indicated that they had to attend to other responsibilities, including being sick, caring for an ill child, work, or caring for older children while their spouse worked, emphasizing the challenges of in-person meetings for mothers. Although we defined sustained participation in terms of the latest intervention module participated in (ie, in-person meeting attended or latest intervention module with visible engagement in the Facebook group), this definition would allow participants with large gaps in participation to pop back into the group at the end of treatment and be counted as having sustained participation. Thus, we recommend not only examining the time to last participation but also the number of treatment sessions or modules to provide a more comprehensive picture of treatment receipt.

In addition to examining sustained participation, we also described engagement in the Facebook-delivered intervention. Over the 6-month intervention, participants in the Facebook condition posted a median of 3 (IQR 1-7; range 0-28) original posts and a median of 88 (IQR 50-142; range 11-309) replies. Previous studies delivered a weight loss intervention based on the DPP lifestyle intervention entirely via a commercial social media platform for 12 weeks; thus, we additionally calculated engagement for the first 12 weeks of this study. Participants posted a median of 3 (IQR 1-5) original posts and a median of 65 (IQR 40-93) replies. Engagement over the first 12 weeks was higher in this study than in our previous 1-arm pilot study of a 12-week version of the Facebook-delivered intervention in which participants contributed a median of 2 (IQR 1-3) original posts and a median of 24 (IQR 15-31) replies [22]. We purposefully revised our intervention posts in response to engagement findings and participant feedback. It may be that the updated intervention posts were more effective at eliciting engagement from participants. Engagement in this study also appeared to be higher than engagement in other Facebook-delivered lifestyle interventions. In our previous research of a 12-week Facebook-delivered weight loss intervention with adults with obesity generally, participants shared a median of 37 (IQR 16-76) original posts or replies [47], less than the median of 68 (IQR 40-93) posts and replies shared during the first 12 weeks in this study (median 92.5, IQR 53-153 over 6 months). Women in this study also appeared to engage more than participants randomized to the comparison condition of a pilot study testing an app to track dietary lapses (median 0 posts, IQR 0-1; median 29.5 replies, IQR 17-61) [68]. The median number of replies over the first 12 weeks of the intervention posted by women in this study was also about twice the median number of replies of other Facebook-delivered weight loss interventions among adults [69,70] and higher than the average engagement in pilot trials of postpartum weight loss interventions that delivered some or all content via Facebook [24].

Despite higher engagement than other Facebook-delivered weight loss interventions, when asked to select reasons why they did not post or reply in the group, 38% (10/26) of participants endorsed "It seemed like nobody in the group was posting so I didn't want to be the only one." This feedback may be related to participants being hesitant to be the first one to respond to an intervention post, as we have heard from participants in previous studies [22]. This feedback may also be partially explained by when these questions were asked. The weight loss counselor posted in the Facebook group twice per day during weeks 1-15 of the intervention and then once a day in weeks 16-25; therefore, participants' reports of their experiences at the 6-month follow-up assessment may be biased toward their more recent experiences in the group compared with the activity in the group over the full 6-month intervention. Future studies could explore whether maintaining a posting schedule of twice daily results in greater participant engagement in the latter weeks of the intervention. Future studies could also experiment with ways to encourage participants to start conversation threads, so that the volume of conversation in the group is less dependent on the weight loss counselor's posts.

Participants also indicated that they did not post or reply in the group because they did not have anything to add to the conversation (22/26, 85%) or the topic was not relevant (6/26, 23%) or interesting (4/26, 15%) to them. In addition, 50% (13/26) of participants noted that they generally preferred lurking over visibly engaging. In one of our previous social media-delivered weight loss interventions, participants who had not used the social media platform before enrolling in the study engaged less in the intervention [43]. In our 1-arm pilot of the Facebook-delivered postpartum weight loss intervention, participants reported in postintervention focus groups that their typical social media habits influenced their level of engagement in the intervention [22]. Thus, we limited enrollment in this study to women who reported posting or replying on Facebook at least weekly. Future research should explore how to convert lurkers into posters as a strategy to boost engagement in the group. Future research could also explore the potential benefits participants experience from reading conversation threads without visibly engaging in them [71].

Another option to increase engagement in digital weight loss groups may be to increase the number of women receiving a postpartum weight loss intervention via a private Facebook group. Pagoto et al [69] recently conducted a proof-of-concept pilot study comparing engagement in a Facebook-delivered lifestyle intervention in which group membership was allowed to grow with engagement in a group in which membership was static. Although total engagement (original posts, replies, poll votes, and reactions) did not differ among the 40 participants initially randomized to the open enrollment group compared with the 40 participants randomized to the closed group, total engagement was higher among all 94 participants in the open enrollment group, and the total volume of engagement contributed by participants and weight loss counselors was associated with participants' weight loss [69]. As Facebook's algorithms prioritize groups with more activity [72], larger groups with more participant posts and replies may be prioritized in women's Facebook feeds, thus increasing the opportunity to engage and subsequently better treatment receipt.

Although engagement in the Facebook condition in this study was higher than that in our previous work with postpartum women or previous studies with adults with obesity, participants appeared to lose less weight. Average weight loss of 3% over 6 months among participants in the Facebook condition was lower than the average weight loss of 4.8% observed in our previous 1-arm pilot study of a 12-week version of the Facebook-delivered intervention [22]. Weight loss among participants in the Facebook condition in this study was similar to the average weight loss achieved in studies of adults generally over shorter periods (ie, 12 or 16 weeks) [47,69,70]. Differences in the samples, including the requirement of being willing to participate in either an in-person or digital intervention and enrollment of women earlier in the postpartum period (average 3.4 months post partum in the 1-arm pilot vs median 6.1 months in this study), may have also contributed to differences in weight loss. Although our weight loss findings should be interpreted with caution, these findings suggest that a deeper investigation into what types of engagement are associated with weight loss during the postpartum period is warranted. Not all utterances in a Facebook weight loss group are associated with weight loss [47], and additional research is needed on how best to engage participants in interactions that lead to successful behavior changes and subsequently weight loss [73].

Taken together, our findings related to the feasibility of recruitment under conditions of randomization to an in-person or Facebook condition and our findings related to participation in in-person intervention meetings indicate that an in-person weight loss program with numerous visits has limited feasibility for many postpartum persons. Further research is needed to develop and test efficacious postpartum weight loss interventions that work with postpartum persons' busy lives. Synchronous video meetings may be an option to foster group cohesion and accountability while overcoming the logistic challenges of in-person meetings. Telehealth or video visits offer the opportunity to connect individuals and groups face-to-face while retaining many of the benefits of in-person interactions. The national telehealth landscape has changed significantly since the initiation of this study. The COVID-19 pandemic has inspired a rapid uptake of telehealth in clinical settings [74,75], including obstetric care [76]. Indeed, in this study, the COVID-19 pandemic necessitated a shift of modality for the last 2 intervention meetings for wave 2 from in-person to videoconferencing. Attendance at these meetings was very similar to attendance at the analogous meetings in wave 1, and women liked not having to travel or arrange childcare. However, some participants mentioned "Zoom fatigue" [77], not surprisingly, as this feedback was provided in April 2020, approximately a month after the COVID-19–related shutdowns. As more and more activities resume in person, "Zoom fatigue" is likely to lessen. Other weight loss trials that transitioned from in-person to video meetings also found this modality acceptable, and participants lost weight [78,79]. Weight loss interventions based on the DPP lifestyle intervention have been successfully delivered via video meetings [80]. Women from a wider geographic range can be enrolled without travel constraints to attend in-person intervention meetings. Video meetings might also alleviate barriers to participation related to childcare, as women could participate in groups while their children sleep, engage in other activities at their home, or participate during lunch or another break during their workday. Another option may be a hybrid approach [81], such as an intervention delivered primarily remotely with a few in-person meetings, an approach that has been shown to be effective in low-income postpartum women [82]. Future research could explore the acceptability and efficacy of delivering a postpartum weight loss intervention via synchronous video group meetings, either in place of in-person meetings or in addition to an intervention delivered primarily digitally.

This feasibility trial also provided an opportunity to pilot and reflect on how contamination was measured. When we designed this study, we defined contamination in 2 ways: participation in other digital or in-person weight loss programs and seeking weight loss information or support on Facebook or other digital social networks. As 3 of the 5 participants who responded affirmatively to a question about concurrent participation in a structured weight loss program reported activities that did not meet our definition of a structured program, in future studies, we will revise the survey question and also have staff call participants to obtain additional details about professional assistance with weight loss outside the study intervention. As the study progressed, we realized that defining seeking weight loss information or support on social media as contamination did not match our study protocols, as we directed women in both conditions to a study Pinterest account where we had compiled online resources helpful for their weight loss journeys (eg, low-calorie recipes, workout videos, and MyFitnessPal tutorial videos) and encouraged women on Facebook to create a healthy Facebook feed by following Pages by public health organizations (eg, AHALiveHealthy and EatRightNutrition). In future studies, we will focus on contamination tracking of participation in structured weight loss programs or weight loss-specific Facebook groups.

Limitations

An additional limitation of this study is its limited racial, ethnic, and socioeconomic diversity. Our sample was more highly educated than US women giving birth overall (53/62, 85% with a bachelor's or higher education vs 34% nationally), more likely to be non-Hispanic White (46/62, 74% vs 51% nationally), and less likely to be unmarried (7/62, 11% vs 40% nationally) [83]. Many behavioral weight loss trials struggle to recruit racially or ethnically diverse samples [84,85]. In future studies, we will use strategies to diversify the participant pool. Targeted recruitment advertisements and strategic placement of such materials can facilitate the recruitment of an ethnically and economically diverse sample into weight loss trials [86]. Effective strategies to recruit low-income and racially or ethnically diverse postpartum women and parents of young children into behavioral trials include working with community partners (eg, Women, Infants, & Children Nutrition Program [WIC]), hiring culturally representative and sensitive research staff, and having multiple contacts with potential participants [87,88]. In future studies in this line of research, we will ask interested individuals to provide some demographic information (eg, Hispanic ethnicity, race, education, and participation in WIC or Supplemental Nutrition Assistance Program) earlier in the eligibility screening process so that we can monitor participant yield and characteristics from different recruitment approaches and refine our strategies to yield a more racially, ethnically, and socioeconomically diverse sample.

Conclusions

Delivering a lifestyle intervention to postpartum women via both in-person and Facebook groups was feasible and acceptable and resulted in weight loss. However, barriers to attending in-person meetings hampered recruitment efforts and attendance at in-person intervention meetings. Although women found the Facebook group convenient and stayed engaged in the group, weight loss appeared lower than that with in-person delivery. Research is needed to further develop care models for postpartum weight loss that balance accessibility with efficacy.

Acknowledgments

This project was supported by National Institutes of Health grant R34HL136979 (principal investigator: MEW). The authors would like to thank the research staff and the students who contributed to this project.

Conflicts of Interest

SLP has received funding from Weight Watchers, International. The other authors have no conflicts of interest to declare.

References

1. Endres LK, Straub H, McKinney C, Plunkett B, Minkovitz CS, Schetter CD, Community Child Health Network of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Postpartum weight retention risk factors and relationship to obesity at 1 year. Obstet Gynecol 2015 Jan;125(1):144-152 [FREE Full text] [doi: 10.1097/AOG.0000000000000565] [Medline: 25560116]
2. Gore SA, Brown DM, West DS. The role of postpartum weight retention in obesity among women: a review of the evidence. Ann Behav Med 2003 Oct;26(2):149-159. [doi: 10.1207/S15324796ABM2602_07] [Medline: 14534032]
3. Gunderson EP, Murtaugh MA, Lewis CE, Quesenberry CP, West DS, Sidney S. Excess gains in weight and waist circumference associated with childbearing: the Coronary Artery Risk Development in Young Adults Study (CARDIA). Int J Obes Relat Metab Disord 2004 Apr;28(4):525-535 [FREE Full text] [doi: 10.1038/sj.ijo.0802551] [Medline: 14770188]
4. Abrams B, Heggeseth B, Rehkopf D, Davis E. Parity and body mass index in US women: a prospective 25-year study. Obesity (Silver Spring) 2013 Aug;21(8):1514-1518 [FREE Full text] [doi: 10.1002/oby.20503] [Medline: 23630108]
5. Archuleta J, Chao SM. Maternal characteristics that impact postpartum weight retention: results from the 2016 Los Angeles Mommy and Baby (LAMB) follow-up study. Matern Child Health J 2021 Jan;25(1):151-161. [doi: 10.1007/s10995-020-03082-3] [Medline: 33188470]
6. Legro NR, Lehman EB, Kjerulff KH. Mode of first delivery and postpartum weight retention at 1 year. Obes Res Clin Pract 2020 May;14(3):241-248. [doi: 10.1016/j.orcp.2020.04.009] [Medline: 32456882]
7. Amorim Adegboye AR, Linne YM. Diet or exercise, or both, for weight reduction in women after childbirth. Cochrane Database Syst Rev 2013 Jul 23;2013(7):CD005627 [FREE Full text] [doi: 10.1002/14651858.CD005627.pub3] [Medline: 23881656]
8. Dodd JM, Deussen AR, O'Brien CM, Schoenaker DA, Poprzeczny A, Gordon A, et al. Targeting the postpartum period to promote weight loss: a systematic review and meta-analysis. Nutr Rev 2018 Aug 01;76(8):639-654. [doi: 10.1093/nutrit/nuy024] [Medline: 29889259]
9. Lim S, Liang X, Hill B, Teede H, Moran LJ, O'Reilly S. A systematic review and meta-analysis of intervention characteristics in postpartum weight management using the TIDieR framework: a summary of evidence to inform implementation. Obes Rev 2019 Jul;20(7):1045-1056. [doi: 10.1111/obr.12846] [Medline: 30942550]

10. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of eHealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
11. Carter-Edwards L, Østbye T, Bastian LA, Yarnall KS, Krause KM, Simmons TJ. Barriers to adopting a healthy lifestyle: insight from postpartum women. BMC Res Notes 2009 Aug 17;2:161 [FREE Full text] [doi: 10.1186/1756-0500-2-161] [Medline: 19686601]
12. Montgomery KS, Bushee TD, Phillips JD, Kirkpatrick T, Catledge C, Braveboy K, et al. Women's challenges with postpartum weight loss. Matern Child Health J 2011 Nov;15(8):1176-1184. [doi: 10.1007/s10995-010-0681-9] [Medline: 20844941]
13. Graham M, Uesugi K, Olson C. Barriers to weight-related health behaviours: a qualitative comparison of the socioecological conditions between pregnant and post-partum low-income women. Matern Child Nutr 2016 Apr;12(2):349-361 [FREE Full text] [doi: 10.1111/mcn.12135] [Medline: 25040706]
14. Ryan RA, Lappen H, Bihuniak JD. Barriers and facilitators to healthy eating and physical activity postpartum: a qualitative systematic review. J Acad Nutr Diet 2022 Mar;122(3):602-13.e2. [doi: 10.1016/j.jand.2021.11.015] [Medline: 34839027]
15. Schaeffer K. Working moms in the U.S. have faced challenges on multiple fronts during the pandemic. Pew Research Center. 2022 May 06. URL: https://www.pewresearch.org/fact-tank/2022/05/06/working-moms-in-the-u-s-have-faced-challenges-on-multiple-fronts-during-the-pandemic/ [accessed 2022-11-01]
16. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
17. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-04-08]
18. Waring ME, Blackman Carr LT, Heersping GE. Social media use among parents and women of childbearing age in the US. Prev Chronic Dis 2023 Feb 16;20:E07 [FREE Full text] [doi: 10.5888/pcd20.220194] [Medline: 36796010]
19. Duggan M, Lenhart A, Lampe C, Ellison NB. Parents and social media: mothers are especially likely to give and receive support on social media. Pew Research Center. 2015 Jul 16. URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2022-08-24]
20. Holtz B, Smock A, Reyes-Gastelum D. Connected motherhood: social support for moms and moms-to-be on Facebook. Telemed J E Health 2015 May;21(5):415-421. [doi: 10.1089/tmj.2014.0118] [Medline: 25665177]
21. An R, Ji M, Zhang S. Effectiveness of social media-based interventions on weight-related behaviors and body weight status: review and meta-analysis. Am J Health Behav 2017 Nov 01;41(6):670-682. [doi: 10.5993/AJHB.41.6.1] [Medline: 29025495]
22. Waring ME, Moore Simas TA, Oleski J, Xiao RS, Mulcahy JA, May CN, et al. Feasibility and acceptability of delivering a postpartum weight loss intervention via Facebook: a pilot study. J Nutr Educ Behav 2018 Jan;50(1):70-4.e1 [FREE Full text] [doi: 10.1016/j.jneb.2017.09.025] [Medline: 29325665]
23. Herring S, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014 Nov;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
24. Silfee VJ, Lopez-Cepero A, Lemon SC, Estabrook B, Nguyen O, Wang ML, et al. Adapting a behavioral weight loss intervention for delivery via Facebook: a pilot series among low-income postpartum women. JMIR Form Res 2018 Sep 10;2(2):e18 [FREE Full text] [doi: 10.2196/formative.9597] [Medline: 30684423]
25. Herring SJ, Bersani VM, Santoro C, McNeil SJ, Kilby LM, Bailer B. Feasibility of using a peer coach to deliver a behavioral intervention for promoting postpartum weight loss in Black and Latina mothers. Transl Behav Med 2021 Jun 17;11(6):1226-1234 [FREE Full text] [doi: 10.1093/tbm/ibaa096] [Medline: 33184667]
26. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
27. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
28. Waring ME, Libby BA, Moore Simas TA, Bracken ML, Bibeau JL, Herrera V, et al. Delivering a post-partum weight loss intervention via Facebook or in-person groups: protocol for a randomized feasibility pilot trial. JMIR Res Protoc 2019 Nov 28;8(11):e15530 [FREE Full text] [doi: 10.2196/15530] [Medline: 31778116]
29. Harris PA, Scott KW, Lebo L, Hassan N, Lightner C, Pulley J. ResearchMatch: a national registry to recruit volunteers for clinical research. Acad Med 2012 Jan;87(1):66-73 [FREE Full text] [doi: 10.1097/ACM.0b013e31823ab7d2] [Medline: 22104055]
30. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
31. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Publishing; 2013.
32. Moseson H, Fix L, Hastings J, Stoeffler A, Lunn MR, Flentje A, et al. Pregnancy intentions and outcomes among transgender, nonbinary, and gender-expansive people assigned female or intersex at birth in the United States: results from a national,

quantitative survey. Int J Transgend Health 2021 Nov 17;22(1-2):30-41 [FREE Full text] [doi: 10.1080/26895269.2020.1841058] [Medline: 34796363]
33. Jake-Schoffman DE, Brown SD, Baiocchi M, Bibeau JL, Daubenmier J, Ferrara A, et al. Methods-motivational interviewing approach for enhanced retention and attendance. Am J Prev Med 2021 Oct;61(4):606-617 [FREE Full text] [doi: 10.1016/j.amepre.2021.04.005] [Medline: 34544560]
34. Revilla M, Ochoa C, Loewe G. Using passive data from a meter to complement survey data in order to study online behavior. Soc Sci Comput Rev 2017 Aug;35(4):521-536 [FREE Full text] [doi: 10.1177/0894439316638457]
35. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: the PROMIS®-Preference (PROPr) scoring system. Med Decis Making 2018 Aug;38(6):683-698 [FREE Full text] [doi: 10.1177/0272989X18776637] [Medline: 29944456]
36. Hanmer J, Dewitt B, Yu L, Tsevat J, Roberts M, Revicki D, et al. Cross-sectional validation of the PROMIS-preference scoring system. PLoS One 2018 Jul 31;13(7):e0201093 [FREE Full text] [doi: 10.1371/journal.pone.0201093] [Medline: 30063733]
37. Lee CF, Hwang FM, Liou YM, Chien LY. A preliminary study on the pattern of weight change from pregnancy to 6 months postpartum: a latent growth model approach. Int J Obes (Lond) 2011 Aug;35(8):1079-1086. [doi: 10.1038/ijo.2010.225] [Medline: 21042322]
38. Onyango AW, Nommsen-Rivers L, Siyam A, Borghi E, de Onis M, Garza C, WHO Multicentre Growth Reference Study Group. Post-partum weight change patterns in the WHO multicentre growth reference study. Matern Child Nutr 2011 Jul;7(3):228-240 [FREE Full text] [doi: 10.1111/j.1740-8709.2010.00295.x] [Medline: 21338469]
39. Otten JJ, Hellwig JP, Meyers LD. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. Washington, DC, USA: The National Academics Press; 2006.
40. Committee on Obstetric Practice. ACOG committee opinion. Exercise during pregnancy and the postpartum period. Number 267, January 2002. American College of Obstetricians and Gynecologists. Int J Gynaecol Obstet 2002 Apr;77(1):79-81. [doi: 10.1016/s0020-7292(02)80004-2] [Medline: 12053898]
41. Chang MW, Nitzke S, Guilford E, Adair CH, Hazard DL. Motivators and barriers to healthful eating and physical activity among low-income overweight and obese mothers. J Am Diet Assoc 2008 Jun;108(6):1023-1028. [doi: 10.1016/j.jada.2008.03.004] [Medline: 18502238]
42. Epifanio MS, Genna V, De Luca C, Roccella M, La Grutta S. Paternal and maternal transition to parenthood: the risk of postpartum depression and parenting stress. Pediatr Rep 2015 May 25;7(2):5872 [FREE Full text] [doi: 10.4081/pr.2015.5872] [Medline: 26266033]
43. Pagoto SL, Waring ME, Schneider KL, Oleski JL, Olendzki E, Hayes RB, et al. Twitter-delivered behavioral weight-loss interventions: a pilot series. JMIR Res Protoc 2015 Oct 23;4(4):e123 [FREE Full text] [doi: 10.2196/resprot.4864] [Medline: 26500186]
44. Wang ML, Waring ME, Jake-Schoffman DE, Oleski JL, Michaels Z, Goetz JM, et al. Clinic versus online social network-delivered lifestyle interventions: protocol for the get social noninferiority randomized controlled trial. JMIR Res Protoc 2017 Dec 11;6(12):e243 [FREE Full text] [doi: 10.2196/resprot.8068] [Medline: 29229591]
45. Food and Nutrition Information Center (U.S.). Interactive DRI for healthcare professionals. United States Department of Agriculture National Agricultural. 2009. URL: https://www.nal.usda.gov/human-nutrition-and-food-safety/dri-calculator [accessed 2023-04-15]
46. American College of Obstetricians and Gynecologists. ACOG committee opinion number 313, September 2005. The importance of preconception care in the continuum of women's health care. Obstet Gynecol 2005 Sep;106(3):665-666. [doi: 10.1097/00006250-200509000-00052] [Medline: 16135611]
47. Pagoto S, Waring ME, Jake-Schoffman DE, Goetz J, Michaels Z, Oleski J, et al. What type of engagement predicts success in a Facebook weight loss group? In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS-51 '18; January 3-6, 2018; Hilton Waikoloa Village, HI, USA p. 3304-3312 URL: https://scholarspace.manoa.hawaii.edu/handle/10125/50307 [doi: 10.24251/HICSS.2018.419]
48. Smith A. Mobile access 2010. Pew Research Center. 2010 Jul 07. URL: https://www.pewresearch.org/internet/2010/07/07/mobile-access-2010/ [accessed 2023-04-15]
49. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
50. Frick KD. Microcosting quantity data collection methods. Med Care 2009 Jul;47(7 Suppl 1):S76-S81 [FREE Full text] [doi: 10.1097/MLR.0b013e31819bc064] [Medline: 19536026]
51. Hernan WH, Brandle M, Zhang P, Williamson DF, Matulik MJ, Ratner RE, Diabetes Prevention Program Research Group. Costs associated with the primary prevention of type 2 diabetes mellitus in the diabetes prevention program. Diabetes Care 2003 Jan;26(1):36-47 [FREE Full text] [doi: 10.2337/diacare.26.1.36] [Medline: 12502656]
52. Ritzwoller DP, Sukhanova A, Gaglio B, Glasgow RE. Costing behavioral interventions: a practical guide to enhance translation. Ann Behav Med 2009 Apr;37(2):218-227. [doi: 10.1007/s12160-009-9088-5] [Medline: 19291342]

53. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-an ISPOR good research practices task force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
54. Mertens IL, Van Gaal LF. Overweight, obesity, and blood pressure: the effects of modest weight reduction. Obes Res 2000 May;8(3):270-278 [FREE Full text] [doi: 10.1038/oby.2000.32] [Medline: 10832771]
55. Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, Look AHEAD Research Group. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011 Jul;34(7):1481-1486 [FREE Full text] [doi: 10.2337/dc10-2415] [Medline: 21593294]
56. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
57. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010 Jan 06;10:1 [FREE Full text] [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]
58. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
59. Groth SW, David T. New mothers' views of weight and exercise. MCN Am J Matern Child Nurs 2008 Nov;33(6):364-370 [FREE Full text] [doi: 10.1097/01.NMC.0000341257.26169.30] [Medline: 18997572]
60. Mailey EL, Huberty J, Dinkel D, McAuley E. Physical activity barriers and facilitators among working mothers and fathers. BMC Public Health 2014 Jun 27;14:657 [FREE Full text] [doi: 10.1186/1471-2458-14-657] [Medline: 24974148]
61. Evenson KR, Aytur SA, Borodulin K. Physical activity beliefs, barriers, and enablers among postpartum women. J Womens Health (Larchmt) 2009 Dec;18(12):1925-1934 [FREE Full text] [doi: 10.1089/jwh.2008.1309] [Medline: 20044854]
62. Krukowski RA, Ross KM. Measuring weight with electronic scales in clinical and research settings during the coronavirus disease 2019 pandemic. Obesity (Silver Spring) 2020 Jul;28(7):1182-1183 [FREE Full text] [doi: 10.1002/oby.22851] [Medline: 32339394]
63. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a web-based obesity treatment program. Telemed J E Health 2011 Nov;17(9):696-699 [FREE Full text] [doi: 10.1089/tmj.2011.0032] [Medline: 21882997]
64. Jerome GJ, Dalcin A, Coughlin JW, Fitzpatrick S, Wang NY, Durkin N, et al. Longitudinal accuracy of web-based self-reported weights: results from the Hopkins POWER Trial. J Med Internet Res 2014 Jul 15;16(7):e173 [FREE Full text] [doi: 10.2196/jmir.3332] [Medline: 25042773]
65. Caldwell AE, Thomas EA, Rynders C, Holliman BD, Perreira C, Ostendorf DM, et al. Improving lifestyle obesity treatment during the COVID-19 pandemic and beyond: new challenges for weight management. Obes Sci Pract 2022 Feb;8(1):32-44 [FREE Full text] [doi: 10.1002/osp4.540] [Medline: 34540266]
66. Cardel MI, Manasse S, Krukowski RA, Ross K, Shakour R, Miller DR, et al. COVID-19 impacts mental health outcomes and ability/desire to participate in research among current research participants. Obesity (Silver Spring) 2020 Dec;28(12):2272-2281 [FREE Full text] [doi: 10.1002/oby.23016] [Medline: 32845582]
67. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
68. Pagoto S, Tulu B, Waring ME, Goetz J, Bibeau J, Divito J, et al. Slip buddy app for weight management: randomized feasibility trial of a dietary lapse tracking app. JMIR Mhealth Uhealth 2021 Apr 01;9(4):e24249 [FREE Full text] [doi: 10.2196/24249] [Medline: 33792547]
69. Pagoto SL, Schroeder MW, Xu R, Waring ME, Groshon L, Goetz JM, et al. A Facebook-delivered weight loss intervention using open enrollment: randomized pilot feasibility trial. JMIR Form Res 2022 May 06;6(5):e33663 [FREE Full text] [doi: 10.2196/33663] [Medline: 35522466]
70. Cavallo DN, Martinez R, Webb Hooper M, Flocke S. Feasibility of a social media-based weight loss intervention designed for low-SES adults. Transl Behav Med 2021 Apr 26;11(4):981-992 [FREE Full text] [doi: 10.1093/tbm/ibaa070] [Medline: 32716040]
71. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
72. Cooper P. How Facebook's algorithm works in 2021 and how to make it work for you. One18media. 2021 May 4. URL: https://one18media.com/facebook-algorithm-works-2021-make-work/ [accessed 2023-04-15]
73. Pagoto S, Waring ME. A call for a science of engagement: comment on Rus and Cameron. Ann Behav Med 2016 Oct;50(5):690-691. [doi: 10.1007/s12160-016-9839-z] [Medline: 27663577]
74. Franciosi EB, Tan AJ, Kassamali B, Leonard N, Zhou G, Krueger S, et al. The impact of telehealth implementation on underserved populations and no-show rates by medical specialty during the COVID-19 pandemic. Telemed J E Health 2021 Aug;27(8):874-880 [FREE Full text] [doi: 10.1089/tmj.2020.0525] [Medline: 33826411]
75. Fischer SH, Uscher-Pines L, Roth E, Breslau J. The transition to telehealth during the first months of the COVID-19 pandemic: evidence from a national sample of patients. J Gen Intern Med 2021 Mar;36(3):849-851 [FREE Full text] [doi: 10.1007/s11606-020-06358-0] [Medline: 33409884]

76. Tozour JN, Bandremer S, Patberg E, Zavala J, Akerman M, Chavez M, et al. Application of telemedicine video visits in a maternal-fetal medicine practice at the epicenter of the COVID-19 pandemic. Am J Obstet Gynecol MFM 2021 Nov;3(6):100469 [FREE Full text] [doi: 10.1016/j.ajogmf.2021.100469] [Medline: 34450341]
77. McClain C, Vogels EA, Perrin A, Sechopoulos S, Rainie L. The internet and the pandemic. Pew Research Center. 2021 Sep 01. URL: https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/ [accessed 2021-11-06]
78. Ross KM, Carpenter CA, Arroyo KM, Shankar MN, Yi F, Qiu P, et al. Impact of transition from face-to-face to telehealth on behavioral obesity treatment during the COVID-19 pandemic. Obesity (Silver Spring) 2022 Apr;30(4):858-863 [FREE Full text] [doi: 10.1002/oby.23383] [Medline: 35037410]
79. Zaman A, Sloggett KJ, Caldwell AE, Catenacci VA, Cornier MA, Grau L, et al. The effects of the COVID-19 pandemic on weight loss in participants in a behavioral weight-loss intervention. Obesity (Silver Spring) 2022 May;30(5):1015-1026. [doi: 10.1002/oby.23399] [Medline: 35118814]
80. Batsis JA, McClure AC, Weintraub AB, Sette D, Rotenberg S, Stevens CJ, et al. Barriers and facilitators in implementing a pilot, pragmatic, telemedicine-delivered healthy lifestyle program for obesity management in a rural, academic obesity clinic. Implement Sci Commun 2020 Sep 30;1:83 [FREE Full text] [doi: 10.1186/s43058-020-00075-9] [Medline: 33015640]
81. Jain B, Bajaj SS, Stanford FC. Randomized clinical trials of weight loss: pragmatic and digital strategies and innovations. Contemp Clin Trials 2022 Mar;114:106687 [FREE Full text] [doi: 10.1016/j.cct.2022.106687] [Medline: 35085830]
82. Phelan S, Hagobian T, Brannen A, Hatley KE, Schaffner A, Muñoz-Christian K, et al. Effect of an internet-based program on weight loss for low-income postpartum women: a randomized clinical trial. JAMA 2017 Jun 20;317(23):2381-2391 [FREE Full text] [doi: 10.1001/jama.2017.7119] [Medline: 28632867]
83. Matthews T, Hamilton BE. Educational attainment of mothers aged 25 and over: United States, 2017. NCHS Data Brief 2019 Feb(332):1-8 [FREE Full text] [Medline: 31112119]
84. Haughton CF, Silfee VJ, Wang ML, Lopez-Cepero AC, Estabrook DP, Frisard C, et al. Racial/ethnic representation in lifestyle weight loss intervention studies in the United States: a systematic review. Prev Med Rep 2018 Mar;9:131-137 [FREE Full text] [doi: 10.1016/j.pmedr.2018.01.012] [Medline: 29616185]
85. Rosenbaum DL, Piers AD, Schumacher LM, Kase CA, Butryn ML. Racial and ethnic minority enrollment in randomized clinical trials of behavioural weight loss utilizing technology: a systematic review. Obes Rev 2017 Jul;18(7):808-817. [doi: 10.1111/obr.12545] [Medline: 28524643]
86. Crane MM, Seburg EM, Levy RL, Jeffery RW, Sherwood NE. Using targeting to recruit men and women of color into a behavioral weight loss trial. Trials 2020 Jun 16;21(1):537 [FREE Full text] [doi: 10.1186/s13063-020-04500-1] [Medline: 32546253]
87. Cui Z, Truesdale KP, Robinson TN, Pemberton V, French SA, Escarfuller J, et al. Recruitment strategies for predominantly low-income, multi-racial/ethnic children and parents to 3-year community-based intervention trials: Childhood Obesity Prevention and Treatment Research (COPTR) consortium. Trials 2019 May 28;20(1):296 [FREE Full text] [doi: 10.1186/s13063-019-3418-0] [Medline: 31138278]
88. Chang MW, Brown R, Nitzke S. Results and lessons learned from a prevention of weight gain program for low-income overweight and obese young mothers: mothers in motion. BMC Public Health 2017 Feb 10;17(1):182 [FREE Full text] [doi: 10.1186/s12889-017-4109-y] [Medline: 28187726]

Abbreviations

DPP: Diabetes Prevention Program
EPDS: Edinburgh Postnatal Depression Scale
LAMB: Los Angeles Mommy and Baby
PROPr: PROMIS-Preference
REDCap: Research Electronic Data Capture
WIC: Women, Infants, & Children Nutrition Program

Edited by L Buis; submitted 29.07.22; peer-reviewed by A Dahl, J Antoun; comments to author 22.08.22; revised version received 22.11.22; accepted 10.03.23; published 27.04.23.

Please cite as:
Waring ME, Pagoto SL, Moore Simas TA, Blackman Carr LT, Eamiello ML, Libby BA, Rudin LR, Heersping GE
Delivering a Postpartum Weight Loss Intervention via Facebook or In-Person Groups: Results From a Randomized Pilot Feasibility Trial
JMIR Mhealth Uhealth 2023;11:e41545
URL: https://mhealth.jmir.org/2023/1/e41545
doi:10.2196/41545
PMID:37103991

Original Paper

The Effect of Periodic Email Prompts on Participant Engagement With a Behavior Change mHealth App: Longitudinal Study

Elena Agachi[1*], MSc; Tammo H A Bijmolt[1*], Prof Dr; Koert van Ittersum[1*], Prof Dr; Jochen O Mierau[2*], Prof Dr

[1]Department of Marketing, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[2]Department of Economics, Econometrics & Finance, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[*]all authors contributed equally

Corresponding Author:
Elena Agachi, MSc
Department of Marketing
Faculty of Economics and Business
University of Groningen
Nettelbosje 2
Groningen, 9747 AE
Netherlands
Phone: 31 50 363 3686
Email: e.agachi@rug.nl

Abstract

Background: Following the need for the prevention of noncommunicable diseases, mobile health (mHealth) apps are increasingly used for promoting lifestyle behavior changes. Although mHealth apps have the potential to reach all population segments, providing accessible and personalized services, their effectiveness is often limited by low participant engagement and high attrition rates.

Objective: This study concerns a large-scale, open-access mHealth app, based in the Netherlands, focused on improving the lifestyle behaviors of its participants. The study examines whether periodic email prompts increased participant engagement with the mHealth app and how this effect evolved over time. Points gained from the activities in the app were used as an objective measure of participant engagement with the program. The activities considered were physical workouts tracked through the mHealth app and interactions with the web-based coach.

Methods: The data analyzed covered 22,797 unique participants over a period of 78 weeks. A hidden Markov model (HMM) was used for disentangling the overtime effects of periodic email prompts on participant engagement with the mHealth app. The HMM accounted for transitions between latent activity states, which generated the observed measure of points received in a week.

Results: The HMM indicated that, on average, 70% (15,958/22,797) of the participants were in the inactivity state, gaining 0 points in total per week; 18% (4103/22,797) of the participants were in the average activity state, gaining 27 points per week; and 12% (2736/22,797) of the participants were in the high activity state, gaining 182 points per week. Receiving and opening a generic email was associated with a 3 percentage point increase in the likelihood of becoming active in that week, compared with the weeks when no email was received. Examining detailed email categories revealed that the participants were more likely to increase their activity level following emails that were in line with the program's goal, such as emails regarding health campaigns, while being resistant to emails that deviated from the program's goal, such as emails regarding special deals.

Conclusions: Participant engagement with a behavior change mHealth app can be positively influenced by email prompts, albeit to a limited extent. Given the relatively low costs associated with emails and the high population reach that mHealth apps can achieve, such instruments can be a cost-effective means of increasing participant engagement in the stride toward improving program effectiveness.

(JMIR Mhealth Uhealth 2023;11:e43033) doi:10.2196/43033

KEYWORDS

mobile health; behavior change; mobile app; digital health; engagement; retention; email; hidden Markov model

Introduction

Background

Following the increasing need for the prevention of noncommunicable diseases [1], behavior change programs have emerged as a widely used support tool for health interventions aimed at improving lifestyle behaviors [2]. Digital behavior change programs have the ability to reach a larger population subset at relatively lower costs than their offline counterparts [3,4] while allowing for tailored material based on individual interactions [5].

In recent years, mobile health (mHealth) apps have gained traction as an increasingly preferred method of delivering digital behavior change interventions [4] by further facilitating access for and interaction with participants [6]. An additional benefit of mHealth apps is their ability to also involve the population segment with lower socioeconomic conditions, which generally shows less interest in preventive health interventions [7-9].

Although digital behavior change programs, especially mHealth apps, are promising tools for improving lifestyle behaviors, in practice, these programs often show low participant engagement (defined as "the extent of usage of the digital behavior change intervention" [10]) and high attrition rates [11-13]. One of the main reasons underlying this phenomenon is the passive nature of behavior change programs, where participants need to act by themselves to benefit [14]. Although higher engagement is crucial for achieving effectiveness of mHealth apps [15-18], inducing higher participant engagement over time is a challenging task [12,19], which requires proactive efforts from the program providers [11,20,21].

Objective

Periodic prompts via emails have been examined as a potential tool that can boost participant engagement with behavior change mHealth apps [22,23]. However, most studies examining the means to increase engagement with an mHealth app are based on small sample sizes and short time spans [24,25]. In a literature review including approximately 35 mHealth apps aimed at increasing physical activity, the sample size varied between 8 and 700 participants, with an average study duration of 8 weeks [26]. Given that small sample sizes and especially short time spans of most interventions can lead to an overestimation of the intervention effects [25,27], it is essential to examine whether periodic prompts via emails can impact participant engagement with a behavior change mHealth app within a longer-term, larger-scale, noncontrolled setup [7,28,29].

This study relied on a large-scale (more than 20,000 participants), open-access mHealth app focused on improving the lifestyle behaviors and wellness of its participants. The analysis in this study used a hidden Markov model (HMM) to examine whether periodic email prompts were able to increase participant engagement with the mHealth app and how this effect evolved over time. By investigating the effect of prompts on continued engagement with the mHealth app, this study hoped to assess (1) whether periodic prompts via email can be a viable tool for increasing participant engagement, (2) how the impact of periodic prompts on engagement evolved over time, and (3) how the observed effects differed among participant subgroups.

Methods

Study Sample

This study was based on data from the mobile app of a digital behavior change program operated in the Netherlands. The program's goal was to improve the wellness and lifestyle behaviors of its participants by promoting physical activity, healthy eating habits, social activity, mental health, good sleep habits, and minimized stress. The mobile app was introduced in October 2017, providing functions such as entering or recording physical activities, reading articles, setting goals, including friends in challenges, answering health questions, being assisted by a web-based coach, and forming a daily "fit-score." On the basis of the individual activities in the mobile app, participants gain points, which can be used to acquire specific products, vouchers for various services, or make charity contributions.

The data analyzed in this study spanned from January 2018, when the mobile app of the health program reached full functionality, to July 2019, when the observation window ended. Data were collected in 2018 and 2019 and analyzed in 2021 and 2022 within a longitudinal, nonexperimental study design. The analyzed data had a weekly frequency, covering 78 weeks and including 22,797 unique participants who enrolled by themselves in the mobile app at any time during the observation window. Of the 33,825 participants who used the mobile app, 22,797 (67.4%) were included, having at least 1 activity during the period between the mobile app introduction and the end of the observation window, indicating an awareness of the mobile app's functionality. All program participants were aged between 18 and 80 years and were residents of the Netherlands.

Enrollment in the mHealth app was open and free, and all the participants involved in this study provided their voluntary and informed consent.

Ethics Approval

Ethics approval for this research project based on the health program was obtained from the institutional research board of the University of Groningen (approval number RDMPFEB20180831-7309).

Measures

Participant Engagement

The main objective of a behavior change mHealth app is to improve the lifestyle behaviors of its participants. Achieving effectiveness in behavior change is highly linked to the degree to which participants engage with the app; only when participants interact with the program and continue use can it have an impact on their behavior [15,16].

Participant engagement has been defined as "the extent of usage of the digital behavior change intervention" [10], being separated into temporal patterns—frequency and duration, and depth—specific intervention content use [30,31]. Participant engagement can be assessed as a subjective measure (ie, self-reported by participants) or an objective measure (ie, measured by the program) [10].

The mHealth app analyzed in this study provided several activities that the participants could perform. For every activity completed, the participants gained points, which varied based on the activity type and duration. Gained points were used as a measure of the activity level of the participants in the mHealth app, which were calculated weekly throughout the observation window. Consequently, this participant engagement measure was objective rather than self-reported, which had the additional benefit of being more robust against reporting bias [32].

The 2 types of activities included in the participant engagement measure were physical activities and web-based coach activities, both of which the participants could access via the mHealth app. Physical activities were activities recorded in the health program with the use of GPS, such as walking, cycling, and running. Web-based coach activities were the interactions that the participants had with regular messages sent via the chat environment programmed by the providers of the mHealth app. The messages in the chat environment were linked to physical activities, health goals that the participants selected, challenges that they joined, or overall health behavior information. For every question answered, the participants received a fixed number of points. Table 1 presents the activity types included in the mHealth app, with the associated number of points.

Table 1. Activity types and number of points gained.

Activity type	Number of points
Web-based coach	1 point gained for any question answered
GPS recorded activity: walking	From 1 point to 696 points, depending on activity duration (mean 17.25, SD 24.46 points)
GPS recorded activity: cycling	From 1 point to 699 points, depending on activity duration (mean 21.37, SD 38.36 points)
GPS recorded activity: running	From 1 point to 695 points, depending on activity duration (mean 39.93, SD 33.06 points)

Periodic Email Prompts

In general, mHealth apps suffer from low participant engagement and high dropout rates [33,34]. Capturing the attention of the participants in an attempt to stimulate their active involvement is crucial for program success [20]. Periodic email prompts are often used as tools for improving participant engagement, with mixed results. Although some studies estimated a positive impact of email prompts on participant engagement [23,28,35], additional research is required in this area [36], with a special focus on overtime effects [10,22].

In this study, we examined the ability of periodic email prompts to improve participant engagement with the mHealth app by measuring the effect of emails on transitions between activity states. For every email sent, a randomly selected subset of the participants did not receive the email in question, which served as the control group for that particular email. The emails sent to the participants of the app were either generic or targeted. The generic emails were sent to all the participants in the same format, independent of their current activity level. The targeted emails were sent out in different versions, depending on the participants' activity level (low activity or high activity); however, it was not possible to identify which participant received which email version. The targeted emails could belong to one of the following categories: welcoming emails, reactivation emails, recruitment emails, newsletters, health campaigns, and special offers.

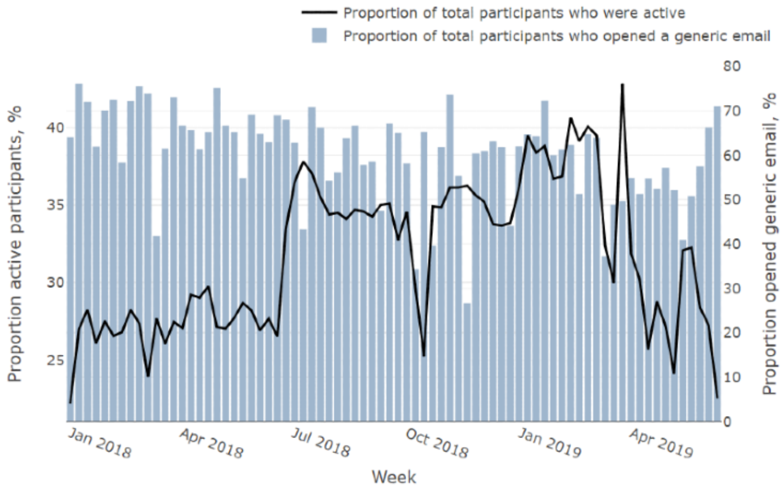
To correct for any potential effect of targeting, generic emails were the main measure used in this study. The emails included in this category had topics such as welcoming participants to the program, sharing general healthy lifestyle information (at regular intervals), presenting topic-specific health information (eg, healthy nutrition and sleep), inviting participants to engage in activities, and presenting special deals in the web shop.

To measure the effect of email prompts, we distinguished between 3 situations: not having received an email, having received an email but not having opened it, and having received an email and opened it. Section 2 in Multimedia Appendix 1 describes in detail the emails' content and their categorization.

Between January 2018 and July 2019, the proportion of participants who received a generic email varied between 0.14% (13/9596 participants in the third week of March 2018) and 39.73% (8943/22,508 participants in the second week of May 2019). The proportion of participants who opened the generic email when they received it varied between 26.65% (271/1017 participants in the third week of November 2018) and 76.1% (159/209 participants in the first week of January 2018). Across the weeks in the observation period, the average proportion of participants who received a generic email in a week was 6.6%, the average proportion of participants who opened a generic email in a week was 3.5%, and the average proportion of participants who opened a generic email when they received one was 60.5%.

Figure 1 displays the evolution of the generic emails and participant engagement over the weeks of the observation window, showing the proportion of participants who received and opened a generic email and the proportion of participants who were active during that week (having gained at least 1 point). Figure 1 implies great variability in both measures, highlighting the need to analyze the connection between email prompts and participant engagement in a dynamic manner.

Figure 1. The proportion of active participants (at least 1 point gained) and the proportion of participants who received and opened a generic email.



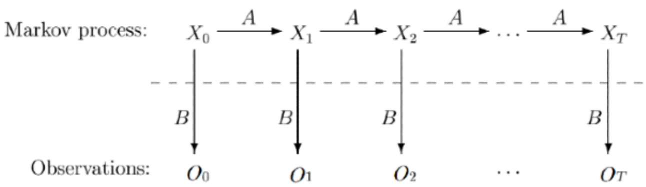
To control for the effect of the individual characteristics of the program participants on their engagement with the behavior change mHealth app [10], gender, age, and neighborhood socioeconomic status (NSES) [37] quintiles were included in the analysis as additional covariates. The NSES quintiles measure follows the methodology outlined in Dekker et al [38], being calculated using nonlinear iterative partial least squares principal component analysis on the following characteristics given on a postcode level: average income, average property value, subsidized renting, share of high-income households, share of owner-occupied properties, share of low-income households, share of population receiving unemployment benefits, share of people receiving disability benefits, and share of people receiving short-term unemployment benefits. A lower NSES quintile corresponds to lower levels of socioeconomic conditions. In addition, to measure whether early adopters of the mHealth app showed higher engagement [39], we included the additional measure of early adoption, which corresponded to the participants who enrolled in the behavior change mHealth app during its first month of existence. In total, 19.06% (4345/22,797) of the mHealth app participants were early adopters.

Statistical Analysis

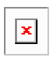
In this study, we used the number of points gained per week to measure engagement, which reflected the level of activity of a participant in the mHealth app. To model the changing levels of activity over time, an HMM was used, where a participant had a specific level of activity each week (latent state) and could transition between the activity states from week to week [40,41]. Using HMMs allows for the disentanglement of the dynamics of participant behavior over time and the analysis of how specific actions can influence these behaviors [42]. Moreover, the HMM model is preferred because such a latent approach allows for the incorporation of the high dropout and inactivity rates that are specific to behavior change programs [43,44]. To understand the drivers of the dynamics of state transitions, nonhomogeneous Markov modeling was used, which allows the transition probabilities to depend on time-varying covariates [45].

A generic HMM is defined as shown in Figure 2, where X_t is the latent activity state at time *t*, with *t* ranging from *0* to *T* (*T* being the last measurement week); *A* is the state transition probability; *B* is the response probability matrix; and O_t contains the observations in the response vector. The Markov process, being separated by the dashed line, was not observed. Instead, only the observations O_t were known; in this study, these were the number of points gained in a week.

Figure 2. A generic hidden Markov model representation. A: state transition probability; B: response probability matrix; O: observations in the response vector; T: last measurement week.



The HMM depicted in Figure 2 consists of 3 main elements (as shown below):



The hidden Markov model formulation includes:

1. Initial state probability *P(X0)*: the probability that participant *i* is in state *X* at time 0.
2. Transition probability *P(Xt|Xt-1)*: the probability that participant *i* is in state *X* at time *t*, given the state membership at time *t-1*.
3. Response probability *P(Oit|Xt)*: the probability that participant *i* displays activity level *O* at time *t*, given the state membership *X* at time *t*.

In the setup of this study, the unobserved states that a participant belonged to were activity states, generating the observed measures of points received in a week from differing activities performed in the mHealth app. The initial state distribution reflected the starting state that a participant belonged to at their moment of joining the mHealth app, which depended on the time-constant covariates that reflected the participant's background (age, gender, NSES, and early adopter). The transitions between the activity states reflected the variability in the participants' behaviors between weeks, which were allowed to depend on both time-constant covariates (age, gender, NSES, and early adopter) and time-varying covariates (the email prompts and time). Including the email prompts in the transition probability model allowed for the examination of the impact of emails on changes in the activity levels of a participant.

When using the HMM for general inference, traditional model selection criteria, such as Akaike information criterion or Bayesian information criterion, often lead to the selection of much larger numbers of states than expected a priori [46-48]. The reason for this is that the neglected data in the model formulation are absorbed into the additional model states, which do not possess a clear interpretation anymore [49]. A recommended approach for dealing with this uncertainty is analyzing a prespecified number of latent states. In this study, following the goal of differentiating between activity states while prioritizing interpretability, the estimated HMM contained 3 states: inactivity, average activity, and high activity. For the estimation of the model, the Latent GOLD software (Statistical Innovations Inc) was used. The Latent GOLD software supports the analysis of latent class models such as HMMs, with the parameter estimates being computed based on a combination of expectation-maximization and Newton-Raphson iterations, where the E step computations use a forward-backward recursion scheme [50].

Results

Characteristics of the Participants

This study analyzed 22,797 participants between January 2018 and July 2019 for a total of 78 weeks. The mHealth app analyzed being an open-access platform, the participants could enroll at any time within the observation window. Table 2 outlines the characteristics of the mHealth app participants. On average, each participant was observed for 50.8 weeks, resulting in a total of 1,129,706 observation points. Every week, approximately one-third of the study population was active, gaining an average of 28 points weekly. A total of 62.82% (14,321/22,797) of the analyzed participants were women, with the most represented age group being between 37 and 46 years and the highest proportion of participants belonging to the second socioeconomic quintile.

Table 2. Characteristics of the study participants (N=22,797).

Key attributes	Values
Total observations, n	1,129,706
Participants, n	22,797
Number of weeks in the mobile app, mean (SD)	50.8 (28.7)
Number of points received weekly, mean (SD)	28.0 (78.8)
Proportion of active participants per week (%), mean (SD)	31.1 (6.3)
Early mobile app adopters, n (%)	4345 (19.06)
Female participants, n (%)	14,321 (62.82)
Participants per age group (years), n (%)	
18-26	1408 (6.18)
27-36	5282 (23.17)
37-46	5406 (23.71)
47-56	5179 (22.72)
57-66	3402 (14.92)
67-80	2120 (9.3)
Participants per NSES[a] quintile (from the lowest to the highest socioeconomic conditions), n (%)	
First	4675 (20.51)
Second	6303 (27.65)
Third	4662 (20.45)
Fourth	3611 (15.84)
Fifth	3546 (15.55)

[a]NSES: neighborhood socioeconomic status.

Model Estimation Results

HMM States and Transitions

Estimating the HMM with 3 states resulted in the outcomes presented in Tables 3 and 4. The 3 states identified by the HMM were labeled as the inactivity state, average activity state, and high activity state. On average, across the weeks of the observation period, 70% (15,958/22,797) of the participants were in the inactivity state, gaining 0 points weekly; 18% (4103/22,797) of the participants were in the average activity state, gaining 27 points weekly; and 12% (2736/22,797) of the participants were in the high activity state, gaining 182 points weekly.

Table 3. Hidden Markov model estimation results: average latent states.

	State		
	1: inactivity	2: average activity	3: high activity
Average state size (%)	70	18	12
Points received, n	0	27	182

Table 4. Hidden Markov model estimation results: average transition probability matrix[a].

State (t-1)	State (t[b])		
	1	2	3
1	0.91	0.07	0.02
2	0.31	0.64	0.05
3	0.09	0.09	0.82

[a]Section 3 in Multimedia Appendix 1 presents detailed model fit criteria and parameter estimates for the hidden Markov model. All the parameter estimates were statistically significant at the 99% confidence level.

[b]t: time point.

The estimated transition matrix (shown in Table 4) reflects the probability of switching between the 3 states across weeks. On average, the inactivity and high activity states were most persistent, for example, a participant who was in the high activity state during week 1 was, on average, 82% likely to remain in that state during week 2. The highest probability of decrease in activity was associated with the transition from the average activity state to the inactivity state: a participant who was in the average activity state during week 1 was 31% likely to transition into the inactivity state during week 2.

HMM Effects of Generic Email Prompts

The relationship of interest in this study is the connection between generic emails and participant engagement. Table 5 shows the estimated posterior probability means of the state distribution depending on whether the participants received and opened a generic email. The posterior probability means indicate the estimated probability that a participant was in each state, given the email prompt. The estimates show that a participant who did not receive a generic email was 68% likely to be in the inactivity state, whereas a participant who received and opened a generic email was 67% likely to be in the inactivity state (a decrease of 1 percentage point). In addition, the participants who received but did not open a generic email were estimated to have a 11 percentage point higher likelihood of inactivity than those who did not receive an email.

Table 5. Hidden Markov model estimation results: posterior probability means associated with the generic email prompts[a].

	1: inactivity	2: average activity	3: high activity
No generic email received	0.68	0.19	0.13
Generic email received but not opened	0.79	0.14	0.06
Generic email received and opened	0.67	0.22	0.11

[a]Section 3 in Multimedia Appendix 1 presents detailed model fit criteria and parameter estimates for the hidden Markov model. All the parameter estimates were statistically significant at the 99% confidence level.

As the HMM allowed for the dynamics of switching between states, Table 6 shows the estimated transition matrices depending on the generic email. On the basis of the estimated transition matrices, the likelihood that a participant remained in the inactivity state between weeks *t* and *t+1* was 91% when no email was received, as opposed to 88% when a generic email was received and opened. This translates into a 3 percentage point decrease in the probability of remaining inactive or, alternatively, a 3 percentage point increase in the probability of moving into one of the activity states after receiving and opening a generic email.

Table 6. Hidden Markov model estimation results: transition matrices accounting for generic email prompts.

State (t[a]-1)	State (t)								
	Transition matrix: no generic email received			Transition matrix: generic email received but not opened			Transition matrix: generic email received and opened		
	1	2	3	1	2	3	1	2	3
1	0.91	0.07	0.02	0.93	0.06	0.01	0.88	0.10	0.02
2	0.31	0.64	0.05	0.41	0.56	0.03	0.29	0.66	0.04
3	0.09	0.09	0.82	0.11	0.13	0.76	0.08	0.10	0.82

[a]t: time point.

HMM Effects of Detailed Email Prompts

To examine whether the effect of the email prompts differed based on email type, the estimated posterior probability means of the state distribution was formulated depending on whether the participants received and opened an email using detailed email categories (Table 7). On the basis of the posterior probability means shown in Table 7, both positive and negative effects could be identified, where a positive effect reflects an increase in participant activity linked to receiving and opening an email, whereas a negative effect reflects the opposite. A positive effect was associated with opening a welcome email (decreased likelihood of inactivity by 12 percentage points) and opening a health campaign email (decreased likelihood of inactivity by 3 percentage points). A negative effect was associated with opening a newsletter or special offer email (increased likelihood of inactivity by 2 percentage points) and opening a reactivation email (increased likelihood of inactivity by 5 percentage points).

Table 7. Hidden Markov model estimation results: posterior probability means accounting for detailed email prompts[a].

	1: inactivity	2: average activity	3: high activity
No email received	0.68	0.19	0.13
Welcome email			
Email received but not opened	0.70	0.24	0.06
Email received and opened	0.56	0.34	0.10
Reactivation email			
Email received but not opened	0.85	0.11	0.04
Email received and opened	0.73	0.18	0.09
Newsletter email			
Email received but not opened	0.80	0.13	0.07
Email received and opened	0.70	0.18	0.12
Health campaign email			
Email received but not opened	0.75	0.19	0.06
Email received and opened	0.65	0.26	0.09
Special offer email			
Email received but not opened	0.83	0.12	0.05
Email received and opened	0.70	0.19	0.11

[a]Section 3 in Multimedia Appendix 1 presents detailed model fit criteria and parameter estimates for the hidden Markov model. All parameter estimates were statistically significant at the 95% confidence level.

HMM Effects of Time and Background Characteristics

Examining the impact of generic email prompts on participant engagement over time revealed the estimated transition matrices and posterior probability means provided in section 4 in Multimedia Appendix 1. During the first half year, 72% (16,413/22,797) of the participants were estimated to be in the inactivity state, which decreased to 65% (14,818/22,797) during the second half year. For the last half year observed, 69% (15,730/22,797) of the participants were estimated to be in the inactivity state. The impact of receiving and opening a generic email on the transition probabilities did not change much over time, being associated with a decreased likelihood of remaining in the inactivity state by 2 percentage points in the first and third half years and 3 percentage points in the second half year.

The background characteristics of the participants were also linked to differences in activity levels. On the basis of the estimated HMM model (transition matrices and posterior probability means shown in section 5 in Multimedia Appendix 1), female participants were more likely to be in the inactivity state and less likely to be in the high activity state compared with male participants (with a difference of 6 percentage points). The lowest socioeconomic group and the youngest age group were associated with a higher likelihood of inactivity, whereas the age group from 47 to 56 years was the most active. Finally, being an early mHealth app adopter was associated with a 5 percentage point decrease in the likelihood of being in the inactivity state (with the same level of increase in the likelihood of being in the high activity state).

The impact of the generic email prompts on participant engagement did not vary substantially between participants depending on their age, gender, or NSES quintile, with the only difference being that the participants in the oldest age group (67 to 80 years) had a higher likelihood of transitioning toward one of the activity states after opening a generic email than the other age groups (an effect of 4 percentage points).

To examine the robustness of the above-discussed results, several additional models were estimated, with the results confirming those presented in this study. Section 6 in Multimedia Appendix 1 contains several alternative specifications of the HMM model and their estimation results, namely using an indicator for any email received (independent of the email type or targeting nature) and incorporating the email prompts as covariates in the response probabilities model. In addition, following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations, the checklist presented in Multimedia Appendix 2 was completed.

Discussion

Principal Findings

Digital behavior change programs are widely implemented as a means of improving lifestyle behaviors and population health. However, such programs often exhibit low participant engagement rates, with additional effort needed from the program providers to stimulate and maintain engagement. This study analyzed the ability of email prompts to increase participant engagement with an mHealth app aimed at supporting behavior change. Although email prompts showed some positive results in stimulating engagement, their effect in a large-scale, nonexperimental setting with a longer time span is still unclear.

The analysis in this study used an HMM to disentangle the dynamics around email prompts and participant engagement. The estimated HMM with 3 latent states revealed that, on average, 70% (15,958/22,797) of the participants were in the inactivity state, gaining 0 points weekly; 18% (4103/22,797) of the participants were in the average activity state—gaining 27 points weekly; and 12% (2736/22,797) of the participants were in the high activity state—gaining 182 points weekly.

Focusing on the effect of generic emails, the estimation results indicated that when allowing for time dependency, receiving and opening a generic email was associated with a 3 percentage point lower likelihood of remaining inactive than when no email was received (equivalent to a 3 percentage point increase in the likelihood of transitioning from inactivity to one of the activity states). By contrast, receiving but not opening a generic email was associated with a higher likelihood of inactivity. This observed negative impact that receiving but not opening an email had on participant engagement can be explained by the higher proportion of inactive participants in the group that did not open the email than in the group that did not receive an email. Given that the average opening rate of the emails sent within the analyzed mHealth app was above 60%, it can be argued that some increase in participant engagement is possible with the use of generic emails; however, additional effort should be directed at ensuring high opening rates for the emails sent.

Allowing the effect of generic emails to vary over time revealed a relatively stable pattern: in all 3 half-year periods analyzed, the likelihood of moving out of inactivity after opening a generic email was between 2 and 3 percentage points, with the strongest effect corresponding to the second half year of the observation period. Estimating the association of the effects observed with participant background characteristics showed that being a male, being older, having a higher socioeconomic status, and being an early adopter of the mHealth app were all factors associated with higher participant engagement. The impact of the generic emails did not significantly depend on the participants' age, gender, or NSES quintile, with the only difference being that the participants in the age group of 67 to 80 years were more likely to move out of inactivity after receiving and opening a generic email than those in the other age groups.

This study further analyzed the impact of detailed email categories on participant engagement, revealing both negative and positive effects. On one hand, it was estimated that emails welcoming participants to the program and emails containing health-related information were associated with an increase in activity levels (12 and 3 percentage points, respectively). On the other hand, emails that contained generic program information, promoted special offers on products in the web shop, or were aimed at reactivating inactive users were linked to a decrease in activity levels (2, 2, and 5 percentage points, respectively). These findings imply that participants were more reactive to health-related information, which was in line with the program's goals and potentially with the participants' motivation for using the mHealth app, while being resistant to emails that deviated from the program's goal of improving health behaviors. Moreover, the negative effect of the reactivation email implies that it is difficult to stimulate activity in participants who have been inactive for long periods, highlighting the importance of focusing on preventing participants from becoming inactive.

Comparison With Previous Work

Digital behavior change programs often exhibit low participant engagement [11-13]. On the basis of a systematic review, Kelders et al [11] estimated that the average adherence to web-based lifestyle interventions is 23%, which is similar to the 30% of active participants identified in the mHealth app in this study. The slightly higher proportion of active participants estimated here can be because of the mobile app format of the behavior change intervention, which is associated with higher flexibility of use and more personalized content [51].

In an attempt to identify means of improving participant engagement, email prompts have been examined in the context of behavior change programs, with studies reporting small to moderate effects [23]. However, the effects of email prompts on participant engagement are often analyzed over a short period [25,27], subsequently diminishing [12,52] or even disappearing [25]. This study estimated that participant engagement increased by approximately 3 percentage points when an email was received and opened. A similar impact was seen in the study of Ryan et al [52], who based on average activity levels, observed an increase of approximately 3% in the steps taken on the days on which an email was sent. A possible reason for the limited effect of email prompts on engagement is that participants can find such reminders annoying [53]. Alternatively, in the case of targeted emails, inadequate personalization is another factor linked to low participant engagement [54]. Finally, inducing higher participant engagement over time is a challenging task [12,19], partially because of the passive nature of behavior change programs, where participants need to act by themselves to benefit [14].

Participant background characteristics are linked to their engagement with the mHealth app [55]. Similar to the findings in this study, previous work has also shown that males [52,56], older age groups [10,56,57], higher socioeconomic status participants [10,52], and early adopters [38] have higher engagement with mHealth apps. The observation that older age groups are more responsive to emails than other age groups can be explained by their appreciation of reminders within mHealth apps [58], indicating that such tools are especially efficient in increasing activity levels among the older population group.

The results of this study show that although email prompts can achieve a small to moderate increase in participant engagement, this tool alone is likely insufficient for increasing activity levels in an mHealth behavior change app. However, given that the costs associated with email prompts are relatively low, they may be a cost-effective means to improve participant engagement with an mHealth app when the program achieves a high population reach [7,59]. In addition, alternative program efforts, such as expert consultation or real-time feedback [7] could be used next to email prompts to further reduce participant dropout and improve the program effectiveness.

Limitations and Future Research

There are several limitations around this study.

First, as a measure of participant engagement, this study solely used the activities recorded within the mHealth app. However, it is likely that the participants performed additional activities that were not recorded in the app environment. To overcome this limitation, one approach could be to combine the currently used objective measure of participant engagement with an additional subjective measure through which participants themselves can report their perceived activity level. Alternatively, a more accurate measure of activity could be achieved by extending the mHealth app to also include a wearable device, which could measure physical activity in a more precise manner.

Second, it is not a given that whenever a participant opens an email, they become aware of its content. It could be the case that some participants briefly open the email only to delete it, without reading any of its elements. In the setup of this study, we consider the action of opening the email to be a sufficient indication that the participant has been exposed to a reminder about the mHealth program analyzed. As a future extension to the current analysis, it could be of interest to examine whether reading the emails can lead to a higher impact on participant engagement. One possible way to measure whether participants read through the emails could be through the use of a clickable follow-up link, which can help distinguish between participants who pay attention to the content of the email and those who do not.

Third, the participants of the mHealth app analyzed in this study were older than 18 years. This excludes children and teenagers, with unclear insights into how the email prompts would work in increasing activity levels among these population groups. Given the importance of developing healthy lifestyle choices from a young age, a further extension to this study could be examining whether and how email prompts help increase activity among the younger population.

Finally, it is highly likely that the specific wording and topics addressed in an email have an impact on its effectiveness in increasing participant engagement. Although data on the detailed elements of the emails were not available in this study, analyzing such information could be a valuable extension. Namely, it would be of interest to examine how varying phrasing of the same topic and different levels of email personalization affect subsequent participant engagement.

Conclusions

In this study, email prompts were examined as a tool for increasing participant engagement with a large-scale, open-access mHealth app with the goal of lifestyle behavior change. On the basis of an HMM allowing for weekly transitions between latent activity states, it was estimated that receiving and opening an email was associated with a small to moderate increase in participant engagement, which persisted over the 78 weeks analyzed. This finding suggests that email prompts can be used for improving participant engagement, albeit to a limited extent. However, given the relatively low costs associated with emails and the high population reach that mHealth apps can achieve, such instruments can be a cost-effective means of improving participant engagement to reduce dropout and improve the effectiveness of behavior change programs.

Acknowledgments

The authors wish to thank Menzis and SamenGezond for providing the data necessary for this project.

Funding for this project was provided by Menzis (the health insurance company that introduced the mHealth app). The funder had no involvement in the study design, data analysis and interpretation, writing of the report, or decision to submit the report for publication.

Authors' Contributions

All authors contributed equally to the conception and design of the study, data acquisition, data analysis and interpretation, writing and revising the paper, and reading and approving the final version of the submitted manuscript.

Conflicts of Interest

EA was funded by Menzis (the health insurance company that introduced the mHealth app) in her position as a Doctor of Philosophy candidate.

Multimedia Appendix 1
Web-based appendix with additional analysis information.
[DOCX File , 1474 KB - mhealth_v11i1e43033_app1.docx]

Multimedia Appendix 2
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[DOCX File , 32 KB - mhealth_v11i1e43033_app2.docx]

References

1. Noncommunicable diseases: mortality. World Health Organization. URL: https://www.who.int/gho/ncd/mortality_morbidity/en/ [accessed 2022-07-10]
2. Bertram MY, Sweeny K, Lauer JA, Chisholm D, Sheehan P, Rasmussen B, et al. Investing in non-communicable diseases: an estimation of the return on investment for prevention and treatment services. Lancet 2018 May 19;391(10134):2071-2078. [doi: 10.1016/S0140-6736(18)30665-2] [Medline: 29627159]
3. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
4. Blandford A. HCI for health and wellbeing: challenges and opportunities. Int J Human Comput Stud 2019 Nov;131:41-51. [doi: 10.1016/j.ijhcs.2019.06.007]
5. Lehto T, Oinas-Kukkonen H. Examining the persuasive potential of web-based health behavior change support systems. AIS Transact Human Comput Interact 2015 Sep 30;7(3):126-140. [doi: 10.17705/1thci.00069]
6. Chen X, Zhu Z, Chen M, Li Y. Large-scale mobile fitness app usage analysis for smart health. IEEE Commun Mag 2018 Apr;56(4):46-52. [doi: 10.1109/mcom.2018.1700807]
7. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
8. Miller DP, Weaver KE, Case LD, Babcock D, Lawler D, Denizard-Thompson N, et al. Usability of a novel mobile health iPad app by vulnerable populations. JMIR Mhealth Uhealth 2017 Apr 11;5(4):e43 [FREE Full text] [doi: 10.2196/mhealth.7268] [Medline: 28400354]
9. Agachi E, Bijmolt TH, Mierau JO, van Ittersum K. Adoption of the website and mobile app of a preventive health program across neighborhoods with different socioeconomic conditions in the netherlands: longitudinal study. JMIR Hum Factors 2022 Feb 02;9(1):e32112 [FREE Full text] [doi: 10.2196/32112] [Medline: 35107433]
10. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
11. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
12. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Feb 09;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
13. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]

14. Prochaska JO, Velicer WF, Fava JL, Rossi JS, Tsoh JY. Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. Addict Behav 2001;26(4):583-602. [doi: 10.1016/s0306-4603(00)00151-9] [Medline: 11456079]
15. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
16. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
17. Kim M, Yang J, Ahn W, Choi HJ. Machine learning analysis to identify digital behavioral phenotypes for engagement and health outcome efficacy of an mHealth intervention for obesity: randomized controlled trial. J Med Internet Res 2021 Jun 24;23(6):e27218 [FREE Full text] [doi: 10.2196/27218] [Medline: 34184991]
18. Li Y, Guo Y, Hong YA, Zeng Y, Monroe-Wise A, Zeng C, et al. Dose-response effects of patient engagement on health outcomes in an mHealth intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jan 04;10(1):e25586 [FREE Full text] [doi: 10.2196/25586] [Medline: 34982724]
19. Kaptein M, Markopoulos P, de Ruyter B, Aarts E. Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. Int J Human Comput Stud 2015 May;77:38-51. [doi: 10.1016/j.ijhcs.2015.01.004]
20. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
21. Atorkey P, Paul C, Bonevski B, Wiggers J, Mitchell A, Byrnes E, et al. Uptake of proactively offered online and telephone support services targeting multiple health risk behaviors among vocational education students: process evaluation of a cluster randomized controlled trial. J Med Internet Res 2021 Jan 06;23(1):e19737 [FREE Full text] [doi: 10.2196/19737] [Medline: 33404504]
22. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
23. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
24. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
25. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 05;7(9):e12956 [FREE Full text] [doi: 10.2196/12956] [Medline: 31489842]
26. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
27. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
28. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013 Mar 25;15(3):e69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]
29. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
30. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
31. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
32. Silfee VJ, Haughton CF, Jake-Schoffman DE, Lopez-Cepero A, May CN, Sreedhara M, et al. Objective measurement of physical activity outcomes in lifestyle interventions among adults: a systematic review. Prev Med Rep 2018 Sep;11:74-80 [FREE Full text] [doi: 10.1016/j.pmedr.2018.05.003] [Medline: 29984142]
33. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
34. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

35. Greaney ML, Sprunck-Harrild K, Bennett GG, Puleo E, Haines J, Viswanath KV, et al. Use of email and telephone prompts to increase self-monitoring in a Web-based intervention: randomized controlled trial. J Med Internet Res 2012 Jul 27;14(4):e96 [FREE Full text] [doi: 10.2196/jmir.1981] [Medline: 22842775]
36. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012 Mar 01;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
37. Zhu Y, Duan M, Riphagen I, Minovic I, Mierau JO, Carrero JJ, et al. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. Int J Epidemiol 2022 Jan 06;50(6):1959-1969 [FREE Full text] [doi: 10.1093/ije/dyab079] [Medline: 34999857]
38. Dekker L, Rijnks R, Mierau J. The health potential of neighborhoods: a population-wide study in the Netherlands. SSM Popul Health 2021 Sep;15:100867 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100867] [Medline: 34377761]
39. Rogers E. Diffusion of Innovations, 4th Edition. Mumbai, India: Free Press; 2010.
40. Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989 Feb;77(2):257-286.
41. Zucchini W, MacDonald I. Hidden Markov Models for Time Series An Introduction Using R, Second Edition. London: Chapman & Hall; 2016.
42. Netzer O, Ebbes P, Bijmolt T. Hidden Markov models in marketing. In: Advanced Methods for Modeling Markets. Cham: Springer; 2017.
43. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
44. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
45. Hughes JP, Guttorp P, Charles SP. A non‐homogeneous hidden Markov model for precipitation occurrence. J Royal Statistical Soc Series C (Applied Statistics) 2002 Jan 06;48(1):15-30. [doi: 10.1111/1467-9876.00136]
46. Langrock R, Kneib T, Sohn A, DeRuiter SL. Nonparametric inference in hidden Markov models using P-splines. Biometrics 2015 Jun;71(2):520-528. [doi: 10.1111/biom.12282] [Medline: 25586063]
47. DeRuiter SL, Langrock R, Skirbutas T, Goldbogen JA, Calambokidis J, Friedlaender AS, et al. A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure. Ann Appl Stat 2017 Mar 1;11(1):362-392. [doi: 10.1214/16-aoas1008]
48. Li M, Bolker BM. Incorporating periodic variability in hidden Markov models for animal movement. Mov Ecol 2017;5:1 [FREE Full text] [doi: 10.1186/s40462-016-0093-6] [Medline: 28149522]
49. Pohle J, Langrock R, van Beest FM, Schmidt NM. Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. J Agric Biol Environ Stat 2017 Jun 5;22(3):270-293. [doi: 10.1007/s13253-017-0283-8]
50. Vermunt J, Tran B, Magidson J. Latent class models in longitudinal research. In: Handbook of Longitudinal Research. Burlington: Elsevier; 2008.
51. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
52. Ryan J, Edney S, Maher C. Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med 2017 Dec;7(4):702-708 [FREE Full text] [doi: 10.1007/s13142-017-0499-8] [Medline: 28523603]
53. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
54. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
55. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022 Jun 08;24(6):e30817 [FREE Full text] [doi: 10.2196/30817] [Medline: 35675111]
56. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
57. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
58. Lilje SC, Olander E, Berglund J, Skillgate E, Anderberg P. Experiences of older adults with mobile phone text messaging as reminders of home exercises after specialized manual therapy for recurrent low back pain: a qualitative study. JMIR Mhealth Uhealth 2017 Mar 30;5(3):e39 [FREE Full text] [doi: 10.2196/mhealth.7184] [Medline: 28360026]

59. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92 [FREE Full text] [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]

Abbreviations

HMM: hidden Markov model
mHealth: mobile health
NSES: neighborhood socioeconomic status
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 28.09.22; peer-reviewed by P Atorkey, YA Hong; comments to author 20.12.22; revised version received 21.01.23; accepted 08.02.23; published 11.05.23.

Please cite as:
Agachi E, Bijmolt THA, van Ittersum K, Mierau JO
The Effect of Periodic Email Prompts on Participant Engagement With a Behavior Change mHealth App: Longitudinal Study
JMIR Mhealth Uhealth 2023;11:e43033
URL: https://mhealth.jmir.org/2023/1/e43033
doi:10.2196/43033
PMID:37166974

Original Paper

How Notifications Affect Engagement With a Behavior Change App: Results From a Micro-Randomized Trial

Lauren Bell[1,2], MSc; Claire Garnett[3], PhD; Yihan Bao[4], MA; Zhaoxi Cheng[5], BSc; Tianchen Qian[6], PhD; Olga Perski[3], PhD; Henry W W Potts[7], PhD; Elizabeth Williamson[1], PhD

[1]Department of Medical Statistics, The London School of Hygiene and Tropical Medicine, London, United Kingdom
[2]Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom
[3]Research Department of Behavioural Science and Health, University College London, London, United Kingdom
[4]Department of Statistics and Data Science, Yale University, New Haven, CT, United States
[5]Department of Biostatistics, Harvard University, Cambridge, MA, United States
[6]Department of Statistics, University of California Irvine, Irvine, CA, United States
[7]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Henry W W Potts, PhD
Institute of Health Informatics
University College London
222 Euston Road
London, NW1 2DA
United Kingdom
Phone: 44 02035495303
Email: h.potts@ucl.ac.uk

Abstract

Background: *Drink Less* is a behavior change app to help higher-risk drinkers in the United Kingdom reduce their alcohol consumption. The app includes a daily notification asking users to "Please complete your drinks and mood diary," yet we did not understand the causal effect of the notification on engagement nor how to improve this component of *Drink Less*. We developed a new bank of 30 new messages to increase users' reflective motivation to engage with *Drink Less*. This study aimed to determine how standard and new notifications affect engagement.

Objective: Our objective was to estimate the causal effect of the notification on near-term engagement, to explore whether this effect changed over time, and to create an evidence base to further inform the optimization of the notification policy.

Methods: We conducted a micro-randomized trial (MRT) with 2 additional parallel arms. Inclusion criteria were *Drink Less* users who consented to participate in the trial, self-reported a baseline Alcohol Use Disorders Identification Test score of ≥8, resided in the United Kingdom, were aged ≥18 years, and reported interest in drinking less alcohol. Our MRT randomized 350 new users to test whether receiving a notification, compared with receiving no notification, increased the probability of opening the app in the subsequent hour, over the first 30 days since downloading *Drink Less*. Each day at 8 PM, users were randomized with a 30% probability of receiving the standard message, a 30% probability of receiving a new message, or a 40% probability of receiving no message. We additionally explored time to disengagement, with the allocation of 60% of eligible users randomized to the MRT (n=350) and 40% of eligible users randomized in equal number to the 2 parallel arms, either receiving the no notification policy (n=98) or the standard notification policy (n=121). Ancillary analyses explored effect moderation by recent states of habituation and engagement.

Results: Receiving a notification, compared with not receiving a notification, increased the probability of opening the app in the next hour by 3.5-fold (95% CI 2.91-4.25). Both types of messages were similarly effective. The effect of the notification did not change significantly over time. A user being in a state of *already engaged* lowered the new notification effect by 0.80 (95% CI 0.55-1.16), although not significantly. Across the 3 arms, time to disengagement was not significantly different.

Conclusions: We found a strong near-term effect of engagement on the notification, but no overall difference in time to disengagement between users receiving the standard fixed notification, no notification at all, or the random sequence of notifications within the MRT. The strong near-term effect of the notification presents an opportunity to target notifications to increase "in-the-moment" engagement. Further optimization is required to improve the long-term engagement.

International Registered Report Identifier (IRRID): RR2-10.2196/18690

(JMIR Mhealth Uhealth 2023;11:e38342) doi:10.2196/38342

KEYWORDS

mobile health; mHealth; digital health; behavior change; behavior change; digital behavior change; engagement; micro-randomized trial; randomized trial; randomization; just-in-time adaptive intervention; adaptive intervention; push notification; notification; excessive alcohol consumption; smartphone app; alcohol; drinking; drinker; mobile phone

Introduction

Background

Hazardous and harmful alcohol consumption is one of the major risk factors for many disease outcomes and poses a major public health burden [1,2]. Delivering brief interventions to reduce hazardous and harmful alcohol consumption is known to be effective [3]; however, such efforts are challenged by the sheer prevalence of harmful drinking and limited capacity of services [4,5]. There is a long-standing recognition of the need to broaden the reach of and access to brief, effective interventions to reduce harmful alcohol consumption for help-seeking individuals [6].

A promising solution is behavior change apps, as these are complex interventions that can capture dynamic patterns in human behavior and deliver support when an individual needs this the most [7-9]. Building on evidence that supports SMS text messaging as interventions to help individuals [10], behavior change apps can provide comprehensive, everyday support within people's homes and diverse communities to maintain healthy behaviors [11]. However, a major concern is that insufficient engagement with an app is likely to hinder behavior change, particularly if a user disengages with the app soon after downloading it [12,13]. Engagement, a construct of both experiential and behavioral aspects [14], fluctuates within and between users over time and is influenced not only by the static content of the intervention but also by internal (eg, the user's momentary mood, cognitive state, and recent patterns of engagement and drinking) and external (eg, the user's current environment) factors [15-18].

Push notifications (reminders or pop-up messages on the screen) are often implemented to increase engagement with a behavior change app [13,19,20] and can have small, positive effects on engagement over a 24-hour period [21]. However, a more immediate causal effect (eg, within the next hour) of a push notification on engagement with behavior change apps has not yet been established [21,22]. We undertook a trial to estimate the causal effect of the notification on near-term engagement in the behavior change app *Drink Less* and to consider how the notification policy could be further optimized to improve engagement.

The *Drink Less* App

Drink Less is a behavior change app that aims to help higher-risk drinkers in the UK adult population reduce their alcohol consumption. The app is freely available to people seeking help with their alcohol consumption, although it has not been advertised or targeted to specific groups of people. *Drink Less* was developed in line with the Medical Research Council guidelines for developing and evaluating a complex intervention [23-25] and the Multiphase Optimisation Strategy (MOST) framework [26,27] and is freely available on the Apple App Store. *Drink Less* is an evidence- and theory-informed intervention with several modules. The overall development and refinement of *Drink Less*, including how the behavior change modules were selected, can be found in previous publications [28,29]. The standard version of the app delivers a local daily notification at 11 AM, asking the user to "Please complete your mood and drinks diary" (Multimedia Appendix 1 provides a visual of the *Drink Less* notification). Daily notifications aim to remind users to self-monitor their drinking habits. The National Institute for Health and Care Excellence for the United Kingdom recommends self-monitoring as an effective technique for the act of noticing recent behavior and how this relates to their goals [30]. However, if a user has already engaged with the app to self-monitor their drinking that day, the notification may be an unnecessary reminder and may ultimately annoy the user over time.

The notification appears on the users' notification center, and tapping the notification opens to the *Drink Less* landing page. The standard version of *Drink Less* sends a daily notification that aims to increase self-monitoring by tracking recent alcohol units consumed (ie, the day before). The delivery time at 11 AM allows users to complete their morning routines before engaging with the app. User feedback was received via the App Store, with the suggestion that a reminder to report drinking diaries in the evenings would be more helpful.

Engagement With Drink Less

We previously reported exploratory research that visualized temporal patterns of engagement with *Drink Less* [31]. The visualizations showed limited depth of engagement, with 85% of sessions occurring within the Self-Monitoring and Feedback module, and a natural peak near 8 PM of both frequency (ie, number of log-ins) and time spent on the app observed in the evenings. This suggested that evenings are opportune moments to engage with *Drink Less* for longer sessions. In the evening, users may be more susceptible to harmful drinking and intervening at this moment of vulnerability and, at an opportune moment to engage, may be more conducive to reducing harmful patterns of drinking. In addition, our exploratory research discovered different trajectories of use, with 50% of users disengaging with the app 22 days after download. We hypothesized that a fixed notification policy may suit some users to maintain engagement, whereas other users may habituate to the daily notification policy and disengage sooner.

Specific Aims and Objectives

We conducted a micro-randomized trial (MRT), a design in which users recruited for the MRT were repeatedly randomized to notifications over time, with outcomes measured after each randomization [32-36]. We aimed to provide evidence of how notifications affect near-term engagement as well as to consider further improvement of the push notification policy. The primary objective was to assess whether sending a notification at 8 PM increases behavioral engagement (opening the app) in the subsequent hour with *Drink Less*. The secondary objectives included the comparison of 2 different types of notifications. notifications. We also explored effect moderation by time and the exploration of effect moderation by user context (with context being a user's dynamic state of engagement or habituation). We also aimed to understand the role of a notification policy more generally for time to disengagement. In addition, we aimed to compare 3 policies on time to disengagement (each policy being the decision rule of delivering notifications used in 1 of the 3 arms). This is the first step in our wider aspiration to optimize the notification policy of *Drink Less*, an aspiration we return to in the *Discussion* section.

Methods

Trial Design

Our study had a 30-day MRT with 2 additional parallel arms. Three different notification policies are implemented in the 2 arms and MRT to address the secondary objectives: (1) a standard policy of sending a daily message of *"Please complete your mood and drinks diary"* sent at 11 AM; (2) the MRT, a random policy that varies the content and sequence of the notifications; and (3) a no notification policy, a policy that does not send notifications. For the secondary objectives, the three policies are referred to as (1) standard notification policy, (2) random notification policy, and (3) no notification policy.

In total, 60% of eligible users were randomized to the MRT, and 40% of eligible users were randomized in equal number to the 2 parallel arms, either receiving the no notification policy or the standard notification policy of *"Please complete your mood and drinking diary"* at 11 AM.

For users randomized to the MRT, each user was randomized daily at 8 PM to receive 1 of the 3 options: no notification (with 40% probability), the standard message (with 30% probability), or a notification randomly selected with replacement from a bank of new messages (with 30% probability).

Following our MRT protocol [37] and the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines [38], we reported the primary and some secondary results here.

Participants

The recruitment period ran from January 2, 2020, to April 1, 2020. *Drink Less* is freely available on the Apple App Store, and individuals who downloaded the app during the recruitment period were eligible to participate in the trial if they self-reported a baseline Alcohol Use Disorders Identification Test (AUDIT) score of ≥8, which indicates excessive alcohol consumption [39]; resided in the United Kingdom; were aged ≥18 years; and reported being interested in drinking less alcohol.

The app prompted eligible users to read the privacy notice (Multimedia Appendix 2) and participant information sheet (Multimedia Appendix 3) before enrolling in the trial. During the informed consent process, users were informed that they could opt out of the trial at any time and that they would receive the standard version of the *Drink Less* app if they withdrew their consent.

The date of download is defined as the date when the onboarding process is completed by each user. The onboarding process involved users completing a registration section where they completed the AUDIT and sociodemographic assessment and then received normative feedback (personalized feedback on how their drinking compares with the behaviors of others). It is only after the completion of the onboarding process that users were then assessed for eligibility and consequently randomized to 1 of the 3 arms.

On enrollment in the study, we turned the permission function off within the app. This was intended to ensure that the participants received the notification policy to which they were randomized. Participants could, however, go into the settings and turn the notification policy off, which is applicable for all apps on the Apple App Store and is beyond the control of any app developer.

Data

Preprocessing of the original use data was required. The raw engagement data are captured by a series of screen views, comprising time stamps of when a new screen is opened in the app. Clearing or *swiping away* the notification is not registered as any use [40]. The length of a session is calculated as the difference (in microseconds) between the first and last screen views, with a new session defined after 30 minutes of inactivity between screen views [41]. This method of calculating the length of sessions means that our measures of the length of time spent on the app are always underestimated because we do not know how long the user observes the last screen view [41]. We did not impose a threshold on our outcome (in terms of the amount or depth of app use), so simply opening the app is measured as engagement. When a user opens *Drink Less*, they are presented with a dashboard with various information about their drinking habits as well as a toolbox of features to access if they want. As such, simply opening the app and viewing the dashboard and toolbox present an opportunity for users to benefit from engaging with *Drink Less*. All time stamps were appropriately adjusted from Coordinated Universal Time to British Summer Time.

Time-Fixed Measures (Baseline)

Time-fixed covariates, measured at baseline, were age, sex, type of employment (manual, nonmanual, or other), and baseline AUDIT score (0-40) [39,42,43]. The AUDIT risk zones were hazardous (8-15), harmful (16-19), and at risk for alcohol dependence (20-40). The participants self-selected the employment status they identified with for the options they were provided. They were not provided with a definition of employment type.

Time-Varying Measures

Time-varying engagement measures within the MRT are the time stamps of when the user opens the app and the length of time (in seconds) spent on the app. This includes the time-varying variables: (1) “Did the user open the app before 8 PM on day of randomization? (yes/no)” and (2) “Did the user open the app any time after 9 PM the day before? (yes/no).”

Time-varying covariates, used as part of post hoc analyses to explore effect moderation, were “habituation” and “already engaged.” “Habituation” was captured using the binary measure “Did the user receive a notification the day before? (yes/no).” “Already engaged” was captured using the binary measure “Did the user open the app between 8 PM-9 PM the day before? (yes/no).”

Interventions

The MRT tested 2 notification types. This trial tested the existing notification with the message of *“Please complete your mood and drinking diary”* and a new notification bank of 30 novel messages (Multimedia Appendix 4). The development of the new notification bank was informed by research with *Drink Less*, which found that the perceived usefulness of the app (the belief that using the app will help the user achieve their goal or goals and an indicator of users’ reflective motivation to engage) was associated with increased engagement for some users. Therefore, the new bank of notifications was designed (with feedback on the content sought from a group of behavioral scientists) to increase users’ reflective motivation to engage with a particular intervention module [44]. All messages contained the phrase “(using a particular module in the app) *can help you drink less*.” Examples include *“Recording if-then plans can help you drink less”* and *“Setting a doable goal can help you drink less. Take a moment to set a doable goal*.” The notification does not lock the user’s screen, and there is no expiry time for notification.

Outcomes

The primary outcome was whether the user opened the app (yes or no) in the hour between 8 PM and 9 PM, following the randomization of receiving a notification at 8 PM. This is a time-varying, binary, and near-term measure of engagement.

We also defined a post hoc outcome of whether the user opened the app between 8 PM on the day of randomization to 8 PM the following day to explore the effect over a 24-hour period.

The secondary outcomes captured across the 3 different policies include the number of days to disengagement, with disengagement defined as the first day in a period of ≥7 consecutive days of no use.

Sample Size

We powered the MRT for the important secondary objective of effect moderation over time, which guarantees at least as much power for the primary objective of detecting a marginal effect. Using a simulation informed by observational *Drink Less* data [31], we determined that a sample of 1200 users was sufficient to provide 80% power, with a type 1 error of 5%, to detect effect moderation over time, assuming a marginal notification effect of 2.16 decaying by 0.911 per day since download (Multimedia Appendix 5). We powered the secondary arms, implementing different notification policies, to detect a minimum absolute difference in time to disengagement of 10%, assuming 55% disengagement by day 22 under the standard policy compared with 65% under the no notification policy. To achieve 80% power with type 1 error of 5%, 372 users were required to receive each notification policy. This was rounded to 400 to simplify the randomization process. Overall, we aimed to recruit 1200 users to the MRT, 400 users to the standard notification policy, and 400 users to the no notification policy.

Previous download trends revealed, on average, an estimate of at least 33 eligible users per day who downloaded *Drink Less* and consented to the privacy notice. We expected the available recruitment window (January 2 to April 1, 2020) to be sufficient to reach our recruitment target of 2000 users. However, we fell short of this target of 2000 users and randomized 598 users in total for three reasons: (1) a large proportion of users (40%) did not give their informed consent to be part of the study; (2) the number of downloads, particularly for March 2020, was less than predicted, based on 2019 trends; and (3) extending the recruitment period to achieve the desired sample size was not possible because of the commencement of a prescheduled National Institute for Health Research–funded randomized controlled trial [45]. Consequently, the primary objective was sufficiently powered; however, we did not achieve the prespecified sample size for the secondary objectives of effect moderation over time and time to disengagement.

Randomization

Simple randomization (unstratified and no blocking) was used. An external engineer generated the randomization sequence and coded it into the app. Together, 2 coauthors (LB and CG) pilot tested the randomization schedule. To further verify the randomization process, 10 volunteers also participated in a pilot test. The 10 volunteers were randomized into 3 different arms and asked to record the notifications received and the use of the app. We confirmed that the randomization process functioned as planned, and all uses were correctly captured.

Statistical Methods Within the MRT

Descriptive statistics (frequency distributions and measures of central tendency) were used to describe the baseline variables of participants.

The primary outcome, within the MRT, was summarized separately for the standard notification, new notification, and no notification by the number of person-days where the app was opened between 8 PM and 9 PM then divided by the number of person-days in the MRT and expressed as a proportion.

The near-term effect of the notification on the primary outcome was expressed on the relative risk (RR) scale and pooled over the longitudinal data across all participants using the Estimator for the Marginal Excursion Effect (EMEE) [46]. The EMEE was developed to estimate the causal effects of time-varying treatments with binary outcomes. The EMEE does not require the correct specification of the marginal mean model (ie, how the time-varying engagement depends on a user’s time-varying contexts), providing robustness to highly complex and stochastic engagement patterns.

The effect of receiving a push notification versus not receiving a notification was estimated overall and then separately for the new message bank and standard notification. All models from the MRT were adjusted for the continuous variables of age, AUDIT score, days since download, the categorical variables of sex and employment type, and the time-varying variables “Did the user open the app before 8 PM that day?” and “Did the users open the app after 9 PM the day before?” The time-varying measures were included to increase the precision of our near-term notification effect, as they were likely highly correlated with the outcome. The covariates of “habituation” and “already engaged” are for the purpose of exploring how these recent states modify the near-term effect of the notification.

Statistical Methods Across Arms

Baseline descriptive statistics and measures of use—the median number of sessions per user and the median length of sessions (seconds)—were reported across the 3 policies. A user was classified as having disengaged on the first day of a period of 7 consecutive days of no use. This outcome was only defined for the first 23 days after follow-up, which lasted 30 days in total. Survival curves were plotted using the Kaplan-Meier estimator [47] and compared using a log-rank test.

Owing to technical glitches, there were some unanticipated missing categorical baseline data. We reported the number of missing values for each arm. We used modal imputation for the baseline variables. To assess the sensitivity of our conclusions to our missing data approach, we imputed the data with the second most common value.

All analyses were conducted using R (version 4.0.5; R Foundation for Statistical Computing) [48] with the *dplyr* [49], *lubridate* [50], *gtsummary* [51], *zoo* [52], *ForImp* [53], and *survminer* [54] packages.

Ethics Approval

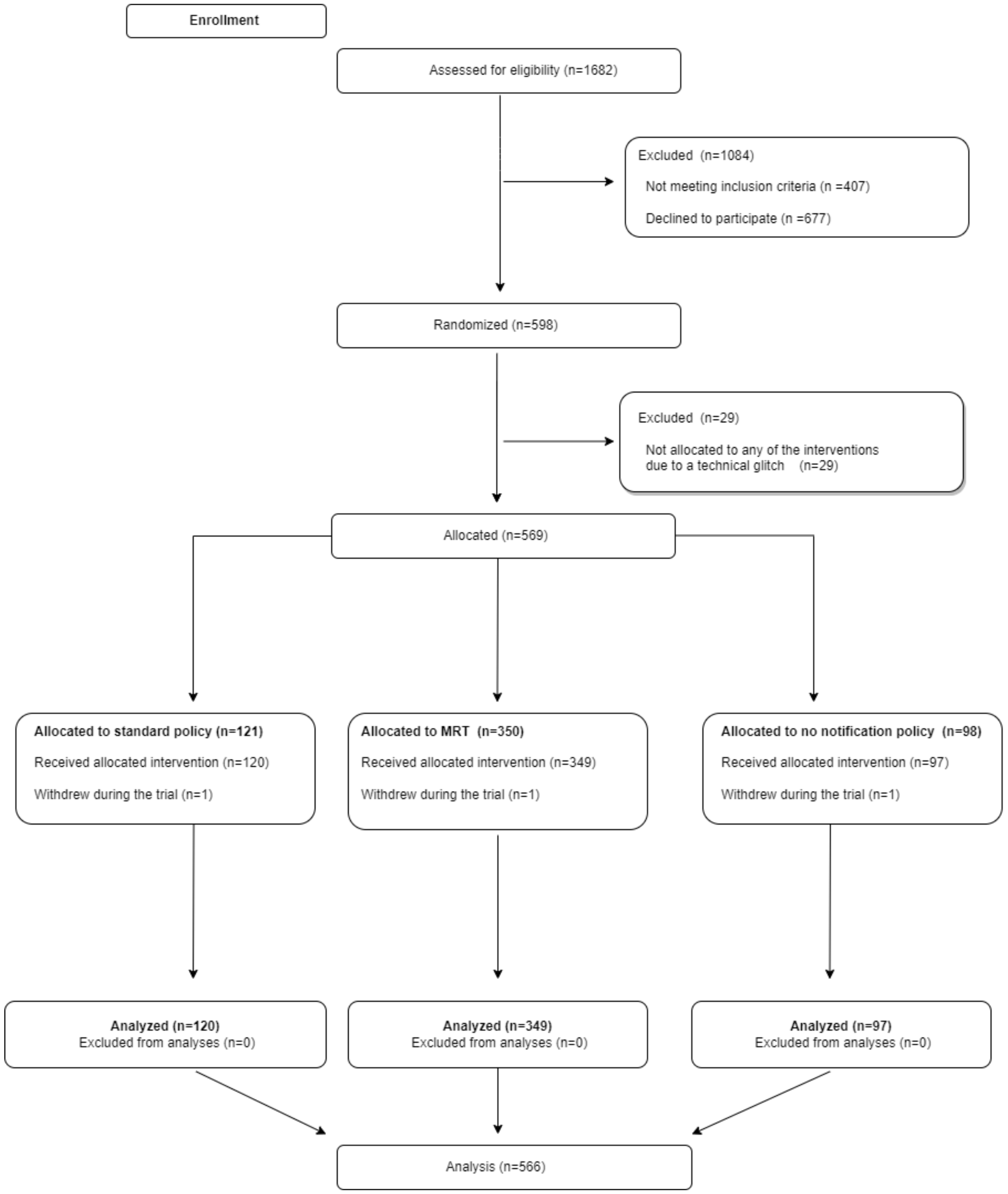
Ethical approval for this study was granted by the University College of London’s Departmental Research Ethics Committee (CEHP/2016/556) on October 11, 2019, and The London School of Hygiene and Tropical Medicine Interventions Research Ethics Committee (17929) on November 27, 2019.

Results

Overview

The anonymized data sets, including data dictionaries, are publicly available [55]. The code for EMEE is openly available on the GitHub account [56]. Figure 1 presents the CONSORT flow diagram of the progress through the MRT.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the trial. MRT: micro-randomized trial.



Recruitment

The trial recruitment period ran from January 2, 2020, to April 1, 2020 (app version 2.0.1). We analyzed a total of 566 users. The mean age was 44 (SD 12) years, with 43.6% (247/566) male and 45.8% (259/566) female users. A total of 62.4% (353/566) of users reported being in nonmanual employment, 12.5% (71/566) reported being in manual employment, and 14.5% (82/566) reported being in other employment types. A total of 48.8% (276/566) of users reported hazardous alcohol consumption, 20.1% (114/566) reported harmful alcohol consumption, and 31.1% (176/566) reported being at risk of alcohol dependence. A total of 68.2% (386/566) of users were disengaged by day 23 or earlier.

Data on sex and employment type were not recorded for the 60 users (standard arm, n=5; MRT, n=40; and no notification arm, n=5). We used modal imputation to set these missing values of sex to female and employment type as nonmanual (Table 1).

Table 1. Description of trial participants by randomized arm.

User characteristics	Baseline summary		
	Standard arm (n=120)	MRT[a] (n=349)	No notification arm (n=97)
Age (years), median (IQR; n=566)	45 (35-55)	43 (34-51)	43 (34-52)
Sex (n=506; standard arm: n=105; MRT: n=309; no notification arm: n=92), n (%)[b]			
Male, n (%)	43 (41)	155 (50.2)	49 (53)
Female, n (%)	62 (59)	154 (49.8)	43 (47)
Employment type (n=506; standard arm: n=105; MRT: n=309; no notification arm: n=92), n (%)[b]			
Nonmanual, n (%)	66 (62.8)	224 (72.5)	63 (68)
Manual, n (%)	19 (18.1)	37 (12)	15 (16)
Other, n (%)	20 (19)	48 (15.5)	14 (15)
AUDIT score (n=566), n (%)			
Hazardous (8-15)	48 (40)	142 (40.7)	49 (51)
Harmful (16-19)	29 (24.2)	84 (24.1)	18 (19)
At risk of alcohol dependence (20-40)	43 (35.8)	123 (35.2)	30 (31)

[a]MRT: micro-randomized trial.

[b]Missing: standard arm: n=15; MRT: n=40; and no notification arm: n=5.

Outcomes and Estimation

In the MRT, 349 users were randomized each day for 30 days, resulting in 10,470 measurements for the primary outcome. There were 30.05% (3146/10,470) of measurements for the new message, 29.72% (3112/10,470) of measurements for the standard notification, and 40.23% (4212/10,470) of measurements for no notification. The proportion of the primary outcome (opening the app between 8 PM and 9 PM) was 0.122 for the new message, 0.131 for the standard message, and 0.036 for no message. For the post hoc 24-hour outcome (from 8 PM to 8 PM the next day), the proportion of opening the app was 0.351 for the new message, 0.342 for the standard message, and 0.280 for no message.

Main Results

Table 2 provides the results for the estimate of the near-term effect of the notifications on engagement. This demonstrates that, on average, the probability of opening *Drink Less* within the hour of receiving a notification increased 3.52-fold (95% CI 2.91-4.25). The 2 different notification types have similar effects, with the probability of opening *Drink Less* within the hour of receiving a standard notification increasing 3.66-fold (95% CI 2.99-4.48) and the probability of opening the *Drink Less* within the hour of receiving a new notification increasing 3.39-fold (95% CI 2.77-4.13).

Table 3 reports the results for how the effect of the notification changes over the first 30 days since download. We did not detect any significant changes over time, with an estimated decay by a factor of 0.993 per day with a 95% CI (0.975-1.012). Although not statistically significant, the point estimate of the decay in effect over time is larger in magnitude for the standard notification compared with the new notification; however, the wide CIs reflect large uncertainties in these estimates.

The summative engagement measures (median number of sessions and median length of sessions) under the 3 policies had a similar number of sessions and length of sessions over the first 30 days since download (Table 4). Users randomized to the no notification policy had, on average, slightly fewer but longer sessions, with a median of 43 seconds in length compared with 36 seconds for the standard policy and 35 seconds for the random policy (MRT).

Table 2. Primary objective—estimated near-term notification effect.

Notification type[a]	Relative risk (95% CI)
Pooled notifications (both standard and new)	3.523 (2.918-4.255)
Standard notification	3.664 (2.993-4.485)
New notification	3.385 (2.774-4.131)

[a]Adjusted for the continuous variables of age, Alcohol Use Disorders Identification Test score, days since download, the categorial variables of sex and employment type, and the time-varying variables “Did the use user the app before 8 PM that day?” and “Did the users use the app after 9 PM the day before?”

Table 3. Change of near-term notification effect over time.

Notification type[a]	Relative risk on the first day after download (95% CI)	Multiplicative change in effect (95% CI)
Pooled notifications (both standard and new)	3.849 (2.811-5.270)	0.993 (0.975-1.012)
Standard notification	4.193 (3.004-5.854)	0.989 (0.970-1.001)
New notification	3.534 (2.536-4.924)	0.997 (0.976-1.017)

[a]Multiplicative change in relative risk per day since download. Adjusted for the continuous variables of age, Alcohol Use Disorders Identification Test score, days since download, the categorical variables of sex and employment type, and the time-varying variables “Did the use user the app before 8 PM that day?” and “Did the users use the app after 9 PM the day before?”

Table 4. Median number of sessions per user and median length of sessions (seconds), across 3 arms for the first 30 days since download.

Policy implemented within arm	Number of sessions, median (IQR)	Length of sessions (seconds), median (IQR)
Standard policy	10.5 (4-23)	36 (9-115)
Random policy (MRT[a])	9 (3-27)	35 (9-113)
No notification policy	6 (3-21)	43 (12-129)

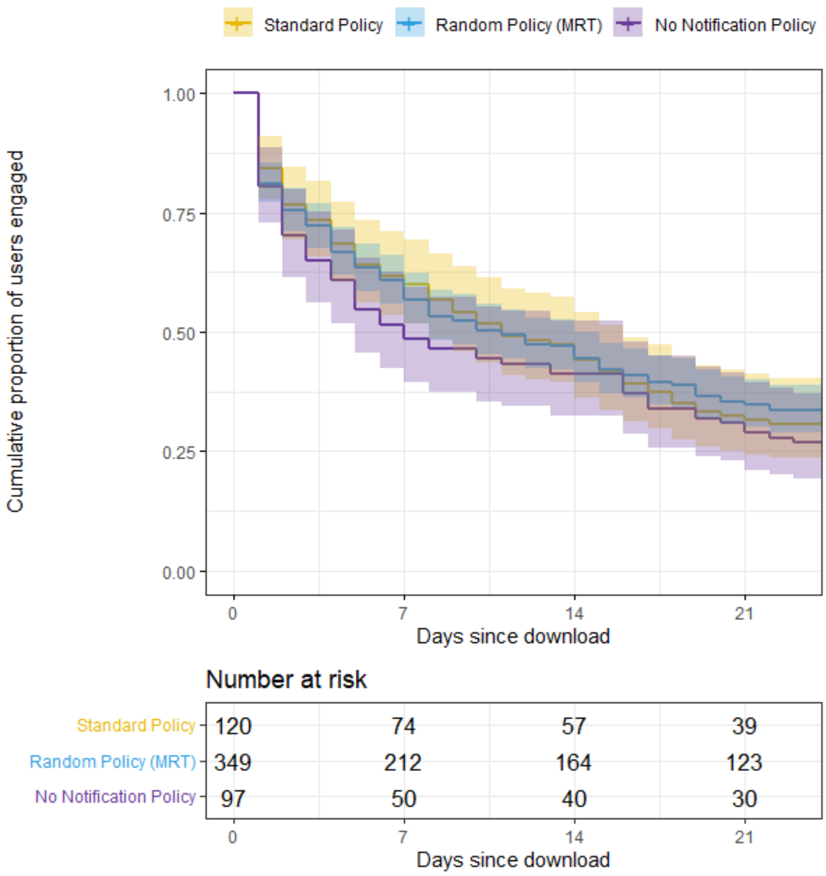
[a]MRT: micro-randomized trial.

Time to Disengagement—Survival Analysis

The median time to disengagement was 11 days for users randomized to the standard policy, 11 days for those randomized to the random policy (MRT), and 7 days for those randomized to the no notification policy. The number of disengaged users was 83 for the standard policy, 232 for the random policy (MRT), and 71 for the no notification policy. The log-rank χ^2_2 test statistic is 1.7, and the corresponding *P* value is .42.

Figure 2 presents the Kaplan-Meier plot of time to disengagement across the 3 notification policies. This plot provides the estimated survival fraction over the first 30 days from the date of download between the 3 policies with 95% CIs. Although the survival fraction of the no notification policy may seem to accelerate at a faster rate than the other policies over the first week, all 3 policies had similar survival rates by day 23.

Figure 2. Kaplan-Meier plot of time to disengagement across the 3 notification policies. MRT: micro-randomized trial.



Ancillary Analyses

Table 5 reports the estimates of the moderation of near-term notification effect, by recent engagement states, calculated from the MRT. We reported the estimated near-term notification effect for both message types pooled, and by notification type, and the effect is separated out by recent engagement states of habituation (yes or no) and recently engaged (yes or no). We reported the multiplicative difference in the notification effect, which demonstrates how the recent states of engagement modify the notification's near-term effect. The near-term effect of both message types remains, and none of the estimates of effect moderation are statistically significant because of a limitation of the study lacking in power. If a user received a notification the day before, the near-term notification effect of a standard message is reduced by 11% (RR 0.889, 95% CI 0.60-1.31), whereas the effect of the new notification remains stable (RR 1.013, 95% CI 0.68-1.52). If a user is "already engaged" (they opened the app between 8 PM and 9 PM the day before), the near-term effect of the standard notification remained relatively stable (RR 0.97, 95% CI 0.65-1.42), whereas the near-term effect of the new notification decreased by 20% (RR 0.80, 95% CI 0.55-1.17).

The reported near-term notification effect for a 24-hour period is presented in Table 6. This demonstrates that notifications increase the probability of opening the app by 1.3-fold over a 24-hour period. Both notification types have a similar magnitude of effect.

Table 5. Estimated effect moderation by recent habituation and engagement.

Notification type[a]	Estimated notification effect		Estimated effect moderation
	Relative risk (95% CI)		Multiplicative difference in effect (95% CI)
	No	Yes	Ratio (yes or no)
Habituation—"Did the user receive a notification the day before?"			
Pooled	3.645 (2.665-4.987)	3.449 (2.761-4.308)	0.946 (0.655-1.368)
Standard	3.935 (2.837-5.458)	3.497 (2.739-4.465)	0.889 (0.601-1.314)
New message	3.357 (2.381-4.732)	3.401 (2.692-4.298)	1.013 (0.677-1.517)
Already engaged—"Did the user open the app between 8 PM and 9 PM the day before?"			
Pooled	3.620 (2.908-4.508)	3.168 (2.314-4.336)	0.875 (0.616-1.242)
Standard	3.720 (2.938-4.710)	3.580 (2.542-5.043)	0.962 (0.652-1.420)
New message	3.521 (2.794-4.439)	2.820 (2.015-3.946)	0.801 (0.548-1.168)

[a]All models are adjusted for the continuous variables of age; Alcohol Use Disorders Identification Test score; days since download; the categorical variables of sex and employment type; the time-varying variables "Did the users open the app before 8 PM that day?" and "Did the users open the app after 9 PM the day before?"; and the effect moderation variable of habituation.

Table 6. Post hoc analysis—estimated near-term notification effect, defined over 24 hours (from 8 PM to 8 PM the next day).

Notification type[a]	Relative risk (95% CI)
Pooled	1.260 (1.187-1.337)
Standard	1.245 (1.161-1.336)
New message	1.274 (1.193-1.360)

[a]Adjusted for the continuous variables of age, Alcohol Use Disorders Identification Test score, days since download, the categorial variables of sex and employment type, and the time-varying variables "Did the use user the app before 8 PM that day?" and "Did the users use the app after 9 PM the day before?"

Discussion

Principal Findings

We have shown that for *Drink Less*, there is a large near-term (3.5-fold) positive effect on engagement. The near-term notification effect for either the standard message type or a message from the new bank has similar effects on increasing engagement in the subsequent hour. Over a 24-hour period, a smaller, significant effect (1.3-fold) remained. We did not detect a significant change in the effects of notifications over time. The effect of receiving a new message that aims to reengage users was not significantly reduced by 20% if the user was already engaged. Furthermore, the effect of receiving a standard message was not significantly reduced by 12% if the user received a notification the previous day. There was no significant difference in (1) the mean number of days to disengagement, (2) the number of sessions, and (3) the length of sessions across the 3 different notification policies. However, a slightly longer median length of time was observed for a session under the no notification policy. One might hypothesize that unprompted behavioral engagement may include more attentive interest and cognitive investment.

In our study, despite evidence of a large positive notification effect on near-term engagement, an overall policy of sending a fixed daily notification or a random mix of notifications did not lengthen the time to disengagement or increase the amount of engagement during the first 30 days of since download. The results of the effect moderation analyses, although requiring confirmation in larger studies, suggest that notifications may be better served as dynamic interventions that adapt to a user's fluctuating patterns of engagement. For example, by sending a notification to users when they are at an increased risk of disengagement, targeting them at that point with a notification intended to increase their perception of the usefulness of the app.

Future Research to Optimize the Notification Policy

Our study demonstrated that for *Drink Less*, notification increases near-term engagement. This finding offers the opportunity for behavior change scientists to directly target the precise momentary states of an individual and to develop and implement dynamic theories for behavior change with *Drink Less*.

Efforts to maintain or increase engagement through consistent notifications could overburden or annoy a user, resulting in a state of disengagement with interventions from a previously motivated user [19]. Our findings suggest that the optimal role of notifications in improving long-term engagement is unlikely to be fixed or random components but better placed as dynamic components (ie, varying not randomly but in response to the user's changing state of engagement and habituation).

The open question now is when do we program notifications to be sent to balance goals of (1) intervening for maximum therapeutic effect based on a user's internal history with *Drink Less* and external environmental factors; and (2) avoiding states of disengagement because of the burden of unhelpful notifications. To begin to answer this question, we will undertake further modeling of this MRT data to explore the within- and between-user effect of the notification over time and the balance of near-term and long-term effects. We will further analyze the data to understand if cue-to-action messages resulted in the task and to determine if the suggested module was engaged with. We imagine that a further optimized policy would (1) keep more users in a state of engagement for longer by sending fewer notifications than the random or fixed notification policies tested here, (2) have a higher near-term notification effect, and (3) ultimately improve the effectiveness of *Drink Less*. A type of machine learning called reinforcement learning may be helpful to personalize and optimizing the sequence of notifications over time [54,57,58]. The available data from our trial can provide a rich source of information to help guide the initial steps (ie, provide a "warm-start") of the learning process of a reinforcement learning algorithm to improve engagement for *Drink Less* or other similar behavior change apps [57,59-61].

Limitations

Overview

Our study was sufficiently powered for the primary objective, to detect a near-term notification effect. However, because we did not achieve our planned sample size, the important secondary objectives of effect moderation over time and time to disengagement between policies were not adequately powered. This resulted in wide CIs and large *P* values for the effect moderation analyses, leaving uncertainties about the existence and magnitude of these effects for the secondary objectives. Further studies with larger sample sizes are required to explore these effects.

There were missing data for a minority of the baseline values for sex and employment type, although our sensitivity analyses showed that the results were not sensitive to how the missing values were imputed.

The values entered for alcohol units consumed as diary entries were deemed too noisy to represent alcohol consumption over time for reasons of bias, extensive missing data, and backfilling (ie, users bulk reporting their drinking outcomes days later). Owing to the priority of not overburdening users with too many notifications sent within a day, our research does not provide a comparison of the near-term effect of the notification for different times of the day.

Generalizability

The recruitment period was from January 2 to April 1, 2020, which began with a typical surge in downloads in the new year and ended during the United Kingdom's first COVID-19 lockdown. Such exogenous shocks to the users' overall environment during the trial are likely to influence the underlying thoughts, emotions, and behaviors of reducing drinking levels (ie, people were mainly housebound) [62] and hence impact the patterns of engagement with *Drink Less*. The interpretation of the results is an average over this period only, with most of the recruitment occurring before the widespread outbreak in the United Kingdom (Multimedia Appendix 6). We also noted that the median time to disengagement in the standard policy arm (11 days) was much sooner than our data visualization cohort experienced (22 days; [31]).

Conclusions

We found a large causal effect of sending notifications on near-term engagement. The probability of opening the app in the immediate hour increased 3.5-fold when receiving a notification compared with not receiving a notification. Notifications are important and effective components of behavior change apps; however, a policy of sending a fixed daily notification or a randomly chosen series of notifications did not increase the amount of engagement or length of time to disengagement for users compared with a policy of no notifications. This suggests that notifications may serve users better when they are implemented as dynamic components, such as sending a notification to increase the perceived usefulness of the app only when the users' engagement patterns show that they are at risk of disengaging.

Further optimization of the notification policy is required to improve long-term engagement. The next stage of research is to explore how our findings would help develop a policy for *Drink Less* to intervene when a user is likely to benefit from support and keep more users engaged in the first 30 days after download.

Acknowledgments

The development of the *Drink Less* app was funded by the National Institute for Health Research (NIHR) School for Public Health Research, Society for the Study of Addiction, Cancer Research UK, and the United Kingdom Centre for Tobacco and Alcohol Studies. The NIHR School for Public Health Research is a partnership between the Universities of Sheffield, Bristol, Cambridge, and Imperial; University College London; The London School for Hygiene and Tropical Medicine; The Liverpool and Lancaster Universities Collaboration for Public Health Research (LiLaC); and Fuse—The Centre for Translational Research in Public Health: collaboration between Newcastle, Durham, Northumbria, Sunderland, and Teesside Universities. The views expressed are those of the authors and not necessarily those of the NIHR or Department of Health and Social Care. The United Kingdom Centre for Tobacco and Alcohol Studies is part of the UK Clinical Research Collaboration, Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council, and NIHR under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged. The authors would like to thank Dr Dave Crane for his important role in the development and factorial screening trial of the *Drink Less* app and the National University of Singapore's Institute of Mathematical Sciences for funding LB and TQ's visit to the program on the Statistical Methods section for Developing Personalized Mobile Health Interventions. The authors also thank the University College London Tobacco and Alcohol Research Group for their helpful feedback on the draft of this paper.

LB is supported by a Doctor of Philosophy studentship funded by the Medical Research Council (MRC) Network of Hubs for Trials Methodology Research (MR/L004933/2-R18). CG is funded by the NIHR (NIHR127651) and the NIHR School for Public Health Research and Cancer Research UK (CRUK: C1417/A22962). TQ is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award R01AA23187, National Institute on Drug Abuse of the National Institutes of Health under award P50DA039838, National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award U54EB020404, and National Cancer Institute of the National Institutes of Health under award U01CA229437. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. OP is funded by Cancer Research UK (CRUK: C1417/A22962). HWWP has received funding from Public Health England, the NIHR, and National Health Service (NHS) England. EW is funded by the Wellcome Trust (grant 224485/Z/21/Z) and MRC project grant (grant MR/S01442X/1). LB would like to thank the MRC Biostatistics Unit and Papworth Hospital for funding time spent on this research.

Conflicts of Interest

CG is part of the team that developed and evaluated the *Drink Less* app as part of a randomized controlled trial (ISRCTN64052601) and is a paid scientific consultant for behavior change and lifestyle organization "One Year No Beer." HWWP has received additional salary support from Public Health England and NHS England. He has a Doctor of Philosophy student who works at and has fees paid by Astra Zeneca and another who works at Better Points. He has research collaborations with Thrive Therapeutic Software Ltd and has one collaboration with Six to Start. He has an engagement project in collaboration with DigitalHealth.London, ZINC, and BMJ Innovations. Other authors declare no conflicts of interest.

Editorial Notice

The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the primary outcome is an engagement measure and not a health outcome. Readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness. However, note that the authors made publicly available the submitted preprint of the trial protocol (through *JMIR Research Protocols*) during recruitment and prior to database lock, clearly stating their trial design, objectives, and outcome measures.

Multimedia Appendix 1

Visual of *Drink Less* notification as it appears on the users' notification center.

[DOCX File , 72 KB - mhealth_v11i1e38342_app1.docx]

Multimedia Appendix 2

Privacy notice.

[DOCX File , 22 KB - mhealth_v11i1e38342_app2.docx]

Multimedia Appendix 3

Information Sheet.

[DOCX File , 14 KB - mhealth_v11i1e38342_app3.docx]

Multimedia Appendix 4

Bank of 30 newly developed messages and their link to the relevant behavior change module.
[DOCX File , 14 KB - mhealth_v11i1e38342_app4.docx]

Multimedia Appendix 5
Sample size calculation.
[DOCX File , 294 KB - mhealth_v11i1e38342_app5.docx]

Multimedia Appendix 6
Recruitment plots.
[DOCX File , 29 KB - mhealth_v11i1e38342_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 96 KB - mhealth_v11i1e38342_app7.pdf]

References

1. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. Br J Cancer 2015 Feb 03;112(3):580-593 [FREE Full text] [doi: 10.1038/bjc.2014.579] [Medline: 25422909]
2. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Br Dent J 2015 Oct 9;219(7):329. [doi: 10.1038/sj.bdj.2015.751]
3. Kaner E, O Dickinson H, Beyer FR, Campbell F, Schlesinger C, Heather N, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Systematic Rev 2007 Apr 18 [FREE Full text] [doi: 10.1002/14651858.cd004148.pub3]
4. Wilson G, Lock CA, Heather N, Cassidy P, Christie MM, Kaner E. Intervention against excessive alcohol consumption in primary health care: a survey of GPs' attitudes and practices in England 10 years on. Alcohol Alcohol 2011;46(5):570-577 [FREE Full text] [doi: 10.1093/alcalc/agr067] [Medline: 21690169]
5. Kaner E, Heather N, McAvoy BR, Lock CA, Gilvarry E. Intervention for excessive alcohol consumption in primary health care: attitudes and practices of English general practitioners. Alcohol Alcohol 1999;34(4):559-566. [doi: 10.1093/alcalc/34.4.559] [Medline: 10456584]
6. Burton R, Henn C, Lavoie D, O'Connor R, Perkins C, Sweeney K, et al. A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. Lancet 2017 Apr;389(10078):1558-1580. [doi: 10.1016/s0140-6736(16)32420-5]
7. Fowler LA, Holt SL, Joshi D. Mobile technology-based interventions for adult users of alcohol: a systematic review of the literature. Addict Behav 2016 Nov;62:25-34. [doi: 10.1016/j.addbeh.2016.06.008] [Medline: 27310031]
8. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res 2014;36(1):9-18 [FREE Full text] [Medline: 26258996]
9. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015 Dec:47. [doi: 10.2147/ahct.s65791]
10. Bendtsen M, McCambridge J, Åsberg K, Bendtsen P. Text messaging interventions for reducing alcohol consumption among risky drinkers: systematic review and meta-analysis. Addiction 2021 May 25;116(5):1021-1033 [FREE Full text] [doi: 10.1111/add.15294] [Medline: 33047865]
11. Satterfield JM. Thinking outside the visit: digitally extending the reach of behavioral health. J Gen Intern Med 2016 Sep 13;31(9):982-984 [FREE Full text] [doi: 10.1007/s11606-016-3736-9] [Medline: 27177912]
12. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
13. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
14. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
15. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
16. Amoakoh HB, Klipstein-Grobusch K, Ansah EK, Grobbee DE, Yveoo L, Agyepong I. How and why front-line health workers (did not) use a multifaceted mHealth intervention to support maternal and neonatal healthcare decision-making in

Ghana. BMJ Glob Health 2019 Mar 08;4(2):e001153 [FREE Full text] [doi: 10.1136/bmjgh-2018-001153] [Medline: 30997162]
17. Chevance G, Perski O, Hekler EB. Innovative methods for observing and changing complex health behaviors: four propositions. Transl Behav Med 2021 Mar 16;11(2):676-685 [FREE Full text] [doi: 10.1093/tbm/ibaa026] [Medline: 32421196]
18. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 01;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
19. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
20. Milward J, Deluca P, Drummond C, Kimergård A. Developing typologies of user engagement with the BRANCH alcohol-harm reduction smartphone app: qualitative study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e11692 [FREE Full text] [doi: 10.2196/11692] [Medline: 30545806]
21. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
22. Williamson C, White K, Rona RJ, Simms A, Fear NT, Goodwin L, et al. Smartphone-based alcohol interventions: a systematic review on the role of notifications in changing behaviors toward alcohol. Subst Abus 2022 Jun 07;43(1):1231-1244. [doi: 10.1080/08897077.2022.2074595] [Medline: 35670777]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
24. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
25. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
26. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
27. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clin Trials 2014 Aug 05;11(4):426-434 [FREE Full text] [doi: 10.1177/1740774514536795] [Medline: 24902922]
28. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019 Mar 01;9(2):296-307 [FREE Full text] [doi: 10.1093/tbm/iby043] [Medline: 29733406]
29. Garnett C, Perski O, Michie S, West R, Field M, Kaner E, et al. Refining the content and design of an alcohol reduction app, Drink Less, to improve its usability and effectiveness: a mixed methods approach. F1000Res 2021;10:511 [FREE Full text] [doi: 10.12688/f1000research.51416.2] [Medline: 34646502]
30. Behaviour change: individual approaches. National Institute for Health and Care Excellence (NICE). 2014 Jan 2. URL: https://www.nice.org.uk/guidance/ph49 [accessed 2023-04-24]
31. Bell L, Garnett C, Qian T, Perski O, Williamson E, Potts HW. Engagement with a behavior change app for alcohol reduction: data visualization for longitudinal observational study. J Med Internet Res 2020 Dec 11;22(12):e23369 [FREE Full text] [doi: 10.2196/23369] [Medline: 33306026]
32. Qian T, Walton AE, Collins LM, Klasnja P, Lanza ST, Nahum-Shani I, et al. Supplemental material for the microrandomized trial for developing digital interventions: experimental design and data analysis considerations. Psychological Method 2022 Jan 13;27(5):874-894. [doi: 10.1037/met0000283.supp]
33. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul 19;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]
34. NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, et al. Assessing real-time moderation for developing adaptive mobile health interventions for medical interns: micro-randomized trial. J Med Internet Res 2020 Mar 31;22(3):e15033 [FREE Full text] [doi: 10.2196/15033] [Medline: 32229469]
35. Carpenter SM, Menictas M, Nahum-Shani I, Wetter DW, Murphy SA. Developments in mobile health just-in-time adaptive interventions for addiction science. Curr Addict Rep 2020 Sep 01;7(3):280-290 [FREE Full text] [doi: 10.1007/s40429-020-00322-y] [Medline: 33747711]
36. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]

37. Bell L, Garnett C, Qian T, Perski O, Potts HW, Williamson E. Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial. JMIR Res Protoc 2020 Aug 07;9(8):e18690 [FREE Full text] [doi: 10.2196/18690] [Medline: 32763878]
38. Schulz KF, Moher D, Altman DG. CONSORT 2010. Lancet 2010 Oct;376(9748):1222-1223. [doi: 10.1016/s0140-6736(10)61879-x]
39. Allen JP, Litten RZ, Fertig JB, Babor T. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcoholism Clin Exp Res 1997 Jun;21(4):613-619. [doi: 10.1111/j.1530-0277.1997.tb03811.x]
40. Use notifications on your iPhone, iPad and iPod touch. Apple. URL: https://support.apple.com/en-gb/HT201925 [accessed 2023-04-24]
41. How a web session is defined in analytics. Google. URL: https://support.google.com/analytics/answer/2731565 [accessed 2023-04-24]
42. Saunders J, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
43. ICD-10 Classification of Mental and Behavioural Disorders (The) Diagnostic Criteria for Research. Geneva: World Health Organization; 1993.
44. Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, et al. Do daily fluctuations in psychological and app-related variables predict engagement with an alcohol reduction app? A series of N-of-1 studies. JMIR Mhealth Uhealth 2019 Oct 02;7(10):e14098 [FREE Full text] [doi: 10.2196/14098] [Medline: 31579022]
45. Garnett C, Oldham M, Angus C, Beard E, Burton R, Field M, et al. Evaluating the effectiveness of the smartphone app, Drink Less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6-month follow-up: protocol for a randomised controlled trial. Addiction 2021 Feb;116(2):412-425 [FREE Full text] [doi: 10.1111/add.15287] [Medline: 33067856]
46. Qian T, Yoo H, Klasnja P, Almirall D, Murphy SA. Estimating time-varying causal excursion effect in mobile health with binary outcomes. Biometrika 2021 Sep;108(3):507-527 [FREE Full text] [doi: 10.1093/biomet/asaa070] [Medline: 34629476]
47. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Statistical Assoc 1958 Jun;53(282):457-481. [doi: 10.1080/01621459.1958.10501452]
48. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
49. Wickham H, François R, Henry L, Müller K, Vaughan D, Posit Software. dplyr: a grammar of data manipulation. CRAN R Project. URL: https://cran.r-project.org/web/packages/dplyr/index.html [accessed 2023-04-24]
50. Grolemund G, Wickham H. Dates and times made easy with lubridate. J Stat Softw 2011;40(3):1-25. [doi: 10.18637/jss.v040.i03]
51. Sjoberg D, Whiting K, Curry M, Lavery J, Larmarange J. Reproducible summary tables with the gtsummary package. R J 2021 Jun;13(1):570. [doi: 10.32614/rj-2021-053]
52. Zeileis A, Grothendieck G. zoo: S3 infrastructure for regular and irregular time series. J Stat Softw 2005;14(6):1-27. [doi: 10.18637/jss.v014.i06]
53. Barbeiro A, Ferrari PA. Manzi, Imputation of Missing Values Through a Forward Imputation Algorithm. Mran Microsoft. 2023. URL: https://mran.microsoft.com/snapshot/2015-02-12/web/packages/ForImp/ForImp.pdf [accessed 2023-04-26]
54. Survminer: survival analysis and visualization. Survminer. URL: https://rpkgs.datanovia.com/survminer/ [accessed 2023-04-24]
55. Garnett C, Bell L, Perski O, Potts H, Qian T. MRT: optimising the notification policy to improve engagement. Open Science Framework. 2022 Jan 31. URL: https://osf.io/w3szp/ [accessed 2023-04-25]
56. DrinkLessMRT. GitHub. 2022 Feb 16. URL: https://github.com/lauren-bell/DrinkLessMRT [accessed 2023-04-25]
57. Zhu F, Liao P, Zhu X, Yao J, Huang J. Effective warm start for the online actor-critic reinforcement learning based mHealth intervention. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2018 Presented at: BCB '18; August 29-September 1, 2018; Washington, DC, USA p. 482-491. [doi: 10.1145/3233547.3233553]
58. Trella AL, Zhang KW, Nahum-Shani I, Shetty V, Doshi-Velez F, Murphy SA. Designing reinforcement learning algorithms for digital interventions: pre-implementation guidelines. Algorithms 2022 Aug 22;15(8):255 [FREE Full text] [doi: 10.3390/a15080255] [Medline: 36713810]
59. Yao J, Brunskill E, Pan W, Murphy S, Doshi-Velez F. Power constrained bandits. Proc Mach Learn Res 2021;149:1-50.
60. Liao P, Dempsey W, Sarker H, Hossain SM, Al'absi M, Klasnja P, et al. Just-in-time but not too much: determining treatment timing in mobile health. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec 27;2(4):1-21 [FREE Full text] [doi: 10.1145/3287057] [Medline: 30801052]
61. Liao P, Greenewald K, Klasnja P, Murphy S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar;4(1):18 [FREE Full text] [doi: 10.1145/3381007] [Medline: 34527853]

62. Wardell JD, Kempe T, Rapinda KK, Single A, Bilevicius E, Frohlich JR, et al. Drinking to cope during COVID-19 pandemic: the role of external and internal factors in coping motive pathways to alcohol use, solitary drinking, and alcohol problems. Alcohol Clin Exp Res 2020 Oct;44(10):2073-2083. [doi: 10.1111/acer.14425] [Medline: 32870516]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CONSORT: Consolidated Standards of Reporting Trials
EMEE: Estimator for the Marginal Excursion Effect
MOST: Multiphase Optimisation Strategy
MRT: micro-randomized trial
RR: relative risk

Edited by G Eysenbach; submitted 29.03.22; peer-reviewed by K Uludag, S Toukhy, D Leightley, E Goldenhersch, S Jia; comments to author 28.06.22; revised version received 08.10.22; accepted 31.03.23; published 09.06.23.

Please cite as:
Bell L, Garnett C, Bao Y, Cheng Z, Qian T, Perski O, Potts HWW, Williamson E
How Notifications Affect Engagement With a Behavior Change App: Results From a Micro-Randomized Trial
JMIR Mhealth Uhealth 2023;11:e38342
URL: https://mhealth.jmir.org/2023/1/e38342
doi:10.2196/38342
PMID:37294612

Original Paper

A Smartphone-Based Implicit Theories Intervention for Health Behavior Change: Randomized Trial

Mike Schreiber[1,2], DPhil; Simone Dohle[2,3], Prof Dr

[1]Faculty of Psychology, University of Vienna, Vienna, Austria

[2]Social Cognition Center Cologne, Department of Psychology, University of Cologne, Cologne, Germany

[3]Institute of General Practice and Family Medicine, University Hospital Bonn, University of Bonn, Bonn, Germany

Corresponding Author:
Mike Schreiber, DPhil
Faculty of Psychology
University of Vienna
Universitätsstr 7
Vienna, 1010
Austria
Phone: 43 1427747353
Email: mike.schreiber@univie.ac.at

Abstract

Background: Implicit theories of health describe individuals' beliefs about the malleability of health. Individuals with an incremental theory of health believe that health, in general, is malleable, whereas individuals with an entity theory of health endorse the idea that health is largely fixed and predetermined. Previous research has shown that an incremental theory of health is associated with beneficial health outcomes and behaviors. A mobile health implicit theories intervention could be an effective way to increase health-promoting behaviors in the general population.

Objective: The aim of this study was to estimate the effect of a smartphone-based intervention designed to promote an incremental theory of health on the frequency of health-promoting behaviors in everyday life. The study used ecological momentary assessment to measure health behavior change.

Methods: This 2-arm, single-blind, delayed intervention design included 149 German participants (mean age 30.58, SD 9.71 years; n=79 female). Participants were asked to report their engagement in 10 health-promoting behaviors throughout the day for 3 weeks. Participants were randomly assigned to either an early intervention group (n=72) or a delayed intervention group (n=77). The intervention materials, designed to promote an incremental theory of health, were provided to participants after 1 week (early intervention group) or 2 weeks (delayed intervention group) of baseline behavior measurement. Data for this study were collected between September 2019 and October 2019.

Results: A paired-samples 2-tailed *t* test revealed that participants reported a stronger incremental theory after responding to the intervention materials (mean 5.58, SE 0.07) compared with incremental theory measured in an entry questionnaire (mean 5.29, SE 0.08; t_{148}=4.07, SE 0.07; *P*<.001; 95% CI 0.15-0.43; *d*=0.33). Multilevel analyses showed that participants reported engaging in health-promoting behaviors more often after being presented with the intervention materials compared with baseline across conditions (*b*=0.14; $t_{146.65}$=2.06, SE 0.07; *P*=.04; 95% CI 0.01-0.28). However, when the analysis was conducted separately for the early and delayed intervention groups, the intervention effect was only significant for the delayed intervention group (*b*=0.27; $t_{1492.37}$=3.50, SE 0.08; *P*<.001; 95% CI 0.12-0.42). There was no significant increase in health-promoting behaviors for the early intervention group (*b*=0.02; $t_{69.23}$=0.14, SE 0.11;*P*=.89; 95% CI −0.2 to 0.23).

Conclusions: This study suggests that a smartphone-based intervention designed to promote an incremental theory of health is a cost- and time-effective approach to increase the frequency of engaging in health-promoting behaviors. However, research is needed to understand the reasons for the difference in intervention effects between the early and delayed intervention groups. The results of this study can guide the development of future digital health interventions that focus on implicit theories to promote health behavior change.

Trial Registration: DRKS – German Clinical Trials Register DRKS00017379; https://drks.de/search/de/trial/DRKS00017379

(JMIR Mhealth Uhealth 2023;11:e36578) doi:10.2196/36578

KEYWORDS

daily diary; ecological momentary assessment; health behavior; implicit theories; lay theories; mindsets; multiple health behavior change; randomized trial; smartphone-based intervention

Introduction

Background

According to the World Health Organization [1], 71% of all worldwide deaths are attributed to noncommunicable diseases like cardiovascular diseases, cancer, respiratory diseases, or diabetes. The risk of having such a disease can decrease by adopting a healthier lifestyle that includes sufficient physical activity, a healthy diet, and avoiding harmful substances like tobacco or alcohol [1,2]. Engagement in such health-promoting behaviors often involves a high level of self-regulatory strategies [3,4]. An essential prerequisite for successful self-regulatory processes is *implicit theories* [5]. Implicit theories (sometimes also framed as mindsets or lay theories) refer to people's beliefs about the changeability of human attributes and characteristics [5,6]. According to Dweck's [6] framework, people differ in the extent to which they hold an *incremental theory*, that is, assuming that a given attribute is developable and malleable, versus an *entity theory*, that is, assuming that an attribute is fixed and stable. Recent research shows that a stronger incremental theory of health has a positive influence on maintaining a healthy lifestyle across multiple health behavior domains [7-10]. Extending these findings, the main aim of this randomized trial was to investigate whether promoting an incremental theory of health increases the frequency of performing health-promoting behaviors in daily life.

Implicit Theories

Early research about implicit theories mainly focused on assumptions about the changeability of intelligence [11,12] or personality [13]. Since this first research, implicit theories have been studied across a wide array of domains like willpower [14,15], morality [16], stereotypes [17], and interpersonal relationships [18]. The majority of studies found that holding a stronger incremental theory in one domain (ie, assuming that the given characteristic is malleable) leads to positive outcomes [5,19]. For example, in a meta-analysis across 113 studies, holding an incremental theory was found to predict successful goal setting, goal monitoring, and goal operating, and, in turn, better self-regulation [5]. Therefore, researchers have developed many interventions to foster an incremental theory to create positive changes for individuals. The modes of delivering such interventions range from single-session approaches [20,21] to multisession approaches [12,22] and large-scale educational programs (eg, the Project for Educational Research That Scales [23]).

In the past decade, research about implicit theories has also become popular in different health domains, like weight management [22,24-26], physical activity [27,28], smoking [29,30], addiction [31,32], and mental health [21,33]. For example, it has been shown that an incremental theory can protect against setback-related weight gain [22], is related to higher motivation and intention to achieve a healthy weight [26], leads to greater motivation to quit smoking [31], and decreases anxiety and depressive symptoms [21].

Implicit theories in different domains are not necessarily interconnected [6,34]. For example, one might believe that one's body weight is rather fixed around a given set point while also thinking that smoking behavior can be changed easily. Therefore, implicit theories have not only been studied in single health domains but also for health in general. Such generalized implicit theories have been examined concerning their impact on multiple health behavior domains [7-10]. In that sense, an incremental theory of (general) health regards the assumption that health is malleable and changeable, whereas an entity theory of health implies that health is perceived as fixed and stable [7-9]. Correlational research has shown that holding an incremental theory of health is related to performing health-promoting [7] and health-protective behaviors [10]. In addition, experimental findings suggest that a strengthened incremental theory of health leads to more positive attitudes toward different health-promoting behaviors [7], stronger intentions to eat healthily [8], and healthier food choices [7]. Previous research is therefore limited. Although correlative research shows evidence for the importance of an incremental theory of health for a healthy lifestyle, it cannot be interpreted causally. Existing experimental studies, on the other hand, generally focus only on one health behavior (eg, eating behavior). Therefore, the main purpose of this study is to investigate whether an intervention that promotes an incremental theory of health influences a variety of health-promoting behaviors. Compared with an intervention that focuses only on implicit theories in a single health domain (eg, body weight or physical activity), focusing on implicit theories of general health may serve as an efficient strategy to encourage multiple health behavior change. In contrast to stationary settings, the delivery of this intervention via digital technologies offers the opportunity to reach a larger audience in a sustainable manner while minimizing implementation costs [35,36]. To increase ecological validity and to minimize recall and retrieval bias, engagement in health-promoting behaviors is measured using ecological momentary assessment in the form of daily diaries [37]. Similarly, it has been shown that stronger incremental theories of health were connected to a higher frequency of performing health-promoting behaviors in daily life measured using experience sampling (study 4) [7]. As these results were only correlational, the present research aims to provide causal insights into whether an intervention to foster incremental theories increases health behaviors in daily life. Therefore, we make the following hypothesis: Being confronted with a smartphone-based intervention to foster incremental theories of health increases the frequency of performing health-promoting behaviors in daily life.

Methods

Study Design

We conducted a 2-arm, randomized trial to investigate whether fostering an incremental theory of health increases the frequency of performing health-promoting behaviors in daily life. The intervention was conceptualized as a delayed-start design [38,39], in which both groups received intervention material at different times. The intervention was delivered via participants' smartphones using Qualtrics (Qualtrics International) questionnaires and included that participants kept a daily diary for 3 weeks. Participants were randomly assigned (single-blind) to an early or delayed intervention group using Qualtrics' randomizer while maintaining an evenly distributed number of participants in each group (1:1 block randomization).

At the beginning of the study (day 0), all participants responded to an entry questionnaire to measure implicit theories of health, health locus of control, health-related self-efficacy, health-related outcome expectancy, health status, health value, health change motivation, and anthropometric (height and weight) and demographic variables (age, gender, education, and occupation). One day after responding to the entry questionnaire, the daily diary phase started. Over the course of 3 weeks (21 days), participants received daily invitations to complete a short questionnaire via texting distributed via SurveySignal [40]. Participants received the invitations daily at 8 PM and had to respond within 4 hours. In these daily questionnaires, participants were asked to indicate whether they performed 10 different health-promoting behaviors throughout the respective day. The number of daily performed health-promoting behaviors served as primary outcome measure. Depending on the assigned condition, participants received intervention materials to foster an incremental theory either after 7 (early intervention group) or 14 days (delayed intervention group) of baseline behavior measurement. After 21 days, we invited participants to participate in a follow-up questionnaire measuring the same constructs—except anthropometric and demographic items—as in the entry questionnaire. Table 1 provides an overview of the study's design.

We chose the delayed-start design as it allows testing for intervention effects between and within both intervention arms [38,39] and helps to disentangle the effects of the intervention itself and the self-monitoring due to the daily diaries. Furthermore, including a baseline in both groups helps participants to get used to the daily diary approach and allows for longitudinal comparisons (before vs after reading the intervention materials).

Table 1. Overview of the intervention flow for the early and delayed intervention group. Links to view intervention materials were sent out at 8 AM on day 8 (early intervention group) or day 15 (delayed intervention group), and the links were valid for 12 hours.

Group	Day 0	Days 1-7	Days 8-14	Days 15-21	Day 22
Early intervention group	Entry questionnaire	Baseline measurement	Postintervention measurement	Postintervention measurement	Follow-up questionnaire
Delayed intervention group	Entry questionnaire	Baseline measurement	Baseline measurement	Postintervention measurement	Follow-up questionnaire

Ethics Approval

Data collection for this study was performed between September 13, 2019, and October 10, 2019. The study was approved by the faculty's ethics commission (ID MSHF0047). It was registered as a randomized trial in the German Clinical Trials Register (trial number DRKS00017379) and was preregistered in the Open Science Framework (OSF [41]).

Sample Size Calculation

Sample size was determined before data collection using G*Power (version 3; Heinrich-Heine-Universität Düsseldorf) [42] based on an expected effect size of f=0.15 (with α=.05 and 1–β=.90). The calculation resulted in a required total sample size of 96 participants. Because we also planned to run multilevel models, a total sample size of 120 was targeted to increase the probability of achieving model convergence. As described further in the Results section, main analyses were performed using data from 149 participants.

Recruitment, Eligibility Criteria, and Compensation

Participants were recruited via the institutes' participant pool and social media postings. Eligibility criteria were a minimum age of 18 years, owning a smartphone with touch display and mobile internet access, and being able to answer daily questionnaires for 21 days. All participants received financial compensation for their participation: €3 (US $3.26) each for completing the entry and follow-up questionnaire, €4 (US $4.35) for responding to the intervention materials, €0.25 (US $0.27) for each completed daily questionnaire, and a bonus of €10 (US $10.87) for responding to more than 17 (80%) of the daily questionnaires (in total, up to €25.25; US $27.46).

Measures

Table 2 provides an overview of the measures included in the different questionnaires of the intervention and informs about the internal consistency of the included scales. Internal consistency of all scales ranged between α=.74 and α=.88 and can be considered good.

Table 2. Overview of variables and Cronbach α of the scales measured in different intervention parts. Check marks indicate that a measure was used in that part of the study.

Measure	Entry	Intervention	Follow-up
Implicit Theories of Health Scale	✓ (α=.88)	✓ (α=.85)	✓ (α=.87)
Internal health locus of control	✓ (α=.76)		✓ (α=.83)
Chance health locus of control	✓ (α=.82)		✓ (α=.88)
Powerful others locus of control	✓ (α=.74)		✓ (α=.80)
Health-related self-efficacy	✓ (α=.85)		✓ (α=.83)
Health-related outcome expectancy	✓ (α=.77)		✓(α=.79)
Health status	✓		✓
Health value	✓		✓
Change motivation (self)	✓		✓
Change motivation (others)	✓		✓
Age	✓		
Gender	✓		
Height	✓		
Weight	✓		
Education	✓		
Occupation	✓		

Implicit Theories

The Implicit Theories of Health Scale (ITHS) [7] was used to measure implicit theories of health. The scale consists of 6 items (eg, "You can substantially change your own health"). Three items represent an incremental theory of health, and 3 items represent an entity theory of health (which were recoded). Answers were given on 7-point Likert scales (1=strongly disagree to 7=strongly agree). A mean across all items was computed, with higher values indicating a stronger incremental theory.

Health-Promoting Behaviors

In the daily diaries, participants were asked every day whether they performed 10 health-promoting behaviors throughout the respective day (see Textbox 1; 0=no, 1=yes). We measured only behaviors (1) that could be performed during a regular day, (2) that were based on national recommendations from public health authorities (eg, Federal Centre for Health Education), and (3) that showed no ceiling or floor effect regarding the frequency of performing these behaviors, determined in a pretest (n=325). Concerning the latter, we did not include behaviors such as brushing one's teeth or washing one's hands because the pretest showed that almost all participants conducted these behaviors daily. The sum of performed health-promoting behaviors per day served as the primary outcome measure.

Textbox 1. Items to measure the frequency of performing health-promoting behaviors.

Nutrition

- I ate at least 2 servings of fruit
- I ate at least 3 servings of vegetables
- I did not eat sweets
- I drank at least 2 liters of water

Physical activity

- I have been physically active for at least 30 minutes, so I started to sweat and/or was slightly out of breath
- I walked or cycled at least 6.5 kilometers
- I exercised

Relaxation

- I took some time to relax
- I slept for at least 7 hours

Hygiene

- I used dental floss

Control Variables

Additionally, we measured health-related locus of control, self-efficacy, outcome expectancy, change motivation, health status, and health value. We included these variables to ensure (1) that the 2 intervention groups did not differ significantly regarding these constructs at baseline and (2) that the intervention only leads to changes in implicit theories and not the other constructs.

Health Locus of Control

The Health- and Illness-Related Locus of Control Questionnaire (Kontrollüberzeugung zu Krankheit und Gesundheit; KKG) [43] was used to measure health locus of control. The KKG consists of 21 items, all answered on 6-point Likert scales (1=strongly disagree to 6=strongly agree). Similar to its English equivalent [44], the KKG consists of 3 subscales (with 7 items each) to measure internal (eg, "If I do not feel well physically, I have to blame myself"), powerful others (eg, "If I feel well physically, then I owe it mainly to the advice and help of others"), and chance health locus of control (eg, "Whether my symptoms last longer depends mainly on chance").

Health-Related Self-Efficacy

The Perceived Health Competence Scale [45] was used to measure health-related self-efficacy. The scale consists of 8 items (eg, "I'm generally able to accomplish my goals with respect to my health") measured on 5-point Likert scales (1=strongly disagree to 5=strongly agree).

Health-Related Outcome Expectancy

Health-related outcome expectancy was measured using 6 statements to assess how much participants agree that specific health behaviors can influence one's own health ("Your health is strongly influenced by...eating behavior,... physical activity and exercising,... consumption of harmful substances,... enough sleep and relaxation,… personal and dental hygiene,... regular doctor visits, and checkups"). Participants' agreement was assessed via 7-point Likert scales (1=strongly disagree to 7=strongly agree).

Further Health-Related Variables

Single items measured current subjective health status ("How would you describe your health status in general?"; 1=bad to 7=excellent), health value ("How important is your health to you?"; 1=not at all important to 7=very important), and the extent to which participants think that they should change their health from their point of view ("It is important to me to change something about my health"; 1=strongly disagree to 7=strongly agree) and from the perspective of others ("From the perspective of others, I should change something about my health"; 1=strongly disagree to 7=strongly agree).

Intervention Materials

Participants received a link to the intervention materials on their smartphones either after 7 (early intervention group) or 14 days (delayed intervention group) of baseline measurement. The links to view the intervention materials were sent out at 8 AM on the specified day (day 8 or 15), and the individual links were available until 8 PM on that day. Like other interventions to promote incremental theories [12,20,22], the intervention materials consisted of informative, exemplary, and reflective components. More precisely, the intervention materials included (1) a (fictitious) newspaper article that described health as mainly influenced by lifestyle and engagement in health-promoting behavior [7], (2) three fictitious blog posts in which individuals reported positive health changes, (3) an essay priming in which participants were asked to describe health changes in their lives, and (4) an article that focused on the benefits of beliefs in changeability in other domains. The materials are available in the OSF repository [41]. After reading the articles and the 3 blog posts, participants answered one question regarding the content of the materials as an attention check. In addition, an independent rater checked the content of the essays to determine whether the participants followed the task description. On the basis of this, we calculated an attention check score, ranging from 1 to 4, with 4 points indicating that all content questions were answered correctly and that the essay fitted the instruction.

Results

Participants

Initially, 393 participants were screened regarding eligibility criteria (see the CONSORT [Consolidated Standards of Reporting Trials] flow diagram. flow diagram in Figure 1). A total of 254 participants were randomly assigned to 1 of the 2 intervention arms (early vs delayed intervention). As some participants discontinued their participation or did not respond to the intervention materials, 162 participants received the allocated intervention (81 participants in both intervention arms). Participants were excluded from data analysis when they did not complete the entry questionnaire (early intervention: n=8; delayed intervention: n=4) or did not respond to the daily diaries during the first week (early intervention: n=1). No participants were excluded based on their attention check scores, as all remaining participants scored 2 points or higher. Further, a regression analysis revealed that the attention check scores did not have an impact on the change in implicit theories measured in the entry questionnaire versus directly after seeing the intervention materials (b=0.19; t_{147}=0.73, SE 0.26; P=.47; 95% CI −0.32 to 0.70). Consequently, main analyses were performed with 149 participants (early intervention: n=72; delayed intervention: n=77). The mean age of the analyzed sample was 30.58 (SD 9.71) years, with 52% (79/149) female and 47% (70/149) male participants. Table 3 includes other demographic characteristics, and the CONSORT flow diagram (Figure 1) provides an overview of participant flow and informs about dropout reasons in each intervention group. For additional follow-up analyses, data of 138 participants were available.

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

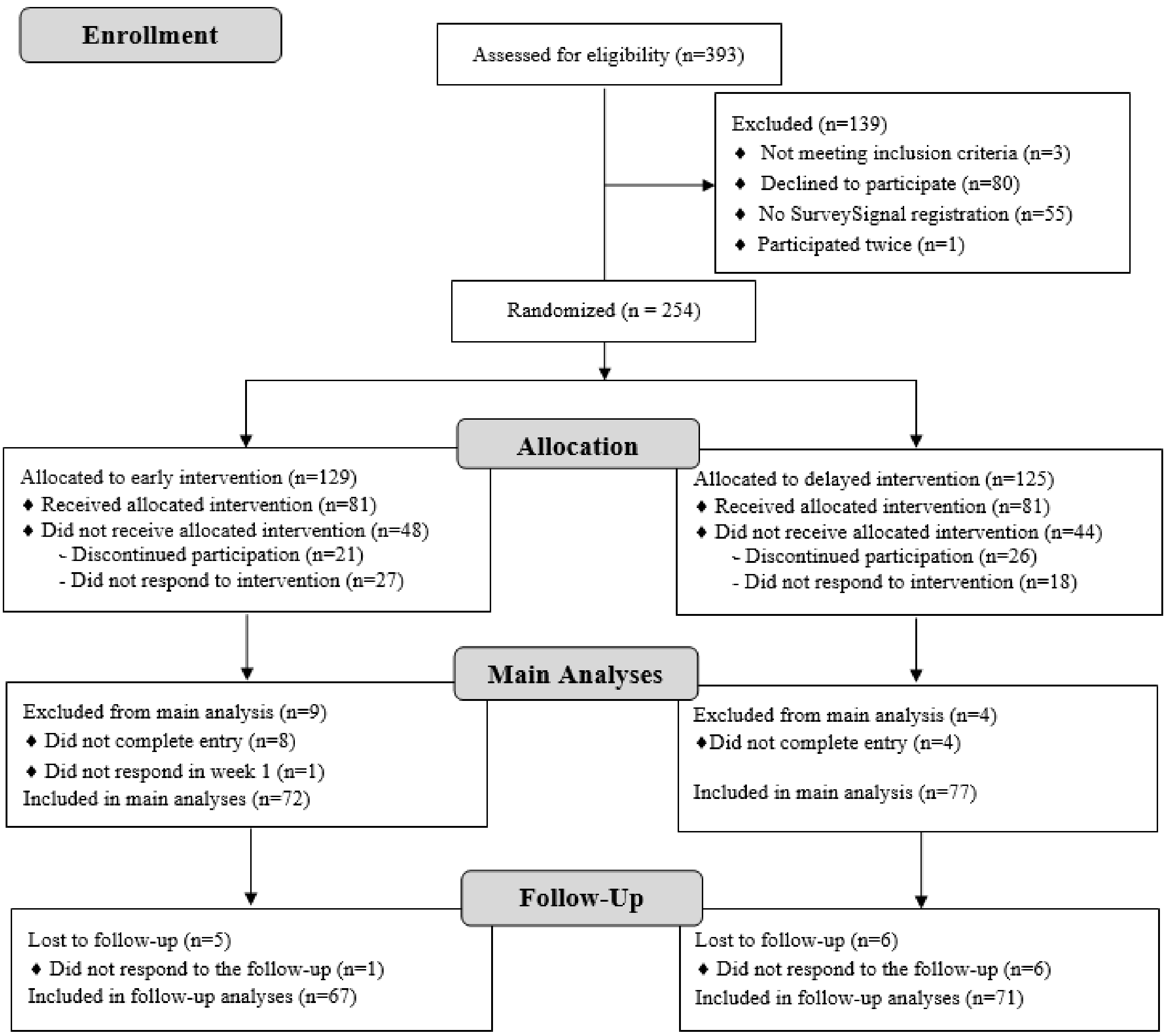


Table 3. Baseline characteristics of participants in total and by intervention group (n=149).

Characteristics	Overall	Early (n=72)	Delayed (n=77)	Condition difference	*P* value
Age (years), mean (SD)	30.58 (9.71)	31.31 (10.44)	29.91 (8.98)	0.88 (147)[a]	.38
Implicit theories, mean (SD)	5.29 (0.99)	5.36 (1.06)	5.22 (0.92)	0.86 (147)[a]	.39
Internal locus, mean (SD)	3.78 (0.70)	3.87 (0.72)	3.69 (0.67)	1.61 (147)[a]	.11
Powerful others locus, mean (SD)	2.90 (0.73)	2.92 (0.73)	2.88 (0.73)	0.34 (147)[a]	.73
Chance locus, mean (SD)	2.35 (0.74)	2.39 (0.79)	2.32 (0.70)	0.60 (147)[a]	.55
Self-efficacy, mean (SD)	3.52 (0.68)	3.54 (0.71)	3.51 (0.64)	0.26 (147)[a]	.80
Outcome-expectancy, mean (SD)	5.56 (0.83)	5.63 (0.81)	5.49 (0.84)	1.02 (147)[a]	.31
Height (in meters), mean (SD)	174.60 (9.63)	175.21 (9.38)	174.04 (9.89)	0.74 (147)[a]	.46
Weight (in kilogram), mean (SD)	77.82 (18.88)	78.18 (21.26)	77.49 (16.51)	0.22 (146)[a]	.82
BMI, mean (SD)	25.37 (5.17)	25.22 (5.43)	25.52 (4.95)	0.35 (146)[a]	.72
Health status, mean (SD)	4.99 (1.15)	4.97 (1.13)	5.00 (1.18)	0.15 (147)[a]	.88
Health value, mean (SD)	6.11 (0.98)	5.97 (1.07)	6.23 (0.87)	1.64 (147)[a]	.10
Change motivation (self), mean (SD)	5.50 (1.22)	5.43 (1.27)	5.56 (1.18)	0.64 (147)[a]	.52
Change motivation (others), mean (SD)	3.37 (1.77)	3.44 (1.68)	3.30 (1.86)	0.50 (147)[a]	.62
Gender, n (%)				2.51 (1)[b]	.11
Male	70 (47)	29 (40)	41 (53)		
Female	79 (53)	43 (60)	36 (47)		
Education, n (%)				3.66 (4)[b]	.45
Lower secondary school	4 (2.7)	2 (3)	2 (3)		
Secondary school	10 (6.7)	4 (6)	6 (8)		
Entitlement to study at a university of applied sciences	3 (2)	3 (4)	0 (0)		
Higher education entrance qualification (“Abitur”)	61 (40.9)	28 (39)	33 (43)		
University degree	71 (47.7)	35 (49)	36 (47)		
Occupation, n (%)				8.90 (6)[b]	.18
Full-time employed	40 (26.8)	24 (33)	16 (21)		
Part-time employed	13 (8.7)	4 (6)	9 (12)		
Studying	81 (54.4)	36 (50)	45 (58)		
Stay-at-home spouse	3 (2)	0 (0)	3 (4)		
Retired	5 (3.4)	3 (4)	2 (3)		
Occupational disability	3 (2)	2 (3)	1 (1)		
Other	4 (2.7)	3 (4)	1 (1)		

[a]These values are the *t* (*df*).

[b]These values are the chi-square.

Precursory Analyses

As depicted in Table 2, there were no significant differences between the 2 intervention groups regarding demographics or other measures included in the entry questionnaire, suggesting that the randomization was successful. Participants answered a total of 3015 daily questionnaires; on average, each participant answered 20.23 (SD 1.42, range 12-21) questionnaires.

As a manipulation check, a paired-samples 2-tailed *t* test with ITHS scores measured in the entry questionnaire and ITHS scores after responding to the intervention materials was performed to test whether the intervention led participants to adopt a stronger incremental theory. This *t* test revealed that participants reported a stronger incremental theory after responding to intervention materials (mean 5.58, SE 0.07), compared with the entry questionnaire (mean 5.29, SE 0.08; t_{148}=4.07, SE 0.07, P<.001, 95% CI 0.15-0.43, d=0.33). Further, a 2 (intervention group: early vs delayed) by 2 (time of assessment: entry questionnaire vs directly after seeing the intervention materials) mixed ANOVA revealed that the intervention led to an increase in incremental theories in both groups indicated by a significant main effect of time of implicit theories assessment ($F_{1,147}$=16.42; P<.001; η_p^2=.10), a nonsignificant main effect of intervention group ($F_{1,147}$=1.19; P=.28; η_p^2=.01), and a nonsignificant interaction ($F_{1,147}$=0.02; P=.89; η_p^2<.01).

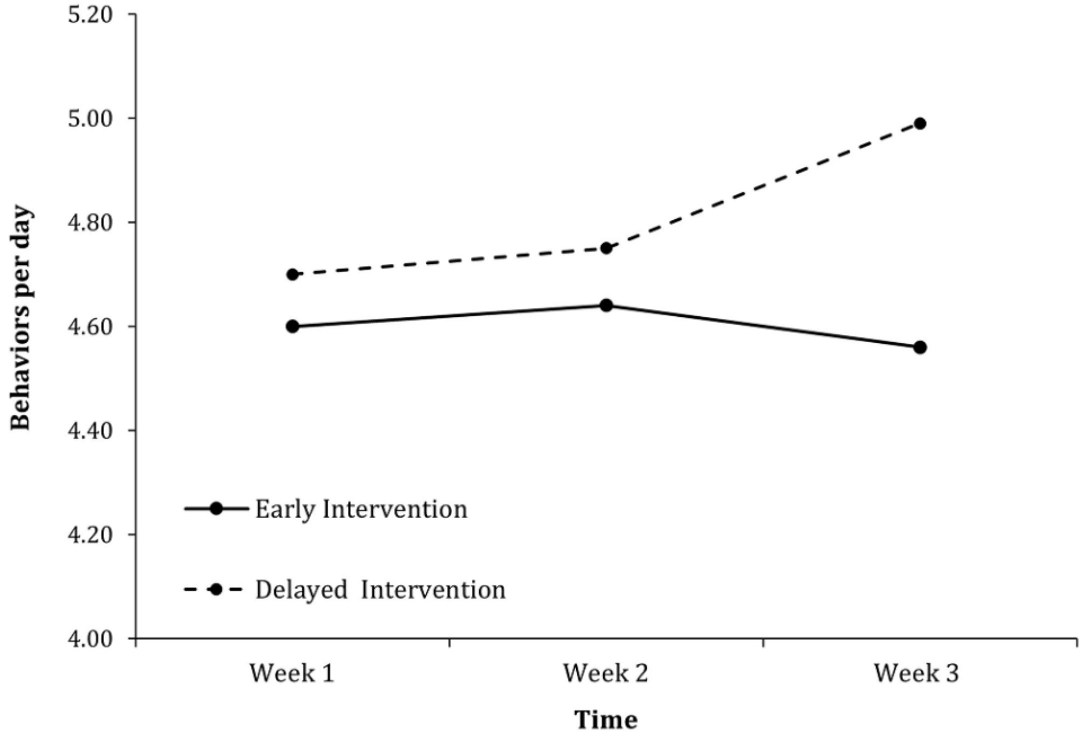
Main Analyses

Based on the preregistration [41], a mixed ANOVA was conducted to test whether the intervention increased the frequency of performing health-promoting behaviors on a weekly level. As within-subject factor, the mean number of health-promoting behaviors per day was aggregated for every week, and intervention group (early versus delayed) was entered as between-subject factor. The results of the mixed ANOVA showed no significant main effect of the intervention group ($F_{1,147}$=0.92; P=.34; η_p^2=.01). There was also no significant difference between mean daily performed behaviors per week ($F_{2,294}$=1.46; P=.23; η_p^2=.01). However, a significant interaction between the intervention group and week emerged ($F_{2,294}$=3.06; P=.048; η_p^2=.02). Table 4 shows marginal means, SEs, and 95% CIs for daily performed behaviors per week for both intervention groups, and Figure 2 illustrates the interaction.

Table 4. Marginal means, SEs, and 95% CIs for mean performed health behaviors per day as a result of the condition × time in a mixed ANOVA.

Condition	Mean (SE)	95% CI
Early intervention		
Week 1	4.60 (0.17)	4.27-4.92
Week 2	4.64 (0.18)	4.29-4.99
Week 3	4.56 (0.18)	4.21-4.91
Delayed intervention		
Week 1	4.70 (0.16)	4.38-5.01
Week 2	4.75 (0.17)	4.42-5.09
Week 3	4.99 (0.17)	4.65-5.33

Figure 2. Mean number of performed health-promoting behaviors per day aggregated on a weekly level for both intervention groups.



As the marginal means and their corresponding confidence intervals did not show significant differences between both groups in any week, the preregistered *t* tests were not conducted. However, as depicted in Figure 2 and indicated by the significant interaction between time and condition, the delayed intervention group may have benefitted from viewing the intervention materials. Therefore, the preregistered multilevel models were performed to test whether the intervention increased the frequency of performing health-promoting behaviors on a daily level. Day was treated as the level 1 unit and participant as the level 2 unit. Intervention status (0=before intervention, 1=after intervention) served as the level 1 predictor, whereas the number of performed health-promoting behaviors served as the level 1–dependent variable. A deviance test was conducted for each analysis to test whether a random-slope or a random-intercept model results in a better model fit. Across both intervention groups, the better fitting random-slope model showed an increase in the number of performed health-promoting behaviors after responding to the intervention materials (b=0.14; $t_{146.65}$=2.06, SE 0.07; P=.04, 95% CI 0.01-0.28). Because of the significant interaction between time and condition found in the mixed ANOVA, additional multilevel models were conducted separately for both intervention groups. These multilevel models revealed that the effect of the intervention only appeared for the delayed intervention group (random-intercept model, b=0.27; $t_{1492.37}$=3.50, SE 0.08; P<.001; 95% CI 0.12-0.42). In contrast, no difference before and after the intervention was detected for the early intervention group (random-slope model, b=0.02; $t_{69.23}$=0.14, SE 0.11; P=.89; 95% CI −0.2 to 0.23). The information criteria and results of the likelihood-ratio tests for comparing the multilevel models are available in Table S1 in Multimedia Appendix 1.

Additional Analyses

Additional analyses revealed that participants did also report a stronger incremental theory in the follow-up questionnaire (mean 5.42, SE 0.08) compared with the entry questionnaire (mean 5.24, SE 0.08; t_{137}=2.42, SE 0.07; P=.02; 95% CI 0.03-0.32; d=0.20). Furthermore, participants reported a stronger internal health locus of control in the follow-up questionnaire (mean 3.92, SE 0.06) compared with the entry questionnaire (mean 3.78, SE 0.06; t_{137}=3.17, SE 0.04; P=.002; 95% CI 0.05-0.23; d=0.28). To test whether our intervention also led to an increase in health-promoting behaviors when controlling for the change in internal health locus of control, we performed additional multilevel models, including the change in internal health locus of control as the level 2 predictor. Table 5 shows that these robustness checks led to the same conclusions as the multilevel models described in the previous section. For all other health-related variables, no significant difference between the entry and follow-up questionnaire emerged.

Table 5. Multilevel models including the change in internal health locus of control as an additional level 2 predictor to determine whether our intervention effects are robust when controlling for the internal health locus of control change.

Variable	*B*	SE	*t (df)*	*P* value	95% CI
Predicting health-promoting behaviors per day (across conditions)					
Intercept	4.70	0.11	39.54 (135.72)	<.001	4.46 to 4.93
Time (0=before intervention; 1=after intervention)	0.15	0.07	2.15 (135.78)	.03	0.01 to 0.29
Change (internal health locus)	0.13	0.22	0.61 (136.14)	.54	−0.29 to 0.56
Predicting health-promoting behaviors per day (early intervention group)					
Intercept	4.69	0.18	26.07 (63.58)	<.001	4.33 to 5.05
Time (0=before intervention; 1=after intervention)	0.04	0.11	0.37 (64.64)	.71	−0.18 to 0.26
Change (internal health locus)	0.40	0.33	1.22 (65.83)	.23	−0.25 to 1.05
Predicting health-promoting behaviors per day (delayed intervention group)					
Intercept	4.78	0.17	27.54 (72.11)	<.001	4.44 to 5.13
Time (0=before intervention, 1=after intervention)	0.26	0.08	3.30 (1374.34)	<.001	0.11 to 0.42
Change (internal health locus)	−0.14	0.30	−0.46 (68.83)	.65	−0.74 to 0.46

Discussion

Principal Findings

The present research aimed to examine whether a smartphone-based intervention to foster incremental theories of health increases the frequency of performing health-promoting behaviors in daily life measured via ecological momentary assessment. Indicated by our manipulation check, we found that the intervention led to stronger incremental theories of health. Furthermore, across conditions, participants showed a significant increase in the frequency of performing health-promoting behaviors after being confronted with the intervention materials. However, this effect was only driven by the delayed intervention group, whereas the early intervention group did not increase in health-promoting behaviors.

One possible explanation for why the effectiveness of the intervention differed between intervention groups may be that incremental theories may only have a beneficial effect in the long run. As depicted in Figure 2, both intervention groups showed a slight increase in health-promoting behaviors between the first and second week. This could be due to the involvement of self-monitoring evoked by the daily diaries, which can have an intervention effect itself [46]. Being confronted with the intervention materials at an early stage (week 1) did not seem to have any additional motivational benefit for the early intervention group. On the other hand, the delayed intervention group was confronted with the fact that they *only* showed half of the measured behaviors every day for 2 weeks. Being shown the intervention materials at this point in time (week 2) had an additional motivating effect beyond the effect of self-monitoring. Instead, the early intervention group shows a decline in health-promoting behaviors in the third week without an additional boost in motivation. This pattern fits Yeager and Dweck's [47] argument that incremental theories are especially helpful when challenges arise (like continually maintaining the motivation to engage in health-promoting behaviors over 3 weeks). As this explanation is only speculative, further research is needed to investigate whether the observed time-dependent effectiveness of the intervention replicates consistently or has resulted by chance.

Although no intervention effect emerged for the early intervention group, the introduced intervention led to changes in health behavior for the delayed intervention group. Thus, this study is the first to show that implicit theories of health can be influenced through an intervention delivered in people's daily lives. It provides further evidence of the relevance of these theories for health behavior change across multiple health domains and extends the existing correlative and experimental findings on implicit theories of health [7-10]. The results show that even a one-shot implicit theory intervention via web-based materials can increase engagement in health-promoting behaviors. Hence, this approach represents a time-, effort-, and cost-efficient way for health promotion. This study also increases the ecological validity of previous findings by measuring health-promoting behaviors using ecological momentary assessment [37]. We show that incremental theories are relevant not only in laboratory research or one-shot web-based questionnaires but also to everyday behavior.

Additional analyses revealed that the intervention-based increase in incremental theories of health is not just short-term, as participants also reported stronger incremental theories in the follow-up questionnaire (compared with the entry questionnaire). In addition, the present intervention led to a stronger internal health locus of control. This is consistent with previous findings showing that a stronger internal health locus of control mediates the effect of an implicit theories of health manipulation on health-promoting outcomes [7]. It also fits findings that an incremental theory of personality intervention increases primary control in the context of mental illness [20,21]. It remains to be tested whether the change in internal health locus of control stems from the intervention materials, the daily diary assessment, or the combination of both.

Limitations and Generalizability

We chose a delayed-start design to test for intervention effects between and within both intervention groups. However, we did not find a difference in health behavior engagement between the 2 groups in the second week of our assessment. Future trials should include a no-treatment control group that does not receive intervention materials. This approach makes it possible to determine whether the pre-post difference in behavior in the delayed intervention group appeared because of the intervention materials or because of the combination of the intervention materials with daily diaries.

We incorporated ecological momentary assessment in which participants were asked daily whether they performed 10 health-promoting behaviors using a simple yes-no format. This format allows for more objective and reliable measures of self-reported health behavior engagement with less recall and retrieval bias compared with standard forms of measurement in which individuals usually have to recall behavior engagement over longer periods (like weeks or months) [37]. However, these self-reports can still be affected by social desirability. Therefore, future studies could incorporate more objective measurements of health behavior engagement, like taking pictures of meals to measure eating behavior or using smartwatches or other devices to measure physical activity (for a physical activity example, see Henderson et al [48]).

Regarding the generalizability of findings, it is essential to note that the surveyed sample differs from the general population, especially regarding age, educational level, and student proportion. Participation in the study required owning a smartphone with internet access, and recruitment was realized via social media and mailing lists. This has limited the studies' accessibility for individuals of older age. Moreover, participants reported high values on other health-relevant measures at baseline (eg, health status, health change motivation, and self-efficacy; see Table 1). It may have been easier for individuals with such characteristics to engage in health-promoting behaviors or adopt new behavioral routines. On the other hand, the present intervention even led to positive changes in health-promoting behaviors for individuals already starting with such advantageous conditions. Thus, individuals lacking these attributes might benefit even more from the intervention introduced.

Recently, the relevance of implicit theories interventions has been seriously tackled in 2 meta-analyses concluding that they only produce weak effects in educational settings [49]. According to classic convictions [50], the reported effect sizes or regression coefficients in the present research also fall in this category. However, it has been argued that these convictions should be used with caution, and effect sizes should be evaluated considering the area or context investigated [50,51]. Especially in health or educational research, even small effects can have far-reaching consequences when evaluated in a broader context [23,51].

This study is in line with the majority of research showing that a stronger incremental theory leads to beneficial outcomes [5]. However, holding an entity theory can be instrumental under specific circumstances. A stronger incremental theory of health not only implies that one's health can improve but also means that one's health might worsen. For this reason, an incremental theory would be less adaptive when a prevention focus is present [52,53], that is, when one is trying to conserve a given health status. For individuals being confronted with the process of aging or suffering from long-lasting diseases, it may be more beneficial to believe in the stability of health.

Theoretical and Practical Implications

As introduced in the present research, addressing implicit theories of health serves as a new approach for achieving positive health behavior change. The expanding research of implicit theories in the health domain [7-10] may guide the development of health interventions or could be integrated into health education. Nevertheless, further steps are needed to test whether the present findings replicate and can be generalized. First, direct replications could test whether the effectiveness of an incremental theories intervention is indeed time-sensitive, as demonstrated in this study. In addition, future studies should also focus on testing the longevity of the intervention effect. Investigating whether the increase in incremental theories can be sustained over longer periods and whether this increase translates into sustained improvements in health behaviors is essential. Therefore, studies with increased follow-up periods are necessary to test the longevity of the intervention effect and the respective impact on health behaviors. Next, conceptual replications could investigate what modes of delivering an implicit theories intervention are most effective and for whom. For example, the effectiveness might be higher when the changeability of health is emphasized several times over an extended course of time. A useful tool in this respect could be an app that tracks and visualizes changes in one's health behavior over time. Finally, the generalizability to different populations and different health behavior measures needs to be ensured.

Research on implicit theories of health would also benefit greatly from examining the antecedents and determinants that lead to adopting an incremental versus entity theory [7]. Auster-Gussman and Rothman [24] found that incremental theories of body weight are more common among young and White individuals with a higher level of income and education. These variables, as well as one's medical history or that of close others, should also play a significant role in the formation of implicit theories [7]. This study shows that even a rather young and educated sample with high self-reported health and high incremental theories at baseline benefits from an incremental theories intervention. Effects might be stronger when studying population groups with higher entity theories of health.

Conclusions

This study is the first randomized trial demonstrating that incremental theories of health can increase because of a single-session smartphone-based intervention. Contrary to our assumptions, the intervention only led to an increase in performing health-promoting behaviors when delivered at a later point in time. Further studies are crucial to assure whether the observed time-dependent variation in effectiveness replicates. Incremental theories interventions might be most effective for individuals holding a stronger entity theory of health. Factors that favor the development of an entity theory of health should be investigated to identify population groups that would benefit most from the interventional approach introduced in this paper.

Acknowledgments

This work was supported by an advanced postdoctoral grant from the University of Cologne awarded to SD. Open access funding was provided by the University of Vienna. The authors would like to thank Marina Hinssen, Anna-Lena Brans, Isabel Skuplik, and Leonie Wiegmann for their practical support. MS and SD were affiliated with the University of Cologne at the time of the trial. MS is now affiliated with the University of Vienna, and SD is now affiliated with the University of Bonn.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1. Regression coefficients and information criteria for comparing the multilevel models to predict health behavior engagement per day.

[DOCX File , 26 KB - mhealth_v11i1e36578_app1.docx]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1354 KB - mhealth_v11i1e36578_app2.pdf]

References

1. Noncommunicable disease: Unhealthy diet. World Health Organization. 2020. URL: http://www.emro.who.int/noncommunicable-diseases/causes/unhealthy-diets.html [accessed 2023-05-16]
2. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. BMJ 2008;337:a1440 [FREE Full text] [doi: 10.1136/bmj.a1440] [Medline: 18796495]
3. Leventhal H, Cameron LD, editors. The Self-Regulation of Health and Illness Behaviour. London, UK: Routledge; 2008.
4. Hofmann W, Friese M, Wiers RW. Impulsive versus reflective influences on health behavior: a theoretical framework and empirical review. Health Psychol Rev 2008;2(2):111-137 [FREE Full text] [doi: 10.1080/17437190802617668]

5. Burnette JL, O'Boyle EH, VanEpps EM, Pollack JM, Finkel EJ. Mind-sets matter: a meta-analytic review of implicit theories and self-regulation. Psychol Bull 2013;139(3):655-701 [FREE Full text] [doi: 10.1037/a0029531] [Medline: 22866678]
6. Dweck CS. Self-theories: Their Role in Motivation, Personality, and Development. Philadelphia, PA: Psychology Press; 2000.
7. Schreiber M, Job V, Dohle S. Is your health malleable or fixed? The influence of implicit theories on health-related attitudes and behaviour. Psychol Health 2020;35(12):1421-1439 [FREE Full text] [doi: 10.1080/08870446.2020.1761975] [Medline: 32400215]
8. Thomas FN, Burnette JL, Hoyt CL. Mindsets of health and healthy eating intentions. J Appl Soc Psychol 2019;49(6):372-380 [FREE Full text] [doi: 10.1111/jasp.12589]
9. Bunda K, Busseri MA. Lay theories of health, self-rated health, and health behavior intentions. J Health Psychol 2019;24(7):979-988 [FREE Full text] [doi: 10.1177/1359105316689143] [Medline: 28810398]
10. Zhang N, Kou Y. Implicit theories of health, consideration of future consequences, and engagement in health protective behaviors during the COVID-19 pandemic in China. J Health Psychol 2022;27(6):1462-1469 [FREE Full text] [doi: 10.1177/13591053211017191] [Medline: 33983050]
11. Hong YY, Chiu CY, Dweck CS. Implicit theories of intelligence: Reconsidering the role of confidence in achievement motivation. In: Kernis MH, editor. Efficacy, Agency, and Self-esteem. Boston, MA: Plenum; 1995:197-216.
12. Blackwell LS, Trzesniewski KH, Dweck CS. Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. Child Dev 2007;78(1):246-263 [FREE Full text] [doi: 10.1111/j.1467-8624.2007.00995.x] [Medline: 17328703]
13. Dweck CS, Leggett EL. A social cognitive approach to motivation and personality. Psychol Rev 1988;95(2):256-273 [FREE Full text] [doi: 10.1037/0033-295x.95.2.256]
14. Job V, Dweck CS, Walton GM. Ego depletion--is it all in your head? implicit theories about willpower affect self-regulation. Psychol Sci 2010;21(11):1686-1693 [FREE Full text] [doi: 10.1177/0956797610384745] [Medline: 20876879]
15. Job V, Walton GM, Bernecker K, Dweck CS. Implicit theories about willpower predict self-regulation and grades in everyday life. J Pers Soc Psychol 2015;108(4):637-647 [FREE Full text] [doi: 10.1037/pspp0000014] [Medline: 25844577]
16. Chiu CY, Dweck CS, Tong JY, Fu JH. Implicit theories and conceptions of morality. J Pers Soc Psychol 1997;73(5):923-940 [FREE Full text] [doi: 10.1037/0022-3514.73.5.923]
17. Levy SR, Stroessner SJ, Dweck CS. Stereotype formation and endorsement: The role of implicit theories. J Pers Soc Psychol 1998;74(6):1421-1436 [FREE Full text] [doi: 10.1037/0022-3514.74.6.1421]
18. Knee CR. Implicit theories of relationships: assessment and prediction of romantic relationship initiation, coping, and longevity. J Pers Soc Psychol 1998;74(2):360-370 [FREE Full text] [doi: 10.1037/0022-3514.74.2.360]
19. Dweck CS. Implicit theories. In: Kruglanski AW, Higgins ET, van Lange PAM, editors. Theories of Social Psychology. Los Angeles, CA: SAGE Publications; 2012:43-61.
20. Schleider JL, Weisz JR. Reducing risk for anxiety and depression in adolescents: effects of a single-session intervention teaching that personality can change. Behav Res Ther 2016;87:170-181 [FREE Full text] [doi: 10.1016/j.brat.2016.09.011] [Medline: 27697671]
21. Schleider J, Weisz J. A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. J Child Psychol Psychiatry 2018;59(2):160-170 [FREE Full text] [doi: 10.1111/jcpp.12811] [Medline: 28921523]
22. Burnette JL, Finkel EJ. Buffering against weight gain following dieting setbacks: an implicit theory intervention. J Exp Soc Psychol 2012;48(3):721-725 [FREE Full text] [doi: 10.1016/j.jesp.2011.12.020]
23. Yeager DS, Hanselman P, Walton GM, Murray JS, Crosnoe R, Muller C, et al. A national experiment reveals where a growth mindset improves achievement. Nature 2019;573(7774):364-369 [FREE Full text] [doi: 10.1038/s41586-019-1466-y] [Medline: 31391586]
24. Auster-Gussman LA, Rothman AJ. Understanding the prevalence and correlates of implicit theories of weight in the United States: Insights from a nationally representative sample. Psychol Health 2018;33(4):483-498 [FREE Full text] [doi: 10.1080/08870446.2017.1373112] [Medline: 28911240]
25. Burnette JL. Implicit theories of body weight: entity beliefs can weigh you down. Pers Soc Psychol Bull 2010;36(3):410-422 [FREE Full text] [doi: 10.1177/0146167209359768] [Medline: 20179317]
26. Taber JM, Klein WMP, Persky S, Ferrer RA, Kaufman AR, Thai CL, et al. Interest in and reactions to genetic risk information: the role of implicit theories and self-affirmation. Soc Sci Med 2017;190:101-110 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.010] [Medline: 28858695]
27. Lyons C, Kaufman AR, Rima B. Implicit theories of the body among college women: implications for physical activity. J Health Psychol 2015;20(9):1142-1153 [FREE Full text] [doi: 10.1177/1359105313508997] [Medline: 24271689]
28. Orvidas K, Burnette JL, Russell VM. Mindsets applied to fitness: growth beliefs predict exercise efficacy, value and frequency. Psychol Sport Exerc 2018;36:156-161 [FREE Full text] [doi: 10.1016/j.psychsport.2018.02.006]
29. Fitz CC, Kaufman A, Moore PJ. Lay theories of smoking and young adult nonsmokers' and smokers' smoking expectations. J Health Psychol 2015;20(4):438-445 [FREE Full text] [doi: 10.1177/1359105313502694] [Medline: 24155189]

30. Thai CL, Coa KI, Kaufman AR. Implicit theories of smoking and association with current smoking status. J Health Psychol 2018;23(9):1234-1239 [FREE Full text] [doi: 10.1177/1359105316648483] [Medline: 27216983]
31. Sridharan V, Shoda Y, Heffner JL, Bricker J. Addiction mindsets and psychological processes of quitting smoking. Subst Use Misuse 2019;54(7):1086-1095 [FREE Full text] [doi: 10.1080/10826084.2018.1555259] [Medline: 30892118]
32. Burnette JL, Forsyth RB, Desmarais SL, Hoyt CL. Mindsets of addiction: implications for treatment intentions. J Soc Clin Psychol 2019;38(5):367-394 [FREE Full text] [doi: 10.1521/jscp.2019.38.5.367]
33. Schleider JL, Weisz JR. Implicit theories relate to youth psychopathology, but how? A longitudinal test of two predictive models. Child Psychiatry Hum Dev 2016;47(4):603-617 [FREE Full text] [doi: 10.1007/s10578-015-0595-2] [Medline: 26443503]
34. Plaks JE, Levy SR, Dweck CS. Lay theories of personality: cornerstones of meaning in social cognition. Soc Pers Psychol Compass 2009;3(6):1069-1081 [FREE Full text] [doi: 10.1111/j.1751-9004.2009.00222.x]
35. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
36. Stark AL, Geukes C, Dockweiler C. Digital health promotion and prevention in settings: scoping review. J Med Internet Res 2022;24(1):e21063 [FREE Full text] [doi: 10.2196/21063] [Medline: 35089140]
37. Shiffman S, Stone AA. Introduction to the special section: ecological momentary assessment in health psychology. Health Psychol 1998;17(1):3-5 [FREE Full text] [doi: 10.1037/h0092706]
38. D'Agostino RB. The delayed-start study design. N Engl J Med 2009;361(13):1304-1306 [FREE Full text] [doi: 10.1056/NEJMsm0904209] [Medline: 19776413]
39. Parkinson Study Group. A controlled, randomized, delayed-start study of rasagiline in early parkinson disease. Arch Neurol 2004;61(4):561-566 [FREE Full text] [doi: 10.1001/archneur.61.4.561] [Medline: 15096406]
40. Hofmann W, Patel PV. SurveySignal: a convenient solution for experience sampling research using participants' own smartphones. Soc Sci Comput Rev 2014;33(2):235-253 [FREE Full text] [doi: 10.1177/0894439314525117]
41. Schreiber M, Dohle S. Implicit Theories Intervention for Multiple Health Behavior Change. 2019. URL: https://osf.io/y7un5/ [accessed 2023-05-16]
42. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191 [FREE Full text] [doi: 10.3758/bf03193146] [Medline: 17695343]
43. Lohaus A, Schmitt GM. Kontrollüberzeugungen zu krankheit und gesundheit (KKG): bericht über die entwicklung eines testverfahrens control beliefs about illness and health (KKG): report on the development of a measure. Diagnostica 1989;35(1):59-72 [FREE Full text]
44. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) scales. Health Educ Monogr 1978;6(2):160-170 [FREE Full text] [doi: 10.1177/109019817800600107] [Medline: 689890]
45. Smith MS, Wallston KA, Smith CA. The development and validation of the Perceived Health Competence Scale. Health Educ Res 1995;10(1):51-64 [FREE Full text] [doi: 10.1093/her/10.1.51] [Medline: 10150421]
46. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (V1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
47. Yeager DS, Dweck CS. What can be learned from growth mindset controversies? Am Psychol 2020;75(9):1269-1284 [FREE Full text] [doi: 10.1037/amp0000794] [Medline: 33382294]
48. John-Henderson NA, Tintzman CS, Counts CJ, Mueller CM. Health mindsets as a predictor of physical activity and body mass index in american indian college students. J Health Psychol 2021;26(12):2098-2105 [FREE Full text] [doi: 10.1177/1359105319901284] [Medline: 31984820]
49. Sisk VF, Burgoyne AP, Sun J, Butler JL, Macnamara BN. To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. Psychol Sci 2018;29(4):549-571 [FREE Full text] [doi: 10.1177/0956797617739704] [Medline: 29505339]
50. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum; 1988.
51. Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. Adv Meth Pract Psychol Sci 2019;2(2):156-168 [FREE Full text] [doi: 10.1177/2515245919847202]
52. Sue-Chan C, Wood RE, Latham GP. Effect of a coach's regulatory focus and an individual's implicit person theory on individual performance. J Manag 2010;38(3):809-835 [FREE Full text] [doi: 10.1177/0149206310375465]
53. Sevincer AT, Kluge L, Oettingen G. Implicit theories and motivational focus: desired future versus present reality. Motiv Emot 2013;38(1):36-46 [FREE Full text] [doi: 10.1007/s11031-013-9359-0]

Abbreviations

BMI: body mass index
CONSORT: Consolidated Standards of Reporting TrialsITHS: Implicit Theories of Health Scale

KKG: Kontrollüberzeugung zu Krankheit und Gesundheit (Health- and Illness- Related Locus of Control Questionnaire)
OSF: Open Science Framework

Edited by L Buis; submitted 27.04.22; peer-reviewed by N Zhang, K Bunda; comments to author 03.06.22; revised version received 31.08.22; accepted 27.04.23; published 15.06.23.

Please cite as:
Schreiber M, Dohle S
A Smartphone-Based Implicit Theories Intervention for Health Behavior Change: Randomized Trial
JMIR Mhealth Uhealth 2023;11:e36578
URL: https://mhealth.jmir.org/2023/1/e36578
doi:10.2196/36578
PMID:37318864

Original Paper

Using Smartphone Survey and GPS Data to Inform Smoking Cessation Intervention Delivery: Case Study

Amanda Luken[1], MHS; Michael R Desjardins[2,3], MA, PhD; Meghan B Moran[4], MA, PhD; Tamar Mendelson[1], MA, PhD; Vadim Zipunnikov[5], PhD; Thomas R Kirchner[6,7], MS, PhD; Felix Naughton[8], MSc, PhD; Carl Latkin[4], PhD; Johannes Thrul[1,9,10], MS, PhD

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[2]Spatial Science for Public Health Center, Johns Hopkins University, Baltimore, MD, United States
[3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[4]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[5]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[6]Department of Social and Behavioral Sciences, New York University School of Global Public Health, New York, NY, United States
[7]Center for Urban Science and Progress, New York University Tandon School of Engineering, New York, NY, United States
[8]Behavioural and Implementation Science Research Group, University of East Anglia, Norwich, United Kingdom
[9]Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD, United States
[10]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia

Corresponding Author:
Amanda Luken, MHS
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
624 N Broadway
Baltimore, MD, 21205
United States
Phone: 1 732 690 2886
Email: aluken95@gmail.com

Abstract

Background: Interest in quitting smoking is common among young adults who smoke, but it can prove challenging. Although evidence-based smoking cessation interventions exist and are effective, a lack of access to these interventions specifically designed for young adults remains a major barrier for this population to successfully quit smoking. Therefore, researchers have begun to develop modern, smartphone-based interventions to deliver smoking cessation messages at the appropriate place and time for an individual. A promising approach is the delivery of interventions using geofences—spatial buffers around high-risk locations for smoking that trigger intervention messages when an individual's phone enters the perimeter. Despite growth in personalized and ubiquitous smoking cessation interventions, few studies have incorporated spatial methods to optimize intervention delivery using place and time information.

Objective: This study demonstrates an exploratory method of generating person-specific geofences around high-risk areas for smoking by presenting 4 case studies using a combination of self-reported smartphone-based surveys and passively tracked location data. The study also examines which geofence construction method could inform a subsequent study design that will automate the process of deploying coping messages when young adults enter geofence boundaries.

Methods: Data came from an ecological momentary assessment study with young adult smokers conducted from 2016 to 2017 in the San Francisco Bay area. Participants reported smoking and nonsmoking events through a smartphone app for 30 days, and GPS data was recorded by the app. We sampled 4 cases along ecological momentary assessment compliance quartiles and constructed person-specific geofences around locations with self-reported smoking events for each 3-hour time interval using zones with normalized mean kernel density estimates exceeding 0.7. We assessed the percentage of smoking events captured within geofences constructed for 3 types of zones (census blocks, 500 ft^2 fishnet grids, and 1000 ft^2 fishnet grids). Descriptive comparisons were made across the 4 cases to better understand the strengths and limitations of each geofence construction method.

Results: The number of reported past 30-day smoking events ranged from 12 to 177 for the 4 cases. Each 3-hour geofence for 3 of the 4 cases captured over 50% of smoking events. The 1000 ft^2 fishnet grid captured the highest percentage of smoking events compared to census blocks across the 4 cases. Across 3-hour periods except for 3:00 AM-5:59 AM for 1 case, geofences contained an average of 36.4%-100% of smoking events. Findings showed that fishnet grid geofences may capture more smoking events compared to census blocks.

Conclusions: Our findings suggest that this geofence construction method can identify high-risk smoking situations by time and place and has potential for generating individually tailored geofences for smoking cessation intervention delivery. In a subsequent smartphone-based smoking cessation intervention study, we plan to use fishnet grid geofences to inform the delivery of intervention messages.

(JMIR Mhealth Uhealth 2023;11:e43990) doi:10.2196/43990

KEYWORDS

adult; application; case study; cessation; delivery; GIS; GPS; health interventions; mHealth; mobile phone; smartphone application; smartphone; smoker; smoking cessation; smoking

Introduction

Cigarette Smoking and Smoking Cessation Among Young Adults

Although cigarette use has declined in recent decades, of a weighted sample of 4200 young adults (ages 19-30 years), 491 (11.7%) young adults in the United States reported current (past 30-day) cigarette smoking in 2019 [1]. Smoking cessation interventions can provide societal and health care cost savings, as well as immediate and long-term health benefits for young adults (eg, decreased risk of cardiovascular diseases, chronic obstructive pulmonary disease, and several types of cancer) [2].

Young adults who smoke rarely use evidence-based smoking cessation strategies in their quit attempts. We need novel interventions that can reach young people and help them quit smoking. A recent study using data from the Population Assessment of Tobacco and Health found that almost all young adults who smoke would like to quit at some point in their lives, but few young adults with a recent quit attempt relied on evidence-based cessation strategies [3]. One explanation for this may be that many interventions for young adults attempt to prevent smoking initiation rather than support smoking cessation [4].

Mobile Phones for Smoking Cessation

The ubiquity of smartphones may help enhance the feasibility, acceptability, and reach of smoking cessation interventions. Almost all 18- to 29-year-olds in the United States own a smartphone [5]. As a result, GPS-enabled smartphones allow researchers to study the behaviors, mobility, and activity spaces of individuals and deliver mobile health (mHealth) interventions that were previously not feasible for potential consumers to access [6]. However, the efficacy of mHealth is understudied in many areas of public health, including smoking cessation.

Few interventions with location information (eg, GPS) include formal spatial science components that may improve intervention delivery. Smoking is often a geographically triggered behavior; people may regularly smoke in the same locations (eg, bars) or have cravings due to an environmental exposure (eg, product or advertisement exposure in a convenience store) [7,8]. GPS-enabled smartphones can register when an individual is near or at a location at-risk for smoking and deliver just-in-time cessation support to resist environmental triggers [9].

Spatial methods are essential for improving smoking cessation by examining the nexus between health and place [8,10,11]. For example, 1 study developed a deep learning model to predict smoking events based on GPS smartphone data [12]. The authors were able to predict smoking events accurately on weekdays and weekends (mean 0.87, SD 0.08) using a 1D convolutional neural network [12]. Overall, the literature using fine-grained geographic information to inform smoking cessation intervention delivery on smartphone apps is scarce, and this proof-of-concept study aims to address this gap.

Geofences for Smoking Cessation

Geofences are virtual perimeters or zones that can trigger smartphone notifications for individuals when entering, exiting, or dwelling within a specified geographic area [13]. Geofences may benefit mHealth interventions since individuals can receive interventions at high-risk locations [14]. Using participants' mobility patterns to generate geofences with the goal of promoting positive behavior change is a growing area of interest in geography and public health research [15,16]. For example, geofencing applications have been developed to support dentist accessibility [17], gambling cessation [18], awareness of air pollution exposure [19], COVID-19 surveillance [20], and tobacco retail exposure for smoking cessation [8], among other uses.

In the context of smoking cessation, Naughton et al [9] studied a cohort of 15 individuals in the United Kingdom and disseminated geofenced-triggered messages to participants when they entered high-risk smoking zones (ie, circular zones with a 100 m radius containing at least 4 self-reported smoking events); the study, however, did not use formal spatial analytical techniques to generate geofences. To extend this previous research, our goal for this study is to generate geofences around high-risk locations for smoking using a kernel density estimation (KDE) approach, which is reliable for analyzing GPS-based activity space data [21]. KDE is a proven spatial method that identifies hot spots of point patterns in space and time. Although other point-pattern techniques are available, KDE is very flexible regarding its parameter customization (eg, bandwidth, output resolution of the density surface, and kernel functions), which is described more in the Methods section [22]. However, the literature on using KDE approaches to generate high-risk geofences is very limited (eg, transportation injury prevention [23,24]) and has been underused for creating geofences tailored to individuals' spatiotemporal patterns of health risk behaviors.

Additionally, the uncertain geographic context problem (UGCoP) poses a challenge for geofence construction [25]. UGCoP represents the concern that the geographic units used for analyses may not represent the "true causally relevant" geographical context [25]. UGCoP is an issue for geofence construction because the geofenced location influences a person's smoking behaviors across space and time, but as presented in UGCoP, the appropriate spatial and temporal dimensions for geofence construction are uncertain. Individuals often report smoking at home, outside, or in the car [26], which could constitute the "true causally relevant" geographic context [25], but few interventions sensitive to capturing the "true causally relevant" geographic context have been developed or tested.

Research Objectives and Anticipated Contributions

This study uses self-reported smartphone surveys and passively tracked GPS data collected from young adults who smoke. The objective is to develop a spatial analytical approach to identify hot spots of self-reported smoking events and to produce KDE-informed geofences for catered smoking cessation intervention delivery in future studies. The proposed method, which incorporates spatial methods, may be applicable to intervene on other health conditions or other substance use as well. In this study, we demonstrate an exploratory method of generating person-specific geofences for high-risk smoking areas by presenting 4 case studies. We chose to test this method on 4 participants as a case study rather than the entire sample to examine the nuances of this methodological approach that would be more difficult to observe within the entire sample. By focusing on 4 participants, we can identify individual-level variation in smoking behaviors and locations that may influence geofence construction and performance. More importantly, we can better understand why this method may not have worked for certain situations or individuals. Although these 4 participants do not capture the full variation of smoking behaviors, they provide an important first step to evaluating this method in light of unanticipated situations and circumstances. For the 4 cases, we create individually tailored geofences that vary temporally, accounting for behavioral changes over the course of a day. A KDE-informed approach can capture the intersection of place and health by taking a person-centered, data-driven approach. To address UGCoP, we experiment with time-specific geofences constructed by various geographical zones and assess how well the different geofences capture smoking events for each case. We operationalize geofence performance as the percent of smoking events captured, such that an ability to capture greater than 80% of smoking events within geofences for a particular time frame was considered good, while greater than 50% was considered adequate.

Methods

Study Design

Young adults from Alameda and San Francisco counties participated in an ecological momentary assessment (EMA) study for 30 days that captured individual-level, spatiotemporal patterns of smoking behaviors. Demographic information, smoking history, and alcohol use were assessed through Qualtrics [27] at baseline. For 30 days, participants reported smoking events and completed 3 daily surveys distributed randomly throughout the day through the PiLR Health app [28]. In the app, participants recorded if they were about to smoke (ie, a smoking event), after which some were asked to complete a smoking survey based on the average number of daily cigarettes smoked at baseline (eg, a baseline rate of 10 cigarettes per day was associated with a 33% chance of receiving a smoking survey). To minimize the burden of participation, smoking surveys were limited to at most 3 per day. Each submitted survey was date-, time-, and GPS-stamped. See previous publications for more information on data collection procedures [29-31].

We selected 4 participants for the case study based on their overall compliance with EMA data collection procedures. The total available data, which includes both smoking and nonsmoking reports, served as a proxy for compliance. We selected 4 case study participants from the total available data quartiles. This sampling strategy was chosen to investigate the feasibility of geospatial analyses for participants with different rates of EMA self-report compliance and to improve the generalizability of findings.

Ethics Approval

All study procedures were approved by the San Francisco Committee on Human Research, University of California (15-18033).

Participants

Eligible participants were recruited using Facebook and Instagram advertisements between 2016 and 2017, were between 18 and 26 years of age, were established smokers (ie, at least 100 cigarettes per lifetime), and reported currently smoking at least one cigarette per day on at least 3 days per week. Study eligibility also required daily smartphone use with GPS capabilities. Women identifying as a sexual minority were oversampled for a nested qualitative study [29]. Study consent was provided electronically on Qualtrics. To confirm their identity, participants were required to send a photo of their ID.

Measures

Smoking Events

The outcome of interest was self-reported cigarette smoking events (yes or no). EMA data included smoking events (eg, cigarette self-reports) and nonsmoking events (eg, random surveys with nonsmoking events) with a linked GPS location. Smoking events reported within 5 minutes of another smoking event were dropped to correct for measurement errors due to technical difficulties with the app. GPS locations were converted to North American Datum 1983 California Zone 3 in US feet.

Time

EMA data included time stamps that were collected in Coordinated Universal Time (UTC) and converted to Pacific Time for analyses. Time was categorized into 3-hour periods (eg, midnight-2:59 AM, 3 AM-5:59 AM, 6 AM-8:59 AM, and 9 AM-11:59 AM).

Baseline Demographics

Demographic data (eg, age, gender, race or ethnicity, and education) were collected at baseline. The frequency of past 30-day cigarette use was also obtained.

Spatial Analyses

Framework of KDE

We employed KDE to identify high-risk zones for each individual. In other words, KDE was run separately for each participant to effectively tailor the resulting geofences to each individual. KDE is a moving window method that calculates the density as the number of events based on their distance to the center of a circle with a radius of the specified bandwidth τ, which determines the degree of spatial smoothing [32]. The window moves and centers along the intersections of a grid and calculates the density at each intersection, which is then considered in unison to provide a weighted average for a location. Events (ie, points) closer to the center of the search radius receive higher weight [33]. The KDE function is defined in Equation 1:

where is the kernel density value at grid point *g*; is the Euclidean distance between grid point *g* and event *i*; and is the weight where its value equals “0” at distance τ [34]. To identify high-risk areas for smoking for each participant, we plotted a kernel density plot of smoking events with a bandwidth of 1320 ft (ie, 0.25 miles), a proxy for comfortable walking distance [35], and a raster cell output size of 150 ft^2. We then extracted high-risk zones using zonal statistics. Zonal statistics calculate a statistic of interest for raster values falling within a “zone” chosen by researchers in another data set [36]. For our study, we averaged KDE raster values within 3 zones: 2020 US census blocks, 500 ft^2 fishnet grid cells, and 1000 ft^2 fishnet grid cells. We then minimum-maximum normalized the mean KDE to the range of 0 to 1 and retained zones with normalized mean KDEs above a threshold chosen during sensitivity analyses. To distinguish risk levels within zones, we categorized zones with normalized mean KDEs above the threshold into terciles (ie, low, medium, and high risk) for each case.

To identify the threshold for “high-risk,” we assessed the performance of geofences constructed for normalized mean KDE thresholds of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 for the 4 case studies. We identified the high-risk normalized mean KDE threshold at the block-level. Performance was assessed by examining the percent of smoking events captured within geofences, irrespective of time of day.

Zonal Statistics by Census Block

We had first chosen census blocks as the “zone” because census blocks are the smallest geographical unit with demographic data and are a federally recognized statistical area [37]. A geographical unit tied to demographic data could be useful for researchers interested in controlling societal-level demographic variables. Further, the census block is a stable geographical unit across the United States with a unique identifier determined by the US Census Bureau. In terms of stability, the census block is reconsidered once every decade [37].

The census block, however, is constructed by both physical (eg, roads, streams, and railroad lines) and nonphysical structures (eg, property lines and city limits), which leads to variability in census block size [37] and thus a normalized mean KDE. Census blocks may be the size of a city block in urban environments or up to hundreds of square miles in rural areas [37]. As a result, differences in census block size may affect block-level normalized mean KDEs (eg, smaller blocks may have fewer raster values to average across than larger blocks). Larger blocks may also have more heterogeneity in raster values than smaller blocks, decreasing the precision of the mean KDE.

Moreover, identifying a large census block as high-risk poses some challenges for intervention delivery. Based on our method, the geofence around the entire census block would trigger the intervention for a large area even though the participant may smoke only in a small subarea of the census block. Census blocks, however, may still offer value in cities where blocks are often equivalent to city blocks and in studies seeking to account for demographic variables [37].

Zonal Statistics by Fishnet Grid

Alternatively, we can create uniform zones by overlaying a fishnet grid over the area of interest (eg, Alameda and San Francisco counties). A fishnet grid is an array of square cells fitted within a geographical area [38]. We repeated the same process of finding the normalized mean KDE and labeling high-risk areas as zones with a normalized mean KDE above the threshold chosen from the census block analyses, except that the zone was a cell in the fishnet grid rather than a census block. We created 2 fishnets, one with 500 ft^2 cells and another with 1000 ft^2 cells.

Geofence Construction

For each 3-hour time interval, high-risk zones were identified as those with (1) normalized mean KDEs above the threshold of 0.3 and (2) at least one smoking event. Normalized mean KDEs were based on all observations of a given participant. For example, each block has the same normalized mean KDE value across all hours of the day and for each day of the week. A block, however, may be high-risk at 1 time of day (eg, evening) yet low-risk at another time of day (eg, morning) only because the participant has no history of smoking during the time of day it is considered low-risk (eg, morning). Once the high-risk blocks were identified, geofences were constructed by generating 100-m buffers around groups of adjacent high-risk blocks or cells. All spatial analyses were conducted in ArcGIS Pro (version 2.8.0) [39].

Results

Sample

The population included 144 participants with a mean age of 22.7 (SD 2.6) years at baseline; 76 of 144 (52.8%) were female; and 57 of 144 (39.6%) identified as non-Hispanic White, 31 of 144 (21.5%) as Hispanic or Latino, and 30 of 144 (20.8%) as Asian (Table 1). Most of the population had either completed some college (54/144, 37.5%) or received an associate or bachelor's degree (51/144, 35.4%). The median number of days with at least one cigarette smoked was 30 (IQR 24-30). On days with at least one smoking event, participants reported smoking a median of 5 (IQR 3-8) cigarettes per day.

Table 1. Baseline characteristics.

Characteristics	Cases sampled (N=4)	Cases not sampled (N=140)	Overall (N=144)
Age (years), mean (SD)	21.3 (2.6)	22.7 (2.4)	22.7 (2.5)
Sex, n (%)			
Male	2 (50)	66 (47.1)	68 (47.2)
Female	2 (50)	74 (52.9)	76 (52.8)
Highest education, n (%)			
Less than or equal to high school	1 (25)	30 (21.4)	31 (21.5)
Some college	2 (50)	52 (37.1)	54 (37.5)
Associate or bachelor's degree	1 (25)	50 (35.7)	51 (35.4)
Master's degree or higher	0 (0)	8 (5.7)	8 (5.6)
Race or Ethnicity, n (%)			
Non-Hispanic White	3 (75)	54 (38.6)	57 (39.6)
Non-Hispanic Black	0 (0)	6 (4.3)	6 (4.2)
Asian	1 (25)	29 (20.7)	30 (20.8)
American Indian or Alaska Native	0 (0)	1 (0.7)	1 (0.7)
Native Hawaiian or Pacific Islander	0 (0)	2 (1.4)	2 (1.4)
Hispanic	0 (0)	31 (22.1)	31 (21.5)
Other or Multiracial	0 (0)	17 (12.1)	17 (11.8)
Cigarettes per smoking day, median (range)	3.5 (3-5)	5 (1-30)	5 (1-30)
Smoking days in the past 30 days, median (range)	30 (25-30)	30 (0-30)	30 (0-30)

Cases Selected for Analysis

Of the 4 cases sampled; their ages ranged from 19 to 25 years, with an average age of 21.3 (SD 2.6) years at baseline. Two were male and 2 were female at birth. Three participants identified as non-Hispanic White and one as Asian. One had less than or equal to a high school education, 2 had some college, and 1 had an associate's or bachelor's degree. Three of the cases reported daily smoking in the past 30 days, while one reported smoking on 25 of the past 30 days. The cases reported smoking between 3 and 5 cigarettes per smoking day in the baseline assessment.

The number of smoking events reported by these 4 cases in EMA data ranged from 16 to 177, while the number of nonsmoking events ranged from 8 to 67 (Table 2). All smoking events were in San Francisco and Alameda counties.

Table 2. Smoking and nonsmoking events of cases.

Quartile	Smoking events, n (%)	Nonsmoking events, n (%)	Total reports, n (%)
25th	16 (66. 7)	8 (33.3)	24 (5.7)
50th	12 (20.3)	47 (79.7)	59 (14.1)
75th	31 (31.6)	67 (31.6)	98 (23.4)
100th	177 (74.7)	61 (74.4)	238 (56.8)

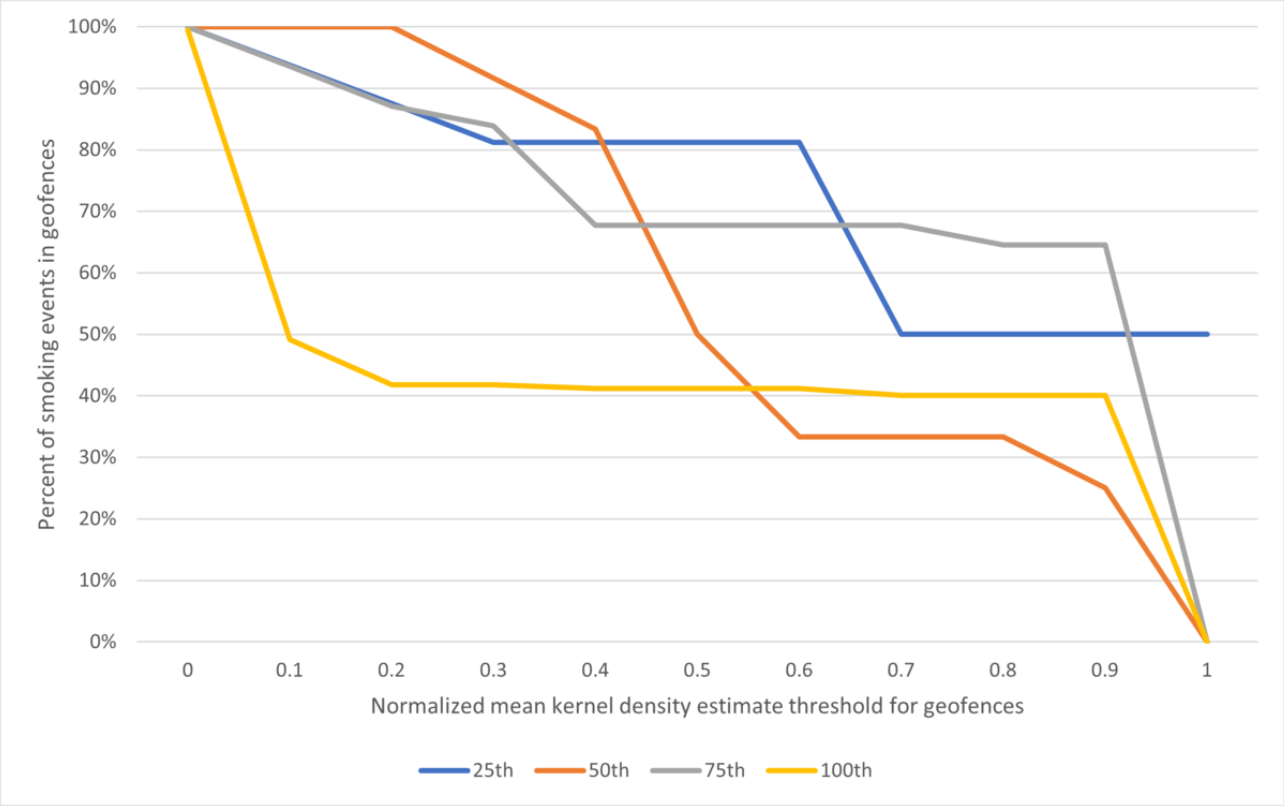
Threshold for Identifying High-Risk Census Blocks

The percent of smoking events captured in time-independent geofences at varying normalized mean KDE thresholds of high-risk for census blocks are presented in Figure 1. The figure displays the percentage of smoking events contained within constructed geofences if zones with normalized mean KDE values above each threshold are retained. For all but 1 participant, a normalized mean KDE threshold of 0.5 captured

50% or more of all possible smoking events within generated geofences.

We proceeded to construct geofences across 3-hour time intervals for census blocks with normalized mean KDEs greater than or equal to 0.3 because they captured at least 80% of smoking events for all but 1 case.

Figure 1. Percent of smoking events within geofences for census blocks among 4 participants at quartiles of ecological momentary assessment self-reported data.



Of note, 3 participants had no geofences generated for census blocks with a normalized mean KDE of 1.0 due to rounding differences (eg, rounding was done to 6 decimal places such that 0.999998 would not be included for this threshold). A 0.0 threshold should capture every block with at least one smoking event reported; however, we had 2 participants that were missing 1 block with at least one smoking event reported in it. This occurred because these smoking events occurred in very small, narrow blocks that were part of roadways, which consequently did not rasterize for a cell output size of 150 ft^2. In other words, these blocks were missing normalized mean KDE values because the blocks were smaller than the cell's output size. In some cases, this may have left out a block at risk for smoking. See Table S1 in Multimedia Appendix 1 for summary statistics of the 4 cases' normalized mean KDEs.

Comparison of Geofence Construction Methods of Census Blocks Compared to Fishnet Grid Across 3-Hour Time Intervals for Each Case

Across 3-hour time intervals, the average percentage of smoking events within geofences ranged from 36.4% to 100%. Geofences contained the highest percentage of smoking events between midnight and 3 AM and between 9 AM and 11:59 AM. Conversely, geofences contained the lowest percentage of smoking events reported between 6 AM and 8:59 AM.

Across the 4 cases, the 1000 ft^2 fishnet grid captured the highest percentage of smoking events for each 3-hour interval (Figure 2), both within cases and averaged across cases. Although there was no difference in the percentage of smoking events captured across geofence construction methods between midnight and noon, the constructed geofences looked very different from each other. Figure 3 compares the geofences constructed by the census block and 500 ft^2 fishnet grid methods for 1 case between noon and 2:59 PM. The census block method's geofence covers a larger area than the 500 ft^2 fishnet grid method's geofence from noon to 2:59 PM, even though they capture the same percentage of smoking events. Further, the census block method generated a wider range of high-risk blocks (normalized mean KDEs between 0.35 and 1.00) across the whole day, whereas the 500 ft^2 fishnet grid method generated a smaller range of high-risk cells (normalized mean KDEs between 0.64 and 1.00). By depicting the tertiles, we can identify how the distribution of normalized mean KDEs changed spatially across time and method within individuals. For example, Figure 3 tells us that in this period of noon to 2:59 PM, the 500 ft^2 fishnet grid method created a geofence only around a high-risk cell, whereas the census block method's geofence encompassed potentially low-risk blocks, yet both capture the same percentage of smoking events for this participant at this time of day. From this, we may visually compare geofence methods and ensure that the highest-risk areas within a geofence are captured, even if the geofence perimeter may differ by method.

Although the 500 ft^2 fishnet grid method captured a slightly higher percentage of smoking events from 9 AM to 6 PM for the case with the most data, the 100th percentile case (Figure 2), it captured the same percentage of smoking events at all other hours relative to the census block method. To understand why this may have happened, we examined the 100th percentile's smoking profile. Figure 4 shows where the 100th percentile participant smoked over 26 noncontiguous blocks generated without any thresholds. The blocks are predominantly yellow, representative of normalized mean KDEs below 0.2. This means that many of these blocks contain fewer smoking events. In fact, 73 of 177 (41.8%) smoking events occurred across 2 block areas, while the remaining 104 of 177 (58.2%) smoking events were dispersed across 24 noncontiguous block areas. This is highlighted in census block normalized mean KDE quintiles, in which the bottom 4 quintiles have normalized mean KDEs below 0.35.

As a result, a majority of the 100th percentile case's zones with smoking events will not result in a geofence to inform intervention delivery because only zones with normalized mean KDEs above 0.3 were retained. For this participant, our method identified that most zones were low-risk relative to other zones, and these low-risk zones would result in a low percentage of smoking events captured in geofences.

Figure 2. Percent of smoking events captured within each geofence method for 4 participants.

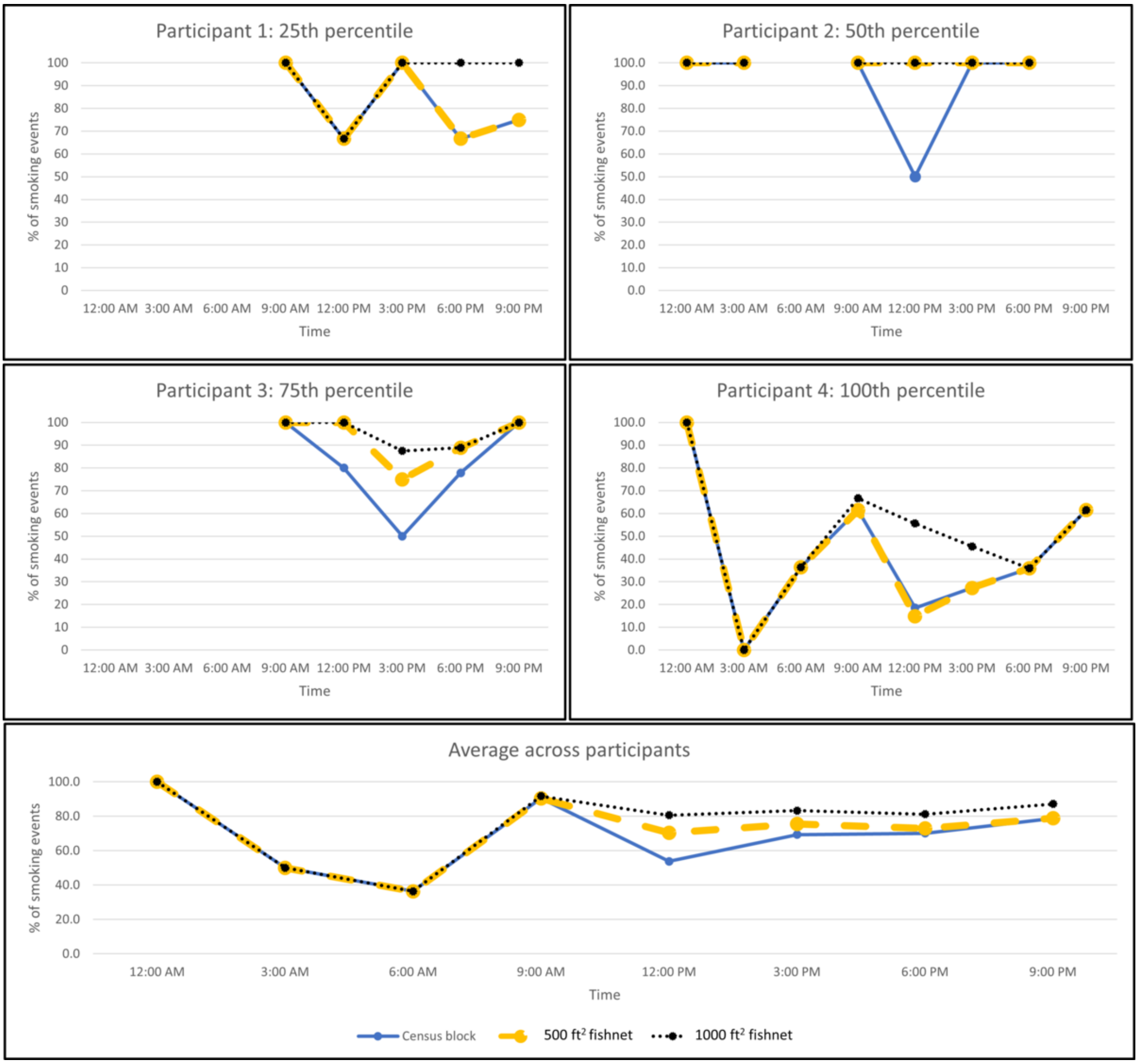


Figure 3. Geofences constructed by the census block versus 500 ft^2 fishnet grid at noon to 2:59 PM for 1 case. KDE: kernel density estimation.

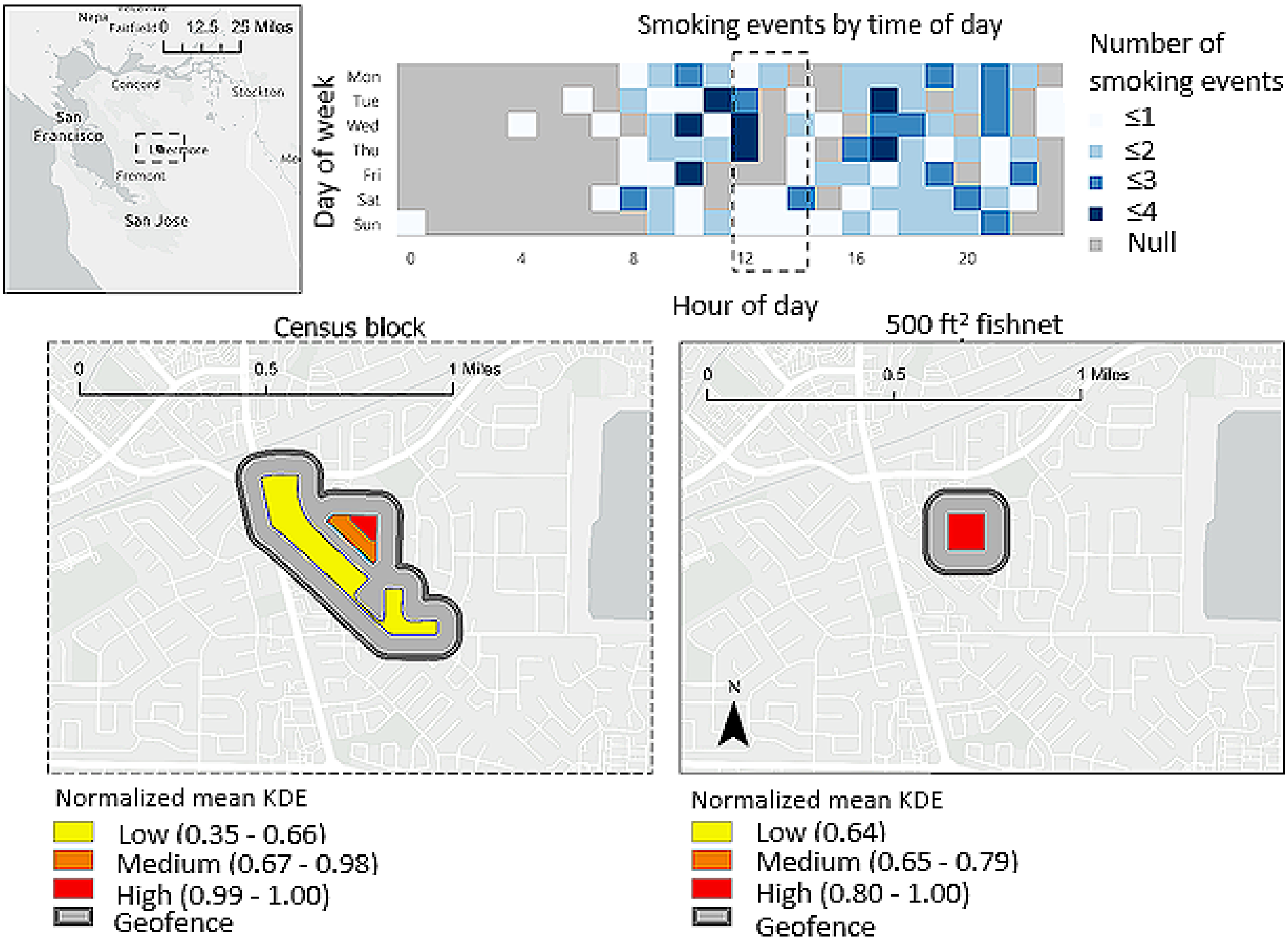
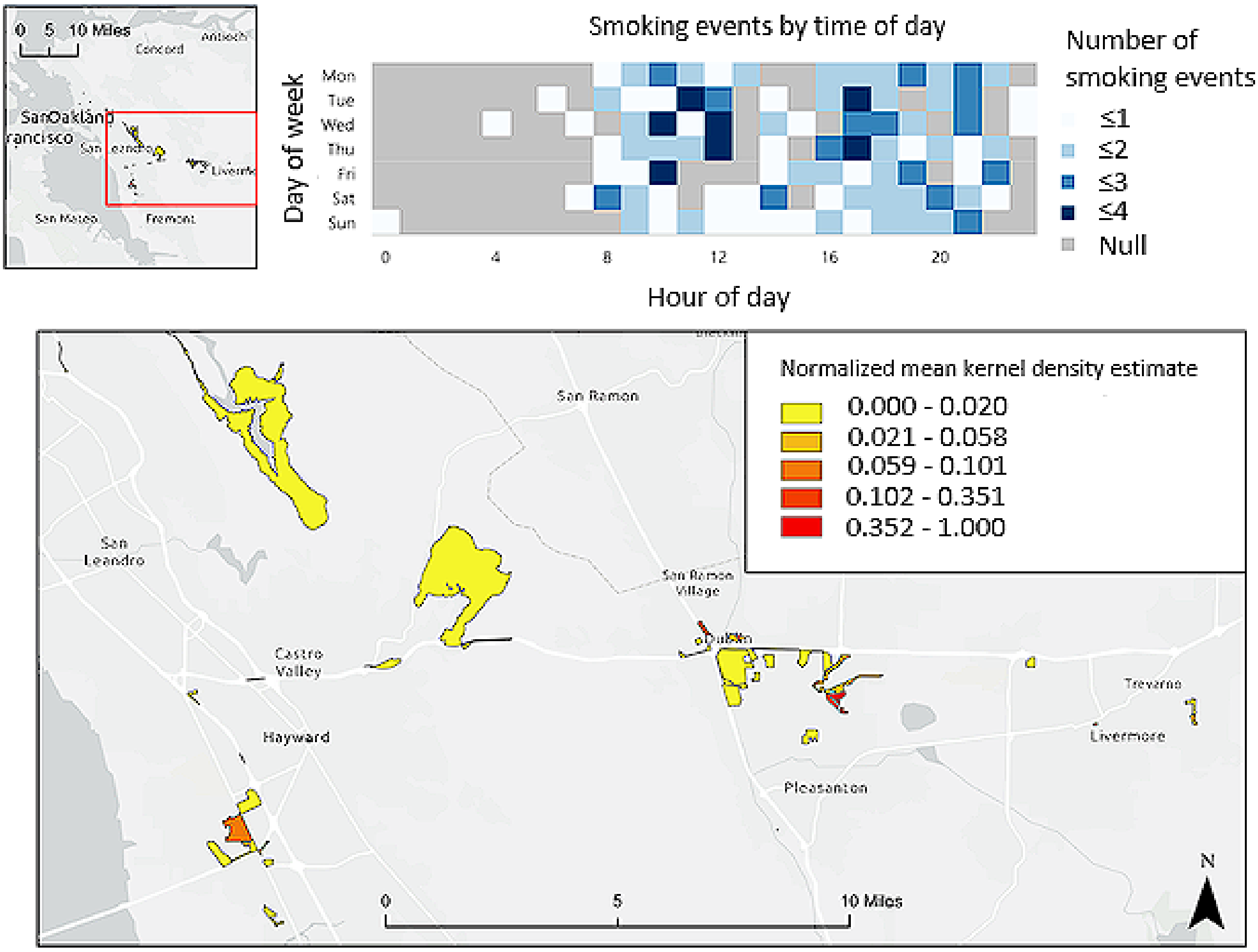


Figure 4. Census blocks with any smoking reports for 100th percentile case across all hours of the day.



Discussion

Overview

This study's objective was to design a spatial approach to identify and construct geofences around person- and time-specific high-risk smoking areas. We collected self-reported, smartphone-delivered surveys on smoking behaviors with passive GPS tracking from young adults who smoke.

Our study found that a kernel density approach for geofence construction could systematically label locations at high risk for smoking. Second, of the 3 methods we used to construct the geofence, the 1000 ft^2 fishnet grid captured the highest percentage of smoking events within and across the 4 cases. Last, we found that although methods may capture the same percentage of smoking events in the early and late hours of the day, the physical geofences appeared different from each other, which may affect intervention delivery.

Identified Locations at High Risk for Smoking Through Kernel Density Methods

To our knowledge, this is the first study to examine an individual's smoking risk profile using KDE methods. Kernel density approaches have already been applied in the tobacco literature to define risk in terms of the tobacco environment (eg, tobacco outlet density) [11,40], and this study demonstrates that KDE methods may also hold value for informing smartphone-based smoking cessation intervention delivery. For our case studies, we found that a normalized mean KDE threshold of 0.3 adequately defined smoking risk.

A KDE approach allowed us to define high-risk locations specific to an individual's smoking profile. Previous studies have examined risks unspecific to the individual, such as the occurrence of more than 4 smoking events within a geographical region [9]. Four smoking events, however, may be considered high-risk for some individuals and low-risk for others, relative to an individual's smoking patterns. If we had generated geofences around blocks with smoking events without classifying person-specific, high- and low-risk locations, the intervention delivery would be triggered at locations where the individual rarely smoked relative to other locations. For an ecological momentary intervention, the cumulation of intervention delivery at both efficient and inefficient times and appropriate and inappropriate locations could lead to intervention burden, which may undermine intervention effectiveness and adherence [41]. Prioritizing these high-risk locations and times may be able to reduce intervention burden and improve intervention delivery effectiveness.

The fact that the KDE methods captured most smoking events also highlights their use for people who smoke, especially if they tend to smoke in the same locations. We saw that our methods captured over 50% of smoking events for 3 of the 4 cases. The fourth case, however, was the heaviest smoker based on the self-reported data and frequently smoked in new locations. This may be due to the fact that more frequent smoking indicates greater nicotine dependence and, hence, a more regular need to maintain blood nicotine levels to avoid withdrawal [42]. In addition, this individual may have a more variable activity space than the other 3 cases, resulting in a more spatially dispersed smoking profile. Other strategies than the 1 employed here may be needed to improve intervention delivery for individuals with spatially dispersed smoking profiles.

Fishnet Grid Geofences Captured More Smoking Events Than Census Block Geofences

We recognized that the UGCoP may result in missing the "true causally relevant" geographic context [25] for smoking cessation geofences, so we generated geofences with 3 different geographical areas—census blocks, 500 ft^2 fishnet grid cells, and 1000 ft^2 fishnet grid cells. Of these 3 geographical areas, the 1000 ft^2 fishnet grid captured the highest percentage of smoking events, followed by the 500 ft^2 fishnet grid and census blocks. Even the most limited method (census blocks) still effectively captured over half of smoking events. Furthermore, the buffer size of 100 m around adjacent high-risk blocks helped capture some smoking events that would not have been captured otherwise.

Census blocks, however, are often delineated by roads [37] and may misrepresent smoking on the road (eg, the road itself is a very small block below the cell output size, or the smoking event may be forced to one of the adjacent blocks). In the transportation literature, a fishnet grid of half a mile predicted how people traveled better than aggregating to census blocks [43]. Thus, it is possible that the 1000 ft^2 fishnet grid may have captured more smoking if our cases were smoking in the car, which may have been missed in the 500 ft^2 fishnet grid and census block methods. Future studies can ask for the smoking location context to discern if this may be the case.

Selected Method for Effective Intervention Delivery

We found that interventions using the census blocks to create geofences will cover a wider area and may trigger more intervention delivery than interventions using 500 ft^2 fishnet grid cells for geofence construction. In another study that constructed geofences around locations with more than 4 smoking events, participants had mixed reviews for the frequency of intervention delivery, such that some reported too many alerts and others reported too few [9]. Aligned with the census block method, some participants may want more proactive alerts slightly further away from their usual smoking location, triggered by larger geofences. Future studies may want to compare these different geofence construction methods and their impact on intervention delivery and participant satisfaction.

As researchers seek to define how spaces are categorized as risky or not risky, the modifiable area unit problem needs to be considered as well [44]. As different geofence construction methods may yield different results in intervention delivery, there is a need for a standardized approach for constructing geofences to improve cross-study comparisons. Census blocks are stable for 10 years [37], while fishnet grid cells can shift based on entered parameters [38]. Clear descriptions of all parameters chosen to construct geofences are needed for reproducibility and to help the field develop standards. Depending on the study goal, studies that want to include demographic census data may use census blocks, while those solely interested in optimal intervention delivery may choose fishnet grids. There are also other methods that can weight noncensus data with demographic information [45] that can be further explored.

While a large number of smoking cessation apps are available on the Apple and Google Play app stores, many lack scientific evidence [46,47]. To support individuals in quitting smoking, we need evidence-based interventions that can promote behavior change through positive engagement and personal relevance (eg, appropriate time and place of intervention delivery) [48,49]. Our approach of identifying hot spots of self-reported smoking events and producing geofences that represent high-risk areas for smoking may be helpful to inform future smartphone-based smoking cessation interventions. The proposed approach for generating geofences will be used in an ongoing smoking cessation intervention study with young adults. Given its ability to capture a good percentage of smoking events across compliance levels and the specificity of the constructed geofences, we plan to automate a version of the 1000 ft^2 fishnet grids to create geofences for participants of an app-based smoking cessation intervention.

Limitations

Our study has several limitations. First, this was an observational study with individuals who were not ready to quit smoking, which may have impacted compliance to report all smoking events. We attempted to address this issue of compliance by selecting cases to study at quartiles of available data to evaluate the method's performance for various compliance scenarios. However, EMA compliance may vary significantly between locations, which could potentially affect the results. Second, we assumed smoking reports within 5 minutes of another report were due to technical issues or double reporting based on expert opinion and dropped these reports. Future studies would benefit from including an app feature that sends a follow-up survey to participants after reporting a high volume of smoking events to confirm the number of cigarettes smoked. Third, some census blocks did not rasterize, meaning they were missing a normalized mean KDE due to the block size being smaller than the cell output size. The fishnet grid method captured any of these points that were not rasterized in the census blocks method. Fourth, the risk threshold was determined solely based on the census blocks to allow for a comparison across methods. The census blocks had the greatest variability in normalized mean KDE, and the fishnet grids captured as many or more smoking events than the census blocks. Fifth, our KDE approach used a fixed kernel with a constant bandwidth (which tends to over smooth), whereas adaptive kernels can better capture the scale at which the point pattern process operates [50]. However, we ran KDE for each individual; therefore, the bandwidths were tailored to the individual. Finally, the temporal snapshots of the geofences may not accurately depict the "true" spatiotemporal point-pattern process since we employed a spatial KDE approach. Therefore, future research will employ spatiotemporal point-pattern methods, such as space-time KDE [50] to fully capture the space-time dynamics of the participants and the subsequent creation of geofences for smoking cessation interventions.

Conclusions

For ecological momentary interventions, it is important to optimize intervention message delivery to minimize intervention burden for participants. Prioritizing intervention delivery to high-risk locations and times may make these interventions relevant for individual participants and consequently improve intervention efficacy. A spatial approach of generating geofences based on high-risk zones (eg, fishnet girds for capturing a greater percentage of events or census blocks for linking spatial data with demographic information) identified through normalized mean KDE surpassing a chosen threshold may assist with prioritizing high-risk locations for intervention delivery, tailored to the needs of an individual participant. By stratifying event occurrence by periods of time, intervention messages can also be appropriately delivered throughout the day. Our study demonstrates that this method can capture a good percentage of smoking events within an urban and suburban environment and illustrates the potential for assessing if it improves person-specific smoking cessation intervention delivery and efficacy. Most importantly, this study highlights that researchers must carefully consider the implications of their chosen geographical unit when designing place-based interventions.

Acknowledgments

This work was supported by the National Institutes of Health (R01 CA246590). AL was also supported by a National Institute on Drug Abuse (NIDA) T32 grant T32 DA007292. The authors would like to thank the reviewers for their constructive feedback.

Authors' Contributions

JT conceptualized this study. AL and MRD created the geofence method. AL performed all analyses with help from MRD, who also created the fishnet grids. AL drafted the manuscript, which all authors reviewed and edited.

Conflicts of Interest

MBM served as a paid expert witness in litigation sponsored by the Public Health Advocacy Institute against RJ Reynolds. This arrangement has been reviewed and approved by Johns Hopkins University in accordance with its conflict of interest policies.

Multimedia Appendix 1
Normalized mean KDE statistics per case.
[DOCX File , 17 KB - mhealth_v11i1e43990_app1.docx]

References

1. Schulenberg J, Johnston L, O'Malley P, Bachman J, Miech R, Patrick M. Monitoring the Future national survey results on drug use, 1975-2019: Volume II, college students and adults ages 19-60. 2020. URL: https://files.eric.ed.gov/fulltext/ED608266.pdf [accessed 2022-09-10]
2. U.S. Department of Health and Human Services. The health consequences of smoking-50 years of progress: A report of the Surgeon General. 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK179276/ [accessed 2022-09-10] [WebCite Cache ID 24455788]
3. Watkins SL, Thrul J, Max W, Ling PM. Cold turkey and hot vapes? A national study of young adult cigarette cessation strategies. Nicotine Tob Res 2020 Apr 21;22(5):638-646 [FREE Full text] [doi: 10.1093/ntr/nty270] [Medline: 30590749]
4. Villanti AC, Niaura RS, Abrams DB, Mermelstein R. Preventing smoking progression in young adults: the concept of prevescalation. Prev Sci 2019 Apr;20(3):377-384 [FREE Full text] [doi: 10.1007/s11121-018-0880-y] [Medline: 29525899]
5. Pew Research Center. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-09-10]
6. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
7. Conklin CA, Robin N, Perkins KA, Salkeld RP, McClernon FJ. Proximal versus distal cues to smoke: the effects of environments on smokers' cue-reactivity. Exp Clin Psychopharmacol 2008 Jun;16(3):207-214 [FREE Full text] [doi: 10.1037/1064-1297.16.3.207] [Medline: 18540780]
8. Kirchner TR, Cantrell J, Anesetti-Rothermel A, Ganz O, Vallone DM, Abrams DB. Geospatial exposure to point-of-sale tobacco: real-time craving and smoking-cessation outcomes. Am J Prev Med 2013 Oct;45(4):379-385 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.016] [Medline: 24050412]

9. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
10. Ajebon MO, Asikhia MO. Geographical analysis of smoking in leeds: a GIS-based location-allocation technique for the optimal location of smoking cessation services. J Geogr Inf Syst 2013;5(3):269-280 [FREE Full text] [doi: 10.4236/jgis.2013.53026]
11. Brooks MM, Siegel SD, Curriero FC. Characterizing the spatial relationship between smoking status and tobacco retail exposure: Implications for policy development and evaluation. Health Place 2021 Mar;68:102530 [FREE Full text] [doi: 10.1016/j.healthplace.2021.102530] [Medline: 33609995]
12. Abo-Tabik M, Costen N, Darby J, Benn Y. Towards a smart smoking cessation app: A 1D-CNN model predicting smoking events. Sensors (Basel) 2020 Mar 17;20(4):1099 [FREE Full text] [doi: 10.3390/s20041099] [Medline: 32079359]
13. Alsaqer M, Hilton B, Horan T, Aboulola O. Performance assessment of geo-triggering in small geo-fences: accuracy, reliability, and battery drain in different tracking profiles and trigger directions. Procedia Engineering 2015;107(4):337-348 [FREE Full text] [doi: 10.1016/j.proeng.2015.06.090]
14. Besoain F, Perez-Navarro A, Jacques Aviñó C, Caylà JA, Barriga NA, Garcia de Olalla P. Prevention of HIV and other sexually transmitted infections by geofencing and contextualized messages with a gamified app, UBESAFE: design and creation study. JMIR Mhealth Uhealth 2020 Mar 17;8(3):e14568 [FREE Full text] [doi: 10.2196/14568] [Medline: 32181752]
15. Kwan MP. From place-based to people-based exposure measures. Soc Sci Med 2009 Nov;69(9):1311-1313. [doi: 10.1016/j.socscimed.2009.07.013] [Medline: 19665828]
16. Kirchner TR, Shiffman S. Spatio-temporal determinants of mental health and well-being: advances in geographically-explicit ecological momentary assessment (GEMA). Soc Psychiatry Psychiatr Epidemiol 2016 Sep;51(9):1211-1223 [FREE Full text] [doi: 10.1007/s00127-016-1277-5] [Medline: 27558710]
17. Wright WG, Rafferty AP, Winterbauer N, Locklear K, Tucker-McLaughlin M. Geofencing: mobile technology as a health promotion tool to raise awareness of a dental clinic in rural North Carolina. J Rural Health 2021 Jun;37(3):667-674. [doi: 10.1111/jrh.12501] [Medline: 32761672]
18. Coral R, Esposito F, Weinstock J. Don't go there: a zero-permission geofencing app to alleviate gambling disorders. Las Vegas, NV: A zero-permission geofencing app to alleviate gambling disorders. Proceedings of the 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC); 2020 Presented at: 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC); 10-13 Jan. 2020; Las Vegas, NV URL: https://ieeexplore.ieee.org/xpl/conhome/9035667/proceeding [doi: 10.1109/ccnc46108.2020.9045251]
19. Garzon SR, Walther S, Pang S, Deva B, Küpper A. Urban air pollution alert service for smart cities. New York, NY: Association for Computing Machinery; 2018 Presented at: Proceedings of the 8th International Conference on the Internet of Things; October 15 - 18, 2018; Santa Barbara, California, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3277593 [doi: 10.1145/3277593.3277599]
20. Jahmunah V, Sudarshan VK, Oh SL, Gururajan R, Gururajan R, Zhou X, et al. Future IoT tools for COVID-19 contact tracing and prediction: A review of the state-of-the-science. Int J Imaging Syst Technol 2021 Jun;31(2):455-471 [FREE Full text] [doi: 10.1002/ima.22552] [Medline: 33821093]
21. Zenk SN, Matthews SA, Kraft AN, Jones KK. How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research? Health Place 2018 May;51:52-60 [FREE Full text] [doi: 10.1016/j.healthplace.2018.02.004] [Medline: 29549754]
22. Carlos HA, Shi X, Sargent J, Tanski S, Berke EM. Density estimation and adaptive bandwidths: a primer for public health practitioners. Int J Health Geogr 2010 Jul 23;9:39 [FREE Full text] [doi: 10.1186/1476-072X-9-39] [Medline: 20653969]
23. McFadyen A, Martin T, Perez T. Low-level collision risk modelling for unmanned aircraft integration and management. 2018 Presented at: Proceedings of the 2018 IEEE Aerospace Conference; 3-10 March 2018; Big Sky, MT, USA. [doi: 10.1109/aero.2018.8396463]
24. Ryan C, Murphy F, Mullins M. Spatial risk modelling of behavioural hotspots: Risk-aware path planning for autonomous vehicles. Transportation Research Part A: Policy and Practice 2020 Apr;134:152-163 [FREE Full text] [doi: 10.1016/j.tra.2020.01.024]
25. Kwan MP. The uncertain geographic context problem. Ann Assoc Am Geogr 2012;102(5):958-968 [FREE Full text] [doi: 10.1080/00045608.2012.687349]
26. Van Gucht D, Van den Bergh O, Beckers T, Vansteenwegen D. Smoking behavior in context: where and when do people smoke? J Behav Ther Exp Psychiatry 2010 Jun;41(2):172-177 [FREE Full text] [doi: 10.1016/j.jbtep.2009.12.004] [Medline: 20060515]
27. Qualtrics. 2020. URL: https://www.qualtrics.com/ [accessed 2023-05-31]
28. PiLR. URL: https://pilrhealth.com/ [accessed 2023-05-31]
29. McQuoid J, Thrul J, Ling P. A geographically explicit ecological momentary assessment (GEMA) mixed method for understanding substance use. Soc Sci Med 2018 Apr;202:89-98 [FREE Full text] [doi: 10.1016/j.socscimed.2018.02.014] [Medline: 29518701]

30. Nguyen N, McQuoid J, Ramo D, Holmes LM, Ling PM, Thrul J. Real-time predictors of smoking among sexual minority and heterosexual young adults: An ecological momentary assessment study. Drug Alcohol Depend 2018 Nov 01;192:51-58 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.07.021] [Medline: 30212756]
31. Thrul J, Gubner NR, Nguyen N, Nguyen C, Goodell EA, Holmes LM, et al. Perceived reward from using cigarettes with alcohol or cannabis and concurrent use: A smartphone-based daily diary study. Addict Behav 2021 Mar;114:106747 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106747] [Medline: 33307406]
32. Delmelle E, Desjardins M. Point pattern analysis. In: Kobayashi A, editor. International Encyclopedia of Human Geography. 2nd ed. Amsterdam: Elsevier Science & Technology; 2020:171-179.
33. Delmelle E, Delmelle EC, Casas I, Barto T. H.E.L.P: A GIS-based Health Exploratory AnaLysis Tool for Practitioners. Appl. Spatial Analysis 2010 Mar 6;4(2):113-137. [doi: 10.1007/s12061-010-9048-2]
34. Bailey T, Gatrell A. Interactive Spatial Data Analysis. Harlow, Essex, UK: Longman Scientific & Technical; 1995.
35. Yang Y, Diez-Roux AV. Walking distance by trip purpose and population subgroups. Am J Prev Med 2012 Jul;43(1):11-19 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.015] [Medline: 22704740]
36. ESRI. How the zonal statistics tools work. ArcGIS Pro. URL: https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/how-zonal-statistics-works.htm [accessed 2022-09-10]
37. Rossiter K. What are census blocks? The United States Census Bureau. 2011. URL: https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html [accessed 2022-09-10]
38. ESRI. How Create Fishnet works. ArcGIS Pro. URL: https://pro.arcgis.com/en/pro-app/2.8/tool-reference/data-management/how-create-fishnet-works.htm [accessed 2022-09-10]
39. ArcGIS Pro. Esri Inc. 2022. URL: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview [accessed 2023-05-31]
40. Valiente R, Escobar F, Urtasun M, Franco M, Shortt NK, Sureda X. Tobacco retail environment and smoking: a systematic review of geographic exposure measures and implications for future studies. Nicotine Tob Res 2021 Aug 04;23(8):1263-1273 [FREE Full text] [doi: 10.1093/ntr/ntaa223] [Medline: 33155040]
41. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 01;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
42. Shiffman S, Paty J. Smoking patterns and dependence: contrasting chippers and heavy smokers. J Abnorm Psychol 2006 Aug;115(3):509-523. [doi: 10.1037/0021-843X.115.3.509] [Medline: 16866591]
43. Zhang M, Kukadia N. Metrics of urban form and the modifiable areal unit problem. Transportation Research Record 2005 Jan 01;1902(1):71-79. [doi: 10.1177/0361198105190200109]
44. Wong DWS. The modifiable areal unit problem (MAUP). In: Janelle DG, Warf B, Hansen K, editors. WorldMinds: Geographical Perspectives on 100 Problems. Dordrecht, Netherlands: Springer; 2004.
45. Mulrooney T, Foster R, Jha M, Beni LH, Kurkalova L, Liang CL, et al. Using geospatial networking tools to optimize source locations as applied to the study of food availability: A study in Guilford County, North Carolina. Applied Geography 2021 Mar;128:102415 [FREE Full text] [doi: 10.1016/j.apgeog.2021.102415]
46. Cook VE, Ellis AK, Hildebrand KJ. Mobile health applications in clinical practice: pearls, pitfalls, and key considerations. Ann Allergy Asthma Immunol 2016 Aug;117(2):143-149. [doi: 10.1016/j.anai.2016.01.012] [Medline: 27499542]
47. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
48. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. Nicotine Tob Res 2005 Apr;7(2):207-216 [FREE Full text] [doi: 10.1080/14622200500055319] [Medline: 16036277]
49. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Mar 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
50. Hohl A, Tang W, Casas I, Shi X, Delmelle E. Detecting space-time patterns of disease risk under dynamic background population. J Geogr Syst 2022;24(3):389-417 [FREE Full text] [doi: 10.1007/s10109-022-00377-7] [Medline: 35463848]

Abbreviations

EMA: ecological momentary assessment
KDE: kernel density estimation
mHealth: mobile health
UGCoP: uncertain geographic context problem
UTC: Coordinated Universal Time

Edited by L Buis; submitted 07.11.22; peer-reviewed by G Martin, G Perez; comments to author 20.12.22; revised version received 23.02.23; accepted 07.04.23; published 16.06.23.

Please cite as:
Luken A, Desjardins MR, Moran MB, Mendelson T, Zipunnikov V, Kirchner TR, Naughton F, Latkin C, Thrul J
Using Smartphone Survey and GPS Data to Inform Smoking Cessation Intervention Delivery: Case Study
JMIR Mhealth Uhealth 2023;11:e43990
URL: https://mhealth.jmir.org/2023/1/e43990
doi:10.2196/43990
PMID:37327031

Original Paper

Incorporating Consumers' Needs in Nutrition Apps to Promote and Maintain Use: Mixed Methods Study

Sandra van der Haar[1], MSc; Ireen Raaijmakers[2], MSc; Muriel C D Verain[2], PhD; Saskia Meijboom[1], BSc

[1]Wageningen Food & Biobased Research, Wageningen University & Research, Wageningen, Netherlands

[2]Wageningen Economic Research, Wageningen University & Research, Wageningen, Netherlands

Corresponding Author:
Sandra van der Haar, MSc
Wageningen Food & Biobased Research
Wageningen University & Research
Bornse Weilanden 9
Wageningen, 6708 WG
Netherlands
Phone: 31 317480171
Email: sandra.vanderhaar@wur.nl

Abstract

Background: Nutrition apps seem to be promising tools for supporting consumers toward healthier eating habits. There is a wide variety of nutrition apps available; however, users often discontinue app use at an early stage before a permanent change in dietary behavior can be achieved.

Objective: The main objective of this study was to identify, from both a user and nonuser perspective, which functionalities should be included in nutrition apps to increase intentions to start and maintain use of these apps. A secondary objective was to gain insight into reasons to quit using nutrition apps at an early stage.

Methods: This study used a mixed methods approach and included a qualitative and a quantitative study. The qualitative study (n=40) consisted of a home-use test with 6 commercially available nutrition apps, followed by 6 focus group discussions (FGDs) to investigate user experiences. The quantitative study was a large-scale survey (n=1420), which was performed in a representative sample of the Dutch population to quantify the FGDs' results. In the survey, several app functionalities were rated on 7-point Likert scales ranging from 1 (very unimportant) to 7 (very important).

Results: A total of 3 different phases of app use, subdivided into 10 user-centric app aspects and 46 associated app functionalities, were identified as relevant nutrition app elements in the FGDs. Relevance was confirmed in the survey, as all user-centric aspects and almost all app functionalities were rated as important to include in a nutrition app. In the starting phase, a clear introduction (mean 5.45, SD 1.32), purpose (mean 5.40, SD 1.40), and flexible food tracking options (mean 5.33, SD 1.45) were the most important functionalities. In the use phase, a complete and reliable food product database (mean 5.58, SD 1.41), easy navigation (mean 5.56, SD 1.36), and limited advertisements (mean 5.53, SD 1.51) were the most important functionalities. In the end phase, the possibility of setting realistic goals (mean 5.23, SD 1.44), new personal goals (mean 5.13, SD 1.45), and continuously offering new information (mean 4.88, SD 1.44) were the most important functionalities. No large differences between users, former users, and nonusers were found. The main reason for quitting a nutrition app in the survey was the high time investment (14/38, 37%). This was also identified as a barrier in the FGDs.

Conclusions: Nutrition apps should be supportive in all 3 phases of use (start, use, and end) to increase consumers' intentions to start and maintain the use of these apps and achieve a change in dietary behavior. Each phase includes several key app functionalities that require specific attention from app developers. High time investment is an important reason to quit nutrition app use at an early stage.

(JMIR Mhealth Uhealth 2023;11:e39515) doi:10.2196/39515

KEYWORDS

mobile health; mHealth; mHealth apps; nutrition apps; diet apps; consumer needs; app; use; nutrition; tool; consumer; eating; habit; users; dietary behavior; reliable; food; database; time; developers; mobile phone

Introduction

Background

Healthy dietary habits play a crucial role in preventing obesity and noncommunicable diseases such as cardiovascular diseases, cancer, and diabetes mellitus [1,2]. Despite its importance, the dietary intake of many consumers in the Western world is suboptimal (eg, excessive intake of sodium and insufficient intake of whole grains and fruits), leading to a high global disease burden [3]. Therefore, there is an urgent need for interventions and tools that stimulate and support a healthy eating pattern. Increased access to smartphones, tablets, and wearables has caused an increase in the popularity of mobile health (mHealth) apps. Currently, over 350,000 mHealth apps are available in the health and fitness category in various app stores worldwide [4]. There is evidence that mHealth apps are effective or likely to be effective in stimulating healthier behaviors, such as increasing physical activity [5,6], reducing sedentary behavior [5], improving dietary habits and intake [5-7], and losing weight [8,9]. Nutrition apps are a part of the mHealth category and specifically focus on tracking food intake and providing dietary advice. Most of these apps function as food diaries in which users log their daily food intake, either by a text search or a barcode scanner [10]. The app subsequently gives the user an overview of daily amounts of calories and other nutrients consumed and provides them with dietary advice. Nutrition apps are considered promising tools for supporting consumers in the transition toward a healthier diet. A study by Wang et al [11] showed that consumers perceive the effectiveness of nutrition apps as rather high. In their study, they included both nutrition and physical activity apps. The use of both types of apps influenced action, consciousness, self-education about nutrition and physical activity, and social life (eg, by sharing dietary experiences in web-based social networks). Furthermore, it facilitated maintaining a healthy diet and exercising more [11].

An important question is which elements are important for nutrition apps to be effective in achieving healthy eating behaviors. Several studies examined the application of behavior change techniques (BCTs) in mHealth apps. BCTs are components of behavior change interventions that can be used alone or in combination with other BCTs, such as goal setting, self-monitoring, and feedback [12]. The inclusion of BCTs seems beneficial to the quality of mHealth apps [13-15] and might in turn influence consumer behavior [16]. This is in line with previous research showing that health behavior interventions are more effective when they integrate such techniques [17,18]. According to several studies, the extent to which BCTs are incorporated into mHealth apps is still insufficient at this point [15,16,19] or is only sufficient in paid versions of the app [20].

Besides the incorporation of science-based components in mHealth interventions, such as BCTs, it is of great importance to focus on the issue of implementation. A key factor in mHealth implementation is the willingness and capability of users to successfully engage with a tool or app [21]. This is an important precondition for both the *efficacy*, the capacity of a given mHealth intervention in a controlled setting, and the *effectiveness*, to have a meaningful effect on users in real life [22]. The user-friendliness of mHealth app use of advertisements, price, and protection of personal data and privacy are examples of aspects that potentially influence the implementation and acceptance of these tools [23-26].

Nowadays, a wide variety of nutrition apps are available; however, only a small group of Dutch consumers (11%) make use of such apps [27]. Users often prematurely discontinue use before a change in dietary behavior can be achieved, indicating possible issues with the implementation of these apps. This was shown by Helander et al [28] in a retrospective study. They concluded that most people who tried out a free mobile app for dietary self-monitoring did not continue using it actively [28]. There might be various barriers for consumers to use a nutrition app or to quit its use at an early stage. A recent systematic literature review by König et al [29] identified 328 barriers and facilitators for nutrition app use. The usability of the app was the most frequently identified barrier in this review [29]. The user burden of nutrition apps is rather high because tracking all eating and drinking moments is a time-consuming activity. Furthermore, there can be several issues with the food tracking feature and the underlying food database in nutrition apps. A study by Ziesemer et al [30] demonstrated that usability issues related to tracking food intake might impact the willingness to record eating events. In addition, Vasiloglou et al [25] showed that consumers would not select nutrition apps that have issues related to their food and nutrient databases, such as an incomplete product list or incorrect estimations of nutrients.

Objectives

To summarize, several factors could contribute to consumers' intentions to use nutrition apps and the early discontinuation of these apps. Therefore, the main objective of this study was to identify, from both a user and nonuser perspective, which functionalities should be included in nutrition apps to increase intentions to start and maintain use of these apps. A secondary objective was to gain insight into reasons to quit using nutrition apps at an early stage.

Methods

Study Design

The study followed a mixed methods approach and consisted of 2 parts: a qualitative and a quantitative study. The qualitative study consisted of a home-use test with commercially available nutrition apps, followed by focus group discussions (FGDs). The results of the FGDs served as a basis for the quantitative part, a survey that aimed to quantify these results in a large, representative sample of the Dutch population.

Ethics Approval and Informed Consent

All participants provided written informed consent for participation in the study. In addition, participants in the qualitative study provided consent for audio recordings of the discussions. Ethics approval for the study was obtained from the Social Ethics Committee of Wageningen University & Research in the Netherlands.

Qualitative Study

Recruitment and Study Procedures

Participants were recruited from the Wageningen Food and Biobased Research consumer database. This database consists of consumers who are interested in participating in nutrition and health research and live in the Wageningen region. An email invitation to participate in this study was sent to a subsample of the panel (750 consumers) using random selection. Inclusion criteria were age between 18 and 60 years and familiarity with smartphone and app use. A total of 62 participants signed up for the study, of which 48 (77%) were invited to participate. There were 3 dropouts during the home-use test and 5 no-shows at the FGDs; therefore, a total of 40 participants completed the study.

Study participants first took part in a home-use test. In total, 6 commercially available, free nutrition apps in which food intake could be tracked were selected for this test: MyFitnessPal, FatSecret, Lifesum, Mijn Eetmeter, FoodProfiler, and SamenGezond. The apps were selected based on a short literature search to identify the prerequisites of successful mHealth apps and their differences in BCTs (goal setting, goal tracking, monitoring, feedback, rewards, social support, identification, game elements, and personalization) and other app functionalities (prompts, synchronization with other apps and devices, and costs). The apps were used as a tool to start the conversation on user experiences and critical app elements; the aim was not to test or rate these specific apps. Participants were asked to download and use 1 of these 6 apps for a period of 3 weeks. Some participants had previous experience with 1 of the apps used in the test. In that case, they were assigned to a different app that was new to them because we also wanted to include their first experience of using the app. After 3 weeks of app use, 6 semistructured FGDs (1 per included app) of 1.5 hours were organized at the Wageningen University & Research campus. The discussions were led by a professional focus group moderator. A focus group guide was designed and used during the discussion to ensure the comparability of the 6 FGDs. The main objective of the discussions was to identify, from a user perspective, what were the critical elements for successfully monitoring and supporting healthy eating behaviors. Another objective was to explore possible reasons to quit using the specific app. In each session, 6 to 8 consumers participated, who all tested the same nutrition app before their session. The 3 main topics addressed in the FGDs were general smartphone use, use of health and nutrition apps, and user experiences with the tested nutrition app. The latter topic was discussed extensively. Participants were asked to describe their positive and negative experiences and how often they had used the app. Subsequently, the app functionalities were discussed and evaluated regarding their usefulness. The functionalities differed per app, but in all focus groups, both predefined intervention components of the app (BCTs such as goal setting, feedback, reward, social support, and knowledge) and the implementation of these components (eg, use aspects such as reminders, chat, synchronization with other devices, and gamification element) were discussed. Audio recordings were conducted, and minutes were recorded for each session. Upon completion of the study, the participants received an incentive of €50 (conversion rate at the time of the study: €1=US $1.12) for their time investment.

Analyses

As a first step, the minutes and recordings of the discussions were analyzed by creating a descriptive matrix per question to identify common themes. Through this matrix, the reasons for discontinuation were identified. The model shown in Figure 1 was built using a bottom-up approach. On the basis of the app evaluation, user quotes from the participants were translated into the app properties. For example, the user quote "I would use the products that were placed under the dinner category at other moments" was translated into the key app property: "logical product categories." Thus, all user quotes related to the tested app were translated, resulting in a list of 46 properties (Table S1 in Multimedia Appendix 1). These app properties were then assigned to 10 different categories (also called "user-centric app aspects"; Table S1 in Multimedia Appendix 1). These categories were partly derived from the predefined list of BCTs and other app properties in the focus group guide (eg, "Monitoring") and partly from topics participants came up with themselves in the discussion (eg, "Database"). In this manner, the model with 10 different user-centric aspects and 46 app properties was built. In the analysis, it was not about the frequency of the quotes but about their unicity, as the aim was to generate a complete overview of user experiences.

The 3 different phases of app use were identified as the final step. Some of the user-centric app aspects particularly occurred when the user installed the app and during the first (few) time(s) of use, the *start phase*. Some aspects typically occurred during daily use, the *usephase*. Most participants reached a point where the app would be abandoned or used permanently, the *end phase*. This was related to a combination of aspects in the 2 earlier phases and the continuous engagement of the user. The phases were placed in the outer ring of the model.

The analyses of all 6 FGDs were performed by the same professional focus group moderator who had led the discussions. At least 1 of the researchers was present at each focus group session, and afterward, the minutes and analyses of the results were reviewed by the researcher.

Figure 1. Overview of the 3 phases of nutrition app use subdivided into 10 “user-centric app aspects.”.

Quantitative Study

Sampling and Study Procedures

A web-based survey was conducted in a representative sample of the Dutch adult population (>18 years), familiar with the use of smartphone apps and either with or without experience regarding the use of nutrition apps. The survey was administered by a professional market research agency (MSI-ACI Europe Ltd). Quota sampling was applied to obtain a representative sample for age, gender, level of education, region, and income. Participants were approached by email to fill out a web-based self-administered survey and received an incentive in the form of credits for a personal saving system.

The survey included questions on nutrition app use, reasons for using and for not using nutrition apps, and the importance of 46 nutrition app functionalities that resulted from the FGDs. In total, 1500 participants completed the survey. During data cleaning, 80 participants were removed from the analyses because they showed no dispersion in their answers on the importance of app functionalities, suggesting that they did not fill out this part of the survey in earnest.

Measures

Nutrition App Use

Respondents were asked to indicate whether they make use of nutrition apps or made use of nutrition apps in the past. On the basis of their responses, they were categorized into former users, previous users, and nonusers. Subsequently, they were asked to indicate their reasons for (not) using nutrition apps, for example, “Because I want to lose weight” or “I never heard of nutrition apps.” Former users were asked what was their reason or were their reasons for quitting the use of the app, for example, “It cost me too much time.” Respondents could select multiple options from a predefined list of reasons or fill in an open answer.

Nutrition App Functionalities

The 46 nutrition app functionalities that resulted from the FGDs (Table S1 in Multimedia Appendix 1) were included in the survey. For each app functionality (eg, “A quick and easy entry of food products”), participants were asked how important they thought this functionality was to include in a nutrition app. The items were randomized into 4 subsets and assessed on 7-point Likert scales ranging from 1 (very unimportant) to 7 (very important).

Demographics

Age, gender, and education level were included in the survey to analyze the sample on sociodemographic characteristics.

Statistical Analysis

Categorical variables were displayed as frequencies and percentages, and numeric variables were displayed as mean (SD). Respondents were categorized into *users* and *nonusers* according to their nutrition app use behavior. The group of users consisted of *current users* and *former users*. The number of participants in each group and the percentage of the total study population were calculated. The mean (SD) scores were calculated for the importance ratings of each nutrition app functionality. The top 3 most important app functionalities per phase were created based on these mean ratings. The 10 user-centric app aspects included multiple app functionalities. Per user-centric app aspect, mean (SD) scores were calculated, combining all app functionalities within the user-centric app aspects (refer to Table S1 in Multimedia Appendix 1 for an overview). One-way ANOVA was used to compare the means of the 3 different groups (nonusers, current users, and former users). The Brown-Forsythe ANOVA with Games-Howell post hoc tests was applied to account for unequal samples and variances. Statistical significance was set at $P<.001$ for all analyses. Statistical analyses were performed using SPSS software (version 25.0; IBM Corp).

Results

Qualitative Study

Phases of App Use and User-Centric App Aspects

A total of 40 participants (8 male individuals and 32 female individuals, with a mean age of 40.9, SD 14.1 years) participated in the FGDs. None of the participants had a low education level, 15% (6/40) of the participants had a medium education level, and 85% (34/40) of the participants had a high education level. The FGDs revealed that users go through 3 different phases when using a nutrition app: start, use, and end. Each phase includes a range of key aspects or categories. In total, 10 such "user-centric app aspects" were identified (Figure 1). Each of these aspects includes a total of 46 different key app functionalities (refer to Table S1 in Multimedia Appendix 1 for a complete overview).

App Functionalities

In the *starting phase*, the purpose (user-centric app aspect 1, Figure 1) of the app should be clear immediately, and a clear introduction (user-centric app aspect 2, Figure 1) to the different functionalities of the app should be present (user: "A tutorial that you can view optionally would be helpful"). The next aspect is personalization (user-centric app aspect 3, Figure 1) of the app by entering personal data and goals (user: "When you create your own list of recipes, you only have to fill it in once, which is convenient"). In addition, certain other app functionalities must be adjustable according to personal wishes, such as how often notifications appear. During the *use phase*, users go through several aspects of the app, either sequentially or simultaneously. First, user-friendliness (user-centric app aspect 4, Figure 1) is of great importance in this phase, especially a quick and easy daily food intake entry (user: "Efficient entry is important. I am very impatient"). Moreover, the product database (user-centric app aspect 5, Figure 1) in the app must be of good quality and not be contaminated with duplicate products or incorrect nutritional information (user: "There is no added value of having so many similar products in the list"). Furthermore, the information (user-centric app aspect 6, Figure 1) or advice provided by the app should educate the user (user: "I didn't know almonds contained so many calories."). In addition, the user must gain sufficient insight into their own dietary behavior and progress by monitoring (user-centric app aspect 7, Figure 1) functions in the app (user: "I adjusted my behaviour based on the daily overview of what I ate"). Visualizing progress toward achieving personal goals can be helpful. Users prefer to receive positive feedback (user: "I liked the encouraging tone of voice of the feedback messages"; user-centric app aspect 8, Figure 1), a variety of different feedback messages, or a game element as a way to provide feedback. In addition, a reward system where, for example, credits can be earned by keeping up with entering food intake daily has a stimulating effect on continued use of the app. Furthermore, some appreciate the possibility of communication (user-centric app aspect 9, Figure 1) with other users (via a community, forum, or social media) or with a web-based coach (user: "I entered a goal to eat more vegetable and the coach would give tips"). In the *end phase*, continuous engagement (user-centric app aspect 10, Figure 1) of the user is particularly important (user: "It took forever to fill the app. That really demotivated me"). In addition, the app should continuously offer new and relevant information to keep users interested and engaged.

Reasons to Quit Use

Entering Food Intake and the Database

Most participants in the home-use test discontinued the use of the nutrition app prematurely. Although some participants did mention a few advantages, such as an increased awareness of personal food intake and dietary habits, most participants experienced too many disadvantages. The poor user-friendliness of the app made the time investment too high for most users. This was mainly linked to the time-consuming task of registering daily food intake. If entering food products could not be achieved quickly and easily, this was a big barrier (user: "I had to type in the same food products over and over again"). In some apps, users could add their own food products to the food product database. This caused contamination of the database, with too many details and options (user: "Full-fat yoghurt had so many entries, with all different amounts of calories."). In contrast, sometimes the database was too limited or incomplete for specific product categories (user: "There were five different options for chocolate milk, but only one type of cheese"). These issues sometimes caused the user to not trust the content of the database (user: "The app showed very different calorie amounts for different apples, it made me question the reliability").

Feedback and Advice

In some cases, the way feedback and advice were framed caused the user's continuous engagement to decline. Some apps used a patronizing tone of voice (user: "I want to be aware of what I eat, but don't be judged through the advice given.") or were too rigid in the feedback provided (user: "One calorie too much and you are in the red zone. That's too abrupt"). For an app to remain relevant and triggering for a longer period, it needs to constantly offer new information to the user. Some users indicated that the advice or tips were too repetitive or already known (user: "The app tells me nuts are healthy. I know that already, so I deliberately entered that I ate nuts just to get rid of this notification."). Some participants indicated that continued use of a nutrition app would be unlikely for them. In the beginning, there can be a steep learning curve because, as a new user, you become aware of your dietary pattern and learn about the macronutrient composition of products and healthier product alternatives. In this phase, a nutrition app can have a lot to offer, and the time and effort to track food intake daily pays off in insights, support, and suggestions to improve the diet. However, after a certain amount of time, when certain behaviors might have been adapted and the learning curve flattens, the necessity and relevance of a nutrition app diminishes (user: "You have to use these type of apps to teach yourself good behaviour. And then you have to put them away"). Still, some users think it is a good idea if the app would notify or email them after a couple of months to remind them of their goals and increase their awareness.

Quantitative Study

Nutrition App Functionalities

A total of 1420 participants were included in the final analyses. The sample was nearly equally distributed in terms of gender, with 49% (696/1420) of male individuals and a mean age of 45.7 (SD 16.5) years, with an age range of 18 to 79 years. The majority had a middle (626/1420, 44%) or high (579/1420, 41%) level of education. The sample was representative for the Dutch population regarding age and gender. Respondents with a low education level were somewhat underrepresented, and respondents with a high education level were overrepresented. Almost all app functionalities (43 out of 46) were rated with a mean score above the neutral score of 4. The possibility of linking the nutrition app with social media (mean 3.8, SD 1.9), a gamification element in the app (mean 3.8, SD 1.9), and the possibility of being in touch with other users through social media platforms (mean 3.5, SD 1.8) were the only functionalities rated with a mean score <4. The complete list of mean ratings per app functionality can be obtained from Table S2 in Multimedia Appendix 1. Table 1 shows the mean importance ratings for the 10 user-centric app aspects, which all include multiple app functionalities (refer to Table S1 in Multimedia Appendix 1 for the associated functionalities). The user-centric app aspects of purpose, introduction, user-friendliness, database, and information received the highest mean ratings (all ≥5).

In Table 2, the top 3 most important app functionalities per use phase and the user-centric app aspect to which they belong are displayed (refer to Table S2 in Multimedia Appendix 1 for the full list of app functionality ratings). In the starting phase, the top 3 app functionalities include a clear introduction and the purpose of the app. Furthermore, flexibility in food tracking is important. In the use phase, app functionalities relating to the product database and the user-friendliness of the app were rated as particularly important. In the end phase, the most important app functionalities were the possibility of setting achievable goals, setting new personal goals, and offering new information to the user continuously.

Table 3 shows the mean importance ratings for the 10 user-centric app aspects per user group (*current users*,*former users*, and *nonusers*). Significant differences between the groups were found in the following app aspects: personalization (*P*<.001), user-friendliness (*P*<.001), database (*P*<.001), and monitoring (*P*<.001). Regarding personalization, user-friendliness, and monitoring, *current users* gave significantly higher ratings than *nonusers*. Regarding the database, *former users* gave significantly higher ratings than *nonusers*. Although the differences between these groups were significant, the mean ratings for all 3 user groups were still very close to each other (ranging from a 0.2 to 0.7 difference on a 7-point Likert scale).

Table 1. Mean importance rating per user-centric app aspect (n=1420).

Phase and user-centric app aspect[a]	Values, mean (SD)
Start	
Purpose	5.1 (1.2)
Introduction	5.2 (1.2)
Use	
Personalization	4.9 (1.1)
User-friendliness	5.2 (1.1)
Database	5.4 (1.2)
Information	5.0 (1.2)
Monitoring	4.7 (1.2)
Feedback	4.4 (1.3)
Communication	4.4 (1.2)
End	
Continuous engagement	4.9 (1.2)

[a]Measured on a 7-point Likert scale (1=very unimportant and 7=very important).

Table 2. Top 3 most important nutrition app functionalities per use phase and the corresponding user-centric app aspect (n=1420).

Phase and app functionality[a]	Values, mean (SD)	User-centric app aspect
Start		
Immediately clear how app should be used	5.45 (1.32)	Introduction
The app has a clear purpose	5.40 (1.40)	Purpose
Possibility to track food intake at own time	5.33 (1.45)	Personalization
Use		
Complete and reliable product database	5.58 (1.41)	Database
Easy navigation through the app	5.56 (1.36)	User-friendliness
Limited advertisements in free version	5.53 (1.51)	User-friendliness
End		
Possibility to set realistic and achievable goals	5.23 (1.44)	Continuous engagement
Possibility to set new personal goals	5.13 (1.45)	Continuous engagement
New and relevant information is continuously offered	4.88 (1.44)	Continuous engagement

[a]Measured on a 7-point Likert scale (1=very unimportant and 7=very important).

Table 3. Mean importance rating per user-centric app aspect, comparing 3 different user groups (current, former, and nonusers).

Phase and user-centric app aspect	Current users (n=276), mean (SD)	Former users (n=38), mean (SD)	Nonusers (n=1106), mean (SD)	*P* value
Start				
Purpose	5.35 (0.90)	5.22 (0.84)	5.09 (1.26)	.004
Introduction	5.31 (1.00)	5.11 (0.97)	5.21 (1.30)	.40
Use				
Personalization	5.21 (0.81)[a]	5.11 (0.67)[a,b]	4.77 (1.16)[b]	<.001
User-friendliness	5.34 (0.87)[a]	5.57 (0.69)[a,b]	5.10 (1.10)[b]	<.001
Database	5.57 (0.95)[a,b]	5.94 (0.76)[a]	5.34 (1.22)[b]	<.001
Information	5.21 (1.00)	5.16 (0.98)	4.91 (1.25)	.001
Monitoring	5.02 (1.01)[a]	4.66 (1.14)[a,b]	4.68 (1.28)[b]	<.001
Feedback	4.65 (1.25)	4.33 (1.32)	4.38 (1.26)	.005
Communication	4.59 (1.24)	4.16 (1.20)	4.35 (1.23)	.007
End				
Continuous engagement	5.17 (0.98)	5.17 (0.90)	4.88 (1.26)	.001

[a,b]Cells with the same letters indicate no significant difference following the post hoc analysis.

Nutrition App Use and Reasons to Quit

Approximately one-fifth (314/1420, 22.1%) of the respondents had experience using a nutrition app, either in the past or at the time of filling out the survey. The majority (1106/1420, 77.9%) never made use of a nutrition app. Within the group of users, a distinction could be made between *current users* (276/1420, 19.4%) and *former users* (38/1420, 2.7%).

Table 4 shows the most important reasons to make use of a nutrition app, as filled out by *current users*. The most important reasons were gaining insight into their own dietary pattern (113/276, 40.9%), losing weight (112/276, 40.6%), and maintaining body weight (96/276, 34.8%). More specific goals such as gaining insight into a specific meal moment (56/276, 20%), gaining insight into healthier alternatives for specific food products (50/276, 18%), or aiming to reduce snacking (48/276, 17%) seem less relevant.

Table 5 shows the most important reasons to stop using a nutrition app, according to *former users*. The following reasons were not selected in the survey and are therefore left out of the table: “App could not be personalized to my needs”; “Too many advertisements”; “It was difficult to keep track of personal progress”; and “It was unclear how to use the app.”

The most frequently mentioned reason was that using the app required too much time (14/38, 36.8%). Other frequently mentioned reasons were that the goal for which the app was installed was not important anymore (6/38, 15.8%) or that the app was not providing new information any longer (5/38, 13.2%). Remarkably, quitting the use because the database was not reliable (1/38, 3%) was mentioned by only 1 participant. The fact that the app could not be personalized to the users' needs or that it was difficult to keep track of personal progress were both not mentioned. The same holds for too many advertisements in the app or that it is unclear how the app works.

Table 6 shows the most important reasons for not using a nutrition app, as answered by the group of *nonusers*. The most frequently mentioned reasons were no need to gain insight into dietary pattern (379/1106, 34%), followed by not seeing the need to use a nutrition app (360/1106, 30%), or having never heard of nutrition apps (235/1106, 21%). Moreover, in this group, the time investment seems to be a barrier because 16.6 (184/1106) of the participants mentioned not having time to use a nutrition app. Privacy does not seem to be an important barrier because being afraid that the data will not be treated confidentially was only mentioned by a relatively small part of the group (108/1106, 10%).

Table 4. Most important reasons to use a nutrition app in current users (n=276)[a].

	Values, n (%)
Gaining insight into own dietary pattern	113 (40.9)
Losing weight	112 (40.6)
Maintaining body weight	96 (34.8)
Gaining insight into macronutrient intake, for example, protein	91 (33.0)
Aiming to eat healthier	85 (30.8)
Improving my health	85 (30.8)
Gaining insight into specific products and nutrients	78 (28.3)
Gaining insight into dietary pattern and physical activity	78 (28.3)
Gaining insight into a specific meal moment	56 (20.3)
Gaining insight into healthier alternatives for specific products	50 (18.1)
Aiming to reduce snacking	48 (17.4)
Other reasons	2 (0.8)

[a]Participants could indicate multiple reasons; therefore, percentages do not add up to 100%.

Table 5. Most important reasons to quit using a nutrition app in former users (n=38)[a].

	Values, n (%)
It costs too much time	14 (36.8)
Goal for which I installed the app was not important anymore	6 (15.8)
The app was not providing new information anymore	5 (13.2)
I reached the goal for which I installed the app	3 (7.9)
App was not user-friendly	3 (7.9)
Too little or inappropriate feedback on dietary intake	2 (5.3)
Functionalities of the app were too limited	1 (2.6)
Database with food products was not reliable	1 (2.6)
Other reasons	5 (13.2)
No space for it on my phone	1 (2.6)
The app triggered unhealthy/obsessed behaviors	2 (5.3)
I had to monitor physical exercise	1 (2.6)
I did not achieve the desired result	1 (2.6)

[a]Participants could indicate multiple reasons; therefore, percentages do not add up to 100%.

Table 6. Most important reasons for not using a nutrition app in nonusers (n=1106)[a].

	Values, n (%)
No need to gain insight in own dietary pattern	379 (34.2)
Do not see the point in using a nutrition app	329 (29.7)
Never heard of it	235 (21.2)
No time to use a nutrition app	184 (16.6)
Not involved in eating differently or healthier	165 (14.9)
Do not feel like learning how a nutrition app works	156 (14.1)
Afraid that data will not be treated confidentially	108 (9.7)
Other reasons	116 (8.2)
User-unfriendliness or other obstacles	35 (2.5)
Sufficient knowledge on healthy nutrition	26 (1.8)
No need to use a nutrition app	31 (2.2)
Not possible owing to illness, age, or specific diet	11 (0.8)
Other	13 (0.9)

[a]Participants could indicate multiple reasons; therefore, percentages do not add up to 100%.

Discussion

Principal Findings

In our study, we found that there are numerous functionalities in nutrition apps that contribute to consumers' intentions to use and maintain using these tools. In total, 3 different phases of app use, 10 user-centric app aspects, and 46 associated app functionalities were identified. We found that consumers encounter several difficulties and barriers in using nutrition apps for a longer period. The qualitative study provided insights into the needs, perceptions, and opinions on app aspects that are important to consider in developing effective nutrition apps. Both by users and nonusers, these aspects were considered as important to include in nutrition apps, and no large differences between the groups were found. Our findings undermine the importance of a participatory approach when designing mHealth interventions, ensuring that the intervention addresses the target user's needs and that the applied technology is easy to use to be successfully implemented in real life. This shows the relevance of not only evaluating the effectiveness of mHealth interventions by assessing health outcomes (eg, what does the use of a nutrition app do with nutritional behavior) but also by including user evaluations of various app aspects for effective implementation (eg, what is important for users to be engaged with nutrition apps) [31].

Comparison With Prior Work and Recommendations

Overview

To the best of our knowledge, this study is the first to identify the different phases of nutrition app use, including user-centric app aspects and key app functionalities. We add to the literature by providing a complete overview of which app functionalities are important in each phase of app use, according to both users and nonusers. Several other studies have examined consumers' preferences and barriers to using diet and nutrition apps. Here, the findings of these studies will be described and compared with our findings.

Findings Per Use Phase

In the starting phase, personalization of different nutrition app features seems to be a promising strategy for user engagement. App features should therefore be customizable and tailored to individual needs and goals, which is also described in the review by König et al [29]. Their findings are in accordance with our survey results, in which we found that especially personalization of the food tracking feature (eg, the possibility to track food intake at a convenient time) was important in the starting phase. Furthermore, we showed the importance of a clear purpose and introduction in the starting phase. A study by Dennison et al [32] evaluating mHealth app use found that participants want to be made fully aware of what the app can do before use. However, users are unlikely to read lengthy instructions and terms and conditions [32]. This emphasizes the importance of a clear and short introduction (eg, a tutorial) when setting up the app.

In the use phase, we found that user-friendliness and the food product database were among the most important app aspects. In a large web-based survey among European consumers, Vasiloglou et al [25] found that one of the primary criteria for selecting nutrition apps was ease of use. An app was less likely to be selected in case of issues with the food product database, such as incorrect nutritional information, a database that does not include local foods, or a database that omits major foods [25]. Several issues relating to the food database were mentioned in the FGDs, ranging from too detailed information and too many options for 1 product to questioning the reliability of the nutritional information provided. Issues with the food database can have consequences for nutrition app selection by consumers [25]. Other studies confirmed that the reliability and quality of the food database are common issues in nutrition apps and that the accuracy of nutrient information is sometimes questionable [33-36].

Although the current state of evidence supports that gamification can have a positive impact on changing health behaviors [37], including a gamification element in a nutrition app received one of the lowest ratings in our survey, meaning that consumers did not see the need to include this functionality in a nutrition app. Another functionality that received a low rating was interaction through social media. This was also found in the review by Snizay et al [38] on engagement with mHealth apps. They found that social support factors (social media and social competition) are not universally useful and might even cause disengagement by triggering negative emotions.

In the end phase, the ability to set new and achievable goals was one of the most important elements of the survey according to users. In the literature, achievable goal setting is identified as a promising facilitator for achieving behavior change [39]. Snizay et al [38] showed that goal setting was related to sustained engagement with mHealth and well-being apps. Incorporating a way to set daily and achievable goals, therefore, seems promising to keep users engaged in the final phase. In addition, nutrition apps should continuously offer new information, facts, and advice to keep users interested and engaged over a longer period.

The user-centric app aspects that arose from our qualitative study included several other validated BCTs such as monitoring and providing feedback. Broadly, the app elements that we identified are on the one hand elements that relate to these BCTs, such as factual information on nutrient intake (information), a visual progress overview (monitoring), and feedback messages (feedback), and on the other hand, the more technical design aspect of the app, such as a tutorial and a reliable database. Our study adds to the literature by showing that, for users, not only these technical aspects but also the way BCTs are implemented are important app elements. This suggests that the BCT mechanisms are not only considered as effective theoretical interventions to include in mHealth apps by health psychologists [13-16] but are, next to technical design aspects, also considered as critical elements from the user perspective.

Differentiating between the phases of use is a relevant approach to changing health behavior. There are several other theoretical models in the field of health behavior change that make a distinction between different stages, such as the Transtheoretical Model. This model describes stage-specific characteristics for behavior change and suggests that behavioral change is a dynamic process, comprising the precontemplation, contemplation, preparation, action, and maintenance stages [40]. Our 3 phases of nutrition app use overlap with the preparation (start), action (use), and maintenance (continuation) stages of the Transtheoretical Model. A user typically goes through the start phase, the use phase, and the end phase, where they reach a point where the app will be used permanently or occasionally or will not be used anymore. The decision for continued use is interlinked with the aspects mentioned in the earlier phases of use. The precontemplation and contemplation phases are lacking in our research. In these phases, awareness and intention are created to start performing a certain behavior. Because in our study participants were required to use a nutrition app as a home-use test and did not start using the app out of their own intrinsic motivation, we cannot make any statements about what app elements are important to create the intention to start using the app. Our starting phase starts when the app is installed and a user starts to navigate through the app, but obviously, in practice, this is preceded by an "intention to start" phase. The Unified Theory of Acceptance and Use of Technology sheds some light on this intention phase and distinguishes 4 factors that are important in this phase: performance expectancy (eg, the belief that a nutrition app will help), effort expectancy (eg, the expectancy that the app will be easy to use), social influence (eg, the beliefs of others who are important), and facilitating conditions (eg, the degree to which a user believes that an infrastructure exists to support use of the system) [41]. Future research is needed to obtain a better understanding of app requirements that are crucial to get users to install and open the app in the first place.

Quit or Start Nutrition App Use

The high time investment that consumers perceived as one of the main barriers was also one of the main reasons for quitting nutrition app use in the survey. The fact that using a nutrition app is perceived as time-consuming can be a result of several issues, such as poor user-friendliness or bad quality of the food product database. Poor user-friendliness is indeed a common issue for nutrition apps [29]. According to a study by Zečević et al [26], technical issues can also be a barrier in this regard. Therefore, nutrition app developers should pay particular attention to a complete and trustworthy food database and the user-friendliness of the app.

Finally, in our qualitative study, we found that nutrition apps can be specifically helpful for new users and that the learning curve is steepest in the beginning. A recent study by Samoggia et al [42] shows that a nutrition-information app is indeed mostly effective among consumers with limited knowledge. One of the main reasons to use a nutrition app that emerged from our survey was to gain more insight into one's dietary pattern. Lowe et al [43] also suggest that nutrition app use can help increase nutrition knowledge and awareness of consumption practices. Losing and maintaining body weight were other frequently mentioned reasons to use a nutrition app. This is in line with the findings of the review by König et al [29]: the main goals of using nutrition apps were related to food tracking, diet improvement, and weight management. Several review studies show the effectiveness of nutrition apps in reducing body weight in different consumer groups [7,44,45]. Therefore, given their magnitude, low cost, and easy accessibility, nutrition apps are promising tools for consumers with limited knowledge to adopt healthier eating behaviors and to lose or maintain body weight.

Limitations and Future Research

Our study has some limitations that need to be addressed. First, in both studies, a relatively large number of higher-educated consumers and, in the qualitative study, consumers interested in nutrition apps participated, which might have biased our results. However, these groups are probably also the ones that make use of nutrition apps in real life; therefore, we expect that the results might in fact be quite representative for implementation in real life. mHealth apps seem to be mostly designed to help a group of higher-educated and motivated consumers, which can be considered a limitation of these types of tools. This means that for unmotivated or lower-educated consumers, other types of interventions might be more suitable, which should be tested in future research. In this qualitative study, free versions of the 6 nutrition apps were included. It might be possible that the paid upgrades of the apps included better or additional features. Furthermore, the consumers who participated in the FGDs had no clear goal when using the app during the home-use test because we asked them to install the app for our study. The fact that they had no clear personal goal with the app, combined with the poor user-friendliness of some of the apps, might have caused early discontinuation by some users. In the survey, all 46 nutrition app functionalities were assessed on 7-point Likert scales and rated on importance. The rating of these 46 functionalities might have caused fatigue; however, we did include motivational messages in between questions, and respondents with no variability in their answers were removed from the analyses. The top 3 most important app functionalities per use phase were based on mean scores. This approach was chosen because we aimed to validate all user-centric app aspects and functionalities that we found in the FGDs. Including all of them in a ranking task or choice-based experiment was not feasible. This method may have caused consumers to rate almost all aspects as important, as they were not forced to make a choice. This made it difficult to draw conclusions on which specific elements are most important to include in nutrition apps. Therefore, a choice-based experiment with a selection of nutrition app functionalities is recommended, making it possible to uncover the trade-offs between different app functionalities. Another limitation is that we examined consumers' intentions and preferences, and we did not study the effect of nutrition app functionalities on actual dietary behavior. Such an intervention would be recommended for future research.

Finally, in both studies, we mainly focused on the elements and functionalities in the nutrition app itself. However, user characteristics and characteristics of the eating context could also influence consumers' intentions to use nutrition apps. The framework of König et al [29] highlights that besides technological reasons, the characteristics of the potential user, the interplay between user and technology, and the social environment also impact whether a nutrition app is used. In addition, Flaherty et al [46] stress the increasing importance of situational involvement and individual characteristics in engagement with mHealth apps.

Conclusions

Nutrition apps should be supportive in all 3 phases of use (start, use, and end) to increase consumers' intentions to start and maintain the use of these apps and achieve a change in dietary behavior. In the starting phase, a clear purpose, introduction, and personalization are important functionalities. In the use phase, a high-quality, credible food product database and user-friendliness are particularly important. In the end phase, the app should continuously offer new information and the possibility of setting new personal goals. High time investment is an important reason to quit the use of a nutrition app. Several other issues with nutrition apps (ie, poor user-friendliness and not offering new information anymore) need to be addressed first before long-term use can be achieved. Because almost all app functionalities in our study were considered as important by both users and nonusers, a choice-based experiment with different nutrition apps is recommended as a next step.

Acknowledgments

This research was supported by the Knowledge Base "Healthy and Safe Food System" program funded by the Ministry of Agriculture, Nature and Food Quality in the Netherlands. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors gratefully acknowledge all the participants in the study. Furthermore, the authors thank Rob Drent (Canvas Concepting) for moderating and analyzing the focus group discussions.

Authors' Contributions

SvdH, IR, and SM designed and executed the qualitative study in collaboration with Canvas Concepting. SvdH, MCDV, and SM designed the quantitative study. SM executed the quantitative study, in collaboration with MSI-ACI Europe Ltd. SM and SvdH analyzed the data of both the qualitative and quantitative study, with support from Canvas Concepting for the analyses of the focus group discussions. SvdH drafted the paper as first author. IR, MCDV, and SM reviewed the paper and contributed to paper discussions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 24 KB - mhealth_v11i1e39515_app1.docx]

References

1. Phillips CM, Chen L, Heude B, Bernard JY, Harvey NC, Duijts L, et al. Dietary Inflammatory Index and Non-Communicable Disease Risk: A Narrative Review. Nutrients 2019;11(8):1873 [FREE Full text] [doi: 10.3390/nu11081873] [Medline: 31408965]
2. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet 2019 Feb 02;393(10170):447-492. [doi: 10.1016/S0140-6736(18)31788-4] [Medline: 30660336]
3. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
4. Aitken, M, Nass, D. Digitial Health Trends 2021: innovation, evidence, regulation, and adoption. IQVIA Institute. 2021 Jul 22. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2022-01-06]
5. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
6. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
7. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
8. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: a systematic literature review. J Telemed Telecare 2015 Jul;21(5):243-253. [doi: 10.1177/1357633X15572203] [Medline: 25680388]
9. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
10. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
11. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: perceived effectiveness by app users. JMIR mHealth uHealth 2016 Apr 07;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
12. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
13. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR mHealth uHealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
14. Antezana G, Venning A, Blake V, Smith D, Winsall M, Orlowski S, et al. An evaluation of behaviour change techniques in health and lifestyle mobile applications. Health Informatics J 2020 Mar;26(1):104-113 [FREE Full text] [doi: 10.1177/1460458218813726] [Medline: 30501364]
15. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
16. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
17. Noar SM, Mehrotra P. Toward a new methodological paradigm for testing theories of health behavior and health behavior change. Patient Educ Couns 2011 Mar;82(3):468-474. [doi: 10.1016/j.pec.2010.11.016] [Medline: 21185144]
18. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
19. Milne-Ives M, Lam C, de Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR mHealth uHealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
20. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
21. Materia FT, Smyth JM. Acceptability of intervention design factors in mHealth intervention research: experimental factorial study. JMIR mHealth uHealth 2021 Jul 26;9(7):e23303 [FREE Full text] [doi: 10.2196/23303] [Medline: 34309563]
22. Enrique B, Marta B. Efficacy, effectiveness and efficiency in the health care: the need for an agreement to clarify its meaning. Int Arch Public Health Community Med 2020 Jan 25;4(1):241-248. [doi: 10.23937/2643-4512/1710035] [Medline: 1710035]

23. Holzmann S, Proell C, Hauner HC, Holzapfel C. Nutrition apps: quality and limitations. Ernaehrungs Umsch 2017;64:80-89. [doi: 10.4455/eu.2017.018]
24. Reinders MJ, Bouwman EP, van den Puttelaar J, Verain MC. Consumer acceptance of personalised nutrition: the role of ambivalent feelings and eating context. PLoS One 2020;15(4):e0231342 [FREE Full text] [doi: 10.1371/journal.pone.0231342] [Medline: 32282841]
25. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. Perspectives and preferences of adult smartphone users regarding nutrition and diet apps: web-based survey study. JMIR mHealth uHealth 2021 Jul 30;9(7):e27885 [FREE Full text] [doi: 10.2196/27885] [Medline: 34328425]
26. Zečević M, Mijatović D, Kos Koklič M, Žabkar V, Gidaković P. User perspectives of diet-tracking apps: reviews content analysis and topic modeling. J Med Internet Res 2021 Apr 22;23(4):e25160 [FREE Full text] [doi: 10.2196/25160] [Medline: 33885375]
27. Helft Nederlanders Gebruikt Gezondheids-Apps. ICT & Health. 2019 Feb 20. URL: https://icthealth.nl/nieuws/helft-nederlanders-gebruikt-gezondheids-apps/ [accessed 2022-05-02]
28. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014 Apr 15;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
29. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR mHealth uHealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
30. Ziesemer K, König LM, Boushey CJ, Villinger K, Wahl DR, Butscher S, et al. Occurrence of and reasons for "Missing events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR mHealth uHealth 2020 Oct 14;8(10):e15430 [FREE Full text] [doi: 10.2196/15430] [Medline: 33052123]
31. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
32. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
33. Bowler K, Hart K. Diet and weight loss apps: a review of user ratings, quality of nutritional information, inclusion of behaviour change techniques and personalised nutrition features. Proc Nutr Soc 2021 Apr 30;80(OCE1):E12. [doi: 10.1017/s0029665121000136]
34. Braz VN, Lopes MH. Evaluation of mobile applications related to nutrition. Public Health Nutr 2019 May;22(7):1209-1214. [doi: 10.1017/S136898001800109X] [Medline: 29734965]
35. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Jun;21(8):1495-1502. [doi: 10.1017/S1368980018000393] [Medline: 29534771]
36. Lin A, Morgan N, Kwon A, Tangney C, Van Horn L, Spring B. Comparison of food composition databases from two popular commercial nutrition apps with a research food and nutrient database. Current Developments Nutrition 2020 Jun;4(2):nzaa060_007. [doi: 10.1093/cdn/nzaa060_007]
37. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
38. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
39. Epton T, Currie S, Armitage CJ. Unique effects of setting goals on behavior change: systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]
40. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
41. Ahmad NN, Tarmidi M, Ridzwan IU, Hamid MA, Roni RA. The application of unified theory of acceptance and use of technology (UTAUT) for predicting the usage of E-Zakat online system. Int J Sci Res 2014 Apr;3(4):63-70 [FREE Full text]
42. Samoggia A, Riedel B. Assessment of nutrition-focused mobile apps' influence on consumers' healthy food behaviour and nutrition knowledge. Food Res Int 2020 Feb;128:108766. [doi: 10.1016/j.foodres.2019.108766] [Medline: 31955740]
43. Lowe B, Fraser I, Souza-Monteiro DM. A change for the better? Digital health technologies and changing food consumption behaviors. Psychol Market 2015 Apr 20;32(5):585-600. [doi: 10.1002/mar.20802]
44. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obesity Rev 2016 Sep;17(9):895-906. [doi: 10.1111/obr.12428] [Medline: 27192162]
45. Besson M, Gurviez P, Carins J. Using digital devices to help people lose weight: a systematic review. J Soc Market 2020 Jul 31;10(3):289-319. [doi: 10.1108/jsocm-07-2019-0115]

46. Flaherty SJ, McCarthy M, Collins AM, McCafferty C, McAuliffe FM. Exploring engagement with health apps: the emerging importance of situational involvement and individual characteristics. Eur J Market 2021 May 06;55(13):122-147. [doi: 10.1108/ejm-06-2019-0531]

Abbreviations

BCT: behavior change technique
FGD: focus group discussion
mHealth: mobile health

Edited by G Eysenbach; submitted 12.05.22; peer-reviewed by B Abaidoo, N Christina, K Turner; comments to author 29.07.22; revised version received 12.10.22; accepted 31.03.23; published 20.06.23.

Please cite as:
van der Haar S, Raaijmakers I, Verain MCD, Meijboom S
Incorporating Consumers' Needs in Nutrition Apps to Promote and Maintain Use: Mixed Methods Study
JMIR Mhealth Uhealth 2023;11:e39515
URL: https://mhealth.jmir.org/2023/1/e39515
doi:10.2196/39515
PMID:37338978

Original Paper

The Effectiveness of a Traditional Chinese Medicine–Based Mobile Health App for Individuals With Prediabetes: Randomized Controlled Trial

Hsueh-Wen Chung[1], MS; Chen-Jei Tai[2,3], MD, PhD; Polun Chang[4], PhD; Wen-Lin Su[5,6], MD, PhD; Li-Yin Chien[7], MPH, ScD

[1]Department of Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[2]Tai's Traditional Chinese Medicine Clinic, Taipei City, Taiwan
[3]Department of Obstetrics and Gynecology, School of Medicine, College of Medicine, Taipei Medical University, Taipei City, Taiwan
[4]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[5]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan
[6]School of Medicine, Tzu Chi University, Hualien, Taiwan
[7]Institute of Community Health Care, College of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan

Corresponding Author:
Li-Yin Chien, MPH, ScD
Institute of Community Health Care
College of Nursing
National Yang Ming Chiao Tung University
No. 155, Sec. 2 Linong Street, Beitou District
Taipei City, 11221
Taiwan
Phone: 886 2 2826 7000 ext 7142
Fax: 886 2 2826 7142
Email: lychien@nycu.edu.tw

Abstract

Background: Traditional Chinese medicine (TCM) theories assert that body constitution and meridian energy lay the foundation for disease prevention. TCM-based health concepts have not yet been incorporated into mobile health (mHealth) apps for individuals with prediabetes.

Objective: The aim of this study was to examine the effectiveness of a TCM mHealth app for individuals with prediabetes.

Methods: This randomized controlled trial recruited 121 individuals with prediabetes at a teaching hospital in New Taipei City between February 2020 and May 2021. The participants were randomly assigned to the TCM mHealth app group (n=42), ordinary mHealth app group (n=41), or control group (n=38). All participants received the usual care that included 15-20 minutes of health education about the disease, along with healthy diet and exercise encouragement. The ordinary mHealth app included physical activity (PA), diet, and disease education, along with individual records. The TCM mHealth app additionally included *qi* and body constitution information, along with constitution-based PA and diet advice. The control group received the usual care alone and did not have access to any app. Data were collected at baseline, at the end of the 12-week intervention, and 1 month after the intervention. Body constitution, including *yang*-deficiency, *yin*-deficiency, and phlegm-stasis, was measured according to the Body Constitution Questionnaire, with higher scores indicating a greater deficiency. Body energy was examined using the Meridian Energy Analysis Device. The Short-Form 36 questionnaire was used to evaluate health-related quality of life (HRQOL), which yielded physical component scores and mental component scores, with higher scores indicating better physical and mental aspects of HRQOL, respectively.

Results: Compared to the control group, the TCM mHealth app group showed greater improvement in hemoglobin A_{1c} (HbA_{1c}), *yang*-deficiency and phlegm-stasis body constitution, and BMI; however, no significant differences were found in these outcomes between the TCM mHealth app and ordinary mHealth app groups. The TCM mHealth app group showed better improvement in body energy and mental component scores than the ordinary mHealth app group. There were no significant differences in fasting plasma glucose, *yin*-deficiency body constitution, Dietary Approaches to Stop Hypertension dietary behavior, and total PA among the three groups after the intervention.

Conclusions: Use of either the ordinary or TCM mHealth app improved HRQOL among individuals with prediabetes. Compared to the outcomes of controls not using any app, use of the TCM mHealth app was effective at improving HbA_{1c}, BMI, *yang*-deficiency and phlegm-stasis body constitution, and HRQOL. Moreover, using the TCM mHealth app seemed to improve the body energy and HRQOL more than when using the ordinary mHealth app. Further studies with a larger sample size and longer follow-up period may be necessary to determine whether the differences favoring the TCM app are clinically meaningful.

Trial Registration: ClinicalTrials.gov NCT04096989; https://clinicaltrials.gov/ct2/show/NCT04096989

(JMIR Mhealth Uhealth 2023;11:e41099) doi:10.2196/41099

KEYWORDS

mHealth app; prediabetes; traditional Chinese medicine; health-related quality of life; body constitution; meridian energy

Introduction

Prediabetes is a subhealth condition characterized by higher than normal blood sugar levels, but not yet at a sufficiently high level to warrant a diagnosis of type 2 diabetes mellitus (T2DM) [1]. The American Diabetes Association proposes a diagnosis of prediabetes according to a fasting plasma glucose (FPG) level in the range of 100-125 mg/dL, hemoglobin A_{1c} (HbA_{1c}) in the range of 5.7%-6.4%, or 2-hour postprandial blood glucose level after the 75-g oral glucose test in the range of 140-199 mg/dL [2]. The prevalence of prediabetes among adults has been estimated to be 34.5% [3], with approximately 5%-10% of those diagnosed with prediabetes ultimately developing T2DM within 1 year [4].

The Centers for Disease Control and Prevention Diabetes Prevention Program (DPP) has been shown to effectively delay or prevent the development of T2DM among individuals diagnosed with prediabetes [5-7]. To reduce the cost and promote DPP-based lifestyle interventions, a technology-assisted DPP intervention is advised to be adopted with mobile and web-based apps and text messaging. Using a smartphone and computer could be ideal techniques to create widely available, easy-to-use diabetes prevention tools [8,9]. A randomized controlled trial (RCT) showed that mobile-delivered DPP achieved significant weight and BMI reductions compared with usual care. However, the intervention does not appear to be effective in controlling HbA_{1c} [10]. Another RCT developed a fully automated algorithm-driven behavioral intervention delivered via the web, internet, mobile phone, and automated phone calls, demonstrating that the intervention group had significantly decreased FPG and HbA_{1c}, and increased physical activity (PA) and vegetable consumption [11,12].

People's lifestyles and behaviors have close associations with their sociocultural background [13]. Body constitution and meridian energy are fundamental concepts in traditional Chinese medicine (TCM). Body constitution represents the individual's body condition that makes them susceptible to certain diseases but not others [14]. Body constitution forms the basis for disease treatment and prevention in TCM [15]. Meridians are channels that form a network in the body through which *qi* and blood (vital energy) flow [16,17]. The energy flow throughout the body via the 12 meridians is referred to as the meridian energy [18]. A high mean meridian energy (the average of the 12 meridian energies or body energy) usually means that *qi* and blood flow are strong and move smoothly throughout the meridians [19].

Several measures use different types of classifications for body constitution [20-23]. Nevertheless, *yang*-deficiency, *yin*-deficiency, and phlegm-stasis body constitution are common features that are prevalent in patients with chronic diseases [24,25]. *Yang*-deficiency refers to an insufficiency of *qi*. Individuals with this deficiency may experience symptoms such as fatigue, shortness of breath, chills, loose stool, and a large volume of urine. *Yin*-deficiency reflects an insufficiency in blood and interstitial fluids, and thus patients with *yin*-deficiency may experience symptoms such as being constantly thirsty, experiencing hot flushes, hard stool, and a low volume of urine. Phlegm is a viscous and turbid pathological factor formed due to an imbalance in body fluid. Phlegm-stasis refers to the accumulation of phlegm in the body as a form of condensation, which results in dizziness, chest tightness, and numbness in the limbs [26]. These TCM concepts of blood, phlegm, and fluids are not equivalent to the Western uses of these terms, but are instead used to represent energetic qualities. For example, in TCM, blood is considered a vehicle for *qi* that carries inherent energy; it nourishes and moistens the body, and it circulates nutritive *qi*. The definitions of relevant TCM terms are presented in Multimedia Appendix 1 [14,16,17,20-22,26-31].

Previous studies showed that individuals with prediabetes or diabetes were more deficient in the body constitutions of *yang*-deficiency, *yin*-deficiency, and phlegm-stasis, and also had lower meridian energy [20,21,32]. In addition, body constitution is related to an unhealthy lifestyle, in which *yang*-deficiency and phlegm-stasis are related to physical inactivity and smoking, respectively [33]. Despite high heterogeneity in the contents of available interventions, TCM lifestyle programs typically involve body constitution–based TCM health education, Chinese dietary therapy, and traditional Chinese exercises [34-38]. These programs were designed to stimulate the *qi*-blood circulation and regulate *Zang-Fu* to enhance quality of life. Previous studies have shown that TCM lifestyle programs improve body constitution [34], dietary behavior [35], and PA [36], while helping to lower blood sugar [37,38]. According to TCM theory, an improved body constitution could decrease the susceptibility to chronic diseases [14]. Increased body energy would manifest in better stamina through TCM lifestyle programs [17]. Therefore, we hypothesized that body constitution and body energy could be improved by body constitution and TCM- based lifestyle modification through *Qigong* (a type of PA) and a healthy dietary regimen, and thus help to achieve blood sugar control and enhance health.

To the best of our knowledge, no study has been conducted to identify whether TCM-based health concepts could be incorporated into a mobile health (mHealth) app for individuals with prediabetes. The need for incorporating TCM body constitutions is based on two key factors: (1) as a sociocultural appropriate method to contextualize PA and diet, and (2) as possible mediation variables for blood sugar control such as HbA_{1c} and FBG. Accordingly, the aim of this study was to develop a TCM mHealth app and examine its effectiveness on blood sugar control, body constitution, body energy, and health-related quality of life (HRQOL) as primary outcomes, as well as on BMI, dietary behavior, and PA as secondary outcomes among individuals with prediabetes. The hypothesis was that the TCM mHealth app would improve overall health through modifying health behavior and BMI. Therefore, the primary outcomes were overall health indicators (including body constitution and meridian energy) and the secondary outcomes were BMI and health behaviors.

Methods

Study Design

This study was an open-label, parallel-group RCT (ClinicalTrials.gov NCT04096989) with a three-group design. We cooperated with the health examination center and outpatient clinics at a teaching hospital in northern Taiwan to recruit individuals diagnosed with prediabetes from February 2020 to May 2021. The inclusion criteria were (1) having been diagnosed with prediabetes (according to an HbA_{1c} of 5.7%-6.4% or an FPG level of 100-125 mg/dL [2]); (2) aged 20 years and above; (3) not having cardiopulmonary disease, cancer, or other major diseases; and (4) provision of informed consent. Those who had used hypoglycemic agents, β-blockers, thiazide diuretics, nicotinic acid, or steroids within the past 3 months were excluded.

Participants were randomly assigned to three groups: TCM mHealth app, ordinary mHealth app, or control group. A statistician drew up a computer-generated randomization list. The allocation sequence was kept in an opaque, sealed, and stapled envelope, and a staff member in the outpatient clinic who was not involved in the study held the sealed envelopes. After the participants agreed to participate in the study, the researcher opened the envelope to reveal their group assignment.

The informed consent form was signed by all participants before enrollment in the study. The informed consent form stated that the risk of participation in this study was low. If the participants felt physically or mentally unwell due to their participation, they had to contact the researchers and had the right to withdraw at any time. No adverse events were reported during the study period.

Ethics Approval

This study was approved by the institutional review board at Taipei Tzu Chi Hospital (approval no. 08-X-026).

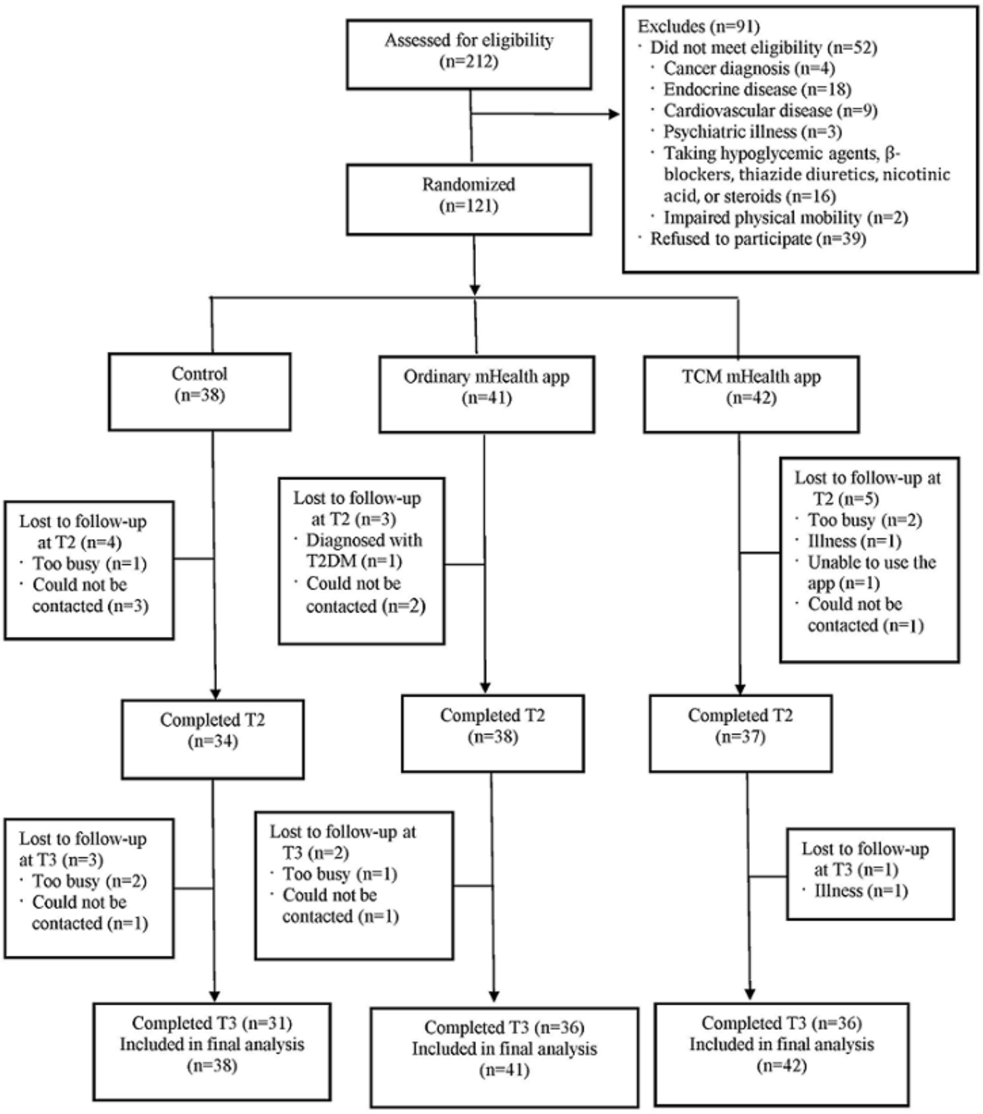
Participants

The required sample size was calculated on the basis of repeated-measures ANOVA (α=.05, power=0.80, effect size=0.3) with three repeated measurements as per a previous study [37]. G-power software indicated that the required sample size was 31 per group. To account for 30% attrition [39], we recruited 121 participants in the study.

Figure 1 presents a flow diagram of participant allocation to the three groups. A total of 212 individuals with prediabetes were assessed for eligibility, 52 of whom did not meet the eligibility criteria and 39 of whom declined to participate. A total of 121 participants were randomly assigned to the TCM mHealth app group (n=42), ordinary mHealth app group (n=41), or control group (n=38).

All participants received the usual care at the study hospital when they received the diagnosis of prediabetes. The usual care was 15-20 minutes of health education by family medicine physicians, including disease explanation, healthy diet advice, and exercise encouragement. The control group received usual care only without the use of any app.

Figure 1. Participant flow diagram. mHealth: mobile health; T1: baseline; T2: end of the intervention; T2DM: type 2 diabetes mellitus; T3: 1 month after the intervention; TCM: traditional Chinese medicine.



Intervention

An expert team that included nursing researchers, TCM doctors, Western medicine doctors, and app developers was formed to guide the development of the intervention app. The team decided that the content embedded in the app would be developed in accordance with the DPP [40,41] and a review of the literature [34,37,42]. The experts sketched, shared, and discussed potential app versions and developed prototypes. To gather user feedback on the prototypes, qualitative interviews with seven experts and five individuals were conducted, with an aim to determine user preference for the prototype and what could be improved in the app. We used the feedback to revise and finalize the mHealth app. We invited five individuals with prediabetes to use the TCM mHealth app for 30 minutes. Subsequently, a short questionnaire was administered to determine app usability and satisfaction. The participants reported positive experiences. They all agreed with the statements “I found the TCM mHealth app easy to use” and “I would recommend the TCM mHealth app to my friends and others.”

The mHealth app (both the ordinary and TCM versions) included four modules: health diary, health education, milestone, and chatroom. The “health diary” tracked the participant’s weight, BMI, blood sugar level, dietary diary, and PA over time. “Health education” provided information about specific topics such as learning about prediabetes, Dietary Approaches to Stop Hypertension (DASH) diet, and PA. The TCM mHealth app additionally included health education topics on body constitution, meridian energy, and advice on a body constitution–based diet (such as foods to avoid and foods that are recommended) and PA such as videos, pictures, and descriptive illustrations of types of Qigong (ie, Baduanjin and belly breath). A text message was sent to the participants to remind them to read the topics every week. Milestones were added for the participants to review their goals so that they could make adjustments to reach their monthly goals. For example, if the participants set a goal of an FPG level of 60-99 mg/dL, when this goal was achieved, a window would pop up to show the achievement. The participants could also check bar and line charts over 1 week or 1 month to compare the discrepancy between the actual and desired values as well as actual and ideal behaviors at different time points so that they could adjust their expected goal as needed.

In the "chatroom," personal and group chat rooms were set up using the LINE app (Naver Corp, Gyeonggi Province, South Korea). The researchers sent text messages to the participants in the personal chat room, provided feedback on the results, and encouraged participants to share their experiences in the group chat room. The participants collected virtual gold by completing questionnaires and quizzes and by achieving the set goals. The participants could use the virtual gold to claim actual prizes from the researchers. Gamification elements were added to encourage engagement with the mHealth app. Screenshots and descriptions of the ordinary and TCM mHealth apps are presented in Multimedia Appendix 2. The taxonomy of behavior change techniques [43] used in the apps is presented in Multimedia Appendix 3.

The participants in the TCM and ordinary mHealth app groups received a face-to-face education session on how to use the mHealth app and create a user account. Both mHealth app groups received information about prediabetes and evidence-based methods to decrease the possibility of progression to T2DM (eg, moderate-intensity PA of ≥150 minutes/week, DASH diet, and disease health education). The TCM mHealth app group additionally received body constitution and *qi* information, as well as constitution-based PA and diet advice (see Multimedia Appendix 2). Participants who were assigned to the TCM mHealth app group filled out the Body Constitution Questionnaire (BCQ) [20-22] at the beginning of the study. Tailored TCM diet and PA advice based on the individual's body constitution as determined by their BCQ results was incorporated into the TCM mHealth app.

The researchers monitored logins and log file analysis at least once every week at the mHealth app backend. If the participants did not use the app, complete the diary, or watch health education, additional text messages were sent to the participants. The CONSORT-EHEALTH guidelines [44] were followed in reporting this study (see Multimedia Appendix 4).

Data Collection

Data were collected at baseline (T1), at the end of the 12-week intervention (T2), and 1 month after the intervention (T3). Sociodemographic characteristics (age, gender, marital status, education level, and employment status), clinical characteristics (history of chronic disease and use of TCM), and lifestyle factors (smoking and alcohol drinking) were collected by a structured questionnaire at baseline. Primary outcome measures, including blood sugar control (FPG and HbA_{1c} levels), body constitution, meridian energy, and HRQOL, were collected at T1, T2, and T3.

Body constitution was assessed by the BCQ [20-22], which measures the presence and severity of symptoms in the most recent 7 days. The BCQ consists of 44 symptom ratings rated on a 5-point Likert-type scale ranging from 1 (not at all) to 5 (very severe). The items were summed and categorized into three types of body constitutions: 19 items for *yang*-deficiency, 19 for *yin*-deficiency, and 16 for phlegm-stasis. A higher score indicates a higher level of deficiency in the body constitution type [45]. The reliability and validity of the BCQ have been supported by the results of previous studies [20-22].

Meridian energy was measured using the Meridian Energy Analysis Device (MEAD) ME-PRO 6.1.1 (Medpex Enterprise Ltd, Taichung, Taiwan). The level of meridian energy was assessed using the MEAD values for the 24 acupoints (Ryodoraku points) along the 12 meridians ranging from 0 to 200 μA [16,46]. The participants were required to take nothing by mouth for at least 8 hours; remove their shoes, socks, and metal materials that may cause a disturbance; and sit in the room for 10-15 minutes before the meridian energy checkup [47]. Body energy was yielded by the average of the energy flows through the 24 acupoints along the 12 meridians, which serves as an indicator for the fluency of *qi* and blood moving throughout the body. Individuals with prediabetes have a lower level of body energy [32]. A low level of body energy indicates that the meridians are blocked and thus the *qi* and blood circulation are not smooth. A lower level of body energy thus indicates worse stamina.

HRQOL was measured by the Medical Outcome Survey Short Form (SF-36) Taiwan version. SF-36 is composed of 36 items, which form two summary scales, namely the physical component score (PCS) and the mental component score (MCS). Higher scores on the PCS and MCS indicate a better physical and mental aspect of HRQOL, respectively [48,49]. The reliability and validity of the SF-36 Taiwan version have been well-established [50,51].

The secondary outcomes in this study were BMI, dietary behavior, and PA. Dietary behavior was assessed by the dietary behavior questionnaire, which consisted of 14 items based on a 4-point Likert scale, ranging from 1 (never) to 4 (always). The scores ranged from 14 to 56, with higher scores indicating a better correspondence to the DASH diet [42]. PA was assessed by the International Physical Activity Questionnaire Taiwan version, which contained 7 questions about the frequency, duration, and intensity of PA in the last 7 days. The results could be further classified into mild, moderate, and vigorous PA and quantified into metabolic equivalents (MET), expressed as MET-minutes/week [52].

Statistical Analysis

The statistical analyses were performed using IBM SPSS Statistics for Windows, version 23.0 (IBM Corp, Armonk, NY, USA). We analyzed the data using an intention-to-treat analysis. For participants with incomplete or missing data, we used the maximum-likelihood method for imputation [53].

Descriptive characteristics are presented as percentages or as means and SD, as appropriate. We used the paired *t*-test to examine the changes in outcome variables within groups. One-way ANOVA was used for comparisons among groups with the Scheffe posthoc test for pairwise comparisons. Finally, we used generalized estimating equations (GEEs) to estimate the intervention effects after adjusting for age, gender, and baseline value of the outcome variables.

Results

Participant Characteristics

The characteristics of the three groups are shown in Table 1. The mean age of the participants was 58.08 (SD 10.21) years (range 29-86 years). Approximately 90% of participants reported a history of chronic disease. There were no significant differences in sociodemographics, disease history, and cigarette and alcohol use among the three groups.

Table 1. Characteristics of the participants in the three groups.

Characteristics	Total (N=121)	CG[a] (n=38)	OMG[b] (n=41)	TCMG[c] (n=42)	χ^2 or F^d (*df*)	*P* value
Age (years), mean (SD)	58.08 (10.21)	60.14 (10.83)	56.93 (10.88)	57.34 (8.81)	1.15 (2)	.32
Sex, n (%)					2.11 (2)	.35
Female	64 (52.9)	18 (47.4)	20 (48.8)	26 (61.9)		
Male	57 (47.1)	20 (52.6)	21 (51.2)	16 (38.1)		
Currently married, n (%)					1.69 (2)	.43
No	28 (23.1)	6 (15.8)	11 (26.8)	11 (26.2)		
Yes	93 (76.9)	32 (84.2)	30 (73.2)	31 (73.8)		
Education level, n (%)					3.61 (4)	.46
Elementary school or below	10 (8.3)	5 (13.2)	2 (4.9)	3 (7.1)		
Junior and senior high school	48 (39.7)	12 (31.6)	16 (39.0)	20 (47.6)		
University or above	63 (52.1)	21 (55.3)	23 (56.1)	19 (45.2)		
Employment status, n (%)					1.18 (2)	.55
Unemployed	49 (40.5)	18 (47.4)	16 (39.0)	15 (35.7)		
Employed	72 (59.5)	20 (52.6)	25 (61.0)	27 (64.3)		
History of chronic disease, n (%)					1.03 (2)	.60
No	13 (10.7)	3 (7.9)	6 (14.6)	4 (9.5)		
Yes	108 (89.3)	35 (92.1)	35 (85.4)	38 (90.5)		
Smoking status, n (%)					6.81 (4)	.15
Never	97 (80.2)	29 (76.3)	30 (73.2)	38 (90.5)		
Quit	20 (16.5)	8 (21.1)	10 (24.4)	2 (4.8)		
Current smoker	4 (3.3)	1 (2.6)	1 (2.4)	2 (4.8)		
Alcohol drinking, n (%)					5.82 (4)	.21
Never	106 (87.6)	32 (84.2)	35 (85.4)	39 (92.9)		
Quit	4 (3.3)	2 (5.3)	0 (0.0)	2 (4.8)		
Current drinker	11 (9.1)	4 (10.5)	6 (14.6)	1 (2.4)		

[a]CG: control group.

[b]OMG: ordinary mobile health app group.

[c]TCMG: traditional Chinese medicine mobile health app group.

[d]F and χ^2 are the respective values of one-way ANOVA and Pearson χ^2 test.

Overall Intervention Effects

The crude effects of the intervention on the outcomes are shown in Figures 2 and 3 (details are shown in Multimedia Appendix 5). Of the outcomes included, there were significant differences in *yang*-deficiency and phlegm-stasis body constitution among the three groups, with the TCM mHealth app group scoring higher than the control group (Figure 2). In addition, the ordinary mHealth app group reported the highest amount of PA among the three groups (Figure 3). For body constitution and PA, GEE results are preferred over the crude results given baseline differences. There were no significant differences in other outcomes between the groups at preintervention. Net effects of the intervention on the outcomes are shown in Table 2 (full model results are shown in Multimedia Appendix 6) using the control group as the reference. To explicitly compare the effects between the TCM and ordinary mHealth groups, Table 3 shows the results using the ordinary mHealth group as the reference (full model results are shown in Multimedia Appendix 7).

Figure 2. Changes in primary outcomes. (A) Fasting plasma glucose. (B) HbA_{1c}. (C) Yang deficiency body constitution. (D) Ying deficiency body constitution. (E) Phlegm stasis body constitution. (F) Body energy. (G) Physical component score. (H) Mental component score. Within-group across-time comparisons were made from a paired *t* test with T1 as the reference. Between-group comparisons were based on one-way ANOVA with the Scheffe posthoc test, with the results presented below graphs. mHealth: mobile health; T1: baseline; T2: end of the intervention; T3: 1 month after the intervention; TCM: traditional Chinese medicine. **P*<.05, #*P*<.001.

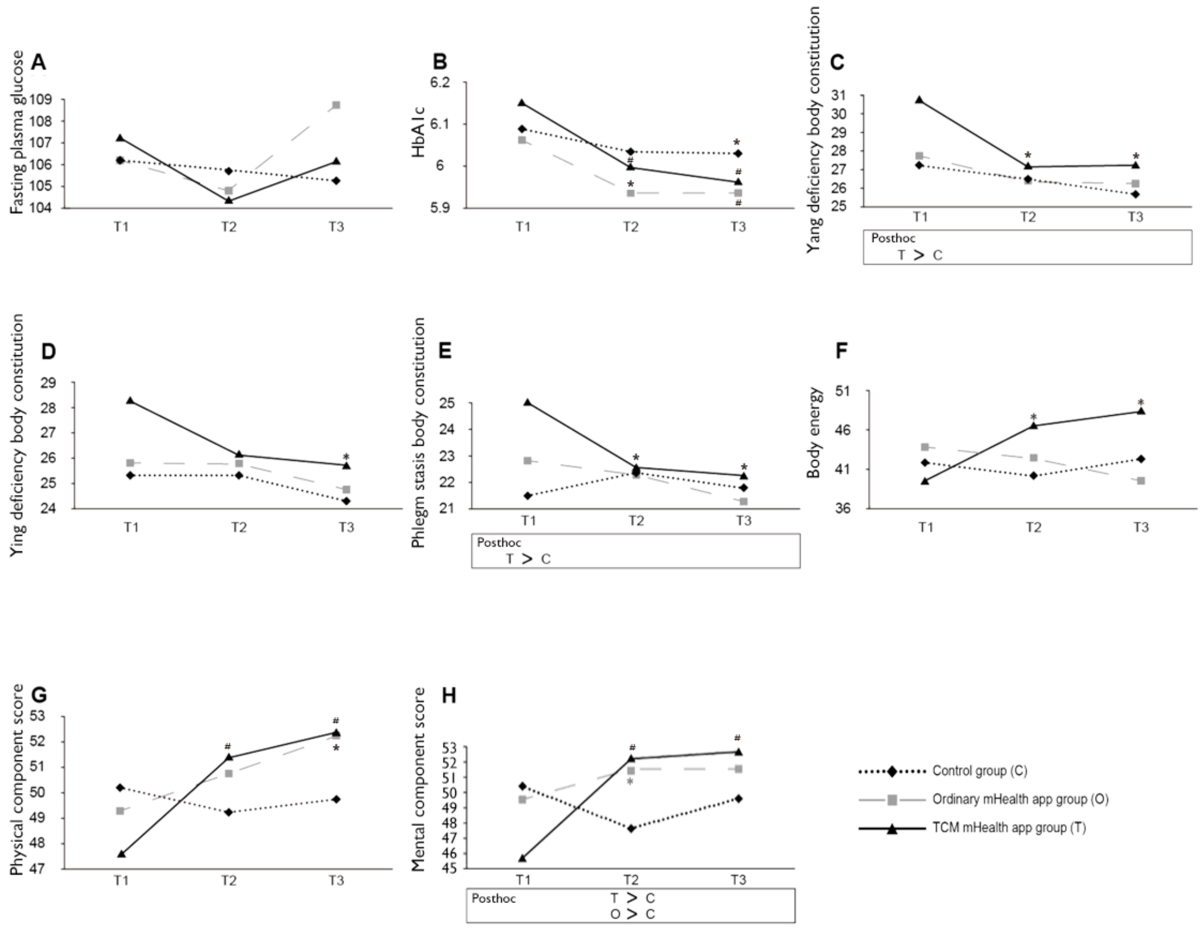


Figure 3. Changes in secondary outcomes. (A) BMI. (B) DASH dietary behavior. (C) Total physical activity. Within-group across-time comparisons were made from a paired *t* test with T1 as the reference. Between-group comparisons were based on one-way ANOVA with the Scheffe posthoc test, with the results presented below the graph in (C). mHealth: mobile health; T1: baseline; T2: end of the intervention; T3: 1 month after the intervention; TCM: traditional Chinese medicine. **P*<.05, #*P*<.001.

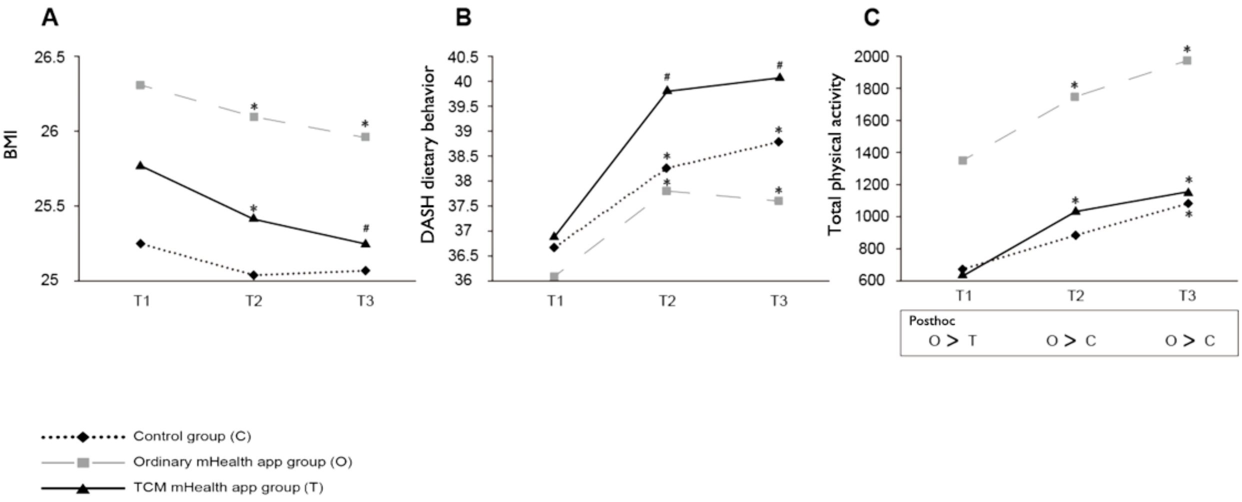


Table 2. Generalized estimating equation models to compare the differences among the three groups, using the control group as the reference.[a]

Parameter	OMG[b]×T2[c], β (95% CI)	OMG×T3[d], β (95% CI)	TCMG[e]×T2, β (95% CI)	TCMG×T3, β (95% CI)
FPG[f]	−1.18 (−5.92 to 3.56)	3.17 (−2.22 to 8.56)	−2.52 (−7.39 to 2.35)	−.37 (−5.61 to 4.86)
HbA_{1c}[g]	−.06 (−.15 to .04)	−.05 (−.14 to .04)	−.08 (−.18 to .02)	−.11 (−.21 to −.01)
Yang-deficiency BC[h]	–.81 (−3.16 to 1.53)	−.46 (−2.73 to 1.81)	−3.15 (−6.09 to −.21)	−2.37 (−5.04 to .29)
Yin-deficiency BC	–.02 (–2.56 to 2.52)	.03 (–2.08 to 2.13)	–2.29 (–5.28 to .70)	–1.62 (–4.21 to .96)
Phlegm-statis BC	−1.36 (−4.16 to 1.44)	−1.88 (−4.38 to .63)	−3.45 (−6.49 to −.42)	−3.30 (−6.01 to −.58)
Body energy	–.16 (−11.23 to 10.91)	−5.84 (−17.68 to 6.01)	8.60 (−1.91 to 19.11)	7.81 (−3.36 to 18.98)
PCS[i]	2.56 (−.44 to 5.56)	2.90 (−.07 to 5.89)	4.93 (1.97 to 7.89)	4.89 (1.92 to 7.87)
MCS[j]	4.63 (.91 to 8.35)	2.68 (−1.16 to 6.51)	8.11 (3.82 to 12.40)	7.26 (3.35 to 11.17)
BMI	−.03 (−.34 to .29)	−.24 (−.61 to .13)	−.12 (−.43 to .20)	−.37 (−.73 to −.02)
DASH[k] dietary behavior	.22 (−1.42 to 1.86)	−.51 (−2.49 to 1.48)	1.21 (−.62 to 3.06)	.99 (−1.05 to 3.04)
Total physical activity	236.33 (−231.76 to 704.43)	213.18 (−360.50 to 786.86)	248.59 (−191.18 to 688.37)	122.74 (−459.22 to 704.71)

[a]Interaction effects were examined after adjustments for age and gender, and the baseline value of the outcome variable; baseline and control group served as references.

[b]OMG: ordinary mobile health app group.

[c]T2: end of the intervention.

[d]T3: 1 month after the intervention.

[e]TCMG: traditional Chinese medicine mobile health app group.

[f]FPG: fasting plasma glucose.

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

[h]BC: body constitution.

[i]PCS: physical component score.

[j]MCS: mental component score.

[k]DASH: Dietary Approaches to Stop Hypertension.

Table 3. Generalized estimating equation models to compare the outcomes between the TCM mHealth app group (n=42) and ordinary mHealth app group (n=41).[a]

Parameter	TCMG[b]×T2[c], β (95% CI)	TCMG×T3[d], β (95% CI)
FPG[e]	−1.82 (−6.30 to 2.66)	3.96 (−9.91 to 1.99)
HbA_{1c}[f]	−.02 (−.12 to .09)	−.06 (−.15 to .04)
Yang- deficiency BC[g]	−2.26 (−5.12 to .61)	−1.96 (−4.83 to .92)
Yin-deficiency BC	−2.16 (−5.00 to .68)	−1.40 (−4.20 to 1.41)
Phlegm-statis BC	−.24 (−5.19 to .70)	−1.55 (−4.03 to .94)
Body energy	8.65 (−2.11 to 19.41)	12.30 (1.31 to 23.30)
PCS[h]	2.32 (−.58 to 5.22)	2.24 (−.87 to 5.35)
MCS[i]	3.14 (−.82 to 7.10)	4.29 (.27 to 8.31)
BMI	−.10 (−.41 to .21)	−.14 (−.49 to .20)
DASH[j] dietary behavior	1.06 (−.48 to 2.60)	1.44 (−.45 to 3.33)
Total physical activity	4.99 (−535.31 to 545.30)	−82.67 (−708.14 to 542.80)

[a]Interaction effects were examined after adjustments for age and gender and the baseline value of the outcome variable; baseline and ordinary mobile health group served as references.

[b]TCMG: traditional Chinese medicine mobile health app group.

[c]T2: end of the intervention.

[d]T3: 1 month after the intervention.

[e]FPG: fasting plasma glucose.

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

[g]BC: body constitution.

[h]PCS: physical component score.

[i]MCS: mental component score.

[j]DASH: Dietary Approaches to Stop Hypertension.

Primary Outcomes

Blood Sugar Control

There were no significant differences in the FPG and HbA_{1c} among the three groups at the three time points. FPG did not change significantly at postintervention and 1 month after the intervention as compared to the preintervention (baseline) levels for all three groups. However, HbA_{1c} decreased significantly over time for all three groups (Figure 2; Multimedia Appendix 5).

The GEE analyses revealed no significant group differences in FPG. The HbA_{1c} levels decreased significantly in the TCM mHealth app group at T3 and were significantly different than those in control group, but there were no significant differences between the ordinary mHealth app and control groups (Table 2; Multimedia Appendix 6).

Body Constitution

The TCM mHealth app group scored the highest in *yang*-deficiency and phlegm-stasis body constitution at T1, with significant differences between the TCM and control groups. There were no significant differences in body constitution of the three groups at T2 and T3. All three types of body constitutions improved significantly at T3 for the TCM group, but no such improvements were observed in the ordinary mHealth app and control groups (Figure 2; Multimedia Appendix 5).

The GEE analyses indicated that the TCM mHealth app group showed significant improvements in *yang*-deficiency body constitution at T2 compared to that in the control group. Furthermore, greater improvement in phlegm-stasis body constitution was found compared to that in the control group, and the effect persisted until 1 month after the intervention. However, no effect on *yin*-deficiency body constitution was observed (Table 2; Multimedia Appendix 6). There were no significant differences in body constitution between the TCM and ordinary mHealth app groups (Table 3; Multimedia Appendix 7).

Body Energy

There were no significant differences in body energy among the three groups at the three time points. Body energy increased significantly from T1 to T2 and from T1 to T3 in the TCM mHealth app group, but remained unchanged in the ordinary mHealth app and control groups (Figure 2; Multimedia Appendix 5).

The GEE results indicated that the TCM mHealth app group showed a significant increase in body energy at T3 compared to that in the ordinary mHealth group (Table 3; Multimedia Appendix 7).

Health-Related Quality of Life

There were no significant differences in the PCS among the three groups at the three time points. The TCM mHealth app group had a significant increase in the PCS from T1 to T2 and from T1 to T3. The ordinary mHealth app group also had a significant increase in the PCS from T1 to T3 (Figure 2; Multimedia Appendix 5).

With regard to the MCS, the TCM mHealth app group had a significantly higher score at T2 compared with that of the control group (*P*=.02), as did the ordinary mHealth app group (*P*=.03; Figure 2; Multimedia Appendix 5). The TCM mHealth app group showed a significant increase in the MCS from T1 to T2 and from T1 to T3. Moreover, the ordinary mHealth app group had a significant increase in the MCS from T1 to T2 and from T1 to T3 (with borderline significance; *P*=.04 and *P*=.05, respectively).

The GEE results indicated that the TCM mHealth app group had a significant increase in the PCS and MCS at T2 and T3 compared with that of the control group (Table 2; Multimedia Appendix 6). In addition, the increase in the MCS in the TCM mHealth app group from T1 to T3 appeared to be higher than that in the ordinary mHealth app group (Table 3; Multimedia Appendix 7). The ordinary mHealth app group had a significant increase in the MCS at T2 compared to that of the control group (Table 2; Multimedia Appendix 6).

Secondary Outcomes

BMI

There were no significant differences in BMI among the three groups at the three time points (Figure 3; Multimedia Appendix 5). The BMI in the TCM and ordinary mHealth app groups decreased significantly from T1 to T2 and from T1 to T3. However, the BMI of the control group remained unchanged.

The GEE results indicated that the TCM mHealth app group showed a significant decrease in BMI at T3 compared to that in the control group (Table 2; Multimedia Appendix 6). However, there were no significant differences in the change in BMI between the TCM and ordinary mHealth app groups (Table 3; Multimedia Appendix 7).

Dietary Behavior

There were no significant differences in the DASH dietary behavior among the three groups at the three time points (Figure 3; Multimedia Appendix 5). The DASH dietary behavior improved significantly in all three groups from T1 to T2 and from T1 to T3. The GEE results showed no significant differences in the DASH dietary behavior among the three groups over time (Tables 2 and 3; Multimedia Appendices 6 and 7).

Physical Activity

The ordinary mHealth app group had a significantly higher PA level than that of the other two groups across all three time points (Figure 3; Multimedia Appendix 5). The PA increased significantly from T1 to T3 in all three groups. The GEE results showed no significant differences in the total PA among the three groups over time (Tables 2 and 3; Multimedia Appendices 6 and 7).

Discussion

Principal Findings

This RCT found that the TCM mHealth group showed better HbA_{1c}, *yang*-deficiency body constitution, phlegm-stasis body constitution, physical aspect of HRQOL, mental aspect of HRQOL, and BMI than the control group. When the TCM and ordinary mHealth app groups were compared, the TCM mHealth app group showed higher body energy and mental aspect scores of the HRQOL at 1 month after intervention than the ordinary mHealth app group. These results suggest that the TCM mHealth app helps to effectively control blood sugar, decrease *yang*-deficiency and phlegm-stasis body constitution, and improve body energy and HRQOL. Incorporating TCM body constitution and meridian energy concepts into an mHealth app seems plausible and can improve health for individuals with prediabetes. According to TCM theory, an improved body constitution could decrease susceptibility to chronic diseases [14] and increased body energy would manifest in better stamina [17]. Therefore, improved body constitution and increased body energy could explain the effectiveness of the TCM mHealth app for improving HbA_{1c} and HRQOL. Further studies are needed to examine the interrelationships among these outcomes.

Previous studies showed that using an mHealth app improved HbA_{1c}, dietary behavior, PA, and BMI compared to those of controls [10-12]. However, we did not find such an effect for the ordinary mHealth app group when compared with the control group in this study. It was noted that all participants in the study had received 15-20 minutes of health education about the disease, healthy diet, and exercise encouragement when they were diagnosed with prediabetes in the study hospital. Possibly owing to this health education, we found that the HbA_{1c}, DASH dietary behavior, and PA improved over time for all three groups. Such universal improvement in health behavior may be the reason for the lack of significant effect on these outcomes when the ordinary mHealth app group was compared to the controls.

The TCM dietary and PA advice used in the TCM mHealth app is mainly based on the types of foods/PA to avoid and those to consume/practice based on the individual's body constitution. Qigong, including belly breathing and Baduanjin, was recommended for all participants since these exercises can improve *qi* and are appropriate for all types of body constitutions. The recommended types of foods consumed differed according to the individual's body constitution. We found that the TCM dietary and PA principles do not conflict with the principles of the DASH diet and recommended PA amount, except that individuals with a *yin*-deficiency body constitution were advised to avoid vigorous PA.

When compared to the ordinary mHealth app group, the TCM mHealth app group did not differ significantly in improving body constitution and HbA_{1c}. However, the TCM mHealth app group showed better improvement in the mental aspect of HRQOL and in increasing body energy than the ordinary mHealth app group. These results imply that using the TCM mHealth app can better improve body energy and HRQOL in individuals with prediabetes compared to the ordinary mHealth app. This may be because people who practice TCM consider that health conditions can improve with sufficient *qi*. The importance of these indicators across cultures needs to be examined in future studies.

Our study showed that both the TCM and ordinary mHealth app groups exhibited a significant improvement in HbA_{1c} and HRQOL with time. This finding is consistent with a previous meta-analysis that reported a positive effect of a lifestyle intervention on HbA_{1c} in individuals with prediabetes [54]. Another meta-analysis reported positive effects of Qigong in improving the HRQOL [55]. Another study demonstrated that individuals with prediabetes who achieved moderate-intensity PA (≥150 minutes/week) have higher levels of HRQOL than inactive people [56]. Therefore, the mHealth app can be used to assist individuals with prediabetes to increase their PA and HRQOL, while decreasing their HbA_{1c}.

This study showed effectiveness of the intervention in HbA_{1c} but not FPG. The insignificant FPG results may be explained as follows. First, various factors can influence FPG, such as emotional state, drugs, and stress hormones [57]. Thus, FPG is not a stable estimate of glycemic exposure [58]. Second, although we encouraged the participants to fast for at least 8 hours, we are unsure if they followed these instructions. In future studies, we suggest that more than two tests should be used to enhance diagnostic accuracy.

In this study, we found that the TCM mHealth app effectively improved the *yang*-deficiency and phlegm-stasis body constitution in individuals with prediabetes, but could not improve the *yin*-deficiency body constitution. More research is needed to develop effective interventions to improve the *yin*-deficiency body constitution among individuals with prediabetes.

Finally, the TCM mHealth app group did show significant improvements in body constitution and body energy in our study, while the other two groups did not show changes in these factors over time. These results suggest that medical practitioners could provide the TCM mHealth app to individuals with prediabetes, through which Qigong and a Chinese dietary regimen can improve body energy, body constitution, and HRQOL. Studies with a larger sample size and longer follow-up period are needed to compare the effectiveness of the two different approaches.

Limitations

This study has several limitations. First, most participants were from an outpatient department and had chronic conditions. Therefore, this sample may have had more complex health problems than present in people with prediabetes alone. The participants may be more homogeneous since they were from one single center. Second, the sample size was small. The limited sample size may be the reason for some statistically insignificant results when comparing between groups. Third, the study was an open-label trial where participants were aware of their group assignments; thus, performance bias was possible. Fourth, the study used a 12-week intervention and a 1-month follow-up period according to previous TCM lifestyle programs [35,37]. A follow-up period from 12 to 24 months may be preferred in future studies. Fifth, the use of the mHealth apps in the follow-up period was not monitored. In addition, during the intervention period, the participants were reminded to use the apps, but the time spent on app use varied, suggesting that the participation level may be different. Lastly, this study included three types of body constitutions. There are many other types of body constitutions that could be considered in future studies.

Conclusion

We developed a TCM mHealth app to incorporate TCM concepts into an mHealth app for individuals with prediabetes. Compared to controls not using the app, the TCM mHealth app appeared to be effective in improving HbA_{1c}, BMI, *yang*-deficiency and phlegm-stasis body constitution, and HRQOL. Compared to individuals using the ordinary mHealth app, individuals using the TCM mHealth app showed higher body energy and mental aspects of the HRQOL 1 month after the intervention. The TCM mHealth app was not effective in improving FPG, *yin*-deficiency body constitution, DASH dietary behavior, and total PA. The study results suggest that individuals with prediabetes could use the TCM mHealth app to improve their body energy and HRQOL. Further studies with a larger sample size and a longer follow-up period are warranted to verify whether the differences favoring the TCM app are clinically meaningful.

Acknowledgments

This study was funded by the Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation (TCRD-TPE-109-05). The publication fee for this manuscript is funded by the National Science and Technology Council, Taiwan (MOST 111-2314-B-A49-011-MY3).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of terms in traditional Chinese medicine.
[PDF File (Adobe PDF File), 96 KB - mhealth_v11i1e41099_app1.pdf]

Multimedia Appendix 2
Screenshots of the ordinary and TCM mHealth apps.
[PDF File (Adobe PDF File), 1813 KB - mhealth_v11i1e41099_app2.pdf]

Multimedia Appendix 3
Taxonomy of the behavior change techniques used in the mHealth app.
[PDF File (Adobe PDF File), 210 KB - mhealth_v11i1e41099_app3.pdf]

Multimedia Appendix 4
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 270 KB - mhealth_v11i1e41099_app4.pdf]

Multimedia Appendix 5
Comparison of the primary and secondary outcomes among the TCM mHealth app, ordinary mHealth app, and control groups (N=121).
[PDF File (Adobe PDF File), 159 KB - mhealth_v11i1e41099_app5.pdf]

Multimedia Appendix 6
Generalized estimating equation models to compare the differences among the three groups, using the control group as the reference.
[PDF File (Adobe PDF File), 253 KB - mhealth_v11i1e41099_app6.pdf]

Multimedia Appendix 7
Generalized estimating equation models to compare the outcomes between the TCM mHealth app (n=42) and ordinary mHealth app groups (n=41).
[PDF File (Adobe PDF File), 239 KB - mhealth_v11i1e41099_app7.pdf]

References

1. Song Y, Wang H, Qin L, Li M, Gao S, Wu L, et al. Efficiency and safety of Chinese herbal medicine in the treatment of prediabetes: a systemic review and meta-analysis of randomized controlled trials. Evid Based Complement Alternat Med 2020;2020:3628036. [doi: 10.1155/2020/3628036] [Medline: 33123206]
2. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2014 Jan;37(Suppl 1):S81-S90. [doi: 10.2337/dc14-S081] [Medline: 24357215]
3. National diabetes statistics report 2020: Estimates of diabetes and its burden in the United States. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2021-03-22]
4. Bansal N. Prediabetes diagnosis and treatment: A review. World J Diabetes 2015 Mar 15;6(2):296-303 [FREE Full text] [doi: 10.4239/wjd.v6.i2.296] [Medline: 25789110]
5. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 07;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
6. Ely EK, Gruss SM, Luman ET, Gregg EW, Ali MK, Nhim K, et al. A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. Diabetes Care 2017 Oct;40(10):1331-1341 [FREE Full text] [doi: 10.2337/dc16-2099] [Medline: 28500215]
7. Mensa-Wilmot Y, Bowen S, Rutledge S, Morgan JM, Bonner T, Farris K, et al. Early results of states' efforts to support, scale, and sustain the National Diabetes Prevention Program. Prev Chronic Dis 2017 Dec 07;14:E130 [FREE Full text] [doi: 10.5888/pcd14.170478] [Medline: 29215975]
8. Atienza AA, Patrick K. Mobile health: the killer app for cyberinfrastructure and consumer health. Am J Prev Med 2011 May;40(5 Suppl 2):S151-S153. [doi: 10.1016/j.amepre.2011.01.008] [Medline: 21521588]
9. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]

10. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the Diabetes Prevention Program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
11. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
12. Block G, Azar KMJ, Romanelli RJ, Block TJ, Palaniappan LP, Dolginsky M, et al. Improving diet, activity and wellness in adults at risk of diabetes: randomized controlled trial. Nutr Diabetes 2016 Sep 19;6(9):e231. [doi: 10.1038/nutd.2016.42] [Medline: 27643726]
13. Sarrafzadegan N, Kelishadi R, Esmaillzadeh A, Mohammadifard N, Rabiei K, Roohafza H, et al. Do lifestyle interventions work in developing countries? Findings from the Isfahan Healthy Heart Program in the Islamic Republic of Iran. Bull World Health Organ 2009 Jan;87(1):39-50 [FREE Full text] [doi: 10.2471/blt.07.049841] [Medline: 19197403]
14. Huang Y, Lin C, Cheng S, Lin C, Lin SJ, Su Y. Using Chinese body constitution concepts and measurable variables for assessing risk of coronary artery disease. Evid Based Complement Alternat Med 2019;2019:8218013. [doi: 10.1155/2019/8218013] [Medline: 31636687]
15. Wang Q. Individualized medicine, health medicine, and constitutional theory in Chinese medicine. Front Med 2012 Mar;6(1):1-7. [doi: 10.1007/s11684-012-0173-y] [Medline: 22460443]
16. Huang S, Chien L, Chang C, Chen P, Tai C. Abnormal gastroscopy findings were related to lower meridian energy. Evid Based Complement Alternat Med 2011;2011:878391. [doi: 10.1155/2011/878391] [Medline: 21052557]
17. Bao GC. The idealist and pragmatist view of qi in tai chi and qigong: a narrative commentary and review. J Integr Med 2020 Sep;18(5):363-368. [doi: 10.1016/j.joim.2020.06.004] [Medline: 32636157]
18. Wang G, Ayati MH, Zhang W. Meridian studies in China: a systematic review. J Acupunct Meridian Stud 2010 Mar;3(1):1-9 [FREE Full text] [doi: 10.1016/S2005-2901(10)60001-5] [Medline: 20633509]
19. Huang S, Chien L, Tai C, Chen P, Lien P, Tai C. Effects of symptoms and complementary and alternative medicine use on the yang deficiency pattern among breast cancer patients receiving chemotherapy. Complement Ther Med 2015 Apr;23(2):233-241. [doi: 10.1016/j.ctim.2015.01.004] [Medline: 25847561]
20. Lin J, Chen L, Lin J, Chang C, Huang Y, Su Y. BCQ-: a body constitution questionnaire to assess Yin-Xu. Part I: establishment of a provisional version through a Delphi process. Forsch Komplementmed 2012;19(5):234-241. [doi: 10.1159/000343580] [Medline: 23128097]
21. Su Y, Chen L, Lin J, Lin J, Huang Y, Lai J. BCQ+: a body constitution questionnaire to assess Yang-Xu. Part I: establishment of a first final version through a Delphi process. Forsch Komplementmed 2008 Dec;15(6):327-334. [doi: 10.1159/000175938] [Medline: 19142042]
22. Lin J, Lin J, Chen L, Chang C, Huang Y, Su Y. BCQs: a Body Constitution Questionnaire to assess stasis in traditional Chinese medicine. Eur J Integr Med 2012 Dec;4(4):e379-e391. [doi: 10.1016/j.eujim.2012.05.001]
23. Wang Q. Classification and diagnosis basis of nine basic constitutions in Chinese medicine. J Beijing Univ Trad Chin Med 2005;28(4):1-8.
24. Tsai C, Su Y, Lin S, Lee I, Lee C, Li T. Reduced health-related quality of life in body constitutions of yin-xu, and yang-xu, stasis in patients with type 2 diabetes: taichung diabetic body constitution study. Evid Based Complement Alternat Med 2014;2014:309403. [doi: 10.1155/2014/309403] [Medline: 25093025]
25. Kung Y, Kuo TBJ, Lai C, Shen Y, Su Y, Yang CCH. Disclosure of suboptimal health status through traditional Chinese medicine-based body constitution and pulse patterns. Complement Ther Med 2021 Jan;56:102607 [FREE Full text] [doi: 10.1016/j.ctim.2020.102607] [Medline: 33220452]
26. Chung H, Chien L, Huang S, Tai C, Tai C. Changes in symptom patterns and health-related quality of life of cancer patients before and after chemotherapy. J Tradit Chin Med 2016 Jun;36(3):326-331 [FREE Full text] [doi: 10.1016/s0254-6272(16)30045-0] [Medline: 27468547]
27. Li L, Yao H, Wang J, Li Y, Wang Q. The Role of Chinese Medicine in Health Maintenance and Disease Prevention: Application of Constitution Theory. Am J Chin Med 2019;47(3):495-506. [doi: 10.1142/S0192415X19500253] [Medline: 31023059]
28. Sancier KM. Electrodermal measurements for monitoring the effects of a qigong workshop. J Altern Complement Med 2003 Apr;9(2):235-241. [doi: 10.1089/10755530360623347] [Medline: 12804076]
29. Weng CS, Hung YL, Shyu LY, Chang YH. A study of electrical conductance of meridian in the obese during weight reduction. Am J Chin Med 2004;32(3):417-425. [doi: 10.1142/S0192415X04002077] [Medline: 15344425]
30. Measuring healthy days: Population assessment of health-related quality of life. Centers for Disease Control and Prevention,. Atlanda, Georgia; 2000. URL: https://www.cdc.gov/hrqol/pdfs/mhd.pdf
31. Ware JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998 Nov;51(11):903-912. [doi: 10.1016/s0895-4356(98)00081-x] [Medline: 9817107]
32. Ji J, Yang S, Lou Q, Liu C. Effect of traditional Chinese medicine meridian massage on prediabetes patients. J Hunan Univ Chin Med 2018;38(2):160-164.

33. Wang Y, Wu X, Wang HHX, Li Y, Fu Y, Wang J, et al. Body constitution and unhealthy lifestyles in a primary care population at high cardiovascular risk: new insights for health management. Int J Gen Med 2021;14:6991-7001 [FREE Full text] [doi: 10.2147/IJGM.S329321] [Medline: 34707390]
34. Sun H, Tian W, Wu S, Chang L, Lung H, Liu C. Traditional Chinese medicine non-drug comprehensive program in the treatment of impaired glucose tolerance: a clinical observation. Hebei J Trad Chin Med. in Chinese 2016;38(8):1162-1165.
35. Liou X, Wang C, Liu Z, Li Y, Chen F, Wu Y. Study on effect of TCM health education on the behavior changes of prediabetic state patients under internet mode. J Chin Med 2016;8(3):93-95.
36. Liu X, Miller YD, Burton NW, Brown WJ. A preliminary study of the effects of Tai Chi and Qigong medical exercise on indicators of metabolic syndrome, glycaemic control, health-related quality of life, and psychological health in adults with elevated blood glucose. Br J Sports Med 2010 Aug;44(10):704-709. [doi: 10.1136/bjsm.2008.051144] [Medline: 18927159]
37. Geng BC, Wu Y, Jin ZF, Dang YY. Effect of health education on TCM blood glucose index and behavior in patients with diabetes mellitus under internet mode. J Clin Nurs 2017;16(6):4-7.
38. Wang Y, Zhu YH, Zhang XT, Cheng F. Clinical observation of comprehensive nondrug interventions of traditional Chinese medicine in community prediabetes population. Shanghai J Trad Chin Med 2017;51(11):45-48.
39. Staite E, Bayley A, Al-Ozairi E, Stewart K, Hopkins D, Rundle J, et al. A wearable technology delivering a web-based diabetes prevention program to people at high risk of type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 15;8(7):e15448 [FREE Full text] [doi: 10.2196/15448] [Medline: 32459651]
40. Ibrahim N, Ming Moy F, Awalludin IAN, Mohd Ali Z, Ismail IS. Effects of a community-based healthy lifestyle intervention program (Co-HELP) among adults with prediabetes in a developing country: a quasi-experimental study. PLoS One 2016;11(12):e0167123 [FREE Full text] [doi: 10.1371/journal.pone.0167123] [Medline: 27935988]
41. Pengpid S, Peltzer K, Skaal L. Efficacy of a church-based lifestyle intervention programme to control high normal blood pressure and/or high normal blood glucose in church members: a randomized controlled trial in Pretoria, South Africa. BMC Public Health 2014 Jun 06;14:568 [FREE Full text] [doi: 10.1186/1471-2458-14-568] [Medline: 24906450]
42. Yang L, Lin C, Lin T, Liu T. Effects of a self-regulation program on dietary behavior, physical activity, and physiological measures among prediabetic adults: a pilot study. J Nurs Healthc Res 2013;9(2):87-95.
43. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
44. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
45. Wong W, Lam CLK, Su Y, Lin SJ, Ziea ET, Wong VT, et al. Measuring body constitution: validation of the Body Constitution Questionnaire (BCQ) in Hong Kong. Complement Ther Med 2014 Aug;22(4):670-682. [doi: 10.1016/j.ctim.2014.05.009] [Medline: 25146072]
46. Chang S, Weng Y, Cheng S, Chang Y, Lee T, Chang C, et al. Application of meridian electrical conductance in the setting of acute ischemic stroke: a cross-sectional study. Evid Based Complement Alternat Med 2019;2019:3098095. [doi: 10.1155/2019/3098095] [Medline: 31485244]
47. Lee Y, Ng HP, Chang Y, Ho W. The development and application evaluation of meridian energy detection system in traditional Oriental medicine: a preliminary study. Evid Based Complement Alternat Med 2018;2018:9469703. [doi: 10.1155/2018/9469703] [Medline: 30174715]
48. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66. [doi: 10.1097/00005650-199401000-00004] [Medline: 8277801]
49. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993 Mar;31(3):247-263. [doi: 10.1097/00005650-199303000-00006] [Medline: 8450681]
50. Fuh JL, Wang SJ, Lu SR, Juang KD, Lee SJ. Psychometric evaluation of a Chinese (Taiwanese) version of the SF-36 health survey amongst middle-aged women from a rural community. Qual Life Res 2000;9(6):675-683. [doi: 10.1023/a:1008993821633] [Medline: 11236857]
51. Tseng H, Lu JR, Gandek B. Cultural issues in using the SF-36 Health Survey in Asia: results from Taiwan. Health Qual Life Outcomes 2003 Nov 26;1:72 [FREE Full text] [doi: 10.1186/1477-7525-1-72] [Medline: 14641915]
52. Bergier J, Kapka-Skrzypczak L, Biliński P, Paprzycki P, Wojtyła A. Physical activity of Polish adolescents and young adults according to IPAQ: a population based study. Ann Agric Environ Med 2012;19(1):109-115 [FREE Full text] [Medline: 22462455]
53. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. Test (Madr) 2009 May 01;18(1):1-43 [FREE Full text] [doi: 10.1007/s11749-009-0138-x] [Medline: 21218187]

54. Sun Y, You W, Almeida F, Estabrooks P, Davy B. The effectiveness and cost of lifestyle interventions including nutrition education for diabetes prevention: a systematic review and meta-analysis. J Acad Nutr Diet 2017 Mar;117(3):404-421 [FREE Full text] [doi: 10.1016/j.jand.2016.11.016] [Medline: 28236962]
55. Wang X, Pi Y, Chen B, Chen P, Liu Y, Wang R, et al. Effect of traditional Chinese exercise on the quality of life and depression for chronic diseases: a meta-analysis of randomised trials. Sci Rep 2015 Nov 03;5:15913. [doi: 10.1038/srep15913] [Medline: 26525282]
56. Taylor LM, Spence JC, Raine K, Plotnikoff RC, Vallance JK, Sharma AM. Physical activity and health-related quality of life in individuals with prediabetes. Diabetes Res Clin Pract 2010 Oct;90(1):15-21. [doi: 10.1016/j.diabres.2010.04.011] [Medline: 20727611]
57. Kyrou I, Tsigos C. Stress hormones: physiological stress and regulation of metabolism. Curr Opin Pharmacol 2009 Dec;9(6):787-793. [doi: 10.1016/j.coph.2009.08.007] [Medline: 19758844]
58. Kaur G, Lakshmi PVM, Rastogi A, Bhansali A, Jain S, Teerawattananon Y, et al. Diagnostic accuracy of tests for type 2 diabetes and prediabetes: a systematic review and meta-analysis. PLoS One 2020;15(11):e0242415 [FREE Full text] [doi: 10.1371/journal.pone.0242415] [Medline: 33216783]

Abbreviations

BCQ: Body Constitution Questionnaire
DASH: Dietary Approaches to Stop Hypertension
DPP: Diabetes Prevention Program
FPG: fasting plasma glucose
GEE: generalized estimating equation
HbA_{1c}: hemoglobin A_{1c}
HRQOL: health-related quality of life
MCS: mental component score
MEAD: Meridian Energy Analysis Device
MET: metabolic equivalent
mHealth: mobile health
PA: physical activity
PCS: physical component score
RCT: randomized controlled trial
SF-36: Medical Outcome Survey Short-Form
T2DM: type 2 diabetes mellitus
TCM: traditional Chinese medicine

Edited by L Buis; submitted 27.07.22; peer-reviewed by R Frost, BY Lai, SB Liang, N Kaufman; comments to author 18.08.22; revised version received 05.01.23; accepted 25.05.23; published 20.06.23.

Please cite as:
Chung HW, Tai CJ, Chang P, Su WL, Chien LY
The Effectiveness of a Traditional Chinese Medicine–Based Mobile Health App for Individuals With Prediabetes: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e41099
URL: https://mhealth.jmir.org/2023/1/e41099
doi:10.2196/41099
PMID:37338977

Original Paper

The Effectiveness of a Mobile Phone–Based Physical Activity Program for Treating Depression, Stress, Psychological Well-Being, and Quality of Life Among Adults: Quantitative Study

Hyungsook Kim[1,2,3], PhD; Kikwang Lee[4], PhD; Ye Hoon Lee[5], PhD; Yoonjung Park[6], PhD; Yonghyun Park[1], PhD; Yeonwoo Yu[4], BA; Jaeyoung Park[5], BA; Sihyeon Noh[5], BA

[1]Hanyang Digital Healthcare Center, Hanyang University, Seoul, Republic of Korea
[2]Department of Cognitive Sciences, School of Intelligence, Hanyang University, Seoul, Republic of Korea
[3]Graduate School of Public Policy, Hanyang University, Seoul, Republic of Korea
[4]Department of Sport, Health, and Rehabilitation, College of Physical Education, Kookmin University, Seoul, Republic of Korea
[5]Division of Global Sport Industry, Hankuk University of Foreign Studies, Gyeonggi-do, Republic of Korea
[6]Department of Health & Human Performance, University of Houston, Houston, TX, United States

Corresponding Author:
Ye Hoon Lee, PhD
Division of Global Sport Industry
Hankuk University of Foreign Studies
81, Oedae-ro, Mohyeon-eup, Cheoin-gu
Gyeonggi-do, 17035
Republic of Korea
Phone: 82 31 330 4986
Email: leeye22@o365.hufs.ac.kr

Abstract

Background: Depression is a substantial global health problem, affecting >300 million people and resulting in 12.7% of all deaths. Depression causes various physical and cognitive problems, leading to a 5-year to 10-year decrease in life expectancy compared with the general population. Physical activity is known to be an effective, evidence-based treatment for depression. However, people generally have difficulties with participating in physical activity owing to limitations in time and accessibility.

Objective: To address this issue, this study aimed to contribute to the development of alternative and innovative intervention methods for depression and stress management in adults. More specifically, we attempted to investigate the effectiveness of a mobile phone–based physical activity program on depression, perceived stress, psychological well-being, and quality of life among adults in South Korea.

Methods: Participants were recruited and randomly assigned to the mobile phone intervention or waitlist group. Self-report questionnaires were used to assess variables before and after treatment. The treatment group used the program around 3 times per week at home for 4 weeks, with each session lasting about 30 minutes. To evaluate the program's impact, a 2 (condition) × 2 (time) repeated-measures ANOVA was conducted, considering pretreatment and posttreatment measures along with group as independent variables. For a more detailed analysis, paired-samples 2-tailed *t* tests were used to compare pretreatment and posttreatment measurements within each group. Independent-samples 2-tailed *t* tests were conducted to assess intergroup differences in pretreatment measurements.

Results: The study included a total of 68 adults aged between 18 and 65 years, who were recruited both through web-based and offline methods. Of these 68 individuals, 41 (60%) were randomly assigned to the treatment group and 27 (40%) to the waitlist group. The attrition rate was 10.2% after 4 weeks. The findings indicated that there is a significant main effect of time ($F_{1,60}$=15.63; P=.003; η_p^2=0.21) in participants' depression scores, indicating that there were changes in depression level across time. No significant changes were observed in perceived stress (P=.25), psychological well-being (P=.35), or quality of life (P=.07). Furthermore, depression scores significantly decreased in the treatment group (from 7.08 to 4.64; P=.03; Cohen d=0.50) but not in the waitlist group (from 6.72 to 5.08; P=.20; Cohen d=0.36). Perceived stress score of the treatment group also significantly decreased (from 2.95 to 2.72; P=.04; Cohen d=0.46) but not in the waitlist group (from 2.82 to 2.74; P=.55; Cohen d=0.15).

Conclusions: This study provided experimental evidence that mobile phone–based physical activity program affects depression significantly. By exploring the potential of mobile phone–based physical activity programs as a treatment option, this study sought to improve accessibility and encourage participation in physical activity, ultimately promoting better mental health outcomes for individuals with depression and stress.

(JMIR Mhealth Uhealth 2023;11:e46286) doi:10.2196/46286

KEYWORDS

depressive symptoms; mobile intervention; exercise; internet-based fitness; mental health

Introduction

Background

Depression is a mental health problem that affects >300 million people worldwide, accounting for 12.7% of all deaths [1]. Depression causes various types of physical and cognitive problems, such as the loss of appetite, sleep disturbances, helplessness, and suicidal thoughts and attempts [2]. Therefore, it has been reported that the mean life expectancy of people with depression is 5 to 10 years shorter than that of the general population [3]. Similar to the global trend, the number of people with depression in South Korea is increasing every year, and medical expenses are also gradually increasing. As the COVID-19 pandemic and social distancing continued in South Korea, 18.9% of the population was classified as a risk group for depression, and the number of people with depression has increased by >30% from approximately 750,000 in 2018 to 830,000 in 2020 [4]. Furthermore, according to the 2018 National Health and Nutrition Examination Survey [5], the stress awareness rate among adults in Korea was 29.1%, and it persisted at approximately 30% over the following 3 years. This means that 3 out of 10 adults feel *very much* or *a lot of* stress in their daily lives. Therefore, interest in intervention methods for managing depression in adults is increasing [6].

The traditional methods of preventing and treating depression include drug therapy, psychotherapy (eg, cognitive therapy and interpersonal therapy), and combinations of both [7]. Antidepressants—for example, selective serotonin reuptake inhibitors—are often used as the first-line treatment for depression because of their advantages of scalability and accessibility [8]. Despite the demonstrable biological underpinnings of the antidepressant mechanisms supporting the use of these drugs [7], it has been reported that only approximately 50% of individuals who take such medication achieve a statistically significant reduction in depressive symptoms [9]. In addition, previous literature has demonstrated the side effects and increased mortality associated with the difficulty of maintaining drug compliance, increased weight and diabetes risk, sexual dysfunction, and various diseases. Moreover, if antidepressants are stopped before recovery, symptoms seem to worsen rapidly, and it is difficult to continue drug treatment because of negative perceptions about drug consumption and dependency [10]. Thus, there is a need for alternative treatment methods that can overcome these issues.

Cognitive behavioral therapy is an alternative treatment that is widely recognized as one of the most effective evidence-based therapies for depression. Studies conducted worldwide have reported small to medium effects on mitigating depression [11]. However, despite the availability of effective treatments, a large proportion of people with depression do not seek or receive professional help, with studies reporting that up to 80% of adolescents with depression do not receive adequate treatment [12]. Barriers to treatment are reported to include prejudice, accessibility issues, and cost issues associated with psychiatric treatment [13]. These issues are not limited to specific regions or countries, as people worldwide often experience reluctance to seek treatment owing to the fear of stigma and high cost of therapy. Furthermore, cognitive behavioral therapy is often a weekly treatment for at least 10 sessions, which can be financially burdensome for many individuals. Therefore, alternative treatment methods that address or avoid these difficulties are needed [14].

Physical activity and exercise can be an effective alternative form of depression treatment [15]. They exert antidepressant effect through multiple biological and psychosocial pathways [16-18]. Kandola et al [7] conducted a review to explore the relationship between physical activity and depression, aiming to understand the mechanisms by which physical activity exerts its antidepressant effect. The review highlights the interdependent changes that occur in the brain owing to exercise, creating a protective environment against depression. These changes include increased levels of neurotrophins, improved brain structure, and enhanced functioning in areas implicated in depression and stress regulation. Exercise produces changes in the brain through various pathways, including neuroplasticity, which is disrupted in depression [19]. Studies have found that exercise can increase the volume of the hippocampus and cortical regions, and high levels of cardiorespiratory fitness are associated with large volumes in these areas [20-22]. In addition, exercise interventions can reduce chronic low-grade inflammation, decrease depressive symptoms, and mitigate oxidative stress, which can contribute to depression. Regular exercise may also help to dampen hypothalamic-pituitary-adrenal axis activity and cortisol sensitivity, leading to increased resilience to stress, and it may have a direct influence on the neuroendocrine system, helping to reduce cortisol levels in people with depression [23]. The review concludes that psychosocial factors, such as self-esteem and self-efficacy, also interact with biological changes to influence depression. Therefore, it is important to note that the mechanisms by which physical activity exerts its antidepressant effect are multifaceted and may vary depending on the individual. Several meta-analyses have found that exercise can reduce the symptoms of depression with a moderate to large effect size and that exercise can be a useful addition to pharmacotherapy and psychotherapy [24-30].

Recently, the COVID-19 pandemic has increased the need for mental health and well-being management in non–face-to-face environments [31]. Depression is strongly associated with social phobias [32], which limit people's ability to exercise in spaces with other people, such as fitness centers. Therefore, easy-to-access and easy-to-use internet-based or mobile phone–based physical activity apps can be good alternatives for people with depression. Mobile phone–based physical activity programs have the potential to provide significant benefits as an adjunct to professional treatment for depression. These benefits include increased convenience, adherence, personalization, social support, and cost-effectiveness [33]. By enabling individuals to engage in physical activity regularly, mobile phone–based programs with reminder and tracking features can increase adherence to physical activity interventions [34]. Furthermore, these programs can be tailored to an individual's specific needs and preferences, potentially increasing motivation and engagement. Social support features, such as web-based communities or peer support, are also common in mobile phone–based programs and may help individuals to stay motivated and engaged in the program [35]. Finally, mobile phone–based programs may be more cost-effective than traditional in-person interventions, making them more accessible to individuals who may not have the financial resources for traditional treatment [36]. Numerous studies have demonstrated the effectiveness of mobile phone–based physical activity programs at home for adults in treating depression [37-40].

In light of the increasing demand for non–face-to-face mental health management during the COVID-19 pandemic, there has been a surge in the development and evaluation of technology-based physical activity interventions [36]. However, despite their potential benefits, there is a lack of studies exploring the effectiveness of web-based physical activity interventions in reducing depression and improving mental health outcomes such as perceived stress [41], psychological well-being [42], and quality of life [36,43]. Moreover, the existing studies have produced mixed outcomes, and many have not included control groups [36], highlighting the need for further studies to examine the mental health outcomes of web-based physical activity interventions.

Objective

To address this issue, we investigated the effectiveness of a mobile phone–based physical activity program among adults in South Korea by examining various mental health indicators such as depression, perceived stress, psychological well-being, and quality of life. Specifically, we hypothesized that the mobile phone–based physical activity program will lead to significant improvement in depression, perceived stress, psychological well-being, and quality of life in the intervention group compared to the control group. We hope that this study of mobile phone–based physical activity programs, which can be easily accessed by people with depression at home, will contribute to mitigating the negative impact of mental health issues. It is our aspiration that this study will provide more comprehensive guidelines for decision makers to promote and execute mobile phone–based exercise interventions for individuals with depression and stress.

Methods

Participants

The participants were recruited through various channels, including the Hanyang Digital Healthcare Center website [44], the Hanyang Happiness Dream Counseling Center, and the related blog [45]. In addition, recruitment efforts included the distribution of posters, flyers, and local newsletters in the community. Participants were recruited from the first week of March 2022 to the third week of April 2022 through the web-based and offline advertisements. Participants completed the baseline questionnaire in the fourth week of April 2022 and the posttest questionnaire in the first week of June 2022. The study enrolled adults who were aged between 18 and 65 years, were fluent in Korean, had basic knowledge about the internet, and could attend the 3 in-person appointments at the project locations (ie, Hanyang Digital Healthcare Center). To be eligible, they needed to have access to the internet and pass the Physical Activity Readiness Questionnaire.

Sample Size Calculation

To calculate the sample size, the G*power 3 program was used, with power set to 95% and significance level of .01 for precise testing. The study measured the degree of depression relief in the fourth week using the sixth week's measurement from a previous study. The effect size was determined to be 1.1957131, and the minimum number of study participants required was calculated to be 54, with 27 participants per group. Given that the study was on the general public who felt depressed for 8 weeks, a dropout rate of 20% was set, and a total of 65 study participants were recruited.

Procedure

Following the randomization process, participants in the intervention group were shown the mobile phone–based physical activity program during their initial appointment and were provided with instructions about how to access and use it at home. To ensure that participants in the treatment group completed the exercise program as intended, we provided them with a set of instructions and guidelines and a calendar to track their progress. We also asked them to complete a Google Form after each exercise session to confirm that they had completed the exercise and to report any issues or concerns. By using the Google Form to collect data about participant compliance, this study minimized any potential bias that could arise from direct contact with participants. Participants in the study were compensated for their time and effort with a monetary reward of KRW ₩100,000 (US $75.76) for the treatment group and KRW ₩50,000 (US $37.88) for the control group at the end of the study. Participants were informed about the compensation at the beginning of the study and were reminded about the amount and timing of payment before the last session. The monetary reward was intended to incentivize participation and improve retention rates.

Ethics Approval and Informed Consent

This study was approved by the Hanyang University (the first authors' institution) institutional review board (HYUIRB-202203-010-2). We obtained the necessary approvals before starting the study to ensure that ethical standards were met, and the rights of the participants were protected. The purpose and procedure of the study were explained before starting of the study, and only those who provided written informed consent and voluntarily participated were included.

Instrument

Overview

The Korean versions of the Patient Health Questionnaire–9 (PHQ-9), Perceived Stress Scale, and World Health Organization (WHO)–5 Well-Being Index were used in this study. These scales have been previously validated in the literature [46]. Translation and cultural adaptation were conducted using established guidelines, including forward and backward translations and cultural adaptation by a team of bilingual experts [47]. The quality-of-life instrument underwent a rigorous translation process, which involved independent translations, review and discussion of discrepancies, back translation, comparison with the original scales, and pilot-testing with Korean-speaking individuals. The translations were revised as needed based on their feedback. The study requested that participants complete paper-and-pencil research surveys at 2 different points in time—at baseline and week 4. The surveys were conducted at the intervention location (ie, Hanyang Digital Healthcare Center).

Depression

As a measurement tool for depression, this study used PHQ-9 [48]. The PHQ-9 is divided into 9 categories of responses to the question, "How often have you suffered from depression-related problems in the past 2 weeks?" The categories are discomfort, depressed feelings, changes in sleep patterns, fatigue, changes in appetite, guilt or worthlessness, poor concentration, restlessness, and suicidal thoughts. High scores calculated by summing the measured scores indicate high degrees of depressive symptoms or high severity of depression.

Perceived Stress

The Korean version of the Perceived Stress Scale [49] was used to measure the perceived stress of the participants. The Perceived Stress Scale has 10 items rated on a 5-point Likert scale (1="not at all" to 5="very much"), and it is composed of 2 subfactors of positive perception and negative perception according to the way stress is perceived. To measure the perceived stress levels, positive perception factors were reverse scored and summed with negative perception factors to calculate the total score on the scale.

Psychological Well-Being

To measure psychological well-being, this study used the WHO-5 Well-Being Index developed by WHO [50], which consists of 5 items. The tool is structured to respond to the following five questions on a 6-point scale ranging from 0="never" to 5="always" in the past 2 weeks: (1) "I felt pleasant and happy," (2) "I was calm and relaxed," (3) "I was active and energetic," (4) "I woke up refreshed in the morning after sleeping," and (5) "My daily life was full of interesting things." Therefore, possible well-being scores range from 0 to 25, with high scores indicating high levels of psychological well-being.

Quality of Life

To measure quality of life, this study used the quality-of-life scale [51]. The tool consists of 4 items, and each item is answered on a 5-point Likert scale ranging from 1="strongly disagree" to 5="strongly agree." High scores indicate high quality-of-life self-ratings. Sample items are "I am satisfied with my life" and "I try to keep developing myself."

Mobile Phone–Based Physical Activity Program

The physical activity program used in this study is mobile based and aims to improve physical fitness factors associated with depression. The program features a series of animated and video demonstrations of exercises that involve repetitive movements, designed to engage and challenge various muscle groups. These movements are performed rhythmically and in a coordinated manner to enhance muscular strength and cardiorespiratory endurance. The program's exercises are repeated several times during each sequence, and 3D animated images display movements designed to enhance coordination and balance by integrating hand and leg movements in different ways. The program's movements challenge the hands, feet, and trunk and are designed to improve muscular strength and cardiorespiratory endurance.

More specifically, the program is composed of 6 therapeutic movement sequences, each incorporating specific fitness elements (as shown in Figure 1). The first and second sequences were designed to proceed more quickly and include mostly arm exercises. Some specific examples of exercises in these sequences include movements such as alternating arm and leg raises while standing on 1 leg. These exercises require participants to use their core muscles to stabilize their bodies while moving their limbs. In addition, the combination of upper and lower body movements can help to improve cardiovascular fitness by increasing the heart rate and respiratory rate.

The third and fourth sequences in the program focus on coordinated movements of both the arms and legs. The exercises are aimed at expanding the range of motion around the body's center of gravity and enhancing muscular endurance and balance. It involves standing with the feet hip-width apart and lifting both arms up and out to the sides of the body, reaching toward the sky. The palms of the hands may face each other or be turned outward. The chest and head are lifted upward to create an expansive posture. This movement is often associated with a feeling of openness and energy, as it stretches and opens the chest, shoulders, and arms. By coordinating the movements of both the arms and legs, participants are required to use multiple muscle groups simultaneously, which can help to improve overall body strength and coordination.

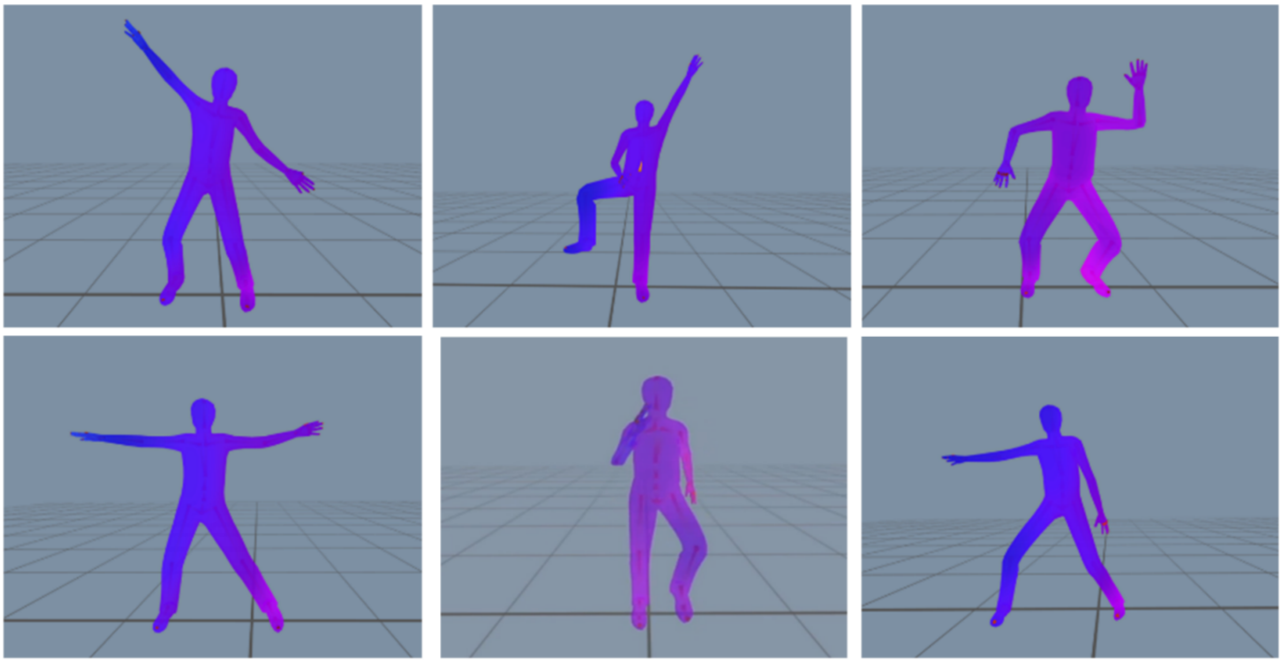
The fifth sequence involves walking and running in place to improve cardiorespiratory capacity. This sequence is specifically designed to improve cardiorespiratory capacity or the ability of the body's cardiovascular and respiratory systems to efficiently deliver oxygen and nutrients to the muscles during physical activity. To perform this sequence, the participant simply walks or runs in place, lifting their feet off the ground and moving their arms back and forth in a natural rhythm. Walking and running in place can be modified to increase the intensity of the exercise by incorporating variations such as high knees, heel kicks, or lateral shuffles. These variations can challenge the cardiovascular and respiratory systems by increasing the intensity and demand of the exercise.

The sixth sequence is designed to enhance coordination between the arms and legs, with the aim of increasing the range of motion around joints. For example, 1 exercise involves lunging forward with 1 leg while simultaneously extending the arm out in front of the body and then alternating sides. Another exercise involves standing with the feet shoulder-width apart and lifting both arms straight up above the head while simultaneously bending the knees and lowering the body into a squat position. Participants may then return to the standing position while lowering the arms to the sides of the body.

In addition, the program's movement exercises consist of actions with characteristics similar to gestures associated with the expression of positive emotions (joy and happiness). The body motions associated with expressions of joy and happiness have the characteristics of expansiveness, wherein the upper extremities and upper body move upward, with the arms extending laterally [52-55]. This typically involves raising both arms and reaching them outward to the sides of the body while also lifting the chest and head upward to create an expansive, uplifting posture. This upward stretching of the hands also has a stress-reducing effect [56,57]. The participants follow the movements of positive emotional expression while upbeat music plays in the background, which can also reduce depression. Moving in rhythm requires the user to engage in active behavior as opposed to psychomotor retardation, which is a behavioral trait associated with depression.

Only participants in the treatment group were instructed to engage in physical activity while watching videos with physical activity content 5 times a week for 4 weeks. The videos consisted of a total of approximately 20 minutes of physical activity content and were delivered to the participants' mobile devices through a YouTube link at the same time every day. The researchers did not provide additional treatment or counseling related to the program content.

Figure 1. Captured images of the mobile physical activity program's 3D human character animation.



Data Analysis

To evaluate the effectiveness of the mobile-based physical activity program for the treatment group and the waitlist group, the self-report questionnaire scores were compared and analyzed using SPSS Statistics for Windows (version 22.0; IBM Corp). To evaluate the impact of the program, we first conducted a 2 (condition) × 2 (time) repeated-measures ANOVA with pretreatment and posttreatment measures and group as independent variables. Following this analysis, we conducted paired-samples 2-tailed *t* tests to compare pretreatment and posttreatment measurements within each group, to gain a more detailed understanding of the data.

Results

Recruitment

Figure 2 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram, which illustrates the participant flow throughout the study. Of the 68 initial participants, 41 (60%) were assigned to the treatment group and 27 (40%) were assigned to the waitlist group. Overall, 88% (36/41) of the participants completed the follow-up assessment in the treatment group, and 93% (25/27) of the participants completed it in the waitlist group.

The participants of this study were 68 adults aged between 18 and 65 years, who voluntarily expressed their intentions to participate. The selected participants were randomly divided into a treatment group (41/68, 60%) and a waitlist group (27/68, 40%). The treatment and waitlist groups were not balanced in terms of participant numbers because most participants expressed their intention to participate voluntarily on the premise that they would be assigned to the intervention group. Overall, 12% (5/41) of participants in the treatment group and 7% (2/27) of participants in the waitlist group dropped out owing to personal reasons. Thus, a total of 61 participants were included in the final data analysis (Table 1). Specifically, of the total 61 participants, the treatment group consisted of 36 (59%) participants, ranging in age from 20 to 54 years, including 17 (47%) male participants and 19 (53%) female participants. Among these 36 participants, 25 (69%) were unmarried and 11 (31%) were married. Regarding educational background, the 36 participants were distributed as follows: 10 (28%) had completed high school, 10 (28%) were college graduates, and 10 (28%) had graduate degrees, with the lowest number of participants having a high school diploma (n=6, 17%). In terms of occupation, of the 36 participants, 16 (44%) were students, 14 (39%) were regular workers, and 3 (8%) were nonregular workers. The waitlist group consisted of a total of 41% (25/61) participants, ranging in age from 19 to 49 years, including 56% (14/25) male participants and 44% (11/25) female participants. Among the 25 participants, 24 (96%) were unmarried and 1 (4%) was divorced. Regarding educational background, of the 25 participants, the most common level had 13 (52%) participants with a high school diploma, followed by 10 (40%) participants with graduate school or higher education, and 2 (8%) participants with a college degree.

Figure 2. Study flow diagram.

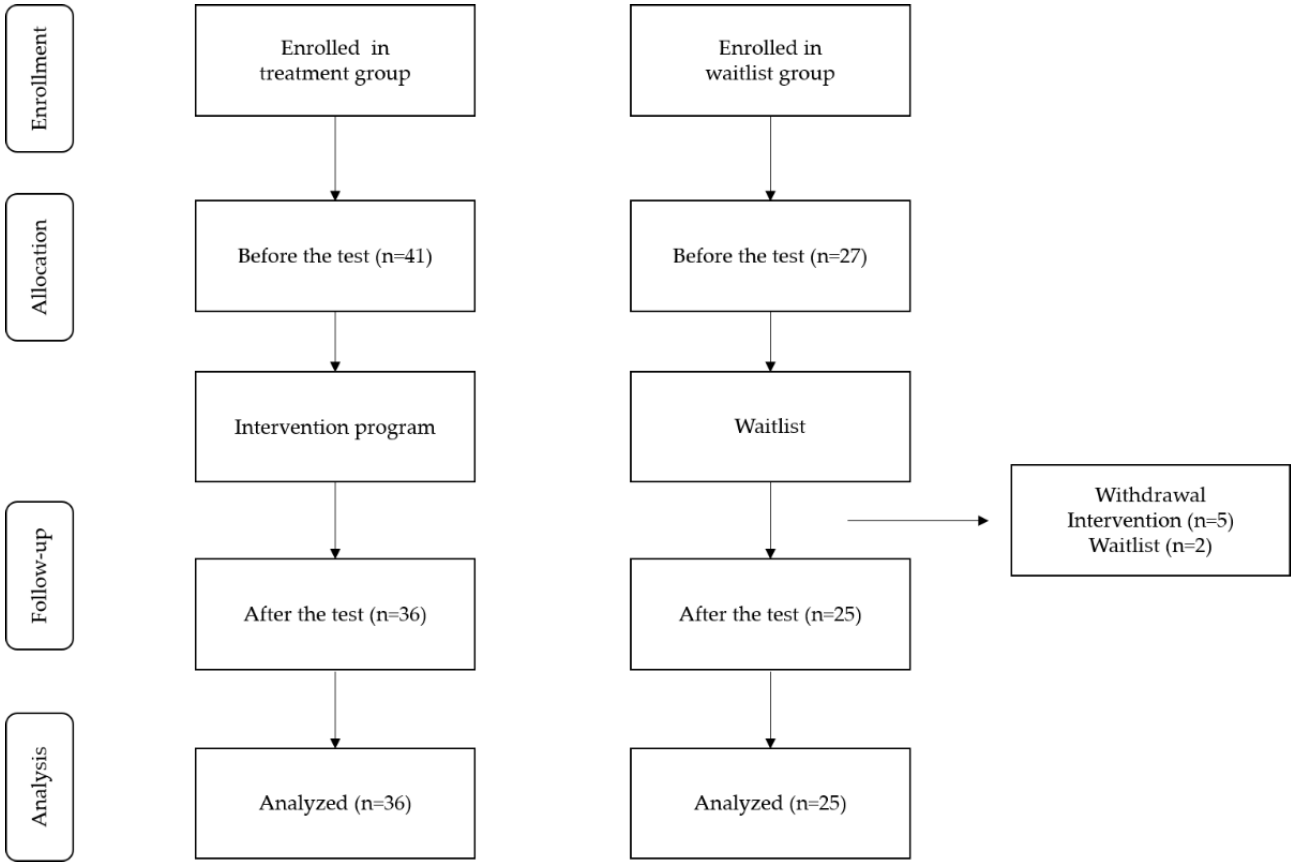


Table 1. Baseline participant characteristics.

Characteristics	All participants (N=68), n (%)	Intervention group (n=41), n (%)	Waitlist group (n=27), n (%)
Sex			
Male	33 (49)	19 (46)	14 (52)
Female	35 (51)	22 (54)	13 (48)
Marital status			
Single	55 (81)	29 (71)	26 (96)
Married	12 (18)	12 (29)	0 (0)
Divorce	1 (1)	0 (0)	1 (4)
Education			
High school graduation	6 (9)	6 (15)	0 (0)
College	27 (40)	12 (29)	15 (56)
Graduation from university	13 (19)	11 (27)	2 (7)
Graduate student	22 (32)	12 (29)	10 (37)
Employment			
Student	35 (51)	19 (46)	16 (59)
Not working	3 (4)	3 (7)	0 (0)
Full time	19 (28)	16 (39)	3 (9)
Part time	11 (16)	3 (7)	8 (30)
Monthly income (KRW ₩ [US $])			
No income	15 (23)	7 (17)	8 (30)
<500,000 (<380.6)	10 (15)	5 (12)	5 (19)
510,000-1,000,000 (388.22-761.21)	6 (9)	3 (7)	3 (11)
1,010,000-1,500,000 (768.82-1141.81)	7 (10)	7 (17)	0 (0)
1,510,000-2,000,000 (1149.43-1522.42)	5 (7)	2 (5)	3 (11)
2,010,000-2,500,000 (1503.03-1903.02)	7 (10)	2 (5)	5 (19)
2,510,000-3,000,000 (1910.63-2283.63)	3 (4)	1 (2)	2 (7)
3,010,000-4,000,000 (2291.24-3044.84)	5 (7)	5 (12)	0 (0)
>4,000,000 (>3052.45)	10 (15)	9 (22)	1 (3)

Changes Observed in the Treatment Group and Waitlist Group

First, Tables 2 and 3 summarize the changes (in the treatment and waitlist groups, respectively) in the investigated variables between baseline and after the implementation of the mobile phone–based physical activity program developed for this study. In the treatment group, the mean depression score significantly decreased from 7.08 (SD 5.49) to 4.64 (SD 4.07; t_{70}=2.14, P=.03; Cohen d=0.50). The mean depression score of the waitlist group also decreased (from 6.72, SD 5.43 to 5.08, SD 3.4), but this change was not statistically significant (t_{48}=1.27, P=.20; Cohen d=0.36). There was significant decrease in the mean perceived stress score of the treatment group (from 2.95, SD 0.51 to 2.72, SD 0.5; t_{70}=1.99, P=.04; Cohen d=0.46). In the waitlist group, the mean perceived stress score decreased from 2.82 (SD 0.58) to 2.74 (SD 0.45), but this change was not statistically significant (t_{48}=0.59, P=.55; Cohen d=0.15).

The mean psychological well-being score of the treatment group increased from 2.63 (SD 0.80) to 2.96 (SD 0.92) between baseline and after the treatment, but this change was not statistically significant (t_{70}=−1.63, P=.10; Cohen d=0.38). In the waitlist group, there was slight increase in the mean psychological well-being score from 2.91 (SD 0.80) to 3.04 (SD 0.62; t_{48}=−0.63, P=.53; Cohen d=0.18). Finally, the mean quality-of-life scores increased from 2.89 (SD 0.92) to 3.27 (SD 0.88) and from 3.12 (SD 0.85) to 3.33 (SD 0.79) in the treatment and waitlist groups, respectively, between baseline and after the treatment. However, neither of these increases was statistically significant (t_{70}=−1.76, P=.08; Cohen d=0.42 for the treatment group; t_{48}=−0.89, P=.37; Cohen d=0.25 for the waitlist group).

Figure 3 shows the changes in the treatment and waitlist groups, between baseline and program completion, according to repeated-measures ANOVA of the depression, perceived stress, psychological well-being, and quality-of-life scores. As shown in Table 3, none of the evaluated changes were statistically significant in association with the mobile phone–based exercise. However, accounting for all participants (irrespective of group), depression levels ($F_{1,60}$=15.63; P=.001; η_p^2=0.21) decreased significantly after treatment relative to baseline. However, according to the ANOVA, overall, there were no significant changes between baseline and after treatment in perceived stress ($F_{1,60}$=7.58; P=.25; η_p^2=0.04), psychological well-being ($F_{1,60}$=6.97; P=.35; η_p^2=0.03), or quality of life ($F_{1,60}$=3.56; P=.06; η_p^2=0.10).

Table 2. Mean differences between pretreatment and posttreatment scores in the treatment group.

	Pretreatment score, mean (SD)	Posttreatment score, mean (SD)	*t* test (*df*)[a]	*P* value	Cohen *d*
Depression	7.08 (5.49)	4.64 (4.07)	2.14 (70)	.03	0.50
Perceived stress	2.95 (0.51)	2.72 (0.49)	1.99 (70)	.04	0.46
Psychological well-being	2.63 (0.80)	2.96 (0.92)	−1.63 (70)	.10	0.38
Quality of life	2.89 (0.92)	3.27 (0.88)	−1.76 (70)	.08	0.42

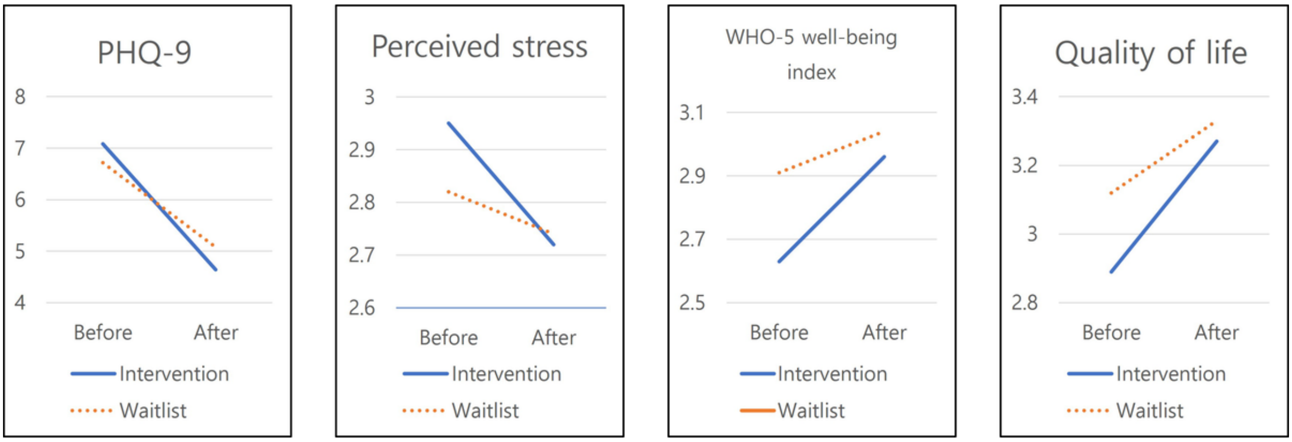
[a]2-tailed *t* test.

Table 3. Mean differences between pretreatment and posttreatment scores in the waitlist group.

	Pretreatment score, mean (SD)	Posttreatment score, mean (SD)	*t* test (*df*)[a]	*P* value	Cohen *d*
Depression	6.72 (5.43)	5.08 (3.4)	1.27 (48)	.20	0.36
Perceived stress	2.82 (0.58)	2.74 (0.45)	0.59 (48)	.55	0.15
Psychological well-being	2.91 (0.80)	3.04 (0.62)	−0.63 (48)	.53	0.18
Quality of life	3.12 (0.85)	3.33 (0.79)	−0.89 (48)	.37	0.25

[a]2-tailed *t* test.

Figure 3. Interaction analysis results. PHQ-9: Patient Health Questionnaire–9; WHO-5: World Health Organization–5.



Second, independent-samples 2-tailed *t* tests were performed on the pretreatment results to evaluate the differences between the treatment group and the waitlist group at baseline. Table 4 summarizes the characteristics of the participants in the treatment group and the waitlist group. Regarding the negative well-being indicators, the level of depression was higher in the treatment group (mean 7.08, SD 5.49) than in the waitlist group (mean 6.72, SD 5.43). Perceived stress was also higher in the treatment group (mean 2.95, SD 0.51) than in the waitlist group (mean 2.82, SD 0.58). However, there were no statistically significant differences between the groups in terms of depression (t_{59}=−0.25, P=.79; Cohen d=0.06) or perceived stress (t_{59}=−0.92, P=.36; Cohen d=0.23). Regarding the pretreatment positive well-being indicators, psychological well-being (mean 2.91, SD 0.80) and quality of life (mean 3.12, SD 0.85) were higher in the waitlist group than in the treatment group (mean 2.63, SD 0.80 for psychological well-being; mean 2.89, SD 0.92 for quality of life), but these differences were not statistically significant (t_{59}=1.33, P=.18; Cohen d=0.35 for psychological well-being; t_{59}=0.96, P=.34; Cohen d=0.26 for quality of life).

Finally, Table 5 summarize the posttreatment intergroup comparisons. The mean scores of the waitlist group for all variables—depression (mean 5.08, SD 3.45), perceived stress (mean 2.74, SD 0.45), psychological well-being (mean 3.04, SD 0.62), and quality of life (mean 3.33, SD 0.79)—were higher than those of the treatment group (mean 4.64, SD 4.07 for depression; mean 2.72, SD 0.49 for perceived stress; mean 2.96, SD 0.92 for psychological well-being; and mean 3.27, SD 0.88 for quality of life). However, none of the posttreatment intergroup differences were statistically significant (t_{59}=0.44, P=.66; Cohen d=0.12 for depression; t_{59}=0.12, P=.90; Cohen

d=0.04 for perceived stress; t_{59}=0.34, P=.73; Cohen d=0.10 for psychological well-being; and t_{59}=0.26, P=.79; Cohen d=0.07 for quality of life).

Table 4. Pretreatment results of the treatment and waitlist groups.

	Treatment group score, mean (SD)	Waitlist group score, mean (SD)	t test (df)[a]	P value	Cohen d
Depression	7.08 (5.49)	6.72 (5.43)	−0.25 (59)	.79	0.06
Perceived stress	2.95 (0.51)	2.82 (0.58)	−0.92 (59)	.36	0.23
Psychological well-being	2.63 (0.80)	2.91 (0.80)	1.33 (59)	.18	0.35
Quality of life	2.89 (0.92)	3.12 (0.85)	0.96 (59)	.34	0.26

[a]2-tailed t test.

Table 5. Posttreatment results of the treatment and waitlist groups.

	Treatment group score, mean (SD)	Waitlist group score, mean (SD)	t test (df)[a]	P value	Cohen d
Depression	4.64 (4.07)	5.08 (3.45)	0.44 (59)	.66	0.12
Perceived stress	2.72 (0.49)	2.74 (0.45)	0.12 (59)	.90	0.04
Psychological well-being	2.96 (0.92)	3.04 (0.62)	0.34 (59)	.73	0.10
Quality of life	3.27 (0.88)	3.33 (0.79)	0.26 (59)	.79	0.07

[a]2-tailed t test.

Discussion

Summary

Owing to the stigma and cost barriers associated with psychiatric treatment, alternative methods that are easily accessible and user-friendly are needed to improve depression and stress management among the South Korean population. This study sought to investigate the effectiveness of a mobile phone–based physical activity program on depression among adults in South Korea. Scales to evaluate the severity of depressive symptoms, perceived stress, psychological well-being, and quality of life were assessed twice (before and after program use) after allocating the adults who participated in the study to a treatment group and a waitlist group. The analysis uncovered disparities in the impact of time on depression between the treatment group and the waitlist group. In addition, the pretreatment and posttreatment comparisons indicated decreased depression and perceived stress scores in the treatment group. The study suggests that digital physical activity applications can be a valuable tool in promoting mental health and well-being, particularly in populations with limited access to traditional mental health services. These findings are particularly relevant in the context of the COVID-19 pandemic, where stress and depression have become more prevalent.

Principal Findings

The strength of this study lies in its demonstration of the effectiveness of a mobile phone–based physical activity program in mitigating depressive symptoms and other mental health indicators among adults. This study adds to the growing body of research exploring the potential of digital applications for depression treatment, which is actively being developed and verified internationally. In recent years, digital physical activity applications have gained popularity, with fitness tracking apps such as MyFitnessPal, Fitbit, and Strava being actively developed. These apps use sensors to track users' physical activity, including steps taken, distance traveled, and calories burned. They also allow users to set goals and track progress over time. Virtual reality fitness games, such as RingFit, Beat Saber, and BoxVR, are another area of ongoing development in digital physical activity applications. These games use virtual reality technology to create immersive environments that users can interact with through physical activity. The gamification of physical activity is also an area of ongoing development through apps such as Pokemon Go and Zombies, Run! This study is particularly important, as it demonstrates the effectiveness of a home-based physical activity program for reducing depressive symptoms and perceived stress among adults in South Korea. The cost-effectiveness of this treatment program adds to its potential as a means of promoting mental health and physical activity in populations with limited access to traditional mental health services. Therefore, this study provides an important foundation for further studies into the efficacy of digital physical activity programs as a means of promoting mental health and physical activity, particularly in populations experiencing high levels of stress and depression.

Recently, especially since the onset of the COVID-19 pandemic, many studies have been conducted to investigate the effectiveness of mobile phone–based health apps in promoting physical activity, and it is evident that such apps are promising tools for promoting physical activity [58]. The mobile phone–based physical activity program evaluated in this study was shown to be effective; thus, our results concur with previous evidence showing that physical activity is associated with improvements in depression and perceived stress. Our mobile phone–based exercise program facilitated a combination of muscular strength training and aerobic training, which has been previously shown to be effective. Although there is robust evidence for the beneficial effects of exercise on mental health, various studies have supported different types and modes of exercise; for example, 16 weeks of aerobic exercise (dance, jump, and traditional games) improved the depressive symptoms of adults with depression [59], and 8 weeks of high-intensity strength training (at 80% of the 1-repetition maximum) was beneficial in the treatment of older adults with depression [60]. However, regardless of the type and mode of exercise, physical activity is beneficial for mental health. Therefore, any type of exercise using mobile phone–based physical activity is recommended for mental health.

Regular physical activity and exercise are effective therapies for most chronic diseases, including mental disorders [15,61]. For example, a recent meta-analysis by Pearce et al [24], which included 15 studies and 2,110,588 person-years, demonstrated that participants accumulating half the recommended amount of physical activity had an 18% lower risk of depression than adults without physical activity. Furthermore, adults accumulating the recommended volume of 8.8 marginal metabolic equivalent task hours per week had a 25% lower risk of diminishing potential benefits. Several systematic reviews have found that exercise can reduce the symptoms of depression with a moderate to large effect size and can be a useful addition to pharmacotherapy and psychotherapy [25-30]. Exercise has been shown to counteract reductions in the secretion of neurotransmitters, such as dopamine and serotonin, thus having a positive effect on emotions and reducing the severity of psychological symptoms, such as anxiety and depression. He et al [62] reported that 10 weeks of voluntary running exercise significantly increased serotonin, dopamine, and norepinephrine levels in the hippocampus, which had been reduced in rats with chronic mild stress.

The results of this study indicate that the mobile phone–based physical activity program developed for this study led to a significant decrease in perceived stress among participants in the treatment group, whereas the waitlist group showed no significant change. The significant reduction in perceived stress observed in the treatment group suggests that mobile phone–based physical activity programs could be a viable option for individuals seeking to reduce stress levels [63]. There are several potential underlying mechanisms that may explain the significant effect of the mobile phone–based physical activity program on perceived stress. First, physical activity has been shown to stimulate the release of endorphins, which are natural chemicals that can improve mood and reduce feelings of stress and anxiety. These endorphins can counteract the negative feelings associated with stress and promote a sense of well-being. Second, engaging in physical activity may help individuals to divert their mind from stressful thoughts or situations, providing a temporary escape and reducing overall levels of perceived stress. In other words, when individuals engage in physical activity, they may shift their attention away from the stressors or negative thoughts that are causing stress. This can provide a temporary break from stress, allowing the individual to clear their mind and return to their stressors with a renewed focus and sense of calmness. The theoretical implications of this study are also important. According to the stress and coping theory [64], stress is a result of the relationship between an individual and their environment and the way in which they perceive and cope with stressors. This theory suggests that stress can be managed through the development of effective coping strategies, such as engaging in physical activity. The findings of this study support the stress and coping theory, as the mobile phone–based physical activity program was found to be effective in reducing perceived stress levels among participants. Furthermore, the results of this study provide support for the use of digital interventions in mental health treatment. The use of mobile phone–based physical activity programs is an example of how technology can be leveraged to improve mental health outcomes. This study adds to the growing body of research on the effectiveness of digital interventions for mental health [65,66]. The use of mobile phone–based physical activity programs could be a useful addition to the mental health treatment landscape, particularly in populations with limited access to traditional mental health services.

On the basis of the results of this study, it appears that the mobile phone–based physical activity program did not have significant effect on perceived stress, psychological well-being, or quality of life among participants. Although there was slight increase in the mean scores for psychological well-being and quality of life in both the treatment and waitlist groups, these increases were not statistically significant. It is possible that the lack of significant findings in this study could be owing to various factors. For example, the duration and frequency of the program may not have been sufficient to produce significant changes in mental health outcomes. In addition, individual differences in adherence to the program and motivation to engage in physical activity could have influenced the results.

Despite the lack of significant findings, the use of mobile phone–based physical activity programs as a cost-effective and accessible alternative for the treatment of depression and perceived stress among adults should not be dismissed [67]. It is important to continue exploring and refining such programs to maximize their potential benefits for mental health. Moreover, the nonsignificant changes in quality of life and psychological well-being in the waitlist group suggest that simply being enrolled in the program could have a positive effect on these outcomes. Future studies should explore the potential placebo effect of simply being enrolled in a program and the potential benefits of combining mobile phone–based physical activity programs with other interventions such as cognitive behavioral therapy or medication.

Practical Implications

On the basis of the findings of this study, practitioners can suggest mobile phone–based physical activity programs as a cost-effective and accessible alternative for the treatment of depression and perceived stress among adults. These programs can be easily implemented at home and can help individuals overcome time and accessibility barriers associated with physical activity.

For example, a practitioner working in a mental health clinic can recommend a mobile phone–based physical activity program to their patients who are experiencing symptoms of depression or perceived stress. The program can be used as a complement to their existing treatment or as a stand-alone intervention. The practitioner can provide guidance about how to use the program, monitor the patient's progress, and adjust the treatment plan accordingly. The expected outcomes of using such a program can be a significant decrease in depressive symptoms and perceived stress levels, leading to improved mental and physical well-being. Patients may also experience increase in motivation, self-efficacy, and overall quality of life. However, it is important to note that the program may not have a significant effect on other variables such as psychological well-being or quality of life, as observed in this study.

Furthermore, a community center or gym can offer a mobile phone–based physical activity program to members who are unable to attend in-person exercise classes. This can help make physical activity more accessible to individuals who may have difficulty in traveling to a gym or attending a class during scheduled times. In addition, the program can be offered at a lower cost than that of in-person classes, making it a more affordable option for those with a tight budget.

Finally, an employer can provide a mobile phone–based physical activity program to their employees as part of a workplace wellness program. This can help employees manage stress and improve their mental health, which could lead to increased productivity and job satisfaction. In addition, offering such a program can demonstrate that the employer values employee well-being and is committed to promoting a healthy work environment.

Limitations and Future Research Directions

The limitations of this study were as follows. First, the treatment period of 4 weeks was short, and it limited the confirmation of the prevention effect over time. Therefore, more rigorous evaluations of the effectiveness of such programs—via randomized controlled trials in which participants are randomly assigned to treatment and waitlist groups—should be conducted, and such trials should be conducted over sufficient treatment periods of at least 3 months.

Second, the participants of this study consisted of members of the general public with low levels of depressive symptoms. Thus, it cannot be generalized that findings apply to adults with high levels of depressive symptoms. Although it is worthwhile to investigate the effectiveness of digital applications for this type of sample in an attempt to prevent severe depression, the program may need to target patients who have mild to severe depressive symptoms for treatment purposes.

Third, although the mobile phone–based physical activity program was modified and supplemented through preliminary studies, many factors need to be corrected in the course of the implementation of the program. Moreover, considering the feedback of the participants who completed 4 weeks of the program, it is possible that the lack of graphical detail in the content composition and errors in the program acted as factors that lowered the motivation of the participants. Therefore, there is a need to improve participation levels by further supplementing and improving the treatment program based on the feedback of the study participants.

Finally, although this study did not find significant changes in psychological well-being or quality of life among participants, the potential benefits of mobile phone–based physical activity programs for mental health should not be overlooked. Further studies are needed to identify optimal program characteristics and implementation strategies to maximize their efficacy in improving psychological well-being and quality of life.

Conclusions

In conclusion, this study investigated the effects of mobile phone–based physical activity programs on depression, perceived stress, psychological well-being, and quality of life. The findings indicated that 4 weeks of training using the program significantly reduced the participants' levels of depression and perceived stress. This study suggests that practitioners can recommend mobile phone–based physical activity programs as a cost-effective and accessible alternative for treating depression and perceived stress among adults. Practitioners can offer guidance to patients about how to use the program and monitor their progress while adjusting their treatment plan accordingly. This program can be used as a complement to existing treatment or as a stand-alone intervention. Expected outcomes may include significant reduction in depressive symptoms and perceived stress levels, increased motivation and self-efficacy, and improved overall quality of life. However, practitioners should be aware that the program may not have significant effect on other variables such as psychological well-being or quality of life. Community centers or gyms can also offer such programs to members who are unable to attend in-person classes, making physical activity more accessible and affordable. Employers can also provide these programs as part of a workplace wellness program, thus promoting a healthy work environment and improving employee productivity and job satisfaction.

Acknowledgments

This study was supported by the Bio and Medical Technology Development Program of the National Research Foundation, funded by the South Korean government (MSIT; NRF-2021M3A9E4080780) and the Hankuk University of Foreign Studies (2023).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1242 KB - mhealth_v11i1e46286_app1.pdf]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/254610 [accessed 2022-04-01]
2. Isacsson G, Reutfors J, Papadopoulos FC, Ösby U, Ahlner J. Antidepressant medication prevents suicide in depression. Acta Psychiatr Scand 2010 Dec;122(6):454-60. Retracted in: Acta Psychiatr Scand 2012 May. [doi: 10.1111/j.1600-0447.2010.01561.x] [Medline: 20384599]
3. Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. Dialogues Clin Neurosci 2011;13(1):7-23 [FREE Full text] [doi: 10.31887/DCNS.2011.13.1/wkaton] [Medline: 21485743]
4. 2021 National Health Insurance Statistical Yearbook. Health Insurance Review & Assessment Service. 2023 Jan 9. URL: https://tinyurl.com/4et6kn6k [accessed 2023-01-18]
5. Korea National Health and Nutrition Examination Survey. Korea Disease Control and Prevention Agency. 2022. URL: https://knhanes.kdca.go.kr/knhanes/eng/index.do [accessed 2021-03-10]
6. Laferton JA, Fischer S, Ebert DD, Stenzel NM, Zimmermann J. The effects of stress beliefs on daily affective stress responses. Ann Behav Med 2020 Mar 24;54(4):258-267. [doi: 10.1093/abm/kaz046] [Medline: 31676902]
7. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019 Dec;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
8. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in prescription drug use among adults in the United States from 1999-2012. JAMA 2015 Nov 03;314(17):1818-1831 [FREE Full text] [doi: 10.1001/jama.2015.13766] [Medline: 26529160]
9. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-1917. [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]
10. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018 Apr 07;391(10128):1357-1366 [FREE Full text] [doi: 10.1016/S0140-6736(17)32802-7] [Medline: 29477251]
11. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014 Apr;171(4):453-462 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13030325] [Medline: 24434956]
12. Mariu KR, Merry SN, Robinson EM, Watson PD. Seeking professional help for mental health problems, among New Zealand secondary school students. Clin Child Psychol Psychiatry 2012 Apr;17(2):284-297. [doi: 10.1177/1359104511404176] [Medline: 21852315]
13. Morawska A, Sultan A. Parental attitudes towards seeking professional help for their children: a community sample from the sultanate of Oman. J Child Fam Stud 2016;25(3):979-987. [doi: 10.1007/s10826-015-0257-9]
14. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017 Feb;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
15. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. Am J Psychiatry 2018 Jul 01;175(7):631-648. [doi: 10.1176/appi.ajp.2018.17111194] [Medline: 29690792]
16. Arem H, Moore SC, Patel A, Hartge P, Berrington de Gonzalez A, Visvanathan K, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. JAMA Intern Med 2015 Jun;175(6):959-967 [FREE Full text] [doi: 10.1001/jamainternmed.2015.0533] [Medline: 25844730]
17. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
18. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med 2014 Jan;44(1):81-121 [FREE Full text] [doi: 10.1007/s40279-013-0090-5] [Medline: 24030837]
19. Krogh J, Rostrup E, Thomsen C, Elfving B, Videbech P, Nordentoft M. The effect of exercise on hippocampal volume and neurotrophines in patients with major depression--a randomized clinical trial. J Affect Disord 2014 Aug;165:24-30. [doi: 10.1016/j.jad.2014.04.041] [Medline: 24882173]
20. Firth J, Stubbs B, Vancampfort D, Schuch F, Lagopoulos J, Rosenbaum S, et al. Effect of aerobic exercise on hippocampal volume in humans: a systematic review and meta-analysis. Neuroimage 2018 Feb 01;166:230-238 [FREE Full text] [doi: 10.1016/j.neuroimage.2017.11.007] [Medline: 29113943]
21. Marosi K, Bori Z, Hart N, Sárga L, Koltai E, Radák Z, et al. Long-term exercise treatment reduces oxidative stress in the hippocampus of aging rats. Neuroscience 2012 Dec 13;226:21-28. [doi: 10.1016/j.neuroscience.2012.09.001] [Medline: 22982624]

22. Zheng H, Liu Y, Li W, Yang B, Chen D, Wang X, et al. Beneficial effects of exercise and its molecular mechanisms on depression in rats. Behav Brain Res 2006 Mar 15;168(1):47-55 [FREE Full text] [doi: 10.1016/j.bbr.2005.10.007] [Medline: 16290283]
23. Heijnen S, Hommel B, Kibele A, Colzato LS. Neuromodulation of aerobic exercise-a review. Front Psychol 2015;6:1890 [FREE Full text] [doi: 10.3389/fpsyg.2015.01890] [Medline: 26779053]
24. Pearce M, Garcia L, Abbas A, Strain T, Schuch FB, Golubic R, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022 Jun 01;79(6):550-559 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
25. Bridle C, Spanjers K, Patel S, Atherton NM, Lamb SE. Effect of exercise on depression severity in older people: systematic review and meta-analysis of randomised controlled trials. Br J Psychiatry 2012 Sep;201(3):180-185. [doi: 10.1192/bjp.bp.111.095174] [Medline: 22945926]
26. D'Amore C, Reid JC, Chan M, Fan S, Huang A, Louie J, et al. Interventions including smart technology compared with face-to-face physical activity interventions in older adults: systematic review and meta-analysis. J Med Internet Res 2022 Oct 31;24(10):e36134 [FREE Full text] [doi: 10.2196/36134] [Medline: 36315229]
27. Josefsson T, Lindwall M, Archer T. Physical exercise intervention in depressive disorders: meta-analysis and systematic review. Scand J Med Sci Sports 2014 Apr;24(2):259-272. [doi: 10.1111/sms.12050] [Medline: 23362828]
28. Kvam S, Kleppe CL, Nordhus IH, Hovland A. Exercise as a treatment for depression: a meta-analysis. J Affect Disord 2016 Sep 15;202:67-86. [doi: 10.1016/j.jad.2016.03.063] [Medline: 27253219]
29. Matias TS, Lopes MV, da Costa BG, Silva KS, Schuch FB. Relationship between types of physical activity and depression among 88,522 adults. J Affect Disord 2022 Jan 15;297:415-420. [doi: 10.1016/j.jad.2021.10.051] [Medline: 34715174]
30. Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Zimmermann AB, Rocha NS, Fleck MP. Exercise and severe major depression: effect on symptom severity and quality of life at discharge in an inpatient cohort. J Psychiatr Res 2015 Feb;61:25-32. [doi: 10.1016/j.jpsychires.2014.11.005] [Medline: 25439084]
31. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
32. Ohayon MM, Schatzberg AF. Social phobia and depression: prevalence and comorbidity. J Psychosom Res 2010 Mar;68(3):235-243. [doi: 10.1016/j.jpsychores.2009.07.018] [Medline: 20159208]
33. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
34. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
35. Hutchesson MJ, Gough C, Müller AM, Short CE, Whatnall MC, Ahmed M, et al. eHealth interventions targeting nutrition, physical activity, sedentary behavior, or obesity in adults: a scoping review of systematic reviews. Obes Rev 2021 Oct;22(10):e13295. [doi: 10.1111/obr.13295] [Medline: 34159684]
36. Vandelanotte C, Duncan MJ, Plotnikoff RC, Rebar A, Alley S, Schoeppe S, et al. Impact of a web-based personally tailored physical activity intervention on depression, anxiety, stress and quality of life: secondary outcomes from a randomized controlled trial. Ment Health Phys Act 2022 Oct;23:100477. [doi: 10.1016/j.mhpa.2022.100477]
37. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
38. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
39. Murray A, Marenus M, Cahuas A, Friedman K, Ottensoser H, Kumaravel V, et al. The impact of web-based physical activity interventions on depression and anxiety among college students: randomized experimental trial. JMIR Form Res 2022 Apr 01;6(4):e31839 [FREE Full text] [doi: 10.2196/31839] [Medline: 35363151]
40. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-based intervention using behavioral activation and physical activity for adults with depression (the eMotion study): pilot randomized controlled trial. J Med Internet Res 2018 Jul 16;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
41. Green J, Huberty J, Puzia M, Stecher C. The effect of meditation and physical activity on the mental health impact of COVID-19-related stress and attention to news among mobile app users in the United States: cross-sectional survey. JMIR Ment Health 2021 Apr 13;8(4):e28479 [FREE Full text] [doi: 10.2196/28479] [Medline: 33788698]
42. Suh A, Li M. How the use of mobile fitness technology influences older adults' physical and psychological well-being. Comput Human Behav 2022 Jun;131:107205. [doi: 10.1016/j.chb.2022.107205]
43. Broekhuizen K, de Gelder J, Wijsman CA, Wijsman LW, Westendorp RG, Verhagen E, et al. An internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016 Apr 27;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]

44. Digital treatment platform. Hanyang University Digital Healthcare. URL: http://www.dtxplatform.kr/ [accessed 2023-06-01]
45. Hanyang Digital Healthcare Center, Hanyang University. URL: https://blog.naver.com/hydhcorg [accessed 2023-05-31]
46. 2019 Standard Guidelines for Mental Health Examination Tools and Use. Ministry of Health and Welfare, Korea. 2020. URL: https://tinyurl.com/28cewbfs [accessed 2023-03-20]
47. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
48. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
49. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
50. Bech P. Measuring the dimension of psychological general well-being by the WHO-5. Qual Life Res 2004;32:15-16.
51. Luo Y, Lanlung (Luke) C, Kim E, Tang LR, Song SM. Towards quality of life: the effects of the wellness tourism experience. J Travel Tour Mark 2018;35(4):410-424. [doi: 10.1080/10548408.2017.1358236]
52. Gunes H, Piccardi M. Fusing face and body gesture for machine recognition of emotions. In: Proceedings of the 2005 IEEE International Workshop on Robot and Human Interactive Communication. 2005 Presented at: ROMAN '05; August 13-15, 2005; Nashville, TN, USA p. 306-311. [doi: 10.1109/roman.2005.1513796]
53. Gunes H, Piccardi M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: Proceedings of the 18th International Conference on Pattern Recognition. 2006 Presented at: ICPR '06; August 20-24, 2006; Hong Kong, China p. 1148-1153. [doi: 10.1109/icpr.2006.39]
54. Gunes H, Shan C, Chen S, Tian Y. Bodily expression for automatic affect recognition. In: Konar A, Chakraborty A, editors. Emotion Recognition: A Pattern Analysis Approach. Hoboken, NJ, USA: John Wiley & Sons; 2015:343-377.
55. Noroozi F, Corneanu CA, Kaminska D, Sapinski T, Escalera S, Anbarjafari G. Survey on emotional body gesture recognition. IEEE Trans Affect Comput 2021 Apr 1;12(2):505-523. [doi: 10.1109/taffc.2018.2874986]
56. Bonura KB. The psychological benefits of yoga practice for older adults: evidence and guidelines. Int J Yoga Therap 2011(21):129-142. [Medline: 22398354]
57. Salmon P, Lush E, Jablonski M, Sephton SE. Yoga and mindfulness: clinical aspects of an ancient mind/body practice. Cogn Behav Pract 2009 Feb;16(1):59-72 [FREE Full text] [doi: 10.1016/j.cbpra.2008.07.002]
58. Srivastav AK, Khadayat S, Samuel AJ. Mobile-based health apps to promote physical activity during COVID-19 lockdowns. J Rehabil Med Clin Commun 2021 Feb 25;4:1000051 [FREE Full text] [doi: 10.2340/20030711-1000051] [Medline: 33884153]
59. Carneiro LS, Fonseca AM, Vieira-Coelho MA, Mota MP, Vasconcelos-Raposo J. Effects of structured exercise and pharmacotherapy vs. pharmacotherapy for adults with depressive symptoms: a randomized clinical trial. J Psychiatr Res 2015 Dec;71:48-55. [doi: 10.1016/j.jpsychires.2015.09.007] [Medline: 26522871]
60. Singh NA, Stavrinos TM, Scarbek Y, Galambos G, Liber C, Fiatarone Singh MA. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. J Gerontol A Biol Sci Med Sci 2005 Jun;60(6):768-776. [doi: 10.1093/gerona/60.6.768] [Medline: 15983181]
61. Jiménez-Pavón D, Carbonell-Baeza A, Lavie CJ. Physical exercise as therapy to fight against the mental and physical consequences of COVID-19 quarantine: special focus in older people. Prog Cardiovasc Dis 2020;63(3):386-388 [FREE Full text] [doi: 10.1016/j.pcad.2020.03.009] [Medline: 32220590]
62. He SB, Tang WG, Tang WJ, Kao XL, Zhang CG, Wong XT. Exercise intervention may prevent depression. Int J Sports Med 2012 Jul;33(7):525-530. [doi: 10.1055/s-0032-1306325] [Medline: 22504906]
63. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in app stores. JMIR Mhealth Uhealth 2018 May 09;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
64. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. Cham, Switzerland: Springer; 1984.
65. Zahrt OH, Evans K, Murnane E, Santoro E, Baiocchi M, Landay J, et al. Effects of wearable fitness trackers and activity adequacy mindsets on affect, behavior, and health: longitudinal randomized controlled trial. J Med Internet Res 2023 Jan 25;25:e40529 [FREE Full text] [doi: 10.2196/40529] [Medline: 36696172]
66. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
67. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

PHQ-9: Patient Health Questionnaire–9
WHO: World Health Organization

Edited by L Buis; submitted 07.02.23; peer-reviewed by O Michael, B Woo, H Cho; comments to author 14.03.23; revised version received 10.04.23; accepted 16.05.23; published 26.06.23.

Please cite as:
Kim H, Lee K, Lee YH, Park Y, Park Y, Yu Y, Park J, Noh S
The Effectiveness of a Mobile Phone–Based Physical Activity Program for Treating Depression, Stress, Psychological Well-Being, and Quality of Life Among Adults: Quantitative Study
JMIR Mhealth Uhealth 2023;11:e46286
URL: https://mhealth.jmir.org/2023/1/e46286
doi:10.2196/46286
PMID:37358905

Original Paper

Feasibility, Acceptability, and Potential Impact of a Novel mHealth App for Smokers Ambivalent About Quitting: Randomized Pilot Study

Jennifer B McClure[1,2], PhD; Jaimee L Heffner[3], PhD; Chloe Krakauer[1], PhD; Sophia Mun[1], MPH; Predrag Klasnja[4], PhD; Sheryl L Catz[5], PhD

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

[2]Kaiser Permanente Bernard J Tyson School of Medicine, Pasadena, CA, United States

[3]Cancer Prevention Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, United States

[4]School of Information, University of Michigan, Ann Arbor, MI, United States

[5]Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, United States

Corresponding Author:
Jennifer B McClure, PhD
Kaiser Permanente Washington Health Research Institute
1730 Minor Ave
Suite 1600
Seattle, WA, 98101
United States
Phone: 1 206 287 2737
Email: Jennifer.B.McClure@kp.org

Abstract

Background: Most smokers are ambivalent about quitting—they want to quit someday, but not now. Interventions are needed that can engage ambivalent smokers, build their motivation for quitting, and support future quit attempts. Mobile health (mHealth) apps offer a cost-effective platform for such interventions, but research is needed to inform their optimal design and assess their acceptability, feasibility, and potential effectiveness.

Objective: This study aims to assess the feasibility, acceptability, and potential impact of a novel mHealth app for smokers who want to quit smoking someday but are ambivalent about quitting in the near term.

Methods: We enrolled adults across the United States who smoked more than 10 cigarettes a day and were ambivalent about quitting (n=60). Participants were randomly assigned to 1 of 2 versions of the GEMS app: standard care (SC) versus enhanced care (EC). Both had a similar design and identical evidence-based, best-practice smoking cessation advice and resources, including the ability to earn free nicotine patches. EC also included a series of exercises called experiments designed to help ambivalent smokers clarify their goals, strengthen their motivation, and learn important behavioral skills for changing smoking behavior without making a commitment to quit. Outcomes were analyzed using automated app data and self-reported surveys at 1 and 3 months post enrollment.

Results: Participants who installed the app (57/60, 95%) were largely female, White, socioeconomically disadvantaged, and highly nicotine dependent. As expected, key outcomes trended in favor of the EC group. Compared to SC users, EC participants had greater engagement (mean sessions 19.9 for EC vs 7.3 for SC). An intentional quit attempt was reported by 39.3% (11/28) of EC users and 37.9% (11/29) of SC users. Seven-day point prevalence smoking abstinence at the 3-month follow-up was reported by 14.7% (4/28) of EC users and 6.9% (2/29) of SC users. Among participants who earned a free trial of nicotine replacement therapy based on their app usage, 36.4% (8/22) of EC participants and 11.1% (2/18) of SC participants requested the treatment. A total of 17.9% (5/28) of EC and 3.4% (1/29) of SC participants used an in-app feature to access a free tobacco quitline. Other metrics were also promising. EC participants completed an average of 6.9 (SD 3.1) out of 9 experiments. Median helpfulness ratings for completed experiments ranged from 3 to 4 on a 5-point scale. Finally, satisfaction with both app versions was very good (mean 4.1 on a 5-point Likert scale) and 95.3% (41/43) of all respondents would recommend their app version to others.

Conclusions: Ambivalent smokers were receptive to the app-based intervention, but the EC version, which combined best-practice cessation advice with self-paced, experiential exercises, was associated with greater use and evidence of behavior change. Further development and evaluation of the EC program is warranted.

Trial Registration: ClinicalTrials.gov NCT04560868; https://clinicaltrials.gov/ct2/show/NCT04560868

(JMIR Mhealth Uhealth 2023;11:e46155) doi:10.2196/46155

KEYWORDS

ambivalence; app; digital health intervention; mHealth intervention; mHealth; motivation; nicotine; smoking; smoking cessation; tobacco

Introduction

Tobacco use is responsible for over 8 million deaths per year worldwide [1]. The health risks of smoking are widely known and well documented, but the addictive properties of nicotine make quitting smoking difficult [2]. This explains why the majority of people who smoke want to quit someday but are not yet ready to commit to giving up tobacco anytime soon. This finding holds true across time, cultures, and countries [3-7].

When people are ready to quit smoking, effective, evidence-based treatment is widely available. This includes a combination of quit advice, supportive counseling, and pharmacotherapy [8,9]. However, to meaningfully reduce population smoking rates worldwide, a broader public health approach is required. Specifically, interventions are needed for those who want to quit smoking someday but are not yet ready to commit to change or to take action. These people are typically not eligible for cessation treatments (such as counseling or pharmacotherapy), which are largely limited to those who are ready to stop smoking.

While some may assume that people who are ambivalent about quitting smoking are not interested in an intervention, research has shown these individuals will enroll in smoking-focused intervention trials [10-14]. This illustrates that they are open to receiving information and assistance, despite their ambivalence about quitting smoking in the near term. A recent meta-analysis of 22 studies also found that interventions in this population can be as effective as interventions targeted to people who are ready to quit; however, the cost of intervention is substantially higher [13]. For example, the pooled cost per quit among smokers who are not yet ready to quit was US $19,510 for pharmacological interventions, US $11,416 for behavioral interventions, and US $14,662 for combined pharmacological and behavioral interventions compared to estimated costs per quit among smokers who were ready to quit, which ranged from US $1807 to US $3326 for behavioral interventions and US $2655 to US $3108 for combination therapy [13]. This cost is a prohibitive barrier for many health care providers, health care systems, and public health agencies, which is why provision of smoking cessation services is typically limited to people who are ready to quit and not offered to those who are ambivalent about quitting. However, this means the majority of smokers are excluded from intervention opportunities, even though they could benefit from them. To further reduce smoking prevalence, new intervention strategies that are both effective and cost-effective are needed for people who are ambivalent about quitting smoking.

We contend that mobile health (mHealth) apps offer a promising platform for intervening with people who are ambivalent about quitting. App-based interventions can have wide population-level reach with relatively little per-person intervention cost. From a user standpoint, they are also convenient and accessible. These benefits have helped drive the ballooning demand for and availability of digital health therapeutics and mHealth apps in recent years [15,16]. Yet, to our knowledge, there are no publicly available mHealth apps at this time that are designed specifically for smokers who are not ready to quit smoking.

It remains an open question whether ambivalent smokers would use an app-based smoking intervention if they are not ready to quit. Although, they would be interested in using an app to help them change their smoking behavior [17-19], especially if the app is responsive to their goals, such as reducing how much they smoke, and if they are not asked to commit to quitting. To date, only 1 published trial has evaluated app-delivered intervention in this population [20]. This study tested a comprehensive intervention that combined daily text messages (motivational support and quizzes) with financial incentives, encouragement to use 1 or more self-selected relaxation and distraction apps, motivational phone support from a tobacco treatment specialist, and precessation use of nicotine replacement therapy (NRT). While the 3-week intervention significantly enhanced quit rates at 6-month follow-up, more research is needed to confirm ambivalent smokers' interest in using app-based smoking interventions and to inform their optimal design.

The primary objective of this randomized pilot study was to evaluate the feasibility and acceptability of using an mHealth app called GEMS to motivate and support smoking behavior change among smokers who want to quit smoking someday but have not yet. Two versions of GEMS were evaluated, each using a similar, but not identical, graphical user interface and content. The standard care (SC) version offered best-practice cognitive behavioral advice and other resources recommended for people who are ready to quit smoking, including access to cessation counseling and pharmacotherapy. The enhanced care (EC) version included this same content plus a series of specific cognitive and behavioral exercises designed to build motivation and enhance self-efficacy to reduce smoking or quit, and to promote quit attempts and cessation. We hypothesized that participants would use both versions of the app, but the EC version would have greater program use and, in turn, better support change in motivation, self-efficacy, and smoking behavior. However, this pilot study was not powered to detect statistically significant differences in cognitive or behavioral outcomes between the 2 app versions. Instead, findings will inform the need for further evaluation of the GEMS app and could inform the design of similar app-based interventions targeting smokers who are ambivalent about quitting smoking.

Methods

Ethics Approval

All research activities were conducted at the Kaiser Permanente Washington (KPWA) Health Research Institute and approved by the KPWA Institutional Review Board (#2020). Data were collected between December 2020 and October 2021. The study is registered with ClinicalTrials.gov (NCT04560868).

Study Design

The study used a parallel, 2-arm design. Participants were randomly assigned to the SC or EC version of the app using an automated, block-stratified randomization scheme (≥15 cigarettes per day vs 10-14 cigarettes). This scheme ensured balanced representation of lighter versus heavier smokers between intervention arms since this could impact users' motivation or ability to change their smoking behavior. Participants were followed for 3 months post enrollment and completed self-report surveys at 1 and 3 months post enrollment. Consistent with the purpose of a pilot trial [21], the goal of this work is to provide proof of concept for the app's feasibility and acceptability.

Recruitment, Eligibility, and Randomization

Participants were recruited through social media ads and screened for eligibility by phone. Individuals were eligible if they met the following criteria: 18 years of age or older; could read and speak in English; smoked at least 100 lifetime cigarettes; smoked in the past week; smoked at least ten cigarettes a day; wanted to quit smoking someday, but not in the next month; reported daily smartphone use; self-reported they could read text on their phone screen; were willing to download and install the app; and used either an Android or Apple smartphone. Individuals were excluded if they reported a lifetime history of dementia, manic depression (bipolar disorder), schizophrenia, contraindications for NRT use (pregnant, nursing, recent heart attack, or uncontrolled arrhythmia); another member of their household was already enrolled in the study; or if we were unable to verify the validity of their phone number or email address.

After completing the baseline survey and being randomized, participants were instructed on how to install the app and access their assigned intervention (EC vs SC). Those who failed to install the app within 3 days were offered assistance. Those who failed to install the app during the 3-month study period were excluded from the analytic sample, ensuring evaluative feedback was only collected from individuals who installed the app.

Intervention Design, Content, and Functionality

General Overview

Both app versions were called GEMS. The name was chosen based on user feedback. Ambivalent smokers liked the name because it did not suggest the app was smoking related, making it more confidential and more appealing than a name implying the app was focused on smoking cessation.

After installing the app and setting up a user account, participants viewed a welcome screen, which explained the program's purpose and a brief tutorial and orientation to the app's features. Content could then be accessed ad-lib until completion of the 3-month follow-up survey, which concluded study participation. At this point, the study team remotely deactivated app access and ceased app usage tracking. Both versions of the app (SC and EC) had a similar design and identical content, except for the addition of the novel experiments in the EC version. This design meant that both groups received an active intervention and provided data useful to assessing the concept of offering an app-based intervention to people who were ambivalent about quitting, while also allowing us to assess differences that might be attributed to the additional features in the EC version. Shared and unique features of each app version are described below.

SC Content and Features Common to Both App Versions

The SC content was based on evidence-based treatment grounded in the US Public Health Service Guidelines for Treatment of Nicotine Dependence [9] and standard cognitive behavioral therapy for smoking cessation [22], with additional content and features informed by user-centered design work conducted by our team [17,18,23]. Messaging acknowledged users were not ready to stop smoking, but content focused on how to stop smoking, as per usual care treatment for smoking cessation. For example, the main feature of this program version was a Quit Guide that included advice on how to quit smoking; didactic information (eg, what is nicotine withdrawal, how does pharmacotherapy work); and a 6-step guide on how to quit (eg, how to choose and use stop-smoking medicines, how to set a quit date, how to prepare for your quit date, what to do on your quit date, and how to stay the course and prevent relapse). Participants could also call a nationwide tobacco quitline from within the app to enroll in free counseling available to all US residents. Other content included a calculator for estimating how much money could be saved by quitting smoking, a daily cigarette tracker, and 2 sets of narrative peer advice presented through short testimonials: 1 set offering motivational encouragement for quitting smoking and 1 modeling how to talk back to common excuses people give for smoking or not quitting. Finally, participants could keep notes on their quitting progress using an in-app journal.

Participants earned badge rewards (gems—hence the app name) for using app features (eg, saving calculator, daily cigarette tracker) and for viewing psychoeducational content (eg, each Quit Guide step). Participants were asked to actively indicate when they read key content by clicking a "Mark as Read" button on each page. Participants using the SC version of the app could earn up to 10 usage badges. After 6 badges were earned, users in both groups could request a free 2-week trial of NRT to help them stop smoking.

Experimental App Content and Unique Features

The EC app version mirrored the SC's design, content, and functionality with 3 key exceptions. First, the home page of the EC version included, as the main content, a series of 9 cognitive and behavioral exercises called experiments. Each experiment was designed to help users clarify their values, build and strengthen their motivation for reducing or quitting smoking, and enhance their self-efficacy for changing smoking behavior by learning specific skills that could help them manage cravings and resist the urge to smoke (Table 1). Unless EC users opted to block text reminders, they also received reminder prompts to initiate or complete experiments. Second, in the EC version, the Quit Guide was in the resource toolbox, accessible from the home page, but it was less prominent than in the SC version, where it occupied the home page. Third, EC participants could earn up to 19 total badges: 9 for completing each of the experiments and 10 for viewing the program content common with the SC version.

The theoretical rationale for the experiments, an overview of their design and flow, and preliminary formative research testing their acceptability and potential impact with smokers ambivalent about quitting have been previously reported [19]. Briefly, the EC intervention is grounded in empirically validated recommendations for treating nicotine dependence [9] and several complementary motivation and behavior change theories (eg, the PRIME theory of motivation [24,25], cognitive behavioral therapy, acceptance and commitment therapy, and social cognitive theory [26-28]). The experiments' design was further informed by Fogg's model of persuasive design [29], which suggests that when people have low motivation for change (as is the case with smokers ambivalent about quitting), the behaviors they are expected to engage in should be simple (ie, require low ability) and coupled with extrinsic triggers to prompt engagement (ie, reminder prompts). The specific behavioral goals and skills targeted in each experiment are summarized in Table 1 and an example is depicted in Figure 1.

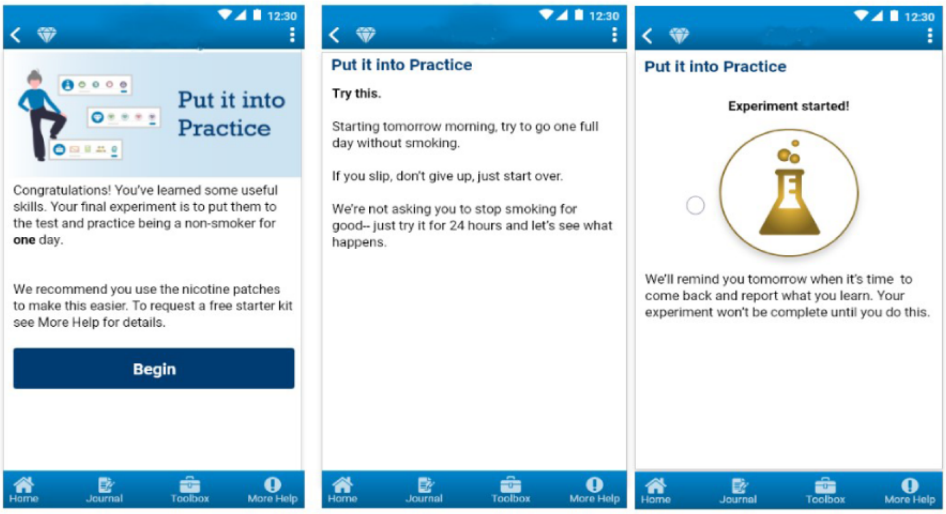
With the exception of the first experiment, which could be completed in a few minutes, each exercise was designed to be practiced for a 24-hour period, after which participants were prompted to return and report what they learned by answering a brief series of reflective questions. Emphasis was placed on trying and learning from each exercise, as opposed to mastery or success, to avoid creating a sense of failure if participants did not complete or master the experiment. The experiments also built on one another, so lessons and skills learned in earlier experiments were designed to support success with later experiments.

To encourage sequential completion and forward progress, each experiment unlocked after completion of the previous experiment. If an experiment was started but not completed, the next experiment automatically unlocked after several days.

Table 1. Enhanced care experiments' targeted skills and goals.

Experiment	Targeted skills and goals
One	Clarify personal values and health goals. Explore how smoking fits with these.
Two	Identify personal reasons for quitting. Build motivation for change.
Three	Identify high-risk situations for smoking. Inform future problem-solving and preparation for quitting.
Four	Learn deep breathing as a tool for stress reduction and craving management. Build self-efficacy for managing cravings and motivation for quitting.
Five	Mindful acceptance. Learn to let urges pass without smoking. Enhance self-efficacy for managing cravings and motivation for quitting. Create positive outcome expectations.
Six	Stimulus control. Learn to reduce the reinforcing effects of smoking. Enhance self-efficacy and motivation for quitting.
Seven	Cognitive restructuring. Reframe not smoking as a positive choice, not a deprivation. Support self-efficacy and create positive outcome expectations for quitting.
Eight	Successive approximation. Reduce daily smoking. Support self-efficacy and positive outcome expectations.
Nine	Put all skills into practice with a 24-hour "practice" quit. Enhance self-efficacy, motivation, and positive outcome expectations.

Figure 1. Example of experiment setup.



Key Outcomes and Baseline Measures

Sources

Self-report surveys were completed on the internet at baseline, 1 month, and 3 months post enrollment. Participants received a US $25 electronic gift card for completing each survey.

App use and qualitative postexperiment ratings were assessed with automated, time-stamped data.

Key Outcomes of Interest

As a pilot study, we examined a range of primary and secondary outcomes to inform the feasibility, acceptability, and potential impact of the EC version relative to the SC version of GEMS. A key outcome was whether people would install the app. Among those who did, primary outcomes of interest, each assessed at 3-month follow-up, were total number of user sessions, presence of a self-reported quit attempt lasting at least 24 hours, and a self-report of no smoking for the past 7 days (7-day point prevalent abstinence [PPA]).

Secondary outcomes used to assess program use and engagement included total duration of app usage calculated as number of days between installation and last use; total number of usage badges earned; number of participants who earned enough usage badges to request free NRT; proportion of those earning NRT who requested it; the proportion of participants who clicked on the Call Now button to access free quitline counseling; and the number of people who used each app feature.

Satisfaction was assessed based on users' overall satisfaction with their assigned program's content and advice. All ratings used a 5-point Likert scale from "not at all" to "extremely." In the EC arm, users also rated the helpfulness of each experiment using a 5-point Likert scale from "not helpful" to "very helpful." Ratings were made in real time following the completion of each experiment.

Additional secondary outcomes included motivation and self-efficacy for both smoking fewer cigarettes a day and quitting smoking, each assessed as cognitive intermediaries of behavior change at the 1-month follow-up using 10-point Likert scales ranging from "not at all" to "extremely." Other secondary indices of behavior change included a self-reported quit attempt lasting at least 24 hours, assessed at 1 month; self-reported 7-day PPA at 1 month; and the proportion of participants who reported a 50% or greater reduction in smoking from baseline to the 3-month follow-up.

Baseline assessment measures included participant demographics, use of smartphones and smoking apps, tobacco and e-cigarette use, and self-reported lifetime diagnosis or treatment for depression, anxiety, bipolar disorder, schizophrenia, alcohol use disorder, or drug use (assessed as a single yes or no for any of the listed conditions). Nicotine dependence was assessed with the Fagerström Test of Nicotine Dependence (FTND) [30]. Problem drinking was assessed with the Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) [31]. Frequency of cannabis use was assessed with a single item from the Cannabis Use Disorder Identification Test-Revised (CUDIT-R) [32]. Response options for this item were never, monthly or less, 2-4 times a month, 2-3 times a week, or 4 or more times a week. Finally, we assessed participants' outcome expectations that the help received from the study would be a key factor in either their smoking less or their quitting smoking. Each was assessed with a 5-point Likert scale ranging from 1="strongly disagree" to 5="strongly agree" and was modified from a similar item previously shown to predict cessation [33].

Data and Programming Issues

Two programming issues are worth noting. First, due to a REDCap programming error, some participants who self-reported 7-day PPA at 3 months were not flagged by the system. As a result, biochemical confirmation was not obtained from these individuals as originally planned. Since these individuals made up a high proportion of individuals who self-reported 7-day PPA at 3 months, only self-reported smoking outcomes were analyzed. Second, due to a code issue, 4 EC participants were allowed to cycle through some of the experiments after a 15-minute practice period instead of the planned 24-hour period. This issue was caught early and corrected, so these individuals were retained in the analyses.

Data Analyses

As defined a priori, outcomes are based on 2 subsets of participants. We first report on the total number of individuals who agreed to join the study and who installed the study app, as an initial indicator of study acceptability. All other analyses used a modified intent-to-treat approach and included all randomized participants who installed the app, regardless of subsequent app usage. Participants who failed to install the app were excluded from this cohort because the goal of these analyses was to assess metrics of feasibility, acceptability, and potential impact of app content among individuals who installed and used the intervention. Everyone in this analytic sample contributed automated app usage data; however, self-reported data at 1 and 3 months post enrollment were subject to missingness. Comparisons of satisfaction ratings for specific app features were restricted to participants who both self-reported use of the feature and whose automated data confirmed this use. Per convention, missing smoking outcomes were conservatively imputed as smoking. For consistency, we used a similar approach when analyzing 24-hour quit attempts (ie, missing data were imputed as not making a quit attempt). As determined a priori, secondary sensitivity analyses were also conducted for these 2 behavioral outcomes using (1) complete cases only and (2) multiple imputation by chained equation with 10 imputed data sets created with logistic regression imputation and Barnard-Rubin adjusted degrees of freedom [34,35]. For all other outcomes, analyses used complete cases without imputation.

Descriptive statistics were used to characterize the baseline sample and outcomes of interest. To compare outcomes across groups at follow-up, regression models were fitted using generalized estimating equations with robust standard errors and an exchangeable working correlation. When applicable, the model was simultaneously fitted to outcomes collected at 1 and 3 months post enrollment. For binary outcomes, we estimated relative risks (RR) of the outcome with the EC version relative to the SC version using a Poisson regression model. When events were too rare to obtain estimates of relative risks, linear regression models were used to estimate risk differences instead. Linear regression models were fitted to continuous outcomes to estimate mean differences between arms.

When applicable, to allow for separate reporting of comparisons at 1 and 3 months post baseline, time of survey collection and the interaction between follow-up time and assigned app version were included as covariates. For precision, we adjusted for the number of cigarettes smoked per day at baseline and, when applicable, baseline values of the outcome. Because groups differed at baseline by the proportion who reported a history of mental health or substance use disorder, risky drinking based on AUDIT-C scores, and household income, and these variables are known to affect cessation outcomes, we also adjusted for these potential confounders in sensitivity analyses. Point estimates are presented with 95% CIs and *P* values are from 2-sided Wald tests. All analyses were conducted in R version 4.0.2 [36].

Sample Size

The total enrolled sample (n=60) and final analytic sample (n=57) exceed the range of 24 to 50 participants commonly recommended for pilot studies [37,38]. Smaller samples are deemed appropriate when the goal is to assess intervention feasibility and acceptability as opposed to intervention efficacy or effectiveness. The study was not powered to detect minimal clinically meaningful differences between groups with statistical significance.

Results

Participants

A total of 60 participants consented and enrolled in the study. Of these, most participants (57/60, 95%) installed the app and were included in the analytic sample (Figure 2). Demographic characteristics of this group are presented in Table 2. These participants were largely female, White, and socioeconomically disadvantaged. One-third of participants (19/57, 33.3%) had previously used health-related apps, but only 7% (4/57) had ever used a smoking cessation app. Participants smoked nearly a pack a day on average (mean 18.1 cigarettes a day) and most (36/57, 63.2%) had FTND scores indicative of “high” or “very high” nicotine dependence. Nearly one-third (18/57, 31.6%) reported using cannabis 2 or more times a week, and a similar proportion (21/57, 36.8%) self-reported previous diagnosis or treatment for either depression, anxiety, bipolar disorder, schizophrenia, alcohol use disorder, or drug use. At baseline, motivation for quitting smoking someday was moderately high (mean 6.2 out of 10, SD 1.2) and self-efficacy for quitting was moderately low (mean 4.1 out of 10, SD 1.8).

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

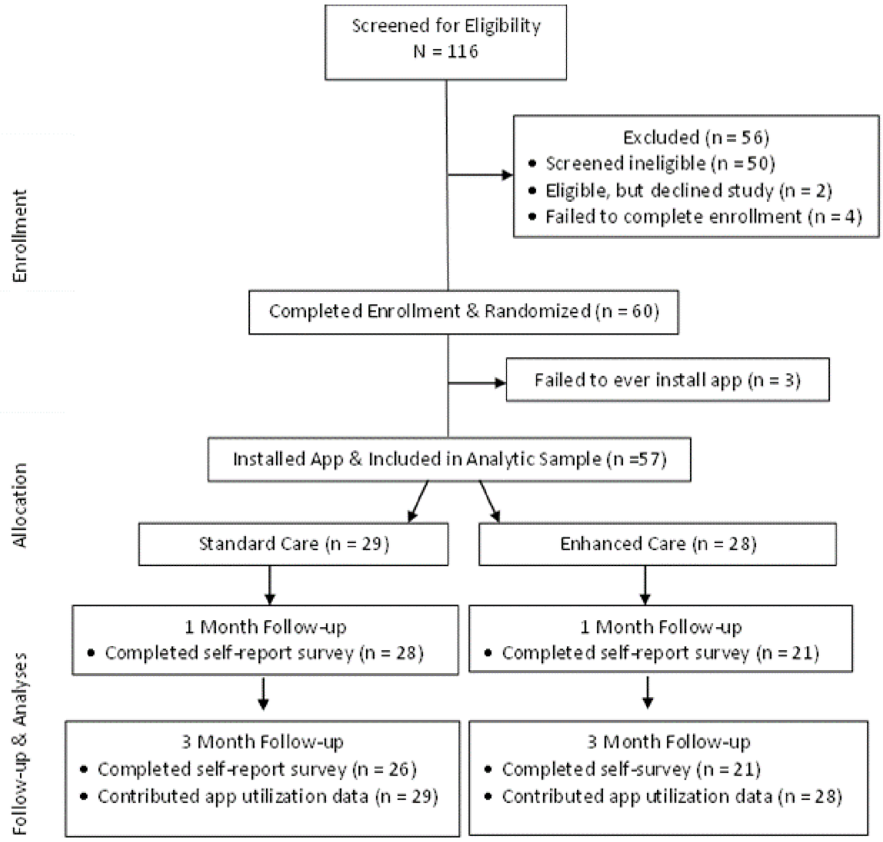


Table 2. Baseline descriptive characteristics.

	Overall (N=57)	Standard care (n=29)	Enhanced care (n=28)
Female, n (%)	41 (71.9)	22 (75.9)	19 (67.9)
White, n (%)	47 (82.5)	24 (82.8)	23 (82.1)
Hispanic, n (%)	0 (0)	0 (0)	0 (0)
Employed, n (%)	29 (50.9)	14 (48.3)	15 (53.6)
Annual household income <US $45,000[a], n (%)	33 (57.9)	13 (44.8)	20 (71.4)
No college degree, n (%)	41 (71.9)	23 (79.3)	18 (64.3)
Mental health or substance use disorder (Yes)[a,b], n (%)	21 (36.8)	9 (31)	12 (42.9)
Age (years), mean (SD)	47.5 (10.3)	47.1 (9.1)	47.9 (11.5)
Cigarettes per day, mean (SD)	18.1 (7.8)	17 (6.6)	19.2 (8.8)
FTND[c] (nicotine dependence)[d], mean (SD)	6.1 (1.9)	5.9 (2)	6.4 (1.7)
Nicotine and tobacco, n (%)			
Use tobacco other than cigarettes (Yes)	6 (10.5)	3 (10.3)	3 (10.7)
Use e-cigarettes (No)	47 (82.5)	23 (79.3)	24 (85.7)
Nicotine dependence: high or very high[d]	36 (63.2)	18 (62.2)	18 (64.3)
Substance and alcohol use, n (%)			
Cannabis use: 2 or more times/week[a]	18 (31.6)	9 (31)	9 (32.2)
Hazardous drinking levels[e]	16 (28.1)	10 (34.5)	6 (21.4)
App use, n (%)			
Ever downloaded health app (Yes)	19 (33.3)	10 (34.5)	9 (32.1)
Ever downloaded smoking app (Yes)	4 (7)	2 (6.9)	2 (7.1)
Motivation[f], mean (SD)			
Reducing smoking	5.5 (1.1)	5.5 (1.1)	5.6 (1.2)
Quitting smoking	6.2 (1.2)	6.1 (1.3)	6.2 (1)
Self-efficacy[f], mean (SD)			
Reducing smoking	3.8 (1.5)	4 (1.5)	3.5 (1.5)
Quitting smoking	4.1 (1.8)	4.2 (1.7)	3.9 (1.9)
Outcome expectation[g], mean (SD)			
Study app will help smoke less (Yes)	3.8 (0.9)	3.8 (0.9)	3.8 (0.8)
Study app will help quit smoking (Yes)	3.7 (0.8)	3.8 (0.9)	3.7 (0.8)

[a]Missing responses: 2 did not provide annual household income; 4 did not answer question about cannabis use; and 2 did not answer questions about mental health or substance use disorders.

[b]Self-reported diagnosis or treatment for depression, anxiety, bipolar disorder, schizophrenia, or alcohol use.

[c]FTND: Fagerström Test of Nicotine Dependence.

[d]Fagerström Test of Nicotine Dependence score. The range is from 0 to 10. Scores 6-7 indicate high dependence and 8-10 indicate very high dependence.

[e]Alcohol Use Disorders Identification Test-Consumption score. Range from 0 to 12. Scores of 4 or above for men and 3 or above for women indicative of drinking levels that are hazardous to one's health and safety.

[f]Likert scale ranging from 1="not at all" to 10="extremely."

[g]Likert scale ranging from 1="strongly disagree" to 5="strongly agree."

Indicators of Feasibility and Acceptability

General App Usage

A total of 3 enrolled participants failed to download the app. Among those who did install the app (57/60, 95%), usage differed significantly by group: EC participants averaged 19.9 (SD 16.2) total sessions compared to an average of 7.3 (SD 6.6) for SC users, yielding an average difference of 12.7 sessions (95% CI 6.23-18.93; *P*<.001).

Overall, duration of app usage was similar between arms but slightly favored the EC group. Mean days of use among EC users were 43 (SD 30.9) compared to 41.8 (SD 34.2) among SC users, yielding an average difference of 1.1 days (95% CI 15.65-17.85 days; *P*=.90).

Most participants (53/57, 93%) earned at least one usage badge. However, EC users earned more badges (mean 10.1, SD 6.0) than SC users (mean 6.1, SD 3.5), yielding an average difference of 4.2 badges (95% CI 1.7-6.7; *P*=.001). EC users were also more likely to earn the requisite 6 badges needed to request free NRT: 78.6% (22/28) of EC users met this bar compared to 62.1% (18/29) of SC users (RR 1.28, 95% CI 0.91-1.79; *P*=.16). Interpretation of the results was unchanged in analyses adjusting for baseline differences (data not shown).

Use of App Content and Features

SC and EC had similar use of the content and features common to both app versions, with 2 notable exceptions: SC users were twice as likely to indicate that they had read all of the content in the Quit Guide subsections compared to EC users, and EC users were 45% more likely to view the journal compared to SC users (see Table 3).

Table 3. Use of features common to both app versions.

App feature	Overall (N=57), n (%)	Standard care (n=29), n (%)	Enhanced care (n=28), n (%)	Relative risk (95% CI)[a]	*P* value[b]
Quit Guide					
Step 1[c]	35 (61.4)	26 (89.7)	9 (32.1)	0.36 (0.21-0.62)	<.001
Step 2[c]	36 (63.2)	26 (89.7)	10 (35.7)	0.40 (0.24-0.66)	<.001
Step 3[c]	34 (59.6)	24 (82.8)	10 (35.7)	0.44 (0.26-0.72)	.001
Step 4[c]	26 (45.6)	19 (65.5)	7 (25)	0.38 (0.20-0.74)	.005
Step 5[c]	25 (43.9)	18 (62.1)	7 (25)	0.41 (0.20-0.81)	.01
Step 6[c]	25 (43.9)	18 (62.1)	7 (25)	0.41 (0.20-0.81)	.01
Cigarette tracker[d]	24 (42.1)	13 (44.8)	11 (39.3)	0.88 (0.48-1.62)	.69
Savings calculator[e]	25 (43.9)	13 (44.8)	12 (42.9)	0.96 (0.54-1.71)	.90
Peer testimonials[f]	21 (36.8)	11 (37.9)	10 (35.7)	0.95 (0.49-1.86)	.88
Peer advice[g]	18 (31.6)	10 (34.5)	8 (28.6)	0.84 (0.39-1.80)	.65
Journal[h]	48 (84.2)	20 (69)	28 (100)	1.45 (1.14-1.85)	.003
More Help[i]	38 (66.7)	19 (65.5)	19 (67.9)	1.05 (0.73-1.49)	.80

[a]Relative risk (95% CI) of using that component of the app in the enhanced care arm relative to the standard care arm, adjusting for cigarettes per day at baseline. Standard care arm is the referent group.

[b]2-sided Wald test for the null of no difference in risk between arms.

[c]Based on completion of quit guide step content as defined by user marking all content in section as read.

[d]Based on use of tracker to log smoking on at least one day.

[e]Based on use of savings calculator to estimate cost savings of quitting smoking.

[f]Based on event data showing each peer testimonial modeling how to talk back to common excuses for not quitting was opened and viewed.

[g]Based on event data showing each vignette providing motivational support and advice was opened and viewed.

[h]Based on event data showing the Journal was opened at least one time, whether or not an entry was created.

[i]Based on opening the More Help page at least one time to access tobacco quitline referral and other information on where to get help quitting smoking.

Overall Satisfaction Ratings

Self-reported satisfaction with participants' assigned app version was similar in both groups. At 3-month follow-up, most respondents reported they would recommend the app to others (19/20, 95% EC users vs 22/23, 95.7% SC users; RR 1.03, 95% CI 0.89-1.19; *P*=.70). Among respondents who earned at least one usage badge by 3 months post enrollment (n=43)—the minimum exposure threshold deemed adequate to evaluate the app content—respondents in both arms reported similarly high satisfaction with their assigned app's overall content and advice. Mean satisfaction ratings in both arms were 4.1 out of 5 (SD 1.1). Similar results were observed at 1-month follow-up; mean satisfaction ratings were 3.6 (SD 1.2) among SC users and 3.8 (SD 1) among EC users (adjusted average difference 0.26, 95% CI –0.35 to 0.87; *P*=.40).

Experiment Engagement and Helpfulness

EC users completed an average of 6.14 (SD 3.31) and 6.89 (SD 3.08) experiments by 1- and 3-month follow-up, respectively. Completion rates across each of the 9 individual experiments ranged from 93% (26/28) to 61% (17/28) (Table 4).

Immediately after completing each experiment, EC users were asked to rate the helpfulness of the experiment. Median helpfulness scores ranged from 3 to 4 on a 5-point Likert scale (Table 4). Experiments receiving the highest median scores focused on learning to identify high-risk situations and triggers for smoking, reducing daily smoking, and making a practice quit attempt (median 4 for each). Exercises focused on learning deep breathing for stress reduction and reframing not smoking as a personal choice (as opposed to a deprivation) also received higher median scores (3.75).

Table 4. Portion of enhanced care participants completing each experiment and median helpfulness ratings.

Experiment	At 1-month follow-up[a], n (%)	At 3-month follow-up[a], n (%)	Helpfulness[b], median (IQR)	Range (minimum-maximum)[b]
One	25 (89)	26 (93)	3 (3-4)	1-5
Two	22 (79)	23 (82)	3 (2-4)	2-5
Three	21 (75)	23 (82)	4 (3-4.25)	1-5
Four	20 (71)	22 (79)	3.75 (2.25-4)	1-5
Five	20 (71)	22 (79)	3 (2.25-4)	1-5
Six	19 (68)	22 (79)	3.5 (3-4.75)	1-5
Seven	18 (64)	20 (71)	3.75 (2-4.08)	1-5
Eight	15 (54)	18 (64)	4 (3-4.83)	3-5
Nine	12 (43)	17 (61)	4 (3-5)	2-5

[a]The number and proportion of all enhanced care participants (n=28) who completed each experiment by 1- and 3-month postenrollment follow-up.

[b]Reflects the median (IQR), and range (minimum-maximum) of helpfulness ratings across all experimental participants who completed each in-app, postexperiment assessment. If a user completed the experiment more than once, the average of their ratings was used. Ratings could range from 1="not at all" to 5="extremely helpful."

Indicators of Intermediate Cognitive Change at 1 Month

Motivation

Among participants still smoking at 1-month follow-up, mean self-reported motivation to quit was 8.2 (SD 2.1) for EC users compared to 7.1 (SD 2.6) for SC users (adjusted mean difference 0.54, 95% CI –0.49 to 1.57; *P*=.30). Motivation to smoke less at the 1-month follow-up averaged 8.3 (SD 1.7) for EC users compared to 7.4 (SD 2.5) among SC users (adjusted mean difference 0.92, 95% CI –0.07 to 1.92; *P*=.07). In both groups, indices of motivation increased from baseline (Table 2) to follow-up.

Self-Efficacy

Among participants still smoking at 1-month follow-up, mean self-efficacy for quitting smoking was 7.4 (SD 2.4) for EC users compared to 7.5 (SD 2.3) for SC users (adjusted mean difference 0.03, 95% CI –1.02 to 1.07; *P*=.96). Self-efficacy for smoking less at the 1-month follow-up averaged 6.3 (SD 2.2) for EC users compared to 7.7 (SD 2.5) for SC users (adjusted mean difference 0.06, 95% CI –0.97 to 1.09; *P*=.06). In both groups, indices of self-efficacy increased from baseline (Table 2) to follow-up.

Indicators of Behavior Change at 3 Months

Requests for Free NRT

Among the 40 participants who earned 6 usage badges and were eligible to request a free trial of NRT, 10 participants requested it (8/22, 36.4% EC users as compared to 2/18, 11.1% SC users; RR 3.18, 95% CI 0.77-13.17; *P*=.11).

Call Now for Free Counseling

At the 3-month follow-up, 17.9% (5/28) of EC participants and 3.4% (1/29) of SC participants had clicked on the Call Now button to connect with a free tobacco quitline counselor (RR 5.17, 95% CI 0.65-41.3; *P*=.12).

Smoking Reduction

At the 3-month follow-up, a similar proportion of participants using both app versions reported a significant reduction in their daily smoking rate: 28.6% (6/21) of EC users compared to 28% (7/25) of SC users (RR 0.01, 95% CI –0.25 to 0.27; *P*=.92) reported a 50% or greater reduction in their baseline daily smoking.

Quit Attempts

At the 3-month follow-up, 39.3% (11/28) of EC users and 37.9% (11/29) of SC users reported making an intentional quit attempt after joining the study, when missing values were imputed as not making a quit attempt (RR 1.01, 95% CI 0.55-1.85; *P*=.98). The interpretation of the results was unchanged in complete case and multiple imputation sensitivity analyses or analyses adjusting for baseline differences (results not shown).

Smoking Abstinence

At the 3-month follow-up, 14.7% (4/28) of EC users and 6.9% (2/29) of SC users reported not smoking, even a puff, in the last 7 days (risk difference 0.08, 95% CI –0.08 to 0.24; *P*=.35). The interpretation of the results was unchanged in complete case and multiple imputation sensitivity analyses or analyses adjusting for baseline differences (data not shown).

Discussion

Principal Findings

The primary objective of this randomized pilot study was to evaluate the feasibility and acceptability of the EC version of the GEMS app and to assess its potential to motivate and support smoking behavior change compared to a similar app that included SC content (SC version) but was not designed specifically for smokers who are ambivalent about quitting smoking. It is encouraging that 95% of the participants who agreed to enroll in the study installed the app. This provides an important initial signal of the intervention's acceptability. However, because of the size and nature of the study, conclusions about the acceptability and impact of the intervention among participants who installed the app cannot be based on the statistical significance of the primary and secondary outcome comparisons. Instead, it is important to look at the trend and pattern of the point estimates at follow-up. Notably, in almost all cases, the observed outcomes trended in favor of the EC app version. This held true for both self-reported outcomes and those based on objective automated data.

As hypothesized, EC participants used the app more often (an average of 19.9 sessions vs 7.3 sessions for SC participants), and a greater proportion reported smoking abstinence at follow-up (14.7% of EC participants vs 6.9% of SC participants). This finding is consistent with previous research showing an association between greater program engagement, or adherence, and higher cessation rates [39,40]. The observed quit rate in the EC arm is similar to that observed from physician advice to quit and low-intensity counseling interventions (average 14%-16%) [9].

Additionally, EC participants earned an average of 4 more usage badges. Notably, EC participants had the potential to earn more badges based on the additional content (experiments) in this version, but both groups had equal opportunity to earn the requisite 6 badges needed to request a free trial of NRT, and a higher proportion of EC users met this bar based on their app usage (78.6% EC users vs 62.1% SC users). Additionally, more EC users who earned the NRT, requested to receive it (36.4% EC vs 11.1% SC). Similarly, at the 1-month follow-up, motivation for quitting smoking trended higher in the EC group (8.3 EC vs 7.4 SC on a 10-point scale), even though groups had similar self-efficacy for quitting. More EC users clicked the Call Now button in the app to access free quitline counseling (17.9% of EC vs 3.4% of SC), although it is not known if these individuals actually enrolled in the free quitline program. Finally, a slightly higher proportion of EC users (39.3%) reported making a quit attempt compared to SC users (37.9%) and 28.6% of EC users reported a meaningful reduction in daily smoking at the 3-month follow-up.

It is also notable that participants in this trial were lower–socioeconomic status heavy smokers, with high levels of concomitant substance use or lifetime substance-related or mental health diagnoses. These groups have been shown to be less likely to engage in treatment and successfully quit smoking [41,42] and, therefore, represent important targets for intervention.

Taken together, these findings confirm the conclusions drawn from our previous formative work that smokers who are ambivalent about quitting, including those who are more socioeconomically disadvantaged, are interested in using an mHealth app to help them reduce or stop smoking [17-19] and that designing this intervention to be sensitive to participants' ambivalence about quitting could increase their likelihood of changing their smoking behavior.

Role of the Experiments

A key question in understanding the feasibility of this app was whether ambivalent smokers would engage with the self-directed, smoking-focused, cognitive, and behavioral experiments if they were not yet ready to commit to quitting smoking. The experiments were designed to teach users specific skills and lessons to help people resist the urge to smoke and encourage a quit attempt. These skills are consistent with the common elements of effective behavioral interventions identified in the US Public Health Service's Tobacco Treatment Guidelines (ie, problem-solving skills, such as identifying high-risk situations, coping skills for managing urges without smoking, basic educational information, and supportive encouragement to make a quit attempt) [9]. Inclusion of these elements has been shown to increase the effectiveness of low-intensity counseling. The results of this pilot suggest these elements may also be useful in self-directed mHealth interventions, though we also acknowledge the importance of the EC reminder prompts. As hypothesized based on Fogg's behavioral model for persuasive design, these prompts appear to have aided continued program engagement [29].

Findings from this study also indicate that most EC participants were willing to try the exercises. Completion rates ranged from 93% of EC users completing the first experiment (clarifying one's values) to 61% completing the last experiment (making a practice quit attempt). These rates are encouraging. Because the addition of the experiments was the key content difference between the 2 app versions, engagement with the exercises likely drove the favorable trends in the outcomes noted above. This is further supported by the fact that EC participants were less likely than SC participants to read the smoking cessation Quit Guide. This was likely an artifact of the differences in the positioning of this content in the app (it was on the home page in the SC app version and located in the Toolbox linked from the home page in the EC app version), but because EC participants were less likely to view this content, exposure to the Quit Guide cannot explain the more favorable outcomes observed in the EC arm.

The graphical user interface for accessing the Quit Guide does not appear to have been an impediment to app usage or behavior change in the EC arm, but it is worth considering if the Quit Guide should be featured more prominently in a future EC version and, if so, whether this would add additional value to users or not.

Limitations and Strengths

The findings from this work must be viewed in the context of the study limitations. Chief among these is the small sample size; the study was not powered to detect minimal clinically meaningful differences with statistical significance, and the sample size limits our ability to draw any firm conclusions about the generalizability of the findings, particularly with regard to smoking behavior change. Additionally, cessation outcomes are based on self-reported data and are subject to social desirability bias. However, biochemical confirmation is not generally recommended in trials with no face-to-face contact and where the demand characteristics to misreport abstinence are low [43], such as this remote trial of smokers who are ambivalent about quitting smoking. Moreover, relying on remote biochemical confirmation of smoking abstinence has been shown to bias outcomes due to low rates of participation [44]. For these reasons, the use of self-reported cessation outcomes is reasonable for this preliminary work. However, self-reporting also has its limitations. In this study, we saw a higher rate of attrition at the 3-month follow-up in the EC arm, resulting in a higher number of imputed smokers in this arm. Despite this, cessation outcomes still favored the EC group.

To our knowledge, this is the first app to have been designed specifically for smokers who are ambivalent about quitting, thus addressing an important intervention gap. Other strengths include the recruitment of a high-risk and low–socioeconomic status sample, a rigorous methodological design that allows the unique effects of the experiments to be tested, reliance on both self-report and automated tracking data for outcomes, and overall strong follow-up participation at 3 months (47/57, 82.5%).

Conclusions

This study provides encouraging evidence that people who are ambivalent about quitting smoking will voluntarily use and remain engaged with an mHealth app that is designed to help them cut back or quit smoking, even if they are not actively planning to change their smoking behavior at program initiation. Further development of app-based interventions targeted at smokers who are ambivalent about quitting is warranted, as is further evaluation of the effectiveness of EC app version.

Acknowledgments

This work was supported by a grant from the National Cancer Institute (R21CA234003; JBM principal investigator). Additional in-kind support for the app development was provided by Kaiser Permanente. We are deeply grateful to Brooks Tiffany for creating the user interface design of the app; Dhaval Patel for developing the app and supporting database; Hiren Desai for his technical support and sponsorship of this work; Tushar Gokhale and Kaiser Permanente’s Telehealth and Mobility team for their support of the app development; Kaiser Permanente’s Mobility Center of Excellence (particularly, Colin Moore, Tonja Williams, and Jocelyn Dela Rosa) for their technical review of the app and ongoing support; Ella Thompson for her assistance managing this study; Yishi Xian for her assistance enrolling participants and collecting follow-up data; Ladia Albertson-Junkans for programming the REDCap surveys; Dr Andrea Cook for her statistical consulting; Sarah Randall for her help preparing this manuscript, and Umesh Singh for his analytic programming support.

Data Availability

This pilot work was not subject to the National Institutes of Health data sharing requirements and did not obtain consent or institutional review board approval for publicly sharing data. As such, the data sets analyzed for this study are not publicly available, but we will consider sharing data upon request and with institutional review board approval.

Conflicts of Interest

JLH has received research support from Pfizer unrelated to this study. None of the other authors have financial or other conflicts of interest to report.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 15360 KB - mhealth_v11i1e46155_app1.pdf]

References

1. Commar A, Prasad V, d'Espaignet ED, Ong L. WHO global report on trends in prevalence of tobacco use 2000-2025, third edition. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/who-global-report-on-trends-in-prevalence-of-tobacco-use-2000-2025-third-edition [accessed 2023-06-05]
2. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2014.
3. Daoud N, Hayek S, Sheikh Muhammad A, Abu-Saad K, Osman A, Thrasher JF, et al. Stages of change of the readiness to quit smoking among a random sample of minority Arab male smokers in Israel. BMC Public Health 2015;15:672 [FREE Full text] [doi: 10.1186/s12889-015-1950-8] [Medline: 26178347]
4. Mbulo L, Murty KS, Husain MJ, Bashir R, Blutcher-Nelson G, Benjakul S, Global Adult Tobacco Survey (GATS) Collaborative Group. Contrasting trends of smoking cessation status: insights from the stages of change theory using repeat

data from the global adult tobacco survey, Thailand (2009 and 2011) and Turkey (2008 and 2012). Prev Chronic Dis 2017;14:E42 [FREE Full text] [doi: 10.5888/pcd14.160376] [Medline: 28570209]
5. Gil KM, Schrop SL, Kline SC, Kimble EA, McCord G, McCormick KF, et al. Stages of change analysis of smokers attending clinics for the medically underserved. J Fam Pract 2002;51(12):1018 [FREE Full text] [Medline: 12540324]
6. Keller CS, McGowan N. Examination of the processes of change, decisional balance, self-efficacy for smoking and stages of change in Mexican American women. South Online J Nurs Res 2001;2(4):1 [FREE Full text]
7. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults — United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
8. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Interventions for tobacco smoking cessation in adults, including pregnant persons: US preventive services task force recommendation statement. JAMA 2021;325(3):265-279 [FREE Full text] [doi: 10.1001/jama.2020.25019] [Medline: 33464343]
9. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, et al. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Department of Health and Human Services; 2008.
10. Vinci C, Lam C, Schlechter CR, Shono Y, Vidrine JI, Wetter DW. Increasing treatment enrollment among smokers who are not motivated to quit: a randomized clinical trial. Transl Behav Med 2022;12(1):ibab114 [FREE Full text] [doi: 10.1093/tbm/ibab114] [Medline: 34424337]
11. McClure JB, Westbrook E, Curry SJ, Wetter DW. Proactive, motivationally enhanced smoking cessation counseling among women with elevated cervical cancer risk. Nicotine Tob Res 2005;7(6):881-889. [doi: 10.1080/14622200500266080] [Medline: 16298723]
12. McClure JB, Peterson D, Derry H, Riggs K, Saint-Johnson J, Nair V, et al. Exploring the "active ingredients" of an online smoking intervention: a randomized factorial trial. Nicotine Tob Res 2014;16(8):1129-1139 [FREE Full text] [doi: 10.1093/ntr/ntu057] [Medline: 24727369]
13. Ali A, Kaplan CM, Derefinko KJ, Klesges RC. Smoking cessation for smokers not ready to quit: meta-analysis and cost-effectiveness analysis. Am J Prev Med 2018;55(2):253-262 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.021] [Medline: 29903568]
14. McClure JB, Catz SL, Ludman EJ, Richards J, Riggs K, Grothaus L. Feasibility and acceptability of a multiple risk factor intervention: the step up randomized pilot trial. BMC Public Health 2011;11:167 [FREE Full text] [doi: 10.1186/1471-2458-11-167] [Medline: 21414216]
15. Olsen E. Digital health apps balloon to more than 350,000 available on the market, according to IQVIA report. Healthcare IT News. 2021. URL: https://www.mobihealthnews.com/news/digital-health-apps-balloon-more-350000-available-market-according-iqvia-report [accessed 2022-06-03]
16. Global digital health market trajectory and analytics report 2022: telemedicine and mobile health apps adoption reaches new heights amid the pandemic - researchandmarkets.com. Business Wire. 2022. URL: https://tinyurl.com/msjbuhek [accessed 2023-05-19]
17. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
18. McClure JB, Heffner J, Hohl S, Klasnja P, Catz SL. Design considerations for mhealth programs targeting smokers not yet ready to quit: results of a sequential mixed-methods study. JMIR Mhealth Uhealth 2017;5(3):e31 [FREE Full text] [doi: 10.2196/mhealth.6845] [Medline: 28283465]
19. Heffner JL, Catz SL, Klasnja P, Tiffany B, McClure JB. Development of a mobile health intervention with personal experiments for smokers who are ambivalent about quitting: formative design and testing. JMIR Form Res 2020;4(8):e21784 [FREE Full text] [doi: 10.2196/21784] [Medline: 32852278]
20. Houston TK, Chen J, Amante DJ, Blok AC, Nagawa CS, Wijesundara JG, et al. Effect of technology-assisted brief abstinence game on long-term smoking cessation in individuals not yet ready to quit: a randomized clinical trial. JAMA Intern Med 2022;182(3):303-312 [FREE Full text] [doi: 10.1001/jamainternmed.2021.7866] [Medline: 35072714]
21. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
22. Perkins KA, Conklin CA, Levine MD. Cognitive-Behavioral Therapy for Smoking Cessation: A Practical Guidebook to the Most Effective Treatments. New York: Routledge; 2008.
23. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016;4(1):e17 [FREE Full text] [doi: 10.2196/mhealth.5181] [Medline: 26872940]
24. West R. Time for a change: putting the transtheoretical (stages of change) model to rest. Addiction 2005;100(8):1036-1039 [FREE Full text] [doi: 10.1111/j.1360-0443.2005.01139.x] [Medline: 16042624]
25. West R. The P.R.I.M.E theory of motivation as a possible foundation for addiction treatment. In: Henningfield JE, Santora PB, Bickel WK, editors. Addiction Treatment: Science and Policy for the Twenty-First Century. Baltimore: Johns Hopkins University Press; 2007:24-34.

26. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
27. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1997.
28. Bandura A. Self-Efficacy: The Exercise of Control. New York: W.H. Freeman and Company; 1997.
29. Fogg B. A behavioral model for persuasive design. 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 1-7. [doi: 10.1145/1541948.1541999]
30. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
31. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 2007;31(7):1208-1217 [FREE Full text] [doi: 10.1111/j.1530-0277.2007.00403.x] [Medline: 17451397]
32. Adamson SJ, Kay-Lambkin FJ, Baker AL, Lewin TJ, Thornton L, Kelly BJ, et al. An improved brief measure of cannabis misuse: the cannabis use disorders identification test-revised (CUDIT-R). Drug Alcohol Depend 2010;110(1-2):137-143 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.02.017] [Medline: 20347232]
33. Fucito LM, Toll BA, Roos CR, King AC. Smokers' treatment expectancies predict smoking cessation success. J Smok Cessat 2016;11(3):143-149 [FREE Full text] [doi: 10.1017/jsc.2014.17] [Medline: 27594921]
34. Barnard J, Rubin D. Miscellanea. Small-sample degrees of freedom with multiple imputation. Biometrika 1999;86(4):948-955. [doi: 10.1093/biomet/86.4.948]
35. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
36. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
37. Sim J. Should treatment effects be estimated in pilot and feasibility studies? Pilot Feasibility Stud 2019;5:107 [FREE Full text] [doi: 10.1186/s40814-019-0493-7] [Medline: 31485336]
38. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005;4(4):287-291. [doi: 10.1002/pst.185]
39. Schmidt CA, Romine JK, Bell ML, Armin J, Gordon JS. User participation and engagement with the see me smoke-free mhealth app: prospective feasibility trial. JMIR Mhealth Uhealth 2017;5(10):e142 [FREE Full text] [doi: 10.2196/mhealth.7900] [Medline: 28993302]
40. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
41. Boland VC, Mattick RP, Siahpush M, Barker D, Doran CM, Martire KA, et al. Factors associated with quitline and pharmacotherapy utilisation among low-socioeconomic status smokers. Addict Behav 2019;89:113-120 [FREE Full text] [doi: 10.1016/j.addbeh.2018.09.029] [Medline: 30290299]
42. Patnode CD, Henderson JT, Thompson JH, Senger CA, Fortmann SP, Whitlock EP. Behavioral counseling and pharmacotherapy interventions for tobacco cessation in adults, including pregnant women: a review of reviews for the U.S. preventive services task force. Ann Intern Med 2015;163(8):608-621 [FREE Full text] [doi: 10.7326/M15-0171] [Medline: 26389650]
43. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002;4(2):149-159 [FREE Full text] [doi: 10.1080/14622200210123581] [Medline: 12028847]
44. Thrul J, Howe CL, Devkota J, Alexander A, Allen AM, Businelle MS, et al. A scoping review and meta-analysis of the use of remote biochemical verification methods of smoking status in tobacco research. Nicotine Tob Res 2022:ntac271. [doi: 10.1093/ntr/ntac271] [Medline: 36449414]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test
CUDIT-R: Cannabis Use Disorder Identification Test-Revised
EC: enhanced care
FTND: Fagerström Test of Nicotine Dependence
KPWA: Kaiser Permanente Washington
mHealth: mobile health
NRT: nicotine replacement therapy
PPA: point prevalent abstinence
RR: relative risk
SC: standard care

Edited by L Buis; submitted 02.02.23; peer-reviewed by J Ferron, M Cozad; comments to author 03.03.23; revised version received 21.03.23; accepted 16.05.23; published 28.06.23.

Please cite as:
McClure JB, Heffner JL, Krakauer C, Mun S, Klasnja P, Catz SL
Feasibility, Acceptability, and Potential Impact of a Novel mHealth App for Smokers Ambivalent About Quitting: Randomized Pilot Study
JMIR Mhealth Uhealth 2023;11:e46155
URL: https://mhealth.jmir.org/2023/1/e46155
doi:10.2196/46155
PMID:37379059

Original Paper

A Mental Health and Well-Being Chatbot: User Event Log Analysis

Frederick Booth[1], BSc, PhD; Courtney Potts[2], BSc, MSc; Raymond Bond[3], BSc, PhD; Maurice Mulvenna[3], BSc, MPhil, PhD; Catrine Kostenius[4], PhD; Indika Dhanapala[5], BSc, MSc, PhD; Alex Vakaloudis[5], PhD; Brian Cahill[5], BEng, MSc; Lauri Kuosmanen[6], PhD; Edel Ennis[2], BSc, PhD

[1]Department of Accounting, Finance & Economics, Belfast, United Kingdom

[2]School of Psychology, Ulster University, Coleraine, United Kingdom

[3]School of Computing, Ulster University, Belfast, United Kingdom

[4]Department of Health, Education and Technology, Luleå University of Technology, Luleå, Sweden

[5]Nimbus Research Centre, Munster Technological University, Cork, Ireland

[6]Department of Nursing Science, University of Eastern Finland, Kuopio, Finland

Corresponding Author:
Courtney Potts, BSc, MSc
School of Psychology
Ulster University
Cromore Road
Coleraine, BT52 1SA
United Kingdom
Phone: 44 28953675
Email: c.potts@ulster.ac.uk

Abstract

Background: Conversational user interfaces, or chatbots, are becoming more popular in the realm of digital health and well-being. While many studies focus on measuring the cause or effect of a digital intervention on people's health and well-being (outcomes), there is a need to understand how users really engage and use a digital intervention in the real world.

Objective: In this study, we examine the user logs of a mental well-being chatbot called ChatPal, which is based on the concept of positive psychology. The aim of this research is to analyze the log data from the chatbot to provide insight into usage patterns, the different types of users using clustering, and associations between the usage of the app's features.

Methods: Log data from ChatPal was analyzed to explore usage. A number of user characteristics including user tenure, unique days, mood logs recorded, conversations accessed, and total number of interactions were used with k-means clustering to identify user archetypes. Association rule mining was used to explore links between conversations.

Results: ChatPal log data revealed 579 individuals older than 18 years used the app with most users being female (n=387, 67%). User interactions peaked around breakfast, lunchtime, and early evening. Clustering revealed 3 groups including "abandoning users" (n=473), "sporadic users" (n=93), and "frequent transient users" (n=13). Each cluster had distinct usage characteristics, and the features were significantly different (P<.001) across each group. While all conversations within the chatbot were accessed at least once by users, the "treat yourself like a friend" conversation was the most popular, which was accessed by 29% (n=168) of users. However, only 11.7% (n=68) of users repeated this exercise more than once. Analysis of transitions between conversations revealed strong links between "treat yourself like a friend," "soothing touch," and "thoughts diary" among others. Association rule mining confirmed these 3 conversations as having the strongest linkages and suggested other associations between the co-use of chatbot features.

Conclusions: This study has provided insight into the types of people using the ChatPal chatbot, patterns of use, and associations between the usage of the app's features, which can be used to further develop the app by considering the features most accessed by users.

(JMIR Mhealth Uhealth 2023;11:e43052) doi:10.2196/43052

KEYWORDS

mental well-being; positive psychology; data analysis; health care; event log analysis; ecological momentary assessment; conversational user interface; user behavior; conversational agent; user interface; user data; digital health application; mental well-being; mobile health app; digital intervention

Introduction

Chatbots, or conversational user interfaces can take diverse roles in supporting mental health. In particular, chatbots are becoming increasingly popular as digital mental health and well-being interventions, with initial evaluations of efficacy showing promise [1-3]. Chatbots may be targeted toward a variety of outcomes such as medication adherence, treatment compliance, aftercare support, delivery of appointment reminders, psychoeducation, user empowerment, and improvement in the self-management of mental health and well-being through monitoring mood or symptom change [1]. They can also be used to promote help-seeking [1]. The potential benefits are recognized by both practitioners and clients [2-5]. In addition to supporting those with mental ill health, digital technologies are also considered to have the potential for preventing mental health problems and for improving the overall mental health of the population [6].

Event logging plays an important role in modern IT systems with many apps logging their events to a local or remote server. These event logs can be used to determine the use of an app and identify user patterns. The analysis of user patterns involves event correlation—a conceptual interpretation procedure where new meaning is assigned to events that occur within a set time frame [7].

Event logging can be easily incorporated into digital products, including apps and websites. The most basic event log data consists of an anonymous unique identifier assigned to each individual, a date-time stamp, and an activity or event, but may also include other contextual variables. These user interaction log data provide useful information on how digital technologies are actually being used, providing valuable insights into user behavior [8]. Event log analysis typically focuses on quantitative data, but it may provide even greater insights when combined with qualitative data such as ecological momentary assessment (EMA) [8-11]. EMA involves asking questions, for example, "how do you feel right now," repeatedly over a period of time in an individual's own environment (ecology). Users answer EMA questions "in the moment," which helps to avoid recall bias [11]. Previous trials with mental health chatbots have used app interaction data [12,13]. For example, participants trialing the mental health chatbot "Wysa" were characterized based on their app usage into more engaged users, termed "high users," and less engaged users, or "low users" [13]. A multilingual mental health and well-being chatbot named "ChatPal" was developed to promote good mental well-being of citizens living in sparsely populated areas across Europe. Once the chatbot was released into the wild, all interactions between the chatbot and users were logged.

The aim of this research is to analyze event log data from the ChatPal chatbot with the objectives of providing insight into the different types of users using k-means clustering, exploring usage patterns, and associations between the usage of the app's features.

Methods

Ethics Approval

This study received ethical approval from the Ulster University Research Ethics Committee (reference numbers REC.21.0021 and FCPSY-21-038-A), the Munster Technological University Research Ethics Committee (reference number MTU21034A), the Ethics Review Authority in Sweden (reference Etikprövningsmyndigheten number 2020-00808), and the University of Eastern Finland Committee on Research Ethics gave a supporting statement (statement 14/2021).

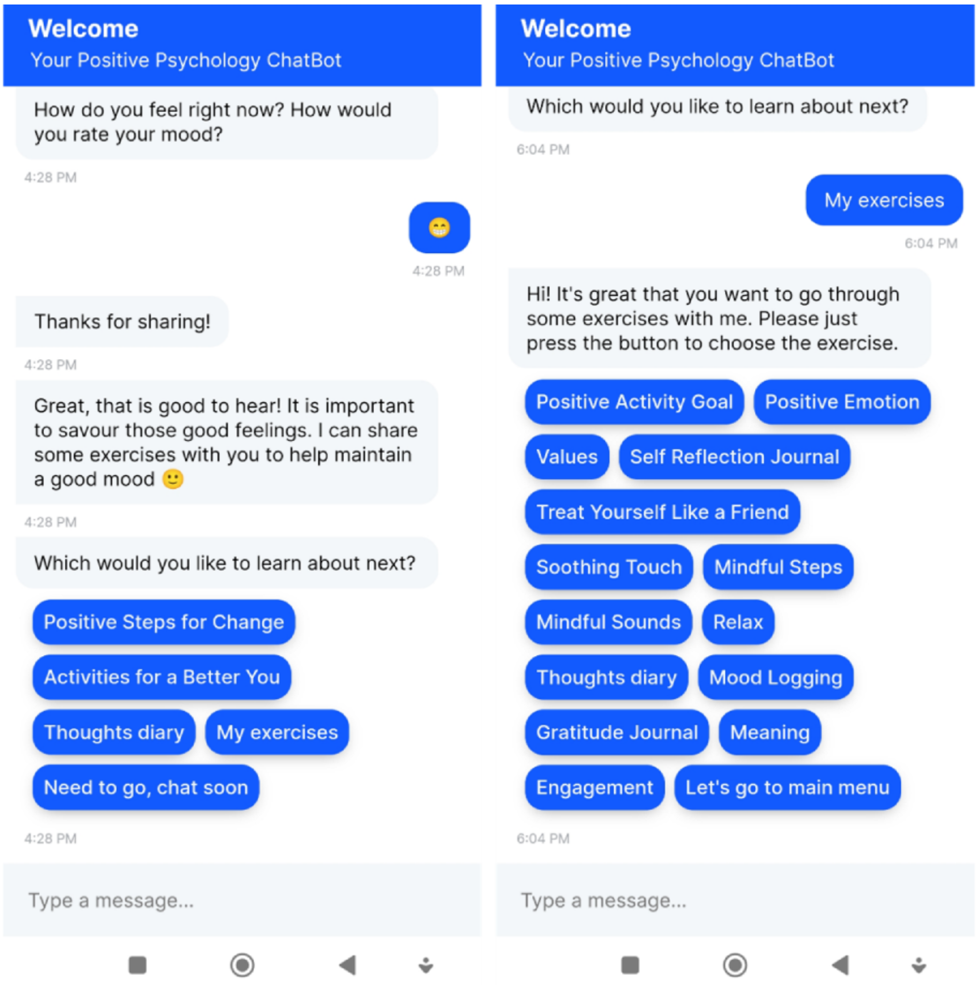
Intervention

The ChatPal chatbot was co-designed with end users and developed as part of the ChatPal project [14], a collaboration between universities and mental health service providers. The remit of the project was to develop and evaluate a chatbot to promote the positive mental well-being of individuals living in rural areas across Europe. This was achieved using an iterative approach similar to a previous study on how users engage and are redirected through a chatbot for depression [15]. A prototype chatbot was released early to support individuals at the beginning of the COVID-19 pandemic, and feedback from this initial phase of the study was used to refine the chatbot [16]. This refined version was then trialed in Northern Ireland, the Republic of Ireland, Scotland, Sweden, and Finland. Individuals across these regions were recruited to take part in a pre-post intervention study, using the ChatPal app as they wished for a period of 12 weeks. The chatbot was also advertised on social media and was freely available on the Apple App Store and Google Play Store. The chatbot was developed based on the concept of positive psychology, with elements of psychological well-being and happiness. Known as the PERMAH model [17], it includes content to encourage positive emotions, engagement, relationships, meaning, accomplishment, and health. Based on mostly scripted conversations with predefined responses, ChatPal can maintain a basic dialogue with a user in order to advise them on how to maintain positive emotions and mental well-being. An overview of the content available in ChatPal can be found in Multimedia Appendix 1.

Chatbot Development

The ChatPal chatbot was developed using the Rasa and PhoneGap frameworks. Rasa (Alan Nichol and Alex Weidauer) is an artificial intelligence (AI)–assisted framework for building contextual chatbots and provides the infrastructure and tools necessary for high-performing, resilient, proprietary contextual assistants. PhoneGap is an open-source framework for developing cross-platform mobile apps, including iPhone and Android. The PhoneGap framework was used to develop the front end of the ChatPal app, resulting in a cross-platform mobile app using HTML5, JavaScript, and CSS (Figure 1). Communication with the ChatPal backend is achieved using HTTP requests or responses. Upon receiving user inputs, the Rasa backend analyzes these inputs using its Natural Language Understanding unit, which extracts user intentions and relevant metadata from the input. Once the intentions and metadata are identified, corresponding actions and responses are decided by the AI Rasa core. ChatPal dialogues required additional functionality such as language selection, onboarding, log entries, and visualized responses (graphs), which were addressed by custom development in Rasa.

Figure 1. Screenshots of the ChatPal chatbot app.



User Log Data Provenance

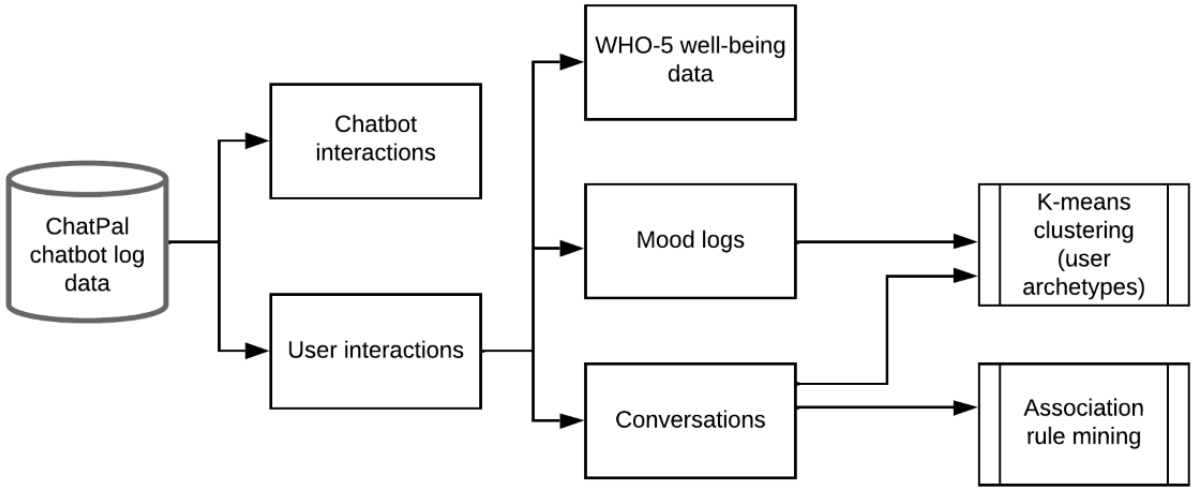
The log data file initially contained all interactions between users and the app (including detailed app events) that occurred during a prepost study period (January 24-June 22, 2022). During data cleaning, it was necessary to identify only the events made by users and extract these user event log details for analysis. While users remained anonymous, each user was assigned a unique ID. Log entries were timestamped to facilitate the tracking of interactions over time. The app afforded users the opportunity to converse with the chatbot to help users manage their mental health, with their current mood being logged each time they used the chatbot. Users were also given the option to complete 5 questions relating to the World Health Organization-Five Well-being Index (WHO-5), a measure of current mental well-being [18].

Data Analysis and Machine Learning

Jupyter notebooks [19] with the programming language Python were used to analyze the log data, with the matplotlib [20] and seaborn [21] libraries used to visualize the results. The scipy [22] and sklearn [23] libraries were used to normalize and perform k-means clustering, and the mlxtend [24] library was used for association rule mining.

Figure 2 shows the methodology for analyzing the ChatPal chatbot event log data. Data on each user were collated, including age, gender, and interactions with the chatbot. As only data pertaining to users older than 18 years was permitted, only users who had completed the age question could be included in the analysis. Of these users, those who did not complete the gender question were assigned an “unknown” status for gender.

Figure 2. Methodology for log analysis. WHO-5: World Health Organization-5 well-being index.



User Interactions

The total number of user interactions with the chatbot was examined across hours of the day to gain insight into daily patterns of use. The tenure (ie, the number of days from first to last use of the app) was calculated for each user along with the number of unique days of use. User retention was then calculated to discover the percentage of users still using the app over time.

User Types

K-means clustering was used to determine the different types of users who interacted with the chatbot, with 6 features relating to the behavioral usage of the chatbot being identified within the log data (Table 1). Before submitting these features to the clustering algorithm, these features were normalized, resulting in the standardization of all variables to ranges between 0 and 1. In order to determine the optimum number of clusters, the k-means algorithm was applied to the data for clusters k=2 to k=15, with the average distance to the centroid from all data points (within-cluster sum of square) being calculated for each iteration. These were then plotted to identify the “pivot” where the resulting graph creates an “elbow,” which corresponds to the optimum k value (ie, the number of user groups or clusters that exist). The k-means algorithm was then used for this “k,” and the resulting output was visualized using principal component analysis which reduced the multiple (high-dimensional) features to just 2 dimensions, enabling the clusters to be plotted for visual inspection and verification. The resulting cluster labels were then mapped back onto the original feature data set to allow each cluster to be analyzed. Independent *t* tests were carried out to evaluate the significance of the difference in the means of the groups.

Table 1. Features used for k-means clustering.

Feature	Description
Unique days	This is the number of unique days the user accessed the chatbot.
Tenure	This is the number of days between the first use and last use of the chatbot.
Mood logs completed	This is the number of times the user recorded their mood over their period of use of the app. This gives an indication of how many times the app was used as the user’s mood was requested each time they used the chatbot, there may be multiple mood logs for each day.
Conversations accessed	This is the number of conversations accessed during their period of use of the chatbot.
Total interactions	This is the total number of interactions with the chatbot.

Feature Usage Analysis

The key feature of the chatbot, a series of scripted “conversations,” was chosen for further analysis. Users could access these conversations via the main menu and were asked to rate the conversation as “good,” “neutral,” or “bad” on completion. The log data was analyzed to examine the number of times each conversation was accessed during the prepost study period, the percentage of users that accessed each conversation, and the rating awarded by users on the completion of each conversation. The date and time when each user accessed any conversation were logged, making it possible to create a daily-ordered set of conversations for each user for analysis. From these sets, it was possible to examine associations between conversations; in other words, examine the pathway from one conversation (the antecedent) to the next (the consequent) using association rule mining.

Results

User Interactions

There were a total of 1403 individual users that accessed the app between January 24, 2022, and June 22, 2022. Only data from users older than 18 years were included in the analysis; thus, the results report on data from 579 adult users, of whom 348 (60.1%) were recruited specifically for a 12-week prepost study [25]. The majority of users identified as female (387/579, 66.8%), male (153/579, 26.4%), or other (6/579, 1%), while other users preferred not to say (12/579, 2.1%) or elected not to answer (21/579, 3.6%). Nonresponses were treated as unknown. The ages of participants were well distributed, with responses of 18-24 (190/579, 32.8%), 25-34 (150/579, 25.9%), 35-44 (95/579, 16.4%), 45-54 (77/579, 13.3%), 55-64 (56/579, 9.7%), and >65 (11/579, 1.9%).

Over the study period, from January 24, 2022, to June 24, 2022, there were a total of 29,298 user interactions with the app, averaging 246 interactions per day. While users interacted with the app throughout the day, peaks in interactions can be seen at key times of the day: breakfast (8 AM-10 AM), lunch (1 PM), and the end of the working day (5 PM) (Figure 3).

A large proportion of users interacted with the app for a short period of time, with 440 users (76%) interacting for less than 10 days. Analysis of the number of unique days of interaction with the app also showed that although there were users who interacted with the app over almost the whole study period, these interactions were sporadic, with the most ardent user interacting with the chatbot over 19 unique days (Figure 4). Overall user retention shows a steady drop-off in users over time (Figure 4). The average tenure of a user was 11.4 days.

A total of 6 features representing the behavioral usage of the app were selected for k-means clustering to identify the different types of users accessing the app. These features represent user characteristics, including the number of unique days the user accessed the app, the tenure of each user, the number of mood logs recorded, the number of conversations accessed, and the total number of interactions with the app.

The optimum number of types of users (clusters) was determined using the elbow method. This method runs k-means clustering on the data set for a range of k values (eg, k=1-15), calculating the average distances to the centroid from all data points, known as the within-cluster sum of squares. These distances are then plotted on a graph, which will show where the distances fall, creating an elbow in the graph. This elbow represents the optimum k value for the clustering solution. Figure 5 shows the results, indicating 2 possible solutions, k=2 and k=3 in this case.

Figure 3. User interactions over the course of the day.

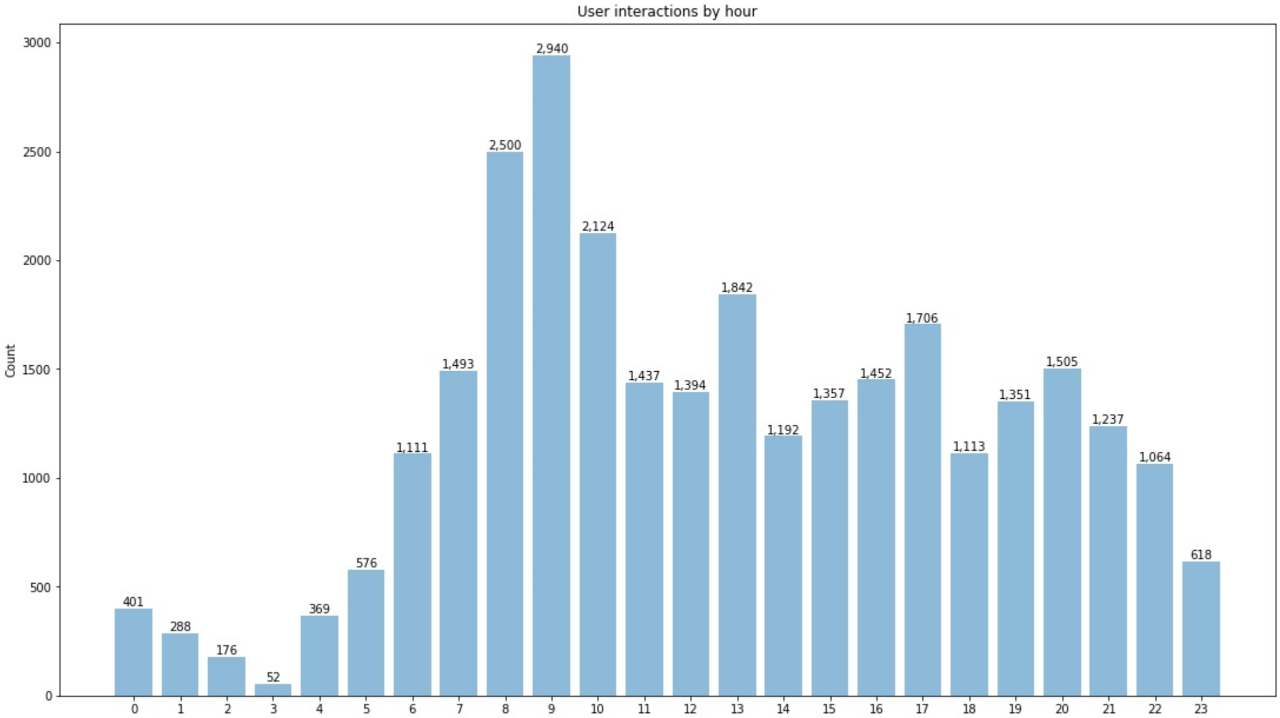


Figure 4. Unique days of use of the chatbot (left) and user retention curve (right).

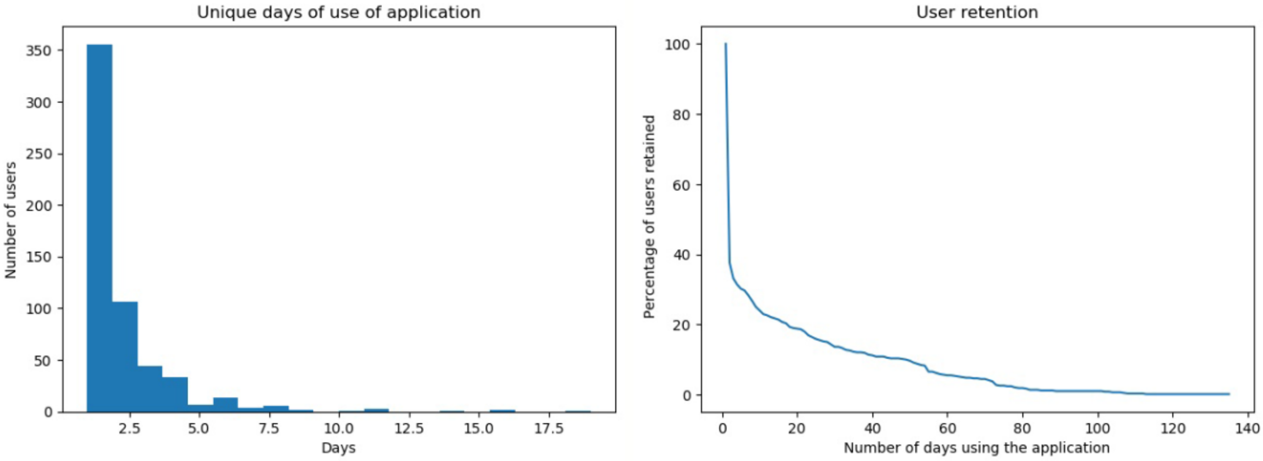
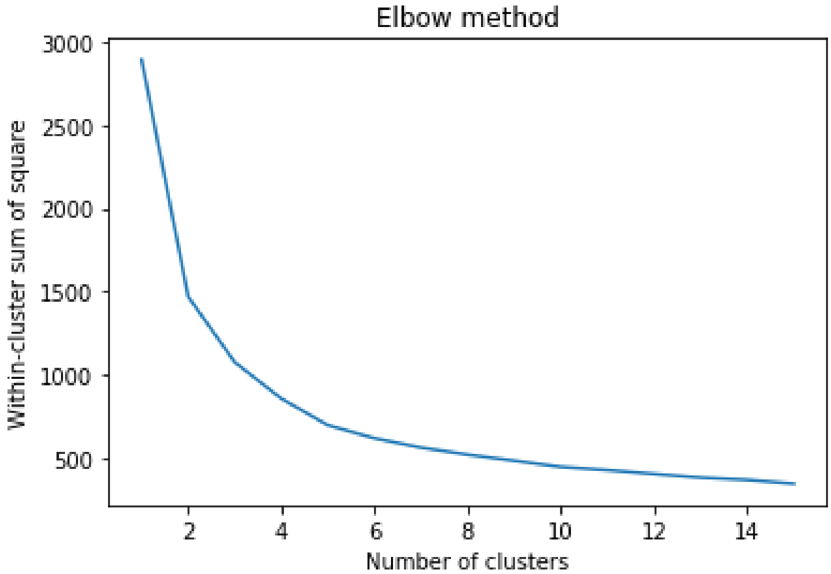


Figure 5. Elbow plot showing the optimum number of clusters.



User Types

Both a 2- and 3-cluster solution were explored. Using k=3 results in a similar principal component analysis plot as in the 2-cluster solution, with 3 well-defined clusters with distinct usage characteristics (Figure 6).

The 3-cluster solution (Figure 7) appeared to be a refinement of the 2-cluster solution, with the largest cluster equating to “abandoning users” and the remaining 2 clusters revealing a more granular look at more invested users. Analysis of the archetypal characteristics of these 3 clusters (Table 2) suggested that “abandoning users” (473/579, 81.7%), “frequent transient users” (13/579, 2.2%), and “sporadic users” (93/579, 16.1%) would be appropriate labels.

Abandoning users generally access the app on 1 or 2 unique days, with an average tenure of 3.6 days (Table 2). They recorded lower numbers of mood logs (1-2), accessed fewer conversations (0-1), and had fewer interactions with the chatbot (25-43) compared to the other 2 clusters. At the other extreme, frequent transient users accessed the app between 8 and 14 unique days, with an average tenure of 68.9 days (Table 2). They recorded the highest number of mood logs (6-12) and conversations (7-12) and had the highest number of interactions with the app (193-258). While sporadic users only used the app for a small number of unique days (3-5), they did so over a longer period (22-61 days) (Table 2). In all other metrics, they exceeded those recorded by abandoning users but did not achieve the numbers attributed to frequent transient users: mood logs (2-5), conversations (2-5), and interactions (20-123). Independent *t* tests on these results found that the 3 archetypes are significantly different statistically across all metrics ($P<.001$).

Daily patterns of usage of ChatPal differed for each archetype (Figure 8). Abandoning users recorded low numbers of interactions with the app over the course of the day, with a small peak of usage in the morning (9 AM). Interactions for sporadic users were higher than those seen for abandoning users, and in general, lower than those for frequent transient users, the exception being in the morning (8 AM), when interactions for sporadic and frequent transient users were the same. Sporadic users showed peak usage times around breakfast (8 AM) and lunch (1 PM). Frequent transient users generally recorded the highest number of interactions with the app over the course of the day, with frequent peaks in usage. These users showed high levels of interactions over most of the morning (4-9 AM), with further peaks in usage at lunchtime (1 PM), afternoon (4 PM), and evening (8 PM).

Figure 6. Principal component analysis plot of output from the k-means algorithm (3-cluster solution).

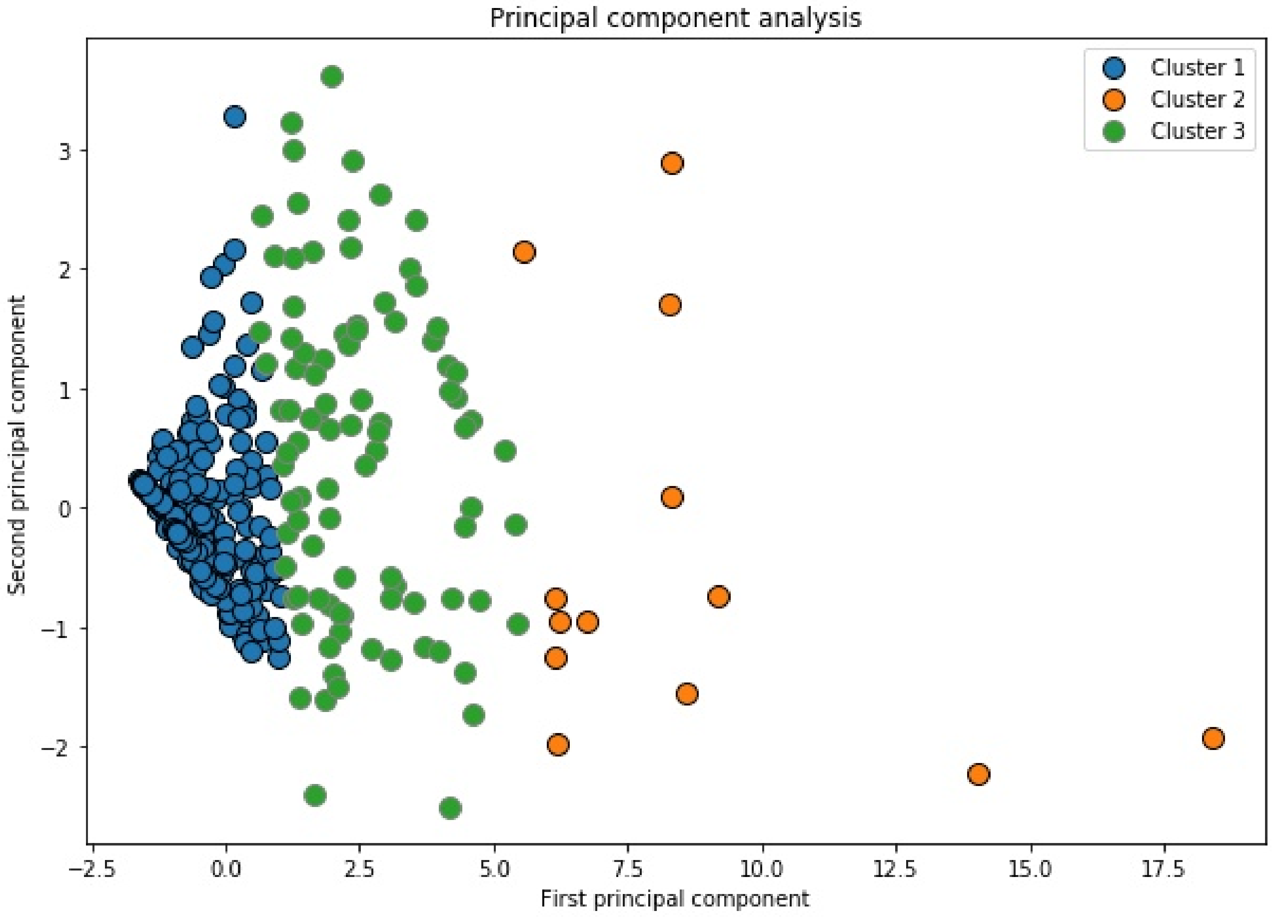


Figure 7. Boxplots of feature values for the different clusters (3-cluster solution).

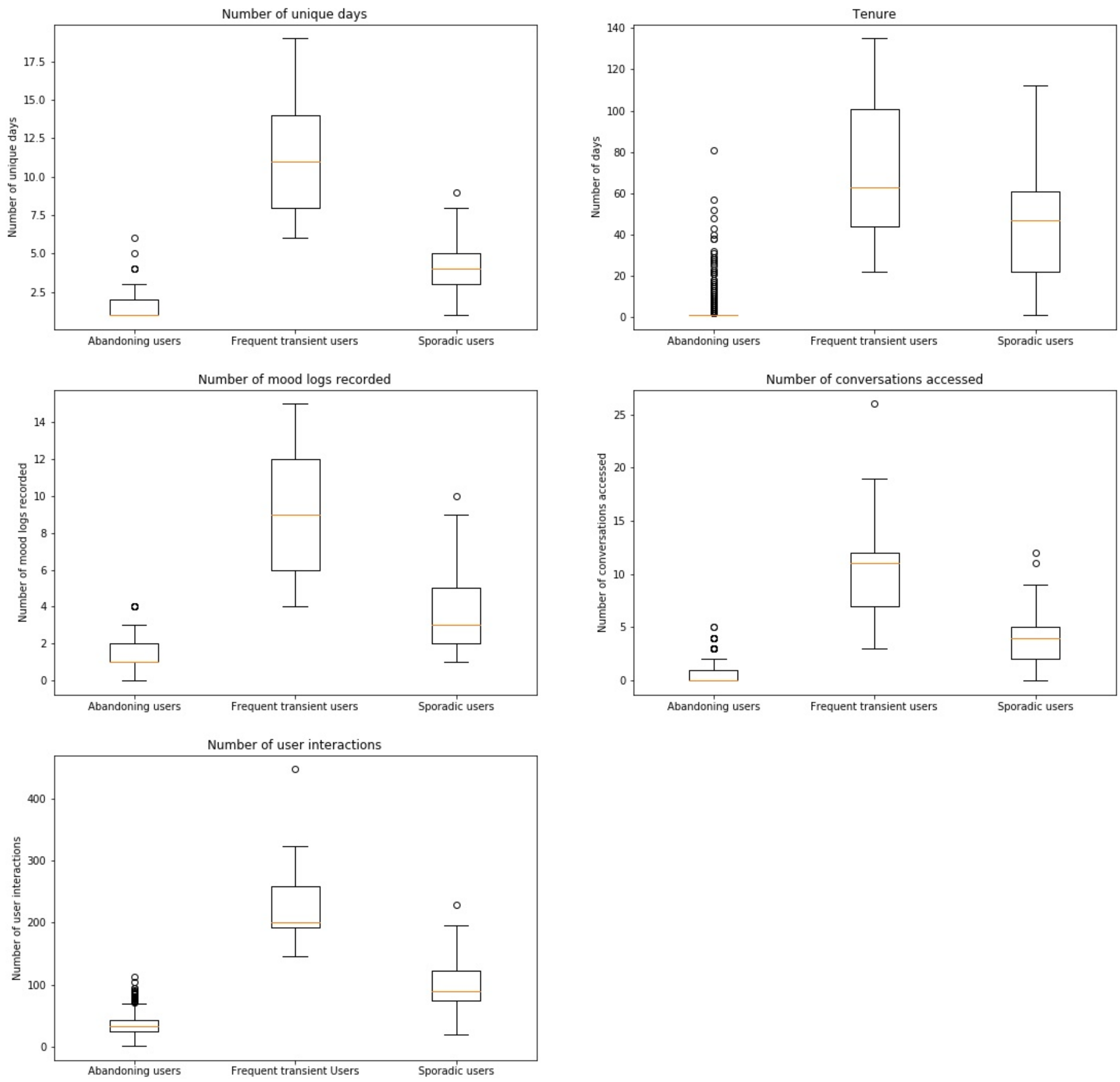
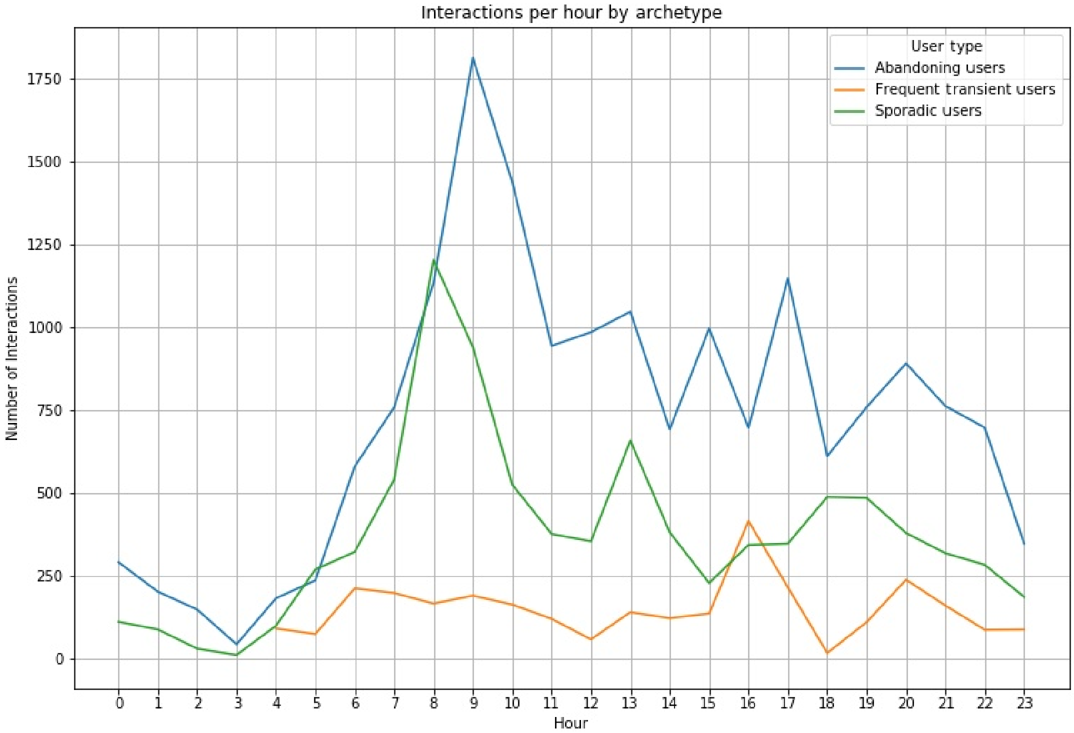


Table 2. Archetypal characteristics for each cluster (3-cluster solution).

Feature	Cluster 1: abandoning users	Cluster 2: frequent transient users	Cluster 3: sporadic users
Users, n (%)	473 (81.7)	13 (2.2)	93 (16.1)
Unique days, mean (SD)	1.3 (0.7)	11.2 (3.9)	4.0 (1.7)
Tenure, mean (SD)	3.6 (8.1)	68.9 (35.4)	43.0 (25.2)
Mood logs completed, mean (SD)	1.3 (0.8)	8.8 (3.8)	3.8 (1.9)
Conversations accessed, mean (SD)	0.8 (1.0)	10.8 (6.3)	4.0 (2.4)
Total interactions, mean (SD)	36.7 (16.8)	229.3 (83.0)	96.1 (36.6)

Figure 8. Average number of interactions per user each hour by archetype.



Feature Usage Analysis

A key element of the app was the provision of mental well-being conversations between the chatbot and the user. Analysis of users' choice of conversations showed that of the 579 users, almost one-third chose to access the "Treat yourself as a friend" conversation (168/579, 29%), with over one-fifth of users accessing the "Relax (140/579, 24.2%)," "Thoughts Diary" (136/579, 23.5%), and "Soothing Touch" (118/579, 20.4%) conversations. Of these, only 68 (11.7%) returned to the "Treat yourself as a friend" conversation on 2 or more occasions, with this number falling to 40 (6.9%) returning on 3 or more occasions. The conversation with the highest return rate was "Thoughts Diary" (109/579, 18.8% for 2 or more occasions; 69/579, 11.9% for 3 or more occasions), while the conversation with the lowest return rate was "How to Help Someone" (2/579, 0.3% for both metrics) (Figure 9). The top conversations received good ratings, reflecting their popularity with users. "Treat yourself as a friend" received the highest percentage of good ratings of the 4 (62.5%), followed by "Relax" (56.3%), "Thoughts Diary" (52%), and "Soothing Touch" (53.7%). Interestingly, the conversations that were used the fewest number of times elicited more positive ratings. "How to help someone," which was only accessed by 0.9% of the total users, resulted in a 100% good rating, with "Goal Setting" and "Goal Quality Check" receiving 80% good ratings (Figure 10).

Figure 9. Percentage of users accessing chatbot conversations. WHO: World Health Organization.

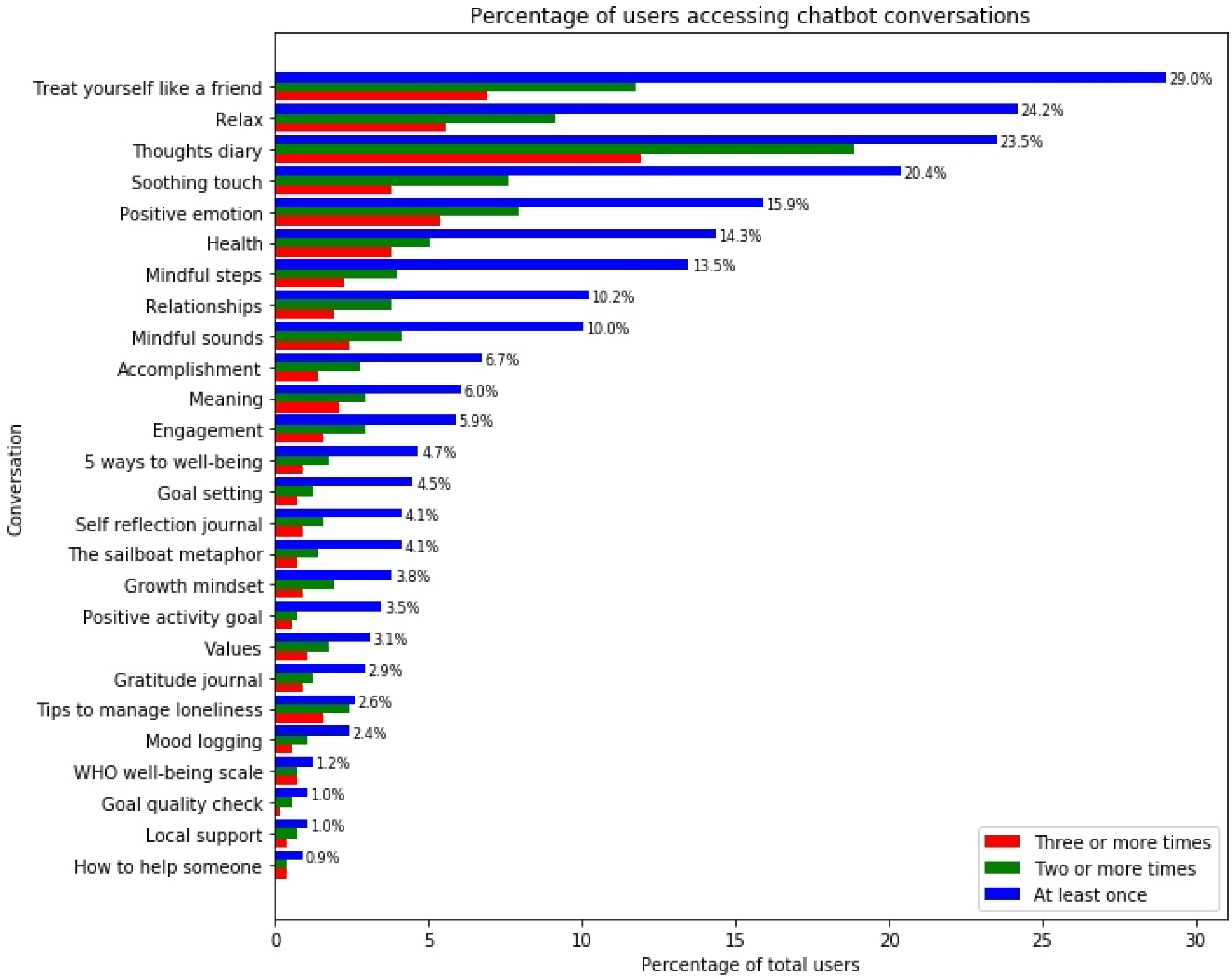
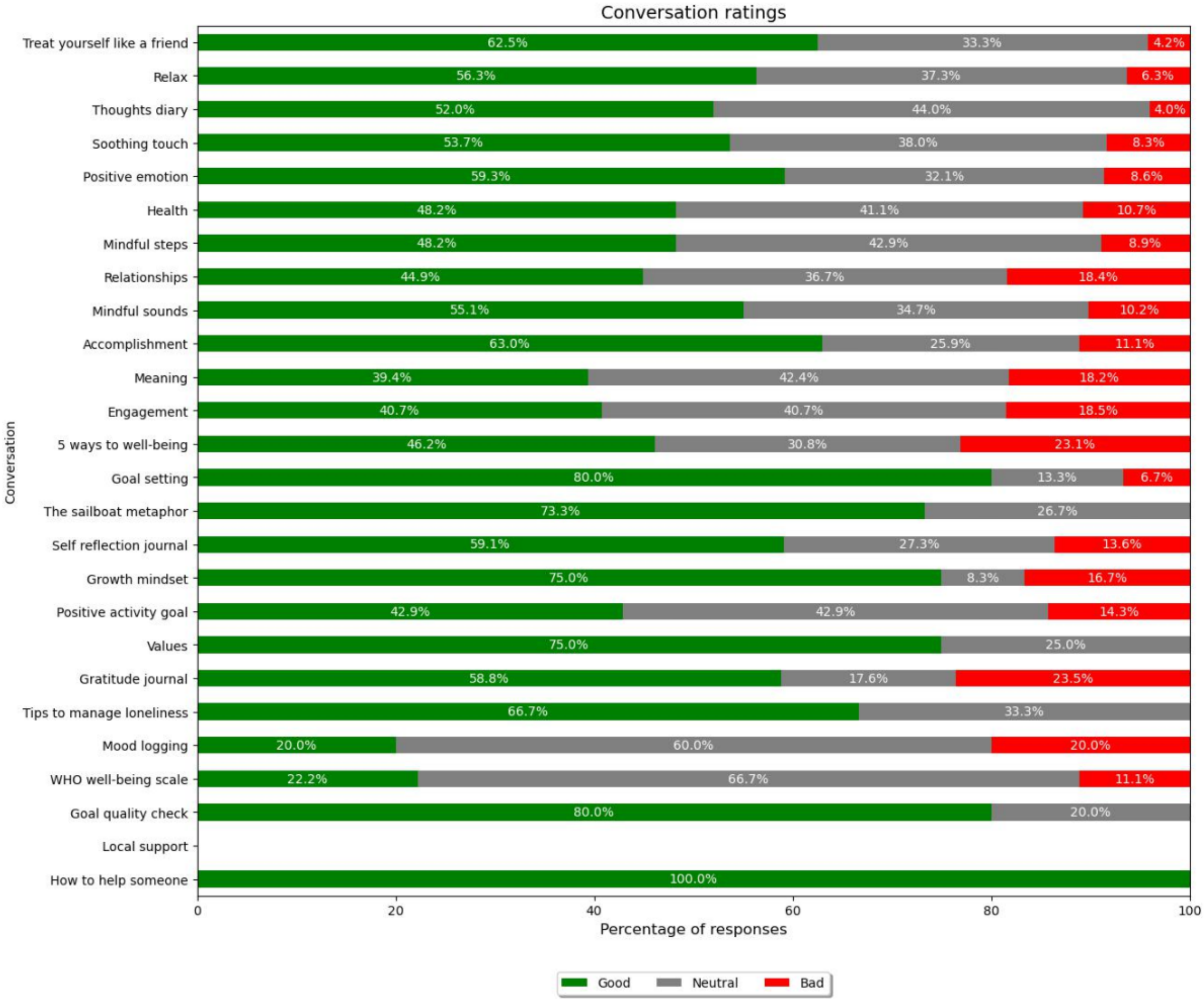


Figure 10. Conversation ratings. WHO: World Health Organization.



Analysis of the number of times each user accessed conversations showed that the majority of users used each conversation between 1 and 3 times. "Thoughts diary," "tips to manage loneliness," and "WHO well-being scale" had a median number of users of 3. A small number of users who accessed these conversations multiple times are represented as outliers on the graph (Multimedia Appendix 2).

The log data relating to "conversations" were examined for each user on a daily basis in order to provide insight into how users accessed conversations. Each time a user moved between 2 conversations, no matter the direction, the transition was counted. Statistical analysis of the resulting 307 transitions revealed the median to be 4, the lower quartile to be 2, and the upper quartile to be 11. Using these statistics, Figure 11 shows the results of the analysis, with low numbers of transitions represented in blue (<2), average represented in green (2-11), high in pink (12-50), and very high in red (>50). The transition between "treat yourself as a friend" and "soothing touch" was the most popular, being performed 168 times. Additionally, high on the scale were the transitions between "positive emotion" or "thoughts diary" (72) and "relax" or "thoughts diary" (62). Association rule mining supports these findings, with the linkage between "treat yourself as a friend," and "soothing touch" receiving the highest support of 0.16 and the highest support of a rule with more than one antecedent being "soothing touch" or "thoughts diary" with "treat yourself like a friend" with support of 0.072 (Table 3).

Figure 11. Associations between conversations (blue=low <2, green=average 2-11, pink=high 12-50, red=very high >50). WHO: World Health Organization.

	Treat yourself like a friend	Relax	Thoughts diary	Soothing touch	Positive emotion	Health	Mindful steps	Relationships	Mindful sounds	Accomplishment	Meaning	Engagement	5 ways to well-being	Goal setting	Self reflection journal	The sailboat metaphor	Growth mindset	Positive activity goal	Values	Gratitude journal	Tips to manage loneliness	Mood logging	WHO well-being scale	Local support	Goal quality check	How to help someone
Treat yourself like a friend		49	47	168	34	31	28	11	22	9	6	6	1	5	1	6	5	5	6	3	11	1	2	0	1	0
Relax	49		62	57	26	32	23	3	33	12	14	8	12	2	12	0	5	4	2	13	1	4	3	5	0	0
Thoughts diary	47	62		19	72	28	18	26	46	18	33	20	8	7	24	24	5	7	22	23	15	2	1	3	4	0
Soothing touch	168	57	19		11	2	30	9	29	3	12	4	1	1	15	4	2	0	1	0	1	1	0	0	2	1
Positive emotion	34	26	72	11		27	17	6	10	8	2	16	8	1	7	2	9	3	8	1	2	0	1	2	10	0
Health	31	32	28	2	27		5	7	13	13	2	9	2	1	1	4	0	0	0	0	0	2	4	0	0	0
Mindful steps	28	23	18	30	17	5		18	10	10	3	3	5	23	0	4	0	0	0	1	1	4	0	0	1	1
Relationships	11	3	26	9	6	7	18		0	8	8	11	3	3	11	2	5	4	0	0	3	0	0	0	1	0
Mindful sounds	22	33	46	29	10	13	10	0		0	3	9	1	2	7	11	2	0	5	2	0	0	4	0	0	11
Accomplishment	9	12	18	3	8	13	10	8	0		3	4	0	0	3	3	0	1	0	5	0	0	0	0	0	0
Meaning	6	14	33	12	2	2	3	8	3	3		1	4	1	2	0	0	2	6	3	0	2	6	0	0	0
Engagement	6	8	20	4	16	9	3	11	9	4	1		5	3	0	0	3	2	1	0	19	5	0	0	13	0
5 ways to well-being	1	12	8	1	8	2	5	3	1	0	4	5		1	7	0	0	0	1	0	0	1	4	4	0	0
Goal setting	5	2	7	1	1	1	23	3	2	0	1	3	1		1	0	4	5	0	0	0	9	1	0	2	10
Self reflection journal	1	12	24	15	7	1	0	11	7	3	2	0	7	1		1	0	1	7	0	0	1	0	0	0	0
The sailboat metaphor	6	0	24	4	2	4	4	2	11	3	0	0	0	0	1		0	0	0	0	3	6	0	0	11	1
Growth mindset	5	5	5	2	9	0	0	5	2	0	0	3	0	4	0	0		0	0	12	1	0	0	0	0	0
Positive activity goal	5	4	7	0	3	0	0	4	0	1	2	2	0	5	1	0	0		1	1	1	0	0	0	0	1
Values	6	2	22	1	8	0	0	0	5	0	6	1	1	0	7	0	0	1		0	0	2	0	1	0	0
Gratitude journal	3	13	23	0	1	0	1	0	2	5	3	0	0	0	0	0	12	1	0		3	4	0	3	0	9
Tips to manage loneliness	11	1	15	1	2	2	1	3	0	0	0	19	0	0	0	3	1	1	0	3	0	0	0	0	0	0
Mood logging	1	4	2	1	0	0	4	0	0	0	2	5	1	9	1	6	0	0	2	4	0		0	0	0	0
WHO well-being scale	2	3	1	0	1	4	0	0	4	0	6	0	4	1	0	0	0	0	0	0	0	0		0	0	0
Local support	0	5	3	0	2	0	0	0	0	0	0	0	4	0	0	0	0	0	1	3	0	0	0		0	0
Goal quality check	1	0	4	2	10	0	1	1	0	0	0	13	0	2	0	11	0	0	0	0	0	0	0	0		10
How to help someone	0	0	0	1	0	0	1	0	11	0	0	0	0	10	0	1	0	1	0	9	0	0	0	0	10	

Table 3. Summary of results of association rule mining on conversations ordered by lift.

Antecedents	Consequents	Support	Confidence	Lift
Soothing touch	Treat yourself like a friend	0.160	0.712	2.069
Treat yourself like a friend	Soothing touch	0.160	0.463	2.069
Thoughts diary	Treat yourself like a friend	0.121	0.391	1.135
Treat yourself like a friend	Thoughts diary	0.121	0.352	1.135
Relax	Treat yourself like a friend	0.104	0.396	1.150
Treat yourself like a friend	Relax	0.104	0.302	1.150
Soothing touch	Thoughts diary	0.094	0.420	1.352
Thoughts diary	Soothing touch	0.094	0.303	1.352
Positive emotion	Thoughts diary	0.093	0.475	1.530
Thoughts diary	Positive emotion	0.093	0.299	1.530
Soothing touch	Relax	0.092	0.410	1.562
Relax	Soothing touch	0.092	0.350	1.562
Relax	Thoughts diary	0.090	0.342	1.101
Thoughts diary	Relax	0.090	0.289	1.101
Positive emotion	Treat yourself like a friend	0.089	0.453	1.314
Treat yourself like a friend	Positive emotion	0.089	0.257	1.314
Health	Thoughts diary	0.073	0.370	1.193
Thoughts diary	Health	0.073	0.236	1.193
Soothing touch, thoughts diary	Treat yourself like a friend	0.072	0.767	2.229
Treat yourself like a friend, thoughts diary	Soothing touch	0.072	0.595	2.654
Soothing touch, treat yourself like a friend	Thoughts diary	0.072	0.452	1.456
Soothing touch	Treat yourself like a friend thoughts diary	0.072	0.322	2.654
Thoughts diary	Soothing touch, treat yourself like a friend	0.072	0.232	1.456
Treat yourself like a friend	Soothing touch, thoughts diary	0.072	0.210	2.229
Mindful sounds	Thoughts diary	0.071	0.512	1.649
Thoughts diary	Mindful sounds	0.071	0.229	1.649

Discussion

Principal Results

The majority of users (348/579, 60%) of the ChatPal chatbot app were recruited as volunteers in a 12-week prepost study [25]. User interactions occurred at all hours of the day, with the majority being during working hours. It was notable that spikes in usage occurred at the start of the working day (9 AM), lunchtime (1 PM), and end of the working day (5 PM). This may simply reflect the fact that users felt they could devote some time to evaluating the app at these times, or it could reflect a need for support during times of the day when users are not pressured by other commitments. User interactions with the app late into the evening and through the night may indicate the need for support at these times. Recent studies found that the COVID-19 pandemic resulted in increased sleep disturbances and worsening mental health in the general population [26-28]. This is one advantage of offering digital solutions in mental health care, as these technologies can be accessed 24/7 when traditional face-to-face services are unavailable.

Analysis of user interaction with the chatbot conversations reflects a high proportion of "abandoning users." The majority of conversations were accessed only once, and only a small percentage of total users accessed conversations more than once. Interestingly, the conversation with the highest return rate was the "thoughts diary," where users could record their thoughts and feelings and review them in subsequent sessions. This indicates that these users felt comfortable enough with the app to share their feelings, and they presumably felt some benefit from doing so as they returned on multiple occasions. Journaling about stressful events has been shown to be beneficial for individuals to understand and make sense of what happened [29].

The majority of conversations received a "good" or "neutral" rating, with "how to help someone" receiving a 100% good rating. While this is encouraging, it is important to recognize that the "how to help someone" conversation received the lowest percentage of overall users. Given these low figures, they may be biased by a small number of returning users. It is interesting that "how to help someone" was the least accessed conversation, as it suggests that people are not using the app to access information that would help others but more for resources to benefit their own well-being.

Analyzing users' transition from one conversation to another revealed that the most frequent transition was the reciprocal journey from "treat yourself like a friend" to "Soothing Touch," followed by transitions between the "positive emotion" or "thoughts diary" and "relax" or "thoughts diary" pairs of conversations. Association rule mining supported these findings as these 3 conversations, "soothing touch," "thoughts diary," and "treat yourself like a friend," were linked, with "thoughts diary" having strong links between the other 2 in either direction. This seems understandable as "treat yourself like a friend" and "positive emotions" are both conversations in which the user is writing something and may want to see it in the "thoughts diary" (saving the written message to the "thoughts diary" is suggested at the end of the "treat yourself like a friend" script, which may prompt the user to go there next). Additionally, "treat yourself like a friend," "soothing touch," and "relax" are all short dialogues or exercises, so users may want to go on chatting and access further conversations. This also demonstrates the value of incorporating efficient scripts into the app, providing quick exercises for users needing support [30].

Further analysis of the chatbot menu structure would provide clarity on whether these conversations naturally complement each other or whether some bias has been introduced. In either case, the associations observed will provide valuable data for the further development of the app, not only in highlighting the conversations that fit naturally with each other and presumably provide the most support for users, but also in highlighting those that do not.

Comparison With Prior Work

The majority of users (76%) accessed the app for a period of less than 10 days with 62.7% accessing ChatPal for 1 day. While the remaining 24% accessed the app for more than a 10-day period, analysis of the number of unique days of usage shows that the app was used for a maximum of 19 unique days by an individual user with the average being 2 unique days. While the app was designed to be beneficial to all users' mental well-being, the results indicate that only a small proportion of participants became invested in using the app. While this lack of user retention is disappointing, it is not unusual. Benchmarking the ChatPal app against a study of 93 mobile apps, 59 of which specialized in mental health [31] shows that ChatPal performed better than average with a drop-off rate in the first 10 days of 77% compared with the average of 80% and a drop off between day 15 and day 30 of 7.1% compared with an average of 20% for other mental health apps. While this indicates that ChatPal has a better than average retention of users over the first 10 days, it also suggests that the remaining users may find benefit in using the app leading to a low drop off of users between days 15 and 30. A recent study [32] found that digital interventions for depression with human guidance yielded better results compared to digital interventions without any external guidance. Perhaps if chatbots such as ChatPal were used in conjunction with health professionals this may encourage usage and result in improved benefits to users, however further work would be needed to confirm this.

Discovering the types of users accessing chatbots and their patterns of usage is essential for the further development and targeting of the services provided by the app. K-means clustering was used to discover 3 different groups of users (abandoning users, frequent transient users, and sporadic users) based on 6 key features extracted from the log data. Further analysis found significant differences between the features for each group of app users.

Abandoning users, making up 87.2% of total users, generally accessed the app on 1 or 2 unique days with an equally low average tenure of 5.7 days. They had significantly less interactions with the app resulting in less moods being logged than invested users and participated in significantly less conversations. In contrast, frequent transient users and sporadic users were more generally invested users, accessing across more unique days with an average tenure of 50.5 days and 68.9 days, respectively. During this time, these users logged their mood on more occasions and generally interacted with the chatbot a lot more, accessing between 5 to 10 different conversations. This gives rise to the possibility that a small number of users found value in the support offered by the chatbot. These results are comparable with previous analyses of the initial trial of the prototype chatbot [16]. Further analysis is needed to explore the demographics of the users in each of these archetypes in order to understand why users belong to the archetype they are associated with and provide insight into users who repeatedly use specific features in order to understand if certain subgroups of users tend to favor certain features over others. The AI could then be used to direct users toward these features within the bot. This may also have the added benefit of increasing adherence and retention.

Policy and Practice Implications

We have included policy and practice implications based on the findings from this study, which may benefit others who are designing digital mental health technologies (Textbox 1).

Textbox 1. Recommendations for policy and practice.

- As the app was used by some during the night, conversations or exercises specifically for treating insomnia could be added to the available conversations and offered when users access the app during the nighttime. However, promoting screen time during the evening or middle of the night may not be appropriate.
- Given the drop-off rate of users over time, strategies could be used to improve user retention. For example, further development of the chatbot, to make it more spontaneous, “remembering” the last conversation or mood, more personalization and added daily reminders.
- Three main groups of app users were identified: “abandoning users,” “sporadic users,” and “frequent transient users.” While there will always be abandoning users due to the free nature of the app, the chatbot could have other engaging features such as peer communication and support to encourage use.
- The most used conversations within the chatbot were not necessarily the best-liked. The app could adapt based on user feedback, possibly changing the order of conversations offered based on ratings or allowing users access to the ratings in order to make informed decisions.
- Users’ moods were only collected at the start of each session. In order to gauge the effectiveness of the app, user moods could be collected at the start and end of each session.
- The World Health Organization-Five Well-being Index (WHO-5) well-being scale within the chatbot gave no feedback to the user. It may be more beneficial to display the score as a time series graph and provide insights to the user on their well-being over time.

Limitations

Despite the detailed event log data captured, tracking the length of individual sessions was difficult due to the lack of “end of session” variable. Almost all users chose not to select the “need to go” option in the chatbot which would have been recorded as “end of session,” and instead must have exited or closed the app which was not recorded in the logs. For this reason, user sessions were tracked on a daily basis. An additional variable to track the end of the session would be useful to explore individual user sessions rather than sessions per day. This omission may also have contributed to the low retention rates as the app was unable to provide little feedback to the user. With the addition of session data, user moods could be logged giving the app the ability to personalize its interactions with the user. For example, returning users could be asked “are you still feeling…” or “Hello again”. In addition, if users’ moods were asked at the start and end of each session, it would be possible to gauge changes in well-being that resulted from using the app. A bias may also have been introduced into the way users interacted with the app due to the rigid structure of the menus presented to the user. These may have caused some users to follow the list of available conversations in the order they are presented in the first instance before deciding which conversations provided the most support for their circumstances. Mood logs were asked each time the app was launched and, while these were used in the discovery of user archetypes, no further analysis was possible due to the following reasons. As users were only asked about their mood at the start of the session, there is nothing to compare the responses to. The addition of session data and requesting a user’s mood at the end of a session would have allowed comparisons to be made and facilitated the analysis of the effect of the app on the users’ mental well-being.

The app recorded interactions based on server time, not local time. This means that log data will be incorrect by 1 hour for Swedish users and 2 hours for Finish users. Not all users reported their country, thus no adjustments were made to account for differing geographical locations. As the WHO-5 scale questions were optional, there was no substantial data collected during the trial period.

Conclusions

The ChatPal app was developed as part of a research project into the use of chatbots to promote positive mental health and well-being. From the log data gathered, 3 main types of users accessed the chatbot: abandoning users, sporadic users, and frequent transient users. Due to the high numbers of abandoning users, it is difficult to evaluate the effectiveness of the app though analysis of the other 2 groups of invested users allows a glimpse into the benefits the app could bring to a user’s mental well-being. It is clear that some users returned to the app on several occasions although there is no evidence that this was linked to improvement in well-being. Improvements incorporated in future versions of the app suggested in this paper would provide data on participants’ moods at the beginning and end of each session and would provide evidence as to what effect the app has on users. Analysis of user transitions from conversation to conversation indicated that some dialogues may complement each other and provide targeted support to users although further analysis may be necessary. Future versions of the chatbot should be enhanced to make each interaction with the user more personalized, learning and adapting from the previous interactions with each user.

Acknowledgments

The authors would like to thank all the clients, participants, project members, supporters, and researchers at Ulster University, University of Eastern Finland, Norrbotten Association of Local Municipalities, Region Norrbotten, Luleå University of Technology, NHS Western Isles, Action Mental Health, and Munster Technological University for their involvement in this research. The ChatPal project is supported by Interreg Northern Periphery and Arctic Programme under grant number 345.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of available content in the ChatPal chatbot.

[PNG File , 307 KB - mhealth_v11i1e43052_app1.png]

Multimedia Appendix 2

Conversations by number of times accessed.

[PNG File , 121 KB - mhealth_v11i1e43052_app2.png]

References

1. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
2. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
3. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
4. Benavides-Vaello S, Strode A, Sheeran BC. Using technology in the delivery of mental health and substance abuse treatment in rural communities: a review. J Behav Health Serv Res 2013;40(1):111-120. [doi: 10.1007/s11414-012-9299-6] [Medline: 23093443]
5. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of Chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
6. Calvo RA, Dinakar K, Picard R, Christensen H, Torous J. Toward impactful collaborations on computing and mental health. J Med Internet Res 2018;20(2):e49 [FREE Full text] [doi: 10.2196/jmir.9021] [Medline: 29426812]
7. Jakobson G, Weissman M. Real-time telecommunication network management: extending event correlation with temporal constraints. In: Sethi AS, Faure-Vincent F, Raynaud Y, editors. Integrated Network Management IV: Proceedings of the Fourth International Symposium on Integrated Network Management. New York, NY: Springer US; 1995:290-301.
8. Mulvenna MD, Bond RR, Grigorash A, O'Neill S, Ryan A. HILDA - a health interaction log data analysis workflow to aid understanding of usage patterns and behaviours. 2018 Presented at: The 2nd Symposium on Social Interactions in Complex Intelligent Systems at Artificial Intelligence and Simulation of Behaviour Convention; April 4-6, 2018; Liverpool, UK URL: https://pure.ulster.ac.uk/ws/portalfiles/portal/11666886/MULVENNA-HILDA-SICIS-20180305.pdf
9. Potts C, Bond R, Ryan A, Mulvenna M, McCauley C, Laird E, et al. Ecological momentary assessment within a digital health intervention for reminiscence in persons with dementia and caregivers: user engagement study. JMIR mHealth uHealth 2020;8(7):e17120 [FREE Full text] [doi: 10.2196/17120] [Medline: 32420890]
10. Bond R, Moorhead A, Mulvenna M, O'Neill S, Potts C, Murphy N. Exploring temporal behaviour of app users completing ecological momentary assessments using mental health scales and mood logs. Behav Inf Technol 2019;38(10):1016-1027. [doi: 10.1080/0144929X.2019.1648553]
11. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
12. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
13. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth uHealth 2018;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
14. ChatPal. 2022. URL: https://chatpal.interreg-npa.eu/ [accessed 2023-06-08]
15. Dosovitsky G, Pineda BS, Jacobson NC, Chang C, Escoredo M, Bunge EL. Artificial intelligence chatbot for depression: descriptive study of usage. JMIR Form Res 2020;4(11):e17065 [FREE Full text] [doi: 10.2196/17065] [Medline: 33185563]
16. Potts C, Bond R, Mulvenna MD, Ennis E, Bickerdike A, Coughlan EK, et al. Insights and lessons learned from trialling a mental health chatbot in the wild. 2020 Presented at: 2021 IEEE Symposium on Computers and Communications (ISCC); September 5-8, 2021; Athens, Greece. [doi: 10.1109/ISCC53001.2021.9631395]
17. Seligman MEP. Flourish: A Visionary New Understanding of Happiness and Well-being. New York, NY: Simon and Schuster; 2011.
18. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]

19. Anaconda Software Distribution. Anaconda Documentation. 2020. URL: https://docs.anaconda.com/free/anaconda/index.html [accessed 2023-06-08]
20. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/MCSE.2007.55]
21. Waskom ML. Seaborn: statistical data visualization. J Open Source Softw 2021;6(60):3021 [FREE Full text] [doi: 10.21105/joss.03021]
22. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
24. Raschka S. MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. J Open Source Softw 2018;3(24):638 [FREE Full text] [doi: 10.21105/joss.00638]
25. Potts C, Lindström F, Bond R, Mulvenna M, Booth F, Ennis E, et al. A multilingual digital mental health and well-being Chatbot (ChatPal): pre-post multicenter intervention study. J Med Internet Res 2023 Jan 31. [doi: 10.2196/43051]
26. Oppenauer C, Burghardt J, Kaiser E, Riffer F, Sprung M. Psychological distress during the COVID-19 pandemic in patients with mental or physical diseases. Front Psychol 2021;12:703488 [FREE Full text] [doi: 10.3389/fpsyg.2021.703488] [Medline: 34475838]
27. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
28. de Castro Moreno CR, Conway SG, Assis M, Genta PR, Pachito DV, Tavares A, et al. COVID-19 pandemic is associated with increased sleep disturbances and mental health symptoms but not help-seeking: a cross-sectional nation-wide study. Sleep Sci 2022;15(1):1-7 [FREE Full text] [doi: 10.5935/1984-0063.20220027] [Medline: 35662970]
29. Ullrich PM, Lutgendorf SK. Journaling about stressful events: effects of cognitive processing and emotional expression. Ann Behav Med 2002;24(3):244-250 [FREE Full text] [doi: 10.1207/S15324796ABM2403_10] [Medline: 12173682]
30. Moore RJ, Arar R. Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework. New York, NY, USA: Association for Computing Machinery; 2019.
31. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
32. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]

Abbreviations

EMA: ecological momentary assessment
WHO-5: World Health Organization—Five Well-being Index

Edited by G Eysenbach, L Buis; submitted 28.09.22; peer-reviewed by J Bullock; comments to author 24.10.22; revised version received 20.12.22; accepted 23.01.23; published 06.07.23.

Please cite as:
Booth F, Potts C, Bond R, Mulvenna M, Kostenius C, Dhanapala I, Vakaloudis A, Cahill B, Kuosmanen L, Ennis E
A Mental Health and Well-Being Chatbot: User Event Log Analysis
JMIR Mhealth Uhealth 2023;11:e43052
URL: https://mhealth.jmir.org/2023/1/e43052
doi:10.2196/43052
PMID:37410539

Original Paper

Patterns and Predictors of Engagement With Digital Self-Monitoring During the Maintenance Phase of a Behavioral Weight Loss Program: Quantitative Study

Nicole Crane[1], MS; Charlotte Hagerman[1], PhD; Olivia Horgan[1], BS; Meghan Butryn[1], PhD

Center for Weight, Eating, and Lifestyle Science, Department of Psychological and Brain Sciences, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Nicole Crane, MS
Center for Weight, Eating, and Lifestyle Science
Department of Psychological and Brain Sciences
Drexel University
3141 Chestnut Street
Stratton Hall
Philadelphia, PA, 19104
United States
Phone: 1 724 740 8648
Email: nvt24@drexel.edu

Abstract

Background: Long-term self-monitoring (SM) of weight, diet, and exercise is commonly recommended by behavioral weight loss (BWL) treatments. However, sustained SM engagement is notoriously challenging; therefore, more must be learned about patterns of engagement with digital SM tools during weight loss maintenance (WLM). In addition, insight into characteristics that may influence SM engagement could inform tailored approaches for participants at risk for poor adherence.

Objective: This study explored patterns of digital SM of weight, diet, and exercise during WLM (aim 1) and examined timing, patterns, and rates of disengagement and reengagement (aim 2). This study also assessed relationships between individual-level factors (weight-related information avoidance and weight bias internalization) and SM engagement (aim 3).

Methods: Participants were 72 adults enrolled in a BWL program consisting of a 3-month period of weekly treatment designed to induce weight loss (phase I), followed by a 9-month period of less frequent contact to promote WLM (phase II). Participants were prescribed daily digital SM of weight, diet, and exercise. At baseline, self-report measures assessed weight-related information avoidance and weight bias internalization. SM adherence was objectively measured with the days per month that participants tracked weight, diet, and exercise. Repeated-measures ANOVA examined differences in adherence across SM targets. Multilevel modeling examined changes in adherence across phase II. Relationships between individual-level variables and SM adherence were assessed with Pearson correlations, 2-tailed independent samples *t* tests, and multilevel modeling.

Results: During WLM, consistently high rates of SM (≥50% of the days in each month) were observed for 61% (44/72) of the participants for exercise, 40% (29/72) of the participants for weight, and 21% (15/72) of the participants for diet. Adherence for SM of exercise was higher than that for weight or diet (P<.001). Adherence decreased over time for all SM targets throughout phase II (P<.001), but SM of exercise dropped off later in WLM (mean 10.07, SD 2.83 months) than SM of weight (mean 7.92, SD 3.23 months) or diet (mean 7.58, SD 2.92 months; P<.001). Among participants with a period of low SM adherence (ie, <50% of the days in a month), only 33% (17/51 for weight, 19/57 for diet) to 46% (13/28 for exercise) subsequently had ≥1 months with high adherence. High weight-related information avoidance predicted a faster rate of decrease in dietary SM (P<.001). Participants with high weight bias internalization had the highest rates of weight SM (P=.03).

Conclusions: Participants in BWL programs have low adherence to the recommendation to sustain daily SM during WLM, particularly for SM of diet and weight. Weight-related information avoidance and weight bias internalization may be relevant indicators for SM engagement. Interventions may benefit from innovative strategies that target participants at key moments of risk for disengagement.

(JMIR Mhealth Uhealth 2023;11:e45057) doi:10.2196/45057

KEYWORDS
weight loss; digital technology; diet; exercise; behavior change; mobile phone

Introduction

Background

Self-monitoring (SM) of weight, diet, and exercise is a cornerstone of behavioral weight loss (BWL) treatment [1], and daily SM of these key weight control behaviors is associated with better weight loss and maintenance [2-4]. Nevertheless, despite its importance, rates of engagement with SM are modest and tend to decrease over time, particularly during weight loss maintenance (WLM) [5,6]. With the rise of digital devices for SM (eg, food tracking apps, digital scales, and Fitbit activity trackers), technology is routinely being incorporated into BWL programs [7,8]. Digital devices may facilitate SM adherence by decreasing burden via time-saving features (eg, nutrition databases and saving frequent foods), portability for real-time monitoring, and passive recording of behavior (ie, Fitbit wristwatch for active minutes) [7]. Research shows that adherence to SM via digital format is higher than that via traditional methods [9], likely for these reasons, and that digital SM facilitates calorie reduction and weight loss [10].

There is a growing body of literature examining rates of engagement with digital SM formats to understand how participants in BWL programs use these tools. A recent systematic review analyzed randomized controlled trials of BWL interventions that incorporated digital SM (at least 12 weeks of treatment and 6-month outcome assessments) and found that 48% (23/48) of studies prescribed daily SM of weight, 69% (37/54) prescribed daily SM of diet, and 71% (39/55) prescribed daily SM of exercise. Across the intervention periods (median 6, range 3-24 months), 58% of studies achieved rates of ≥50% for SM engagement, and only 9% reached ≥75% engagement [7]. SM rates decreased over time, with only 45% of the studies showing SM rates of ≥50% by 6 months (n=33), and 38% had rates of ≥50% by ≥12 months (n=8). Engagement was the highest for SM of weight, followed by SM of diet and exercise. Synthesis of these results suggests that, although higher than rates of engagement for traditional SM methods, rates of engagement with digital SM tools are modest, and difficulties with sustaining adherence over time remain prevalent. Within this body of work, however, few prior studies have included assessment beyond 6 months; therefore, the dynamics of digital SM during the WLM phase are still unclear. Further work is needed to conceptualize digital SM engagement during this critical period (eg, 6 months and beyond when SM rates decline).

To explore this question, this project is focused on SM behavior during the WLM phase of a previously published clinical trial that assessed whether providing weight loss coaches with access to participants' digital SM data enhances outcomes during lifestyle modification (LM) [11]. A previous publication from this parent study found that participants enrolled in a WLM intervention were more likely to self-monitor weight and eating behavior when coaches remotely monitored their data and used the data to drive treatment contacts (SMS text messages and telephone calls) versus when coaches did not have access to their data. Data sharing was also associated with less weight regain over time (although total weight loss did not differ by condition), and the frequency of dietary SM mediated the effect of treatment condition on weight loss [11]. This prior report supports the central role of SM for weight loss and maintenance and suggests that providing coaches with access to data via digital SM tools may maximize the efficacy of these tools for long-term weight control. Nevertheless, no previous analyses have been conducted to understand the nuanced patterns of SM behavior during the WLM phase when participant adherence becomes more variable owing to the difficulty of sustaining behavior change over long periods.

Furthermore, the majority of prior work on digital SM tools attempts to understand their use through the lens of percentage of adherence based on prescribed frequency, with most studies reporting the mean percentage of days that participants successfully self-monitor or the percentage of participants who maintained a certain level of SM [7]. A more nuanced exploration of SM would be helpful, including the timing of *disengagement* and rates of complete disengagement (ie, 0 days tracked) as well as how patterns of engagement vary across participants; for example, Robertson et al [12] used profile analyses to assess different patterns of SM engagement. The results showed 4 distinct profiles of use for digital SM tools among participants enrolled in a 6-month workplace weight loss intervention: minimal users (29% of the sample), activity trackers (55%), dedicated all-around users (11%), and dedicated all-around users with exceptional food logging (5%) [12]. Weight outcomes were only substantially better among the dedicated all-around users with exceptional food logging, aligning with other work that highlights the importance of dietary SM for weight loss [9,13]. Another study focused on self-weighing behavior during a 12-month BWL treatment and found 3 profiles: high/consistent (75% of the sample; SM of weight >6 days per week regularly), moderate/declined (16.2%; SM of weight 4-5 days per week, then declined to 2 days gradually) and minimal/declined (8.8%; SM of weight 5-6 days per week, then declined to 0 days suddenly), with the high/consistent group losing more weight at 6 and 12 months [14]. These findings show intriguing preliminary evidence for between-persons variation in patterns of engagement with digital SM tools, which has implications for weight loss success and needs further research, particularly during the WLM period.

Another gap in the literature on digital SM tools for weight control surrounds predictors and moderators of engagement. Little is known about what individual-level variables relate to strong adherence to SM prescriptions in BWL programs. Some previous work has found that higher initial weight loss [15], enhanced social support [16,17], and heightened binge eating severity [18] were associated with higher rates of SM during BWL programs (uncontrolled eating and emotional eating were explored as predictors of SM engagement but were not substantially related) [18]. There is also theoretical support for the idea that previous SM behavior is likely a strong predictor of consistent long-term use of SM, given that past behavior is a strong indicator of future behavior [14,19]. To our knowledge, no studies have explored individual-level predictors of long-term use of digital SM tools during WLM.

Weight-related information avoidance, which is the tendency to prevent or delay acquisition of potentially unwanted weight-related information [20], has strong theoretical support for a relationship with digital SM use. Digital SM tools provide BWL program participants immediate detailed information on progress and goal attainment, which should increase awareness of current eating or exercise patterns [16,21]. This may be differentially helpful (vs distressing) for SM engagement based on an individual's level of weight-related information avoidance. Those with low weight-related information avoidance may be eager to engage with digital SM data and find value in reflecting on patterns of behavior [22]. However, for those with high weight-related information avoidance, viewing SM data may be distressing and reduce willingness to engage in future SM [23]. Research on health information avoidance suggests that people avoid health information for three reasons as follows: (1) it may cause unpleasant emotions (ie, guilt and shame), (2) it may dictate undesired action (ie, seeing weight gain on the scale dictates a reduction in calories and change in eating habits), and (3) it may dictate a change in beliefs (ie, seeing the calories associated with one's favorite menu item at a restaurant may dictate changing beliefs about the feasibility of incorporating it into a weight loss diet) [20]. For all these reasons, the data provided via digital SM tools have the potential to be highly upsetting for those with high health information avoidance. This can have long-term implications for SM engagement because avoidance is likely to continue owing to negative reinforcement (ie, avoidance decreases distress associated with confronting weight-related information). Previous work from the initial BWL phase of the parent study for this analysis shows a relationship between higher weight-related information avoidance and poorer SM of exercise and weight but not diet [24]. Other studies have found that confronting information that can be perceived as a failure (eg, high calorie intake and weight gain) is associated with a higher likelihood of avoiding subsequent SM (eg, self-weighing) [25,26], supporting a relationship between health information avoidance and SM engagement. Further work is needed to replicate these results and explore the relationships during WLM to see whether health information avoidance continues to predict decreased SM engagement as time progresses.

There is also a theoretical and empirical rationale for a relationship between weight bias internalization and digital SM engagement during WLM. Weight bias internalization happens when individuals are aware of negative stereotypes associated with weight, apply these stereotypes to themselves, and engage in self-critical dialogue because of their body size [27]. This negative self-concept (eg, being lazy and lacking willpower) can be associated with lower confidence and self-efficacy [28], which is a consistent predictor of poorer engagement in weight control behaviors during LM [29] and lower weight loss success [30,31]. As participants view data from their digital SM tools and reflect on progress, the lack of goal attainment may contribute to feelings of failure or frustration. For those with high internalized weight bias, these perceived failures may be associated with more intense experiences of shame, guilt, and self-blame than for those with low internalized weight bias, further worsening their confidence and self-efficacy and leading to decreased engagement in SM [28,32]. In addition, in some cases (including the parent study), digital SM information is addressed by BWL coaches who monitor participant progress and provide personalized feedback (eg, SMS text messages) based on data [11]. Although this is meant to enhance supportive accountability, allow coaches to provide more tailored feedback, and increase motivation, this type of surveillance may deter those with high weight bias internalization from engaging in digital SM because they may have heightened sensitivity to the shame surrounding potential negative evaluations that may occur while others are monitoring their data [33,34]. A small body of prior work shows a relationship between high internalized weight bias and lower rates of SM engagement among those attempting weight loss [27,35], but this has not been examined in WLM.

Objectives

In line with precision medicine initiatives [36], insight into individual characteristics (weight-related information avoidance and weight bias internalization) that influence SM engagement will help to drive more tailored intervention approaches for those who may be at risk for poor adherence to this key weight control behavior. This study aimed to address current gaps in the literature by exploring patterns of adherence to daily SM of weight, diet, and exercise via digital tools during the WLM phase of a BWL program (aim 1). Among BWL program participants with low SM adherence, this study also examined timing and patterns of disengagement and explored the extent to which participants reengaged with SM after low rates of earlier engagement (aim 2). Finally, this study also sought to determine how individual-level factors (weight-related information avoidance and weight bias internalization) were associated with SM of weight, exercise, and diet during WLM (aim 3).

Methods

Overview

This study is a secondary analysis of data from a completed randomized controlled trial (ClinicalTrials.gov NCT03337139) [11] assessing whether coach contact with access to digital SM data enhanced WLM outcomes compared with coach contact without access to digital SM data. Participants (N=77) were adults (aged 18-70 years) with overweight or obesity (BMI 25-45 kg/m^2) who had access to a smartphone and internet and could safely engage in exercise. The exclusion criteria of the parent study included a medical or psychiatric condition that posed a risk for program adherence or safety; pregnancy or plan to become pregnant or move from study area; the use of pacemaker; a history of bariatric surgery; recent start of, or change to, a medication that can affect weight; and weight loss of ≥10% in the past 3 months.

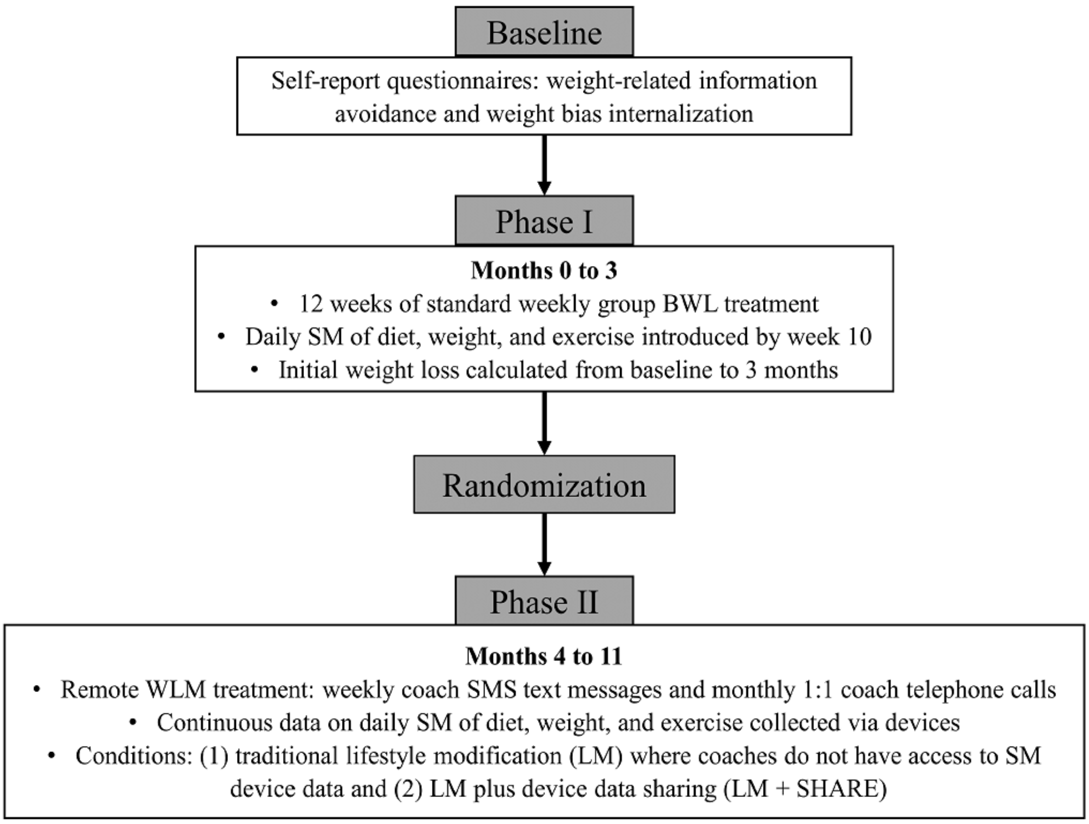
Ethics Approval and Informed Consent

The parent study was approved by the Drexel University Institutional Review Board (institutional review board protocol number: 1611004954), and all participants provided written informed consent before participation.

Study Flow and Description of BWL Treatment

For a summary of the flow of the study, refer to Figure 1. During months 0 to 3 (phase I), all participants received 12 weeks of standard weekly group BWL treatment, which included tailored calorie goals, traditional behavioral skills adapted from the Look AHEAD (Action for Health in Diabetes) and Diabetes Prevention programs [37,38] (eg, goal setting and problem-solving), and progressive exercise goals of up to 250 minutes per week. Coaches had training in BWL treatment and degrees in psychology or a related field. All participants were provided with digital tools to track weight, exercise, and diet: Yunmai smart scale (weight), Fitbit Flex (exercise; Fitbit Inc), and Fitbit app (diet). Participants were instructed to (1) weigh themselves weekly during weeks 1 to 10 and then weigh daily, (2) wear their Fitbit Flex daily to monitor active minutes, and (3) log all food and drink intake daily. Coaches did not have access to device data during phase I.

Figure 1. Summary of study flow, including data collection and treatment details. BWL: behavioral weight loss; LM: lifestyle modification; LM+SHARE: lifestyle modification plus device data sharing; SM: self-monitoring.



At the end of phase I, participants were randomly assigned (matched for phase I weight loss) to 1 of 2 remote WLM treatment conditions for months 4 to 11. Both conditions included weekly SMS text messages with a coach and monthly one-on-one coach telephone calls (15 minutes) that reviewed ≥1 core behavioral skills taught during group sessions. All participants were prescribed continued daily SM of weight, diet, and exercise throughout phase II. Monthly coach calls were focused on positive reinforcement and self-reflection when participants were succeeding with behavior changes and weight loss goals and focused on problem-solving barriers or fostering motivation when participants were struggling with goal attainment. The conditions differed in terms of coaches' access to participants' digital SM data. In the standard LM condition, coaches did not have access to SM device data. Instead, participants self-reported goal progress during monthly calls, and coaches used that self-report to drive discussion of behavioral skills. In the LM condition, weekly SMS text messages were standardized across participants and were not personalized by a coach. In the LM plus device data sharing (LM+SHARE) condition, coaches viewed participants' SM outcomes on a web-based portal and used these data to personalize telephone calls and SMS text messages. Coaches were trained in how to use data to enhance a sense of supportive accountability and to use data to drive more tailored personalized feedback and goal setting. SM adherence was a key discussion point during monthly coach telephone calls, given the critical role of SM behavior for WLM. Whether coaches were viewing participants' SM data themselves (LM+SHARE condition) or reacting to participants' self-reported SM adherence (LM condition), they were trained to identify barriers to success, facilitate effective problem-solving and tailored goal setting, and engage in motivational enhancement when adherence was poor. Coaches handled the lack of goal achievement and lack of weight loss progress in similar ways. The results from the parent study indicated that participants in the LM+SHARE condition had higher rates of weight and dietary SM [11]; therefore, treatment condition will be controlled for in analyses.

Measures

Data Collection

Assessments were completed at baseline, month 3 (the end of phase I and beginning of phase II), month 6, and month 12 (the end of phase II). This study used self-report questionnaire data from baseline assessments as well as continuous SM data collected daily from participants' devices throughout phase II (months 4-11). The participants' SM devices (Fitbit Flex, wireless scale, and Fitbit dietary SM app) automatically uploaded data remotely to a research portal. Thus, once SM data were recorded by the participants' SM devices, there was no additional burden on participants to transfer these data to the research team.

Weight-Related Information Avoidance

An adapted version of the Information Avoidance Scale (IAS) [39] was created for the parent study to assess the level of weight-related information avoidance surrounding key weight control behaviors. The 10 items on this self-report measure include statements about attitudes or tendencies to seek out versus avoid information about calorie intake, physical activity, and weight. At baseline, participants responded to each statement on a 7-point Likert scale (ranging from 1=*strongly disagree* to 7=*strongly agree*). Total scores were calculated as the average across the 10 items. The measure created showed strong internal consistency (Cronbach α=.85) [24].

Weight Bias Internalization

The Weight Bias Internalization Scale (WBIS) [40] is an 11-item self-report measure that assesses the extent to which the respondent believes that negative stereotypes or self-statements about weight apply to them. At baseline only, participants were presented with certain statements (eg, "As an overweight person, I feel that I am just as competent as anyone") and asked to rate their agreement on a 7-point Likert scale (ranging from 1=*strongly disagree* to 7=*strongly agree*). Total scores are calculated as the average rating across the 11 items. The questionnaire has high internal consistency and construct validity [40].

SM Adherence (Phase I and Phase II)

For all months 1 to 11, the percentage of days per month that participants successfully self-monitored weight, diet, and exercise was calculated to create an average monthly adherence score for each month. For each participant, average overall phase I (months 1-3) and phase II (months 4-11) adherence scores were also calculated for each SM target (separate variables for phase I vs phase II). A valid day of exercise SM was defined as logging ≥500 steps, and a valid day of dietary SM was defined as logging ≥5 foods, both of which have precedent in the literature [24,41,42].

Patterns of SM Engagement and Adherence

Several metrics were calculated to understand patterns of SM engagement throughout phase II (calculated separately for SM of exercise, weight, and diet). A cutoff of 50% of days was chosen as a threshold to define *low* versus *high* adherence to SM and was selected for several reasons. First, previous work has used 50% as a cutoff for defining SM adherence versus nonadherence [43], and moderate adherence has been defined as 12 to 16 days per month (approximately 50%) [14]. In addition, the systematic review of digital SM within BWL interventions showed that, by months 6 to 12 of the intervention period, engagement rates of ≥50% were achieved in only a minority of studies (38%-45%) [7], suggesting that adherence of >50% is relatively difficult to achieve. Finally, within this sample, exploratory analyses were conducted to ensure that a 50% cutoff indicated a meaningful shift in adherence rates rather than adherence hovering right around 50% (eg, changing from 52% to 48%), which would not necessarily be clinically relevant. For each participant, the first month in which average adherence dropped to <50% (separate for each SM target) was identified as their drop-off month. For SM of weight, diet, and exercise, 2-tailed paired sample *t* tests confirmed that average monthly adherence in the month before drop-off was significantly higher than adherence during the drop-off month, and average monthly adherence in the month after drop-off was significantly lower. This further supports the use of 50% as a clinically relevant metric for high versus low adherence because adherence meaningfully shifts before and after reaching this cutoff.

The percentage of participants who maintained high adherence (≥50%) throughout all of phase II was calculated, as well as the typical time during phase II where low adherence first occurs (ie, early in WLM during months 4-7 or late in WLM during months 8-11). For those participants whose adherence dropped to <50%, rates of reengagement were also calculated by establishing the number of participants who successfully rebounded to adherence rates of ≥50% at some point during the rest of phase II. The total number of months that participants exhibited complete disengagement (0% adherence) and the number of consecutive months that participants completely disengaged during phase II were also calculated for each SM target. Adherence to SM during phase II was also compared with participants' phase I SM adherence. The month in which adherence for each SM target dropped by ≥10% compared with average adherence during phase I was identified, as well as the number of months that participants maintained adherence equal to the average of phase I. In addition, we investigated whether participants' adherence successfully rebounded back to phase I levels once it dropped in phase II.

Data Analysis

All data analyses were conducted in SPSS (version 28; IBM Corp) and SAS (version 9.4; SAS Institute Inc) software, and α was set to .05. All data were screened before statistical testing to assess for outliers and normality. The distributions for weight-related information avoidance scores and for SM adherence during phase II were nonnormally distributed.

Although most parametric tests are robust to skewness [44], nonparametric tests and bootstrapping were conducted for analyses using these variables as sensitivity analyses. Given the results from the parent study showing differences in weight and dietary SM between the LM and LM+SHARE conditions [11], our results are reported separately by condition where appropriate to help illustrate any differences between the groups.

For *aim 1*, descriptive statistics for adherence to SM of weight, diet, and exercise throughout phase II were calculated, within each month and across all of phase II (months 4-11). The percentages of participants who maintained high adherence and those who maintained low adherence were also calculated. Repeated-measures ANOVA (robust to assumptions of nonnormality [45]) assessed for differences in average phase II adherence rates across SM of diet versus weight versus exercise. Multilevel modeling was used to examine changes in adherence rates for each type of SM time in phase II (ie, month in study; level 1) while accounting for between-person variance (level 2). Analyses also controlled for study condition (level 2), and the time × condition interaction was explored too. Chi-square likelihood tests examined whether the inclusion of random participant slope effects improved model fit. The results of the best-fitting model are presented.

For *aim 2*, all aforementioned SM engagement variables (eg, month adherence dropped to <50% and month adherence dropped to <phase I average adherence) were calculated for each SM target, and descriptive statistics were calculated. Repeated-measures ANOVA assessed for differences in average month where adherence dropped to <50% and <phase I average across SM of diet, weight, and exercise (separate models).

For *aim 3*, Pearson correlations were used to assess relationships between phase II SM adherence for weight, diet, and exercise and weight-related information avoidance and weight bias internalization (bootstrapping with 1000 samples was conducted as sensitivity analysis to confirm results in nonnormally distributed variables). Two-tailed independent samples *t* tests (Mann-Whitney *U* tests for nonnormally distributed variables) were used to assess group differences in weight-related information avoidance and weight bias internalization between participants with low phase II adherence and those with high phase II adherence on each SM target. Using iterative multilevel model building procedures, we also tested cross-level interactions between the hypothesized person-level predictors (ie, weight bias internalization and weight-related information avoidance; level 2) and time (ie, month in study; level 1) for each type of SM adherence. Between-person variables were grand-mean centered. Cross-level interactions were compared with random slope models for fit. For significant cross-level interactions, simple slopes were calculated and graphed to depict SM adherence across time at the mean of, as well as 1 SD above and 1 SD below, the between-person predictor, which allows for better visualization of interaction effects and in a way that is more interpretable and clinically meaningful [44].

Results

Descriptive Statistics

Of the 77 participants, 72 (94%) provided phase II data and were included in these analyses. Participants were on average aged 51.27 (SD 13.47) years, predominantly female (58/72, 81%), and non-Hispanic/Latino (69/72, 96%). Approximately half of the participants (37/72, 51%) identified as White, 38% (27/72) as Black/African American, 7% (5/72) as other or >1 race, 3% (2/72) as Asian, and 1% (1/72) as American Indian/Alaska Native. On average, participants lost 5.89% (SD 4.31%) of their body weight during phase I. Higher percentage of weight loss during phase I was correlated with higher engagement in SM of weight (r=−0.28; P=.02), diet (r=−0.41; P<.001), and exercise (r=−0.26; P=.03) in phase I. The relationships between previous SM behavior (during phase I) and SM engagement during phase II can be seen in Table 1.

Table 1. The correlation matrix of individual-level variables and phase II adherence for each self-monitoring (SM) target.

Variable, mean (SD)	Average adherence to SM of weight in phase II	Average adherence to SM of diet in phase II	Average adherence to SM of exercise in phase II	Baseline IAS[a] Score	Baseline WBIS[b] Score	Average adherence to SM of weight in phase I	Average adherence to SM of diet in phase I	Average adherence to SM of exercise in phase I
Average adherence to SM of weight in phase II, 53.2% (3%)								
r	—[c]	—	—	—	—	—	—	—
P value	—	—	—	—	—	—	—	—
Average adherence to SM of diet in phase II, 49.34% (2.9%)								
r	*0.638* [d]	—	—	—	—	—	—	—
P value	*<.001*	—	—	—	—	—	—	—
Average adherence to SM of exercise in phase II, 80.01% (2.3%)								
r	*0.581*	*0.604*	—	—	—	—	—	—
P value	*<.001*	*<.001*	—	—	—	—	—	—
Baseline IAS Score, 2.13 (0.96)								
r	0.048	−0.108	−0.067	—	—	—	—	—
P value	.68	.36	.57	—	—	—	—	—
Baseline WBIS Score, 3.58 (1.09)								
r	0.001	−0.085	0.073	0.228	—	—	—	—
P value	.99	.48	.55	.06	—	—	—	—
Average adherence to SM of weight in phase I, 88.25% (1.5%)								
r	*0.443*	*0.257*	*0.381*	*−0.240*	0.094	—	—	—
P value	*<.001*	*.03*	*<.001*	*.04*	.44	—	—	—
Average adherence to SM of diet in phase I, 86.63% (1.5%)								
r	*0.337*	*0.543*	*0.592*	*−0.203*	0.103	*−0.537*	—	—
P value	*.004*	*<.001*	*<.001*	*.09*	.40	*<.001*	—	—
Average adherence to SM of exercise in phase I, 94.13% (1.4%)								
r	*0.320*	0.213	*0.421*	*−0.316*	−0.026	*−0.765*	*−0.587*	—
P value	*.006*	.07	*<.001*	*.007*	.83	*<.001*	*<.001*	—

[a]IAS: Information Avoidance Scale.

[b]WBIS: Weight Bias Internalization Scale.

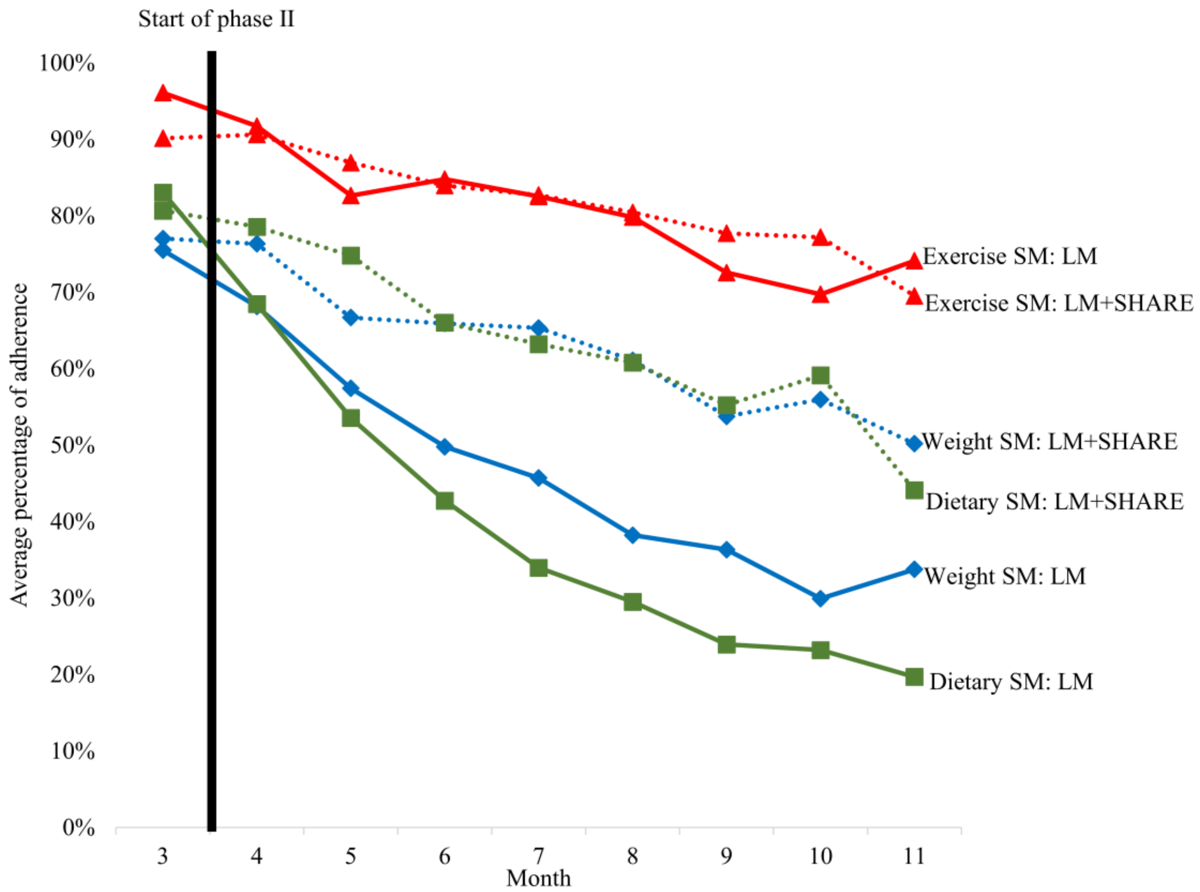
[c]Not applicable.

[d]Italics denotes significance (meeting threshold of *P*<.05).

Aim 1

At the end of phase I (month 3), 86% (62/72) of the participants had high adherence (≥50%) to SM of weight, 88% (63/72) had high adherence to SM of diet, and 97% (70/72) had high adherence to SM of exercise, indicating that most of the participants were still actively engaged with SM at the end of month 3 and presumably entered phase II with goals to maintain that behavior. During the WLM phase, consistently high rates (≥50% for every month) of SM were observed for 61% (44/72) of the participants for exercise, 40% (29/72) of the participants for weight, and 21% (15/72) of the participants for diet. Throughout phase II, the average percentage of adherence for SM of exercise (mean 80.01%, SD 2.3%) was significantly higher than the percentage of adherence to self-weighing (mean 53.2%, SD 3%) or food logging (mean 49.34%, SD 2.9%; $F_{2,142}$=64.95; *P*<.001; η^2_p=0.48). Adherence to SM of weight and diet were not significantly different (*P*=.63). Only 13% (9/72) of the participants consistently exhibited ≥50% adherence on all 3 SM targets. The best-fitting multilevel models examining the effect of time on all types of SM adherence retained a random slope. The average rates of SM adherence for weight (*b*=−0.05, SE 0.01; t_{71}=−8.75; *P*<.001), diet (*b*=−0.06, SE 0.00; t_{71}=−12.30; *P*<.001), and exercise (*b*=−0.03, SE 0.00; t_{71}=−6.53; *P*<.001) significantly decreased over time throughout phase II (months 4-11) when controlling for study condition. Refer to Figure 2 for a depiction of monthly adherence for each type of SM across time for the LM versus LM+SHARE conditions. The models tested a significant time × condition interaction, but none of the interactions were significant (weight SM: *P*=.16; dietary SM: *P*=.13; and exercise SM: *P*=.71).

Figure 2. Average percentage adherence by month for self-monitoring (SM) of weight, diet, and exercise separated by condition (lifestyle modification [LM] vs LM plus device data sharing [LM+SHARE]); this includes SM adherence at the end of phase I (month 3) and throughout phase II (months 4-11).



Aim 2

Average Time to <50% Adherence by SM Type

Average adherence dropped to <50% at 10.07 (SD 2.83) months for SM of exercise, at 7.92 (SD 3.23) months for SM of weight, and at 7.58 (SD 2.92) months for SM of diet. The average month of drop-off for adherence to SM of exercise was significantly later in WLM than drop-off for adherence to SM of weight or diet ($F_{2,142}$=33.22; P<.001; η^2_p=0.32), but they did not differ from each other (P=.97). In the LM condition, average adherence dropped to <50% at 9.86 (SD 2.92) months for SM of exercise, at 7.03 (SD 2.95) months for SM of weight, and at 6.31 (SD 2.36) months for SM of diet. In the LM+SHARE condition, average adherence dropped to <50% at 10.27 (SD 2.76) months for SM of exercise, at 8.76 (SD 3.30) months for SM of weight, and at 8.78 (SD 2.92) months for SM of diet. Among those with low engagement at some point during phase II, the majority disengaged with SM early (months 4-7) rather than late (42/57, 74%, disengaged early for SM of diet; 35/51, 69%, disengaged early for SM of weight; and 16/28, 57%, disengaged early for SM of exercise).

Average Time to Drop of ≥10% From Original Adherence by SM Type

When comparing participants' SM adherence during phase II to their phase I average adherence, adherence dropped by ≥10% compared with phase I adherence at 7.97 (SD 3.16) months for SM of exercise, at 5.88 (SD 2.63) months for SM of weight, and at 5.82 (SD 2.52) months for SM of diet. This occurred significantly later for SM of exercise than for SM of diet or weight ($F_{2,142}$=19.53; P<.001; η^2_p=0.22). Throughout WLM, participants achieved rates of adherence that were at, or above, their phase I average during more months for SM of exercise (mean 5.35, SD 2.49 months) than for SM of weight (mean 2.42, SD 2.69 months) and diet (mean 2.35, SD 2.56 months; $F_{2,142}$=56.73; P<.001; η^2_p=0.44).

Reengagement

Analyses examined the likelihood of participants returning to high adherence (ie, ≥50% of the days in any month) after a period of low adherence (ie, <50% of the days in a month). For SM of exercise, 46% (13/28) of the participants rebounded back to high adherence, whereas only one-third of the participants rebounded for SM of weight or diet (17/51, 33%, for weight and 19/57, 33%, for diet). Among those who successfully reengaged with SM of weight, the first month of rebounded rates of ≥50% tended to be month 7.94 (SD 2.05) compared with month 8.16 (SD 1.80) for SM of diet and month 7.15 (SD 2.19) for SM of exercise. When rates of adherence fell by ≥10% below the phase I average, only 30% (19/64) of the participants went on to achieve weight SM adherence rates at or above phase I levels compared with 29% (19/65) of the participants for dietary SM and 69% (36/52) of the participants for exercise SM. When rates dropped below the phase I average, those who reengaged tended to do so at 7.53 (SD 1.84) months for weight SM, at 6.53 (SD 1.65) months for dietary SM, and at 7.56 (SD 1.86) months for exercise SM.

Patterns of Complete Disengagement (0% Adherence)

When looking at complete disengagement (0% monthly adherence), 43% (31/72) of the participants had at least 1 full month of complete disengagement from SM of diet, and 32% (23/72) had at least 1 full month with complete disengagement from SM of weight, whereas only 19% (14/72) totally disengaged from SM of exercise for a full month. Participants who completely disengaged from self-weighing did so for an average of 2.96 (SD 1.52) months. Those who completely disengaged from SM of diet did so for an average of 3.42 (SD 2.08) months, and those who disengaged from SM of exercise did so for an average of 3.00 (SD 1.66) months. For all 3 SM targets, the months of total disengagement tended to occur consecutively for most of the participants (15/31, 48% to 9/14, 64%), rather than as a pattern where adherence increased and then decreased back down to zero.

Aim 3

Weight SM

The patterns of the Pearson correlation analyses with bootstrapping and those without bootstrapping remained the same; therefore, for ease of interpretation, the results of Pearson correlations without bootstrapping are reported (Table 1). Phase II SM of weight was significantly correlated with past SM behavior during phase I (weight: *r*=0.44; *P*<.001; diet: *r*=0.34; *P*=.004; and exercise: *r*=0.32; *P*=.006) such that participants who had higher engagement on any of the SM targets throughout phase I engaged in more self-weighing in phase II. The average adherence to SM of weight throughout phase II was not correlated with baseline weight-related information avoidance or weight bias internalization.

When dichotomizing the sample into 2 groups based on self-weighing adherence (those who maintained high adherence to weight SM throughout all of months 4-11 and those who did not; Table 2), individuals who maintained consistently high self-weighing adherence in months 4 to 11 had higher baseline weight bias internalization scores than those who did not maintain high adherence to self-weighing (*U*=700.50*; P*=.03).

Table 2. Results of group comparisons (high vs low self-monitoring [SM] adherence) on baseline levels of weight-related information avoidance and weight bias internalization.

	Weight SM		Dietary SM				Exercise SM			
	High adherence[a], median[b]	Low adherence, median	High adherence		Low adherence		High adherence		Low adherence	
			Median	Mean (SD)[c]	Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)
Baseline IAS[d] score	2.11	1.80	1.80	N/A[e]	2.10	N/A	2.10	N/A	1.80	N/A
Baseline WBIS[f] score	*4.27* [g]	3.36	N/A	3.42 (1.21)	N/A	3.62 (1.07)	N/A	3.75 (1.18)	N/A	3.30 (0.89)

[a]The high-adherence group maintained rates of ≥50% throughout all of months 4 to 11.

[b]Mann-Whitney *U* tests were used for nonnormally distributed variables, and medians are reported (due to skewness).

[c] Two-tailed independent samples *t* tests were used for normally distributed variables and mean (SD) values are reported.

[d]IAS: Information Avoidance Scale.

[e]N/A: not applicable.

[f]WBIS: Weight Bias Internalization Scale.

[g]*P*=.03 (significant difference from low-adherence group).

Cross-level interaction models examined whether baseline weight-related information avoidance and weight bias internalization moderated the influence of time on weight SM. Cross-level interaction models were not significant (*P*=.16 and *P*=.29, respectively) and did not improve model fit compared with the random slope models examining the influence of time on weight SM tested in aim 1.

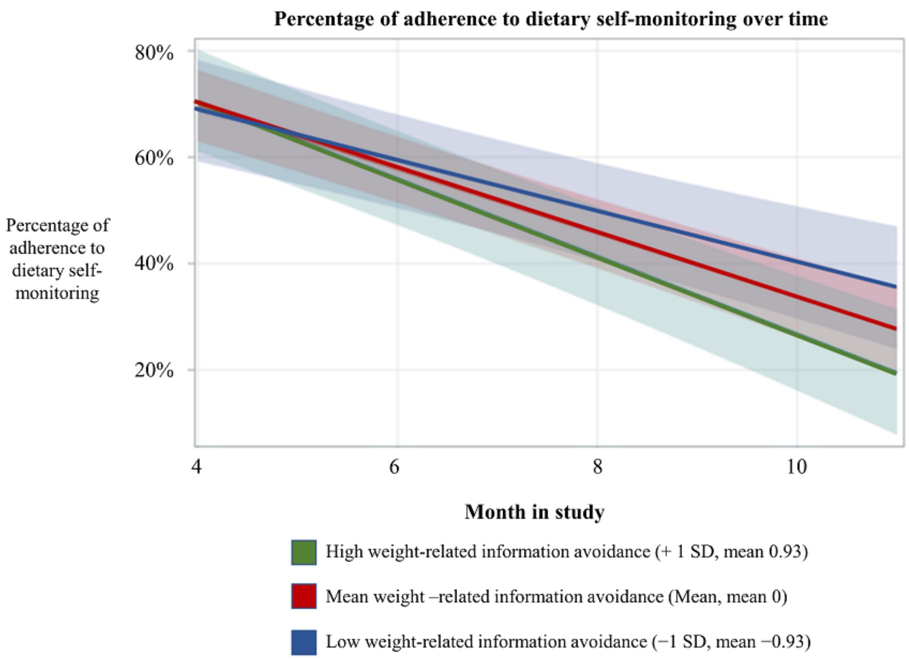
Dietary SM

As seen in Table 1, average phase II adherence for SM of diet was unrelated to baseline weight-related information avoidance and weight bias internalization scores. Phase II dietary SM was significantly correlated with phase I SM of weight (*r*=0.26; *P*=.03) and diet (*r*=0.54; *P*<.001) but not exercise *(r*=0.21; *P*=.07). Participants who had higher engagement with dietary and weight SM previously tended to engage in more food logging during phase II. Group comparisons between those who maintained high adherence to dietary SM during months 4 to

11 and those who did not can be seen in Table 2. The groups did not differ on baseline weight-related information avoidance or weight bias internalization (none of the *P* values met the threshold for statistical significance).

Cross-level interaction models examined whether baseline weight-related information avoidance and weight bias internalization moderated the influence of time on dietary SM. There was a significant interaction between weight-related information avoidance and time on dietary SM (b=−0.01, SE 0.01; $t_{73.7}$=−2.13; *P*=.04). The cross-level interaction model fit was significantly improved from the random slope model of time on dietary SM tested in aim 1. Simple slope analyses found that participants with high (b=−0.07, SE 0.01; $t_{74.4}$=−8.61; *P*<.001) and moderate weight-related information avoidance (b=−0.06, SE 0.01; $t_{69.8}$=−10.38; *P*<.001) had a steeper decline in dietary SM over time than those with low weight-related information avoidance (b=−0.05, SE 0.01; $t_{69.0}$=−5.38; *P*<.001). Refer to Figure 3 for a depiction of this interaction. The cross-level interaction between weight bias internalization and time on dietary SM was not significant and did not improve model fit *P*=.08).

Figure 3. Visualization of the cross-level interaction between weight-related information avoidance and time on dietary self-monitoring adherence (weight-related information avoidance was centered within person; more positive scores indicate more avoidance).



Exercise SM

Average phase II adherence for exercise SM was not correlated with baseline weight-related information avoidance or baseline weight bias internalization scores (Table 1). Average phase II adherence for exercise SM was significantly correlated with phase I SM engagement for weight (r=0.38; *P*<.001), diet (r=0.59; *P*<.001), and exercise (r=0.42; *P*<.001) such that participants who had higher engagement on any of the SM targets throughout phase I engaged in more consistent exercise tracking during phase II. As seen in Table 2, participants who achieved consistently high adherence to exercise SM in phase II did not differ from those with low adherence on weight bias internalization or health information avoidance.

Cross-level interaction models examined whether between-person factors moderated the effect of time on exercise SM. The cross-level interaction between weight bias internalization and time (and health information avoidance and time) on exercise SM were not significant and did not improve model fit *P*=.87 and *P*=.56, respectively).

Discussion

Overview of Study Objective

Daily SM of weight, diet, and physical activity is a common prescription in BWL programs [1] because this practice is highly predictive of participant success [2-4]. However, adherence to SM tends to wane over time, especially during the WLM phase [5,6]. Very few studies have examined patterns of adherence to different SM tools over long periods of time, and almost none have examined how individual differences predict different types of SM adherence. This study explored changes in SM of different tools across time, the unique timing and patterns of disengagement and reengagement, and whether theory-based individual-level factors could predict SM adherence among participants in a year-long BWL program. The findings can inform attempts to prevent disengagement with SM tools and ultimately lead to greater weight control success.

Rates of SM Adherence

For all 3 SM targets, the rates of adherence declined across months 4 to 11. These data mirror results from previous analyses with these data during phase I (months 0-3), in which adherence to SM of weight and diet but not exercise decreased over 12 weeks [41]. SM adherence during phase II was strongly related to previous SM engagement during phase I, an expected association owing to the consistently strong association between past and current behavior [19]. Throughout phase II, the rates of SM of weight (53%, of the days) and diet (49%, of the days) were significantly lower than those of exercise (80%, of the days). These rates of SM of diet and weight are comparable with the median rates seen in past interventions, whereas the rates of SM of exercise were higher than usual [7]. Of the 72 participants, only 9 (13%) of the participants had strong (ie, ≥50%) adherence for all 3 SM targets throughout phase II, and total disengagement was common: almost half of the participants (31/72, 43%) had at least 1 month with no dietary SM, and almost one-third (23/72, 32%) had at least 1 month with no weight SM. SM engagement was modest in both conditions; however, the rates of participants achieving high SM adherence tended to be higher in the LM+SHARE condition, and more LM+SHARE participants achieved strong adherence across all 3 targets than those in the LM condition. These results parallel findings from the parent study on the potential benefit of coach surveillance of SM data [11]. The results suggest that nonadherence to SM, particularly of diet and weight, is a major problem in BWL treatment, although coach monitoring of SM data may provide 1 avenue for improvements.

The comparatively low rates of dietary SM are unsurprising because calorie tracking is a high-burden behavior that requires ample time and patience. Participants continually struggle with this behavior during BWL trials [46], and past research shows that this type of *active* SM (ie, calorie tracking) has lower engagement than *passive* SM (eg, wearing a Fitbit band) [7]. Efforts should be made to help participants track their food more easily, either by creating more user-friendly food tracking technology or by identifying tracking strategies that reduce participant burden without sacrificing effectiveness (eg, tracking only dietary lapses [47]). However, the low rates of self-weighing compared with the higher rates of exercise SM are notable, given that self-weighing, much like exercise tracking, is a low-burden behavior (ie, participants simply need to step on a scale). Therefore, the discrepancy between exercise SM and weight SM may be due to deliberate health information avoidance [48]. Weighing can be highly distressing for participants in BWL programs, and many people with overweight or obesity report avoiding the scale in fear of experiencing the negative feelings it may evoke [49]. Past studies show that people are less likely to weigh themselves when they have recently gained weight [50] or eaten more calories than usual [24], suggesting that the declining adherence for self-weighing may be a result of avoidance of the scale as eating and exercise behavior become less stringent than they were at the start of the program. The results of this study too suggest that efforts to enhance participant engagement with self-weighing may require addressing participant reactions to weight information (eg, with self-compassion training) rather than logistical efforts related to decreasing burden of the SM behavior.

Patterns of Disengagement and Reengagement

Average adherence for diet and weight SM fell to <50% around 7 months into the program, whereas average adherence for exercise SM fell off later, 10 months into the program. Among those who did have low engagement, most dropped off fairly early (ie, months 4-7) in phase II. Among participants who had a meaningful decrease in adherence (ie, >10%) compared with their phase I frequency of SM, the drop-off tended to occur just before month 6 (ie, 2 months into phase II) for diet and weight and around month 8 for SM of exercise. These results suggest that participants are at risk for SM disengagement, particularly with diet and weight, around the 6-month mark of BWL programs; thus, this may be an optimal time for a potential intervention.

Only approximately one-third of the participants whose weight and dietary SM adherence dropped to <50% ever reengaged (ie, restored levels of SM to ≥50% by the end of the program;17/51, 33% for weight SM and 19/57, 33% for dietary SM). However, almost half of the participants (13/28, 46%) whose exercise SM adherence dropped to <50% were able to reinstate those levels later during WLM. Therefore, when participants disengage with SM of weight and diet in particular, they are highly unlikely to reengage. It seems to be more likely that participants will pick back up with SM of exercise even if they have had low levels of adherence previously, suggesting that this behavior is more resilient against prior difficulties. Both rates of disengagement and reengagement were more promising when data were shared with coaches (LM+SHARE), suggesting that remote coach monitoring may be 1 way to help protect participants against dropping the key weight control behavior of SM.

Explaining SM Adherence and Adherence Trajectories

Although nonadherence to weight SM may be evidence of deliberate avoidance, self-reported weight-related information avoidance at baseline was not predictive of SM of weight or exercise. It is possible that the preference to deliberately avoid weight-related information predicts SM less strongly than expected because avoidance is a dynamic factor that changes owing to situational factors; for example, 1 study found that the preference to avoid weight-related information was associated with state variables (such as shame and negative mood) among adult women with overweight or obesity but not with trait variables (such as BMI, age, or past stigma regarding weight) [51]. Nevertheless, higher baseline weight-related information avoidance was associated with a steeper decline in *dietary* SM over time. Participants with a stronger tendency to avoid negative weight-related information may find it difficult to confront their calorie intake when they expect that the numbers will elicit shame [49]. As time progresses in WLM, and more participants drift from their calorie goals, recording that information seems to be more challenging for those who enter BWL treatment with higher weight-related information avoidance tendencies. Future work should confirm the dynamics of this relationship with more frequent assessment of weight-related information avoidance throughout WLM.

Contrary to expectations, in this study, baseline weight bias internalization was associated with higher adherence to weight SM. This finding is surprising because past research shows that weight bias internalization is associated with body image avoidance [52] and the avoidance of health care entirely [27]. It is possible that weight bias internalization could lead to *greater* motivation for weight loss to reduce weight-related guilt and shame; however, research consistently shows that such internalization is ultimately maladaptive [27]. Further work is needed to clarify these conflicting results and elucidate the underlying mechanism by which weight bias internalization predicts higher levels of weight SM. Research is also needed to determine whether the value of self-weighing is different between those with high weight bias internalization and those with low weight bias internalization.

Implications for BWL Treatment

Overall, the results emphasize the fact that SM of diet and weight is a challenge during BWL treatment (even when using digital tools) and should be prioritized as intervention targets. Long-term SM adherence was associated with higher engagement at earlier points of the BWL program (phase I), suggesting that individuals who can establish a consistent, regular SM routine early in treatment will find it easier to maintain it during WLM. Deliberate weight-related information avoidance may occur, as evidenced by the low rates of SM of weight despite its being a low-burden behavior in comparison with SM of diet. Thus, strategies to increase rates of self-weighing among participants in BWL programs may be best designed to target participants' reaction to SM (eg, self-compassion training) versus logistical problem-solving to decrease burden. This study is the first to identify potential individual-level factors related to use of digital SM tools during WLM. The findings suggest that participants in BWL programs entering treatment with higher rates of weight-related information avoidance and lower rates of weight bias internalization may be at higher risk for low long-term engagement with dietary and weight SM, respectively. This has clinical utility because these individuals can then be identified at baseline and targeted throughout treatment with specific strategies to facilitate sustaining SM as a key weight control behavior. The findings point to the utility of just-in-time adaptive interventions (JITAIs) to promote reengagement with SM tools among participants whose SM starts to decline. JITAIs are dynamic, identifying critical moments for intervention and providing tailored support [53]. Such interventions may be especially effective 6 to 8 months into treatment because, in this study, this was a critical period with high rates of disengagement from SM. Without this intervention, participants may have a difficult time resuming adherence to these crucial behaviors because the current data suggest that few participants who disengage will ever reengage. Future research should identify what psychological or practical support participants need at these times to inspire them to reengage.

Strengths and Limitations

This study had several strengths. It included long-term assessment of the use of digital SM tools after the intensive phase of a BWL program, which was a noted gap in the literature. SM data were collected objectively from wireless scales and passive Fitbit sensors, which is a particularly valid method of data collection. Given the design of the parent study, these analyses also provided a chance to look at SM patterns with remote coach monitoring of participant SM data and those without. Data sharing with coaches is a new development within BWL treatment innovation that is not included in most interventions. Thus, it is helpful to clarify what long-term digital SM behavior looks like with coach data surveillance and without. A limitation of the study is its sample size of 77 participants, limiting power to detect small effects. Additional research in a larger, more diverse (specifically, sex diverse) sample would increase the generalizability of, and confidence in, the findings. There was also attrition throughout phase I, where 10 (11%) of the 87 participants dropped out before randomization into phase II and 5 (6%) of the 77 enrolled in phase II did not provide data for this analysis. It is possible that these individuals were more likely to have disengaged from SM and thus would have exhibited poor rates of adherence. Thus, phase II SM adherence rates may have been lower (and the results may have differed) if these analyses were conducted using a data set that included all participants. Furthermore, definitions of SM adherence (eg, high adherence: ≥50% and valid days of calorie tracking: logging ≥5 foods) were based on past literature but are still somewhat arbitrary; for example, it is unclear whether using a threshold of 800 calories per day or at least 2 eating episodes per day is a better determinant of valid calorie days [43].

Conclusions

This study found that weight, diet, and exercise SM declined over time during the maintenance phase of a BWL intervention. Adherence to dietary SM was the poorest, followed by adherence to weight SM, whereas adherence to exercise SM was comparatively higher. Few participants maintained high levels of SM across the full study, and total disengagement from SM was common, with low rates of reengagement. These rates of adherence are particularly troubling, given the trial's strong emphasis on SM. Higher baseline health information avoidance and lower baseline weight bias internalization were associated with poorer SM. The findings suggest that future BWL interventions may benefit from JITAIs that identify when participants are at risk for disengagement and provide adaptive support to promote better SM adherence.

Acknowledgments

The authors would like to thank the study participants and study staff for their contributions to this study. This work was supported by the National Institutes of Health (R21DK112741).

Authors' Contributions

NC and MB conceptualized the study aims. NC completed data preparation. NC and CH completed data analyses. NC, OH, and CH wrote the initial draft of the manuscript. All authors contributed to revisions and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
2. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
3. Laitner MH, Minski SA, Perri MG. The role of self-monitoring in the maintenance of weight loss success. Eat Behav 2016 Apr;21:193-197 [FREE Full text] [doi: 10.1016/j.eatbeh.2016.03.005] [Medline: 26974582]
4. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
5. Wolin KY, Steinberg DM, Lane IB, Askew S, Greaney ML, Colditz GA, et al. Engagement with eHealth self-monitoring in a primary care-based weight management intervention. PLoS One 2015;10(10):e0140455 [FREE Full text] [doi: 10.1371/journal.pone.0140455] [Medline: 26469065]
6. Acharya SD, Elci OU, Sereika SM, Music E, Styn MA, Turk MW, et al. Adherence to a behavioral weight loss treatment program enhances weight loss and improvements in biomarkers. Patient Prefer Adherence 2009 Nov 03;3:151-160 [FREE Full text] [doi: 10.2147/ppa.s5802] [Medline: 19936157]
7. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
8. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: a systematic review and meta-analysis. Nutrients 2020 Jul 03;12(7):1977 [FREE Full text] [doi: 10.3390/nu12071977] [Medline: 32635174]
9. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. Obesity (Silver Spring) 2016 Aug;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
10. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
11. Butryn ML, Martinelli MK, Crane NT, Godfrey K, Roberts SR, Zhang F, et al. Counselor surveillance of digital self-monitoring data: a pilot randomized controlled trial. Obesity (Silver Spring) 2020 Dec;28(12):2339-2346 [FREE Full text] [doi: 10.1002/oby.23015] [Medline: 33098278]
12. Robertson MC, Raber M, Liao Y, Wu I, Parker N, Gatus L, et al. Patterns of self-monitoring technology use and weight loss in people with overweight or obesity. Transl Behav Med 2021 Aug 13;11(8):1537-1547 [FREE Full text] [doi: 10.1093/tbm/ibab015] [Medline: 33837792]
13. Harvey J, Krukowski R, Priest J, West D. Log often, lose more: electronic dietary self-monitoring for weight loss. Obesity (Silver Spring) 2019 Mar;27(3):380-384 [FREE Full text] [doi: 10.1002/oby.22382] [Medline: 30801989]
14. Zheng Y, Burke LE, Danford CA, Ewing LJ, Terry MA, Sereika SM. Patterns of self-weighing behavior and weight change in a weight loss trial. Int J Obes (Lond) 2016 Sep;40(9):1392-1396. [doi: 10.1038/ijo.2016.68] [Medline: 27113642]
15. Tronieri JS, Wadden TA, Walsh O, Berkowitz RI, Alamuddin N, Chao AM. Measures of adherence as predictors of early and total weight loss with intensive behavioral therapy for obesity combined with liraglutide 3.0 mg. Behav Res Ther 2020 Aug;131:103639 [FREE Full text] [doi: 10.1016/j.brat.2020.103639] [Medline: 32450367]
16. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
17. Verheijden MW, Bakx JC, van Weel C, Koelen MA, van Staveren WA. Role of social support in lifestyle-focused weight management interventions. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S179-S186. [doi: 10.1038/sj.ejcn.1602194] [Medline: 16052189]
18. Martinelli MK, D'Adamo L, Butryn ML. Binge eating predicts adherence to digital self-monitoring during behavioral weight loss. Eat Behav 2020 Dec;39:101448 [FREE Full text] [doi: 10.1016/j.eatbeh.2020.101448] [Medline: 33157520]

19. Conner M, Armitage CJ. Extending the theory of planned behavior: a review and avenues for further research. J Appl Social Pyschol 1998 Aug;28(15):1429-1464. [doi: 10.1111/j.1559-1816.1998.tb01685.x]
20. Sweeny K, Melnyk D, Miller W, Shepperd JA. Information avoidance: who, what, when, and why. Rev General Psychol 2010 Dec 01;14(4):340-353. [doi: 10.1037/a0021288]
21. Carver C, Scheier M. On the Self-Regulation of Behavior. Cambridge, United Kingdom: Cambridge University Press; 2001.
22. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and 'punishy': exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554 [FREE Full text] [doi: 10.1177/2055207618797554] [Medline: 30202544]
23. Webb TL, Chang BP, Benn Y. 'The ostrich problem': motivated avoidance or rejection of information about goal progress. Soc Personality Psychol Compass 2013 Nov 04;7(11):794-807. [doi: 10.1111/spc3.12071]
24. Schumacher LM, Martinelli MK, Convertino AD, Forman EM, Butryn ML. Weight-related information avoidance prospectively predicts poorer self-monitoring and engagement in a behavioral weight loss intervention. Ann Behav Med 2021 Mar 16;55(2):103-111 [FREE Full text] [doi: 10.1093/abm/kaaa034] [Medline: 32491152]
25. Reyes NR, Oliver TL, Klotz AA, Lagrotte CA, Vander Veur SS, Virus A, et al. Similarities and differences between weight loss maintainers and regainers: a qualitative analysis. J Acad Nutr Diet 2012 Apr;112(4):499-505. [doi: 10.1016/j.jand.2011.11.014] [Medline: 22709701]
26. Tanenbaum ML, Ross KM, Wing RR. Overeat today, skip the scale tomorrow: an examination of caloric intake predicting nonadherence to daily self-weighing. Obesity (Silver Spring) 2016 Nov;24(11):2341-2343 [FREE Full text] [doi: 10.1002/oby.21650] [Medline: 27619935]
27. Pearl RL, Puhl RM. Weight bias internalization and health: a systematic review. Obes Rev 2018 Aug;19(8):1141-1163 [FREE Full text] [doi: 10.1111/obr.12701] [Medline: 29788533]
28. Pearl RL, Puhl RM, Dovidio JF. Differential effects of weight bias experiences and internalization on exercise among women with overweight and obesity. J Health Psychol 2015 Dec;20(12):1626-1632. [doi: 10.1177/1359105313520338] [Medline: 24505060]
29. Linde JA, Rothman AJ, Baldwin AS, Jeffery RW. The impact of self-efficacy on behavior change and weight change among overweight participants in a weight loss trial. Health Psychol 2006 May;25(3):282-291. [doi: 10.1037/0278-6133.25.3.282] [Medline: 16719599]
30. Burke LE, Ewing LJ, Ye L, Styn M, Zheng Y, Music E, et al. The SELF trial: a self-efficacy-based behavioral intervention trial for weight loss maintenance. Obesity (Silver Spring) 2015 Nov;23(11):2175-2182 [FREE Full text] [doi: 10.1002/oby.21238] [Medline: 26381151]
31. Warziski MT, Sereika SM, Styn MA, Music E, Burke LE. Changes in self-efficacy and dietary adherence: the impact on weight loss in the PREFER study. J Behav Med 2008 Feb;31(1):81-92. [doi: 10.1007/s10865-007-9135-2] [Medline: 17963038]
32. Corrigan PW, Larson JE, Rüsch N. Self-stigma and the "why try" effect: impact on life goals and evidence-based practices. World Psychiatry 2009 Jun;8(2):75-81 [FREE Full text] [doi: 10.1002/j.2051-5545.2009.tb00218.x] [Medline: 19516923]
33. Kahan S, Puhl RM. The damaging effects of weight bias internalization. Obesity (Silver Spring) 2017 Feb;25(2):280-281. [doi: 10.1002/oby.21772] [Medline: 28124505]
34. Hain B, Langer L, Hünnemeyer K, Rudofsky G, Zech U, Wild B. Translation and validation of the German version of the weight self-stigma questionnaire (WSSQ). Obes Surg 2015 Apr;25(4):750-753. [doi: 10.1007/s11695-015-1598-6] [Medline: 25663147]
35. Pearl RL, Puhl RM, Himmelstein MS, Pinto AM, Foster GD. Weight stigma and weight-related health: associations of self-report measures among adults in weight management. Ann Behav Med 2020 Nov 01;54(11):904-914 [FREE Full text] [doi: 10.1093/abm/kaaa026] [Medline: 32333673]
36. Severin R, Sabbahi A, Mahmoud AM, Arena R, Phillips SA. Precision medicine in weight loss and healthy living. Prog Cardiovasc Dis 2019 Jan;62(1):15-20 [FREE Full text] [doi: 10.1016/j.pcad.2018.12.012] [Medline: 30610881]
37. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]
38. The Diabetes Prevention Program Research Group. The diabetes prevention program. Design and methods for a clinical trial in the prevention of type 2 diabetes. Diabetes Care 1999 Apr;22(4):623-634 [FREE Full text] [doi: 10.2337/diacare.22.4.623] [Medline: 10189543]
39. Howell JL, Shepperd JA. Establishing an information avoidance scale. Psychol Assess 2016 Dec;28(12):1695-1708. [doi: 10.1037/pas0000315] [Medline: 27099980]
40. Durso LE, Latner JD. Understanding self-directed stigma: development of the weight bias internalization scale. Obesity (Silver Spring) 2008 Nov;16 Suppl 2:S80-S86 [FREE Full text] [doi: 10.1038/oby.2008.448] [Medline: 18978768]
41. Butryn ML, Godfrey KM, Martinelli MK, Roberts SR, Forman EM, Zhang F. Digital self-monitoring: does adherence or association with outcomes differ by self-monitoring target? Obes Sci Pract 2020 Apr;6(2):126-133 [FREE Full text] [doi: 10.1002/osp4.391] [Medline: 32313670]

42. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
43. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining adherence to mobile dietary self-monitoring and assessing tracking over time: tracking at least two eating occasions per day is best marker of adherence within two different mobile health randomized weight loss interventions. J Acad Nutr Diet 2019 Sep;119(9):1516-1524 [FREE Full text] [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
44. Field A. Discovering Statistics Using IBM SPSS Statistics. Thousand Oaks, California, United states: SAGE Publications; 2013.
45. Berkovits I, Hancock GR, Nevitt J. Bootstrap resampling approaches for repeated measure designs: relative robustness to sphericity and normality violations. Educational Psychological Measurement 2016 Jul 02;60(6):877-892. [doi: 10.1177/00131640021970961]
46. Yu Z, Sealey-Potts C, Rodriguez J. Dietary self-monitoring in weight management: current evidence on efficacy and adherence. J Acad Nutr Diet 2015 Dec;115(12):1931-1938. [doi: 10.1016/j.jand.2015.04.005] [Medline: 26028176]
47. Pagoto S, Tulu B, Waring ME, Goetz J, Bibeau J, Divito J, et al. Slip buddy app for weight management: randomized feasibility trial of a dietary lapse tracking app. JMIR Mhealth Uhealth 2021 Apr 01;9(4):e24249 [FREE Full text] [doi: 10.2196/24249] [Medline: 33792547]
48. Howell J, Lipsey N, Shepperd J. Health information avoidance. In: The Wiley Encyclopedia of Health Psychology. Hoboken, New Jersey: Wiley; 2020.
49. Hartmann-Boyce J, Boylan A, Jebb SA, Aveyard P. Experiences of self-monitoring in self-directed weight loss and weight loss maintenance: systematic review of qualitative studies. Qual Health Res 2019 Jan;29(1):124-134. [doi: 10.1177/1049732318784815] [Medline: 29984630]
50. Goldstein SP, Goldstein CM, Bond DS, Raynor HA, Wing RR, Thomas JG. Associations between self-monitoring and weight change in behavioral weight loss interventions. Health Psychol 2019 Dec;38(12):1128-1136 [FREE Full text] [doi: 10.1037/hea0000800] [Medline: 31556659]
51. Hagerman C. Interrupting the Cycle of Weight Stigma and Unhealthy Behavior: The Potential of Self-affirmation and Self-compassion Exercises. Washington, D.C: The George Washington University; Aug 31, 2021.
52. Marshall RD, Latner JD, Masuda A. Internalized weight bias and disordered eating: the mediating role of body image avoidance and drive for thinness. Front Psychol 2019;10:2999 [FREE Full text] [doi: 10.3389/fpsyg.2019.02999] [Medline: 32038383]
53. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

Abbreviations

AHEAD: Action for Health in Diabetes
BWL: behavioral weight loss
IAS: Information Avoidance Scale
JITAI: just-in-time adaptive intervention
LM: lifestyle modification
LM+SHARE: lifestyle modification plus device data sharing
SM: self-monitoring
WBIS: Weight Bias Internalization Scale
WLM: weight loss maintenance

Edited by L Buis; submitted 08.01.23; peer-reviewed by J Bibeau, D West; comments to author 06.03.23; revised version received 17.04.23; accepted 18.05.23; published 18.07.23.

Please cite as:
Crane N, Hagerman C, Horgan O, Butryn M
Patterns and Predictors of Engagement With Digital Self-Monitoring During the Maintenance Phase of a Behavioral Weight Loss Program: Quantitative Study
JMIR Mhealth Uhealth 2023;11:e45057
URL: https://mhealth.jmir.org/2023/1/e45057
doi:10.2196/45057
PMID:37463017

Original Paper

Long-Term Outcomes of a Comprehensive Mobile Smoking Cessation Program With Nicotine Replacement Therapy in Adult Smokers: Pilot Randomized Controlled Trial

Jennifer D Marler[1], MD; Craig A Fujii[1], BS, MIDS; MacKenzie T Utley[1], BS; Daniel J Balbierz[1], BS; Joseph A Galanko[2], PhD; David S Utley[1], MD

[1]Pivot Health Technologies, Inc, San Carlos, CA, United States

[2]Department of Pediatrics, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Jennifer D Marler, MD
Pivot Health Technologies, Inc
1010 Commercial St.
Suite C
San Carlos, CA, 94070
United States
Phone: 1 4082145545
Email: marler@pivot.co

Abstract

Background: Increased smartphone ownership has led to the development of mobile smoking cessation programs. Although the related body of evidence, gathered through the conduct of randomized controlled trials (RCTs), has grown in quality and rigor, there is a need for longer-term data to assess associated smoking cessation durability.

Objective: The primary aim was to compare smoking cessation outcomes at 52 weeks in adult smokers randomized to a mobile smoking cessation program, Pivot (intervention), versus QuitGuide (control). The secondary aims included comparison of other smoking-related behaviors, outcomes and participant feedback, and exploratory analyses of baseline factors associated with smoking cessation.

Methods: In this remote pilot RCT, cigarette smokers in the United States were recruited on the web. Participants were offered 12 weeks of free nicotine replacement therapy (NRT). Data were self-reported via a web-based questionnaire with videoconference biovalidation in participants who reported 7-day point-prevalence abstinence (PPA). Outcomes focused on cessation rates with additional assessment of quit attempts, cigarettes per day (CPD), self-efficacy via the Smoking Abstinence Self-Efficacy Questionnaire, NRT use, and participant feedback. Cessation outcomes included self-reported 7- and 30-day PPA, abstinence from all tobacco products, and continuous abstinence. PPA and continuous abstinence were biovalidated using witnessed breath carbon monoxide samples. Exploratory post hoc regression analyses were performed to identify baseline variables associated with smoking cessation.

Results: Participants comprised 188 smokers (n=94, 50% in the Pivot group and n=94, 50% in the QuitGuide group; mean age 46.4, SD 9.2 years; n=104, 55.3% women; n=128, 68.1% White individuals; mean CPD 17.6, SD 9.0). Several cessation rates were higher in the Pivot group (intention to treat): self-reported continuous abstinence was 20% (19/94) versus 9% (8/94; P=.03) for QuitGuide, biochemically confirmed abstinence was 31% (29/94) versus 18% (17/94; P=.04) for QuitGuide, and biochemically confirmed continuous abstinence was 19% (18/94) versus 9% (8/94; P=.046) for QuitGuide. More Pivot participants (93/94, 99% vs 80/94, 85% in the QuitGuide group; P<.001) placed NRT orders (mean 3.3, SD 2.0 vs 1.8, SD 1.6 for QuitGuide; P<.001). Pivot participants had increased self-efficacy via the Smoking Abstinence Self-Efficacy Questionnaire (mean point increase 3.2, SD 7.8, P<.001 vs 1.0, SD 8.5, P=.26 for QuitGuide). QuitGuide participants made more mean quit attempts (7.0, SD 6.3 for Pivot vs 9.5, SD 7.5 for QuitGuide; P=.01). Among those who did not achieve abstinence, QuitGuide participants reported greater CPD reduction (mean −34.6%, SD 35.5% for Pivot vs −46.1%, SD 32.3% for QuitGuide; P=.04). Among those who reported abstinence, 90% (35/39) of Pivot participants and 90% (26/29) of QuitGuide participants indicated that their cessation program helped them quit.

Conclusions: This pilot RCT supports the long-term effectiveness of the Pivot mobile smoking cessation program, with abstinence rates durable to 52 weeks.

Trial Registration: ClinicalTrials.gov NCT04955639; https://clinicaltrials.gov/ct2/show/NCT04955639

(JMIR Mhealth Uhealth 2023;11:e48157) doi:10.2196/48157

KEYWORDS

smoking cessation; digital health; smartphone; digital sensor; carbon monoxide; breath sensor; biofeedback; mobile apps; health promotion; app; mobile phone

Introduction

Background

In 2020, a total of 12.5% of adults in the United States (approximately 30.8 million people) smoked cigarettes [1]. Although this represents a marked decrease in the peak prevalence of 42.6% in 1964, smoking remains a formidable public health problem as the leading preventable cause of death, illness, and disability in the United States [2,3].

Behavioral interventions and pharmacotherapy have proven to be effective smoking cessation tools. Traditionally, behavioral intervention via counseling has been delivered in person in one-on-one or group sessions or remotely through phone quitlines. Food and Drug Administration (FDA)–approved pharmacotherapy such as nicotine replacement therapy (NRT) is available over the counter and by prescription; bupropion and varenicline are available by prescription. The use of combined behavioral support and pharmacotherapy increases the chance of smoking cessation by 70% to 100% compared with just brief advice or support [4]. The use of monotherapy is also effective, with NRT, bupropion, varenicline, and individual counseling increasing the chance of success by 50% to 60%, 50% to 80%, 100% to 140%, and 40% to 80%, respectively, compared with minimal support or placebo [4,5].

Unfortunately, the impact of these interventions is limited; less than one-third of adult cigarette smokers who make a quit attempt use any type of cessation counseling or FDA-approved pharmacotherapy [6]. Most quit attempts are unassisted, with resultant success rates of approximately 7% to 8% [3].

The rise of mobile smoking cessation interventions seeks to address this shortcoming by leveraging the ubiquitous nature of smartphones; in 2021, a total of 85% of adults in the United States owned a smartphone [7]. Assessment of earlier smartphone app–based smoking cessation programs has reported focus on ease-of-use features and simplistic tools, with low inclusion rates of evidence-based approaches such as the 5 A's (Ask, Advise, Assess, Assist, and Arrange follow-up), counseling, and pharmacotherapy [8-10].

More recently, there has been both an increased presence of evidence-based cessation tools in mobile app–based smoking cessation programs and an increase in the rigor with which these programs are assessed, with the publication of several relevant randomized controlled trials (RCTs). BinDhim et al [11] compared an interactive decision aid app (intervention) with a static information app (control) for smoking cessation in an automated double-blind RCT. At 6 months, self-reported continuous abstinence was achieved in 7.3% (25/342) of participants in the intervention arm and 3.2% (11/342) in the control arm (intention to treat [ITT]; relative risk [RR]=2.27, 95% CI 1.09-4.86; *P*=.03) [11]. Garrison et al [12] compared mobile mindfulness training through the Craving to Quit app plus experience sampling (intervention) with experience sampling only (control) in a researcher-blind, parallel RCT. They reported biovalidated 7-day point-prevalence abstinence (PPA) in 9.8% (14/143; intervention) versus 12.1% (22/182; control) of patients (ITT; χ^2_1=0.4, *P*=.51) at 6 months [12]. Etter and Khazaal [13] conducted a 2-arm, parallel-group, individually randomized, double-blind RCT, in which the Stop-tabac app (intervention comprising an app with informational text, personal calculators, ecological momentary intervention for challenging situations, a quiz, contact information for telephone quitlines, coaching via automated messages, a discussion forum, and modules focused on NRT and electronic cigarettes) was compared with a control app (5 brief information pages and the aforementioned calculators). At 6 months, 7-day PPA was reported in 11.29% (298/2639) of intervention arm participants and 11.98% (318/2654) of control arm participants (ITT; odds ratio [OR] 0.94, 95% CI 0.79-1.11; *P*=.43) [13]. Finally, Bricker et al [14] compared iCanQuit (intervention; an acceptance and commitment therapy–based app) with QuitGuide (control; a United States Clinical Practice Guideline [USCPG]–based app from the National Cancer Institute [NCI]) in a 2-group, stratified, double-blind, individually randomized clinical trial in smokers. At 6 months, 29.57% (359/1214) of participants in the intervention arm and 21.57% (259/1201) of participants in the control arm self-reported 7-day PPA (ITT; OR 1.62, 95% CI 1.34-1.95; *P*<.001) [14].

Although the aforementioned RCTs evaluated behavioral intervention–focused mobile smoking cessation programs, a few RCTs have assessed similar mobile interventions with an enhanced evidence-based component, specifically, the provision of NRT. Hébert et al [15] conducted a 3-armed pilot RCT in which participants were randomized to Smart-T2 (a "just-in-time adaptive intervention that uses ecological momentary assessments to assess the risk for imminent smoking lapse and tailor treatment messages based on the risk of lapse and reported symptoms"), the NCI QuitGuide app, or usual tobacco cessation clinic care. Participants received 2 weeks of NRT. At 12 weeks, biovalidated 7-day PPA was reported in 22% (6/27) of Smart-T2 participants, 15% (4/27) of QuitGuide participants, and 15% (4/27) of usual care participants (ITT; *P*>.05) [15]. Webb et al [16] led a 2-arm, single-blinded, parallel-group RCT; participants were randomized to Quit Genius (intervention; a smartphone app delivering cognitive behavioral therapy content, one-to-one coaching, craving tools, and tracking capabilities) or very brief advice (control; in-person advice informed by the Ask, Advise, and Act model). Participants had the option to receive 12 weeks of NRT. Biovalidation via carbon monoxide (CO) breath sampling was performed in slightly less than half of the participants. At 6 months, 35.9% (95/265) of Quit Genius participants and 27.6% (73/265) of very brief advice participants achieved 7-day PPA (ITT; RR=1.32, 95% CI 1.03-1.69; *P*=.03) [16].

Similarly, we performed a remote pilot RCT in which participants were randomized to Pivot (intervention; an app-based smoking cessation program comprising activities and lessons, a personal CO breath sensor, in-app text-based human-provided coaching, NRT, and a moderated web-based community) or the NCI QuitGuide app (control). Participants had access to 12 weeks of NRT. At 6 months, 36% (34/94) of Pivot participants and 27% (25/94) of QuitGuide participants self-reported 7-day PPA (ITT; RR=1.3, 95% CI 0.8-1.9; *P*=.27), and 28% (26/94) of Pivot participants and 15% (14/94) of QuitGuide participants achieved biovalidated abstinence (ITT; RR=1.9, 95% CI 1.1-3.5; *P*=.02) [17]. In the context of these RCTs, the natural next step in the evolution of data pertaining to mobile smoking cessation programs is the establishment of longer-term outcomes through smoking cessation durability.

Objectives

The primary aim of this study was to assess longer-term smoking cessation outcomes in the aforementioned RCT, in which Pivot was compared with QuitGuide [17]. Specifically, we aimed to compare self-reported and biovalidated PPA and continuous abstinence at 52 weeks after enrollment. Secondary aims included the assessment of other aspects of smoking-related behavior, such as quit attempts, cigarettes per day (CPD), and self-efficacy, and the performance of exploratory analyses of predictors of smoking cessation.

Methods

Design

In this 2-arm, parallel-group, noncrossover, single-center RCT, participants were randomized to 1 of 2 app-based smoking cessation programs: Pivot (intervention) or QuitGuide (control). All participants had access to 12 weeks of free NRT. The 6-month outcomes, including comparison of user engagement and retention, attitudes toward quitting, smoking behavior, and participant feedback, have been previously reported [17]. In this paper, we report the outcomes at 52 weeks. This study was performed remotely on an ambulatory basis. All participants provided electronic informed consent before taking part.

Participants

Eligibility criteria were being aged ≥21 years, being a daily cigarette smoker (≥5 CPD) for the previous 12 months, having plans to quit smoking in the next 30 days, being a resident of the United States, being able to read and comprehend English, owning and using a smartphone compatible with the study app (iPhone 5 and above with operating system iOS 12 and above or Android 7.0 and above with operating system Android 7.0 and above), having daily internet access on smartphone, and self-reporting comfort with downloading and using smartphone apps.

Exclusion criteria were pregnancy (self-reported); health contraindications to NRT use (irregular heartbeat, high blood pressure not controlled with medication, myocardial infarction or stroke within the last 2 months, breastfeeding, skin allergies to adhesive tape or serious skin problems, stomach ulcers, and history of seizures); use of other smoking cessation support, including apps or actively taking medication to quit smoking; daily marijuana use; residence with another study participant; an immediate family member being a study participant; failure to provide contact information or verify email address; and participation in a previous study sponsored by Pivot Health Technologies Inc (formerly Carrot Inc).

Recruitment

Participants were recruited in the United States through web media (Facebook and Google Ads). Potential participants were asked to provide contact information and answer questions on demographics using a web-based screening form. Study staff reviewed each web-based screening form.

Using nonproportional quota sampling, potential participants were called on a first-come, first-served basis, with the aim of enrolling 40% to 60% men, no more than 50% of participants from any decade-spanning age group (eg, 30-39 years), no more than 70% of participants in the non-Hispanic White race category, and up to 20% of participants not employed. The goals of these nonproportional quota sampling ranges were to ensure representation among men, racial and ethnic minority groups, age groups, and individuals of varying socioeconomic status.

During the screening phone call, potential participants were asked questions to confirm study eligibility. During this call, study personnel informed the potential participants of the study details and answered any questions. Potential eligible participants who wanted to proceed with the study were emailed an electronic Health Insurance Portability and Accountability Act (HIPAA) authorization form and an electronic informed consent form, which they signed before participating in the study.

Randomization and Blinding

Participants were randomly assigned in a computer-generated 1:1 ratio to either QuitGuide or Pivot using randomly permuted blocks of sizes 2 and 4. The allocation sequence was provided by the Study Randomizer software application (2017) [18]. Participants were stratified by daily smoking frequency (≤14 vs ≥15 CPD), employment status (full-time or part-time employment vs not employed), race (minority race vs White), and expected difficulty to stay quit (DTQ; scale of 1-10; self-reported score of ≤5 vs ≥6). These 4 factors were chosen as they have been associated with cessation outcomes in previous studies [19-25]. The researchers were blinded to treatment allocation until after randomization was performed.

Intervention: Pivot

Pivot is a 12-month digital smoking cessation program based on the USCPG for tobacco cessation. Pivot includes the Pivot Breath Sensor and Pivot app (Pivot Health Technologies Inc).

The Pivot Breath Sensor is a portable, personal mobile breath sensor that measures the level of CO in the exhaled breath. The user submits a breath sample by exhaling into the sensor mouthpiece. The sensor displays the exhaled breath CO value in parts per million (ppm) to the user directly on the device. When paired with the user's smartphone, the user's CO values also populate the user's Pivot app. Displayed CO values are color coded and categorized as most consistent with not smoking (green; 0-6 ppm), possibly smoking (orange; 7-9 ppm), or smoking (red; ≥10 ppm).

The self-guided Pivot app leverages evidence-based principles and clinical best practices. This includes the USCPG-recommended 5 A's; tailoring on readiness to quit [4]; the provision of FDA-approved NRT with accompanying education on use and adherence [4,26,27]; the incorporation of effective methods for smoking cessation based on cognitive behavioral therapy and self-determination theory [3,28,29]; and cognitive behavioral therapy–based counseling through a live, dedicated coach [3,4,30]. Pivot app functions include interactive educational activities and the ability to log cigarettes, set a quit date, create a quit plan, complete practice quits (1-24 hours in duration), play educational games, watch educational videos, interact with one's dedicated human coach via in-app SMS text messaging, view CO breath sample values and trends, learn about and then order NRT, access the moderated web-based Pivot community discussion forum, share goals and progress with the web-based Pivot community discussion forum or one's social network via SMS text messaging or email, and complete daily check-ins after the quit date.

The educational journey in the Pivot app comprises 4 tracts—Learn, Reduce, Prepare to Quit, and Maintain My Quit—and is designed to accommodate smokers along the spectrum of readiness to quit. Participants may navigate between tracts as desired to access content most relevant to their goals and needs.

Pivot users are assigned a human coach with whom they work one-on-one over the duration of their use of Pivot (up to 1 year). Communication between the coach and Pivot user is via asynchronous in-app SMS text messaging. Pivot coaches are tobacco treatment specialists. The coach reaches out periodically, approximately once per week, during the participant's active use of Pivot. Participants may reach out to their coach whenever and however often they like.

Pivot users may access the moderated web-based discussion community through the Pivot app. The forum is moderated by a tobacco treatment specialist. The web-based community forum is a place to give and receive support and advice from others going through the Pivot program.

Control: QuitGuide

QuitGuide is a product of Smokefree [31], a smoking cessation resource created by the Tobacco Control Research Branch at the NCI in collaboration with tobacco control professionals and smoking cessation experts and with input from ex-smokers [32]. A well-established smoking cessation app, QuitGuide has been used in previous RCTs in which digital smoking cessation programs were compared [14,15,33]. The app focuses on helping users understand their smoking patterns and build the skills needed to become and stay smoke-free [32]. Specifically, QuitGuide helps users focus on motivations to quit; prepare to quit by developing a quit plan, identifying and planning how to address triggers and moods, teaching about FDA-approved smoking cessation medications, and identifying and providing access to social support; quit smoking by acknowledging user progress and teaching skills to address cravings; and stay quit by presenting tips and motivations to stay smoke-free and address slips if they occur. The QuitGuide app functions include educational reading activities, including focus on FDA-approved cessation medications and associated adherence. Additional QuitGuide app functions include tracking and reviewing cigarettes, moods, triggers, and cravings; setting tip message notifications for locations and times when one is prone to smoke; setting a quit date; creating a quit plan; completing journal entries; sharing goals and progress with one's social network via SMS text messaging or email; accessing additional chat and phone support; and providing updates on quit status after the quit date.

QuitGuide was used as the control for the following reasons: the content follows the USCPG for tobacco cessation; it is an app-based smoking cessation program, thereby enabling intrastudy comparison of same-modality interventions; the app is nonproprietary and free to the public; and its use in previous well-designed RCTs [14,15,33] provides context and enables interstudy comparison with earlier data.

NRT Provision

Participants had access to free, FDA-cleared, over-the-counter NRT. Participants were provided with on-label information about the NRT and were able to order it on the web (QuitGuide) or in their study app (Pivot). The types of NRT offered included nicotine patches (7, 14, or 21 mg), nicotine gum (2 or 4 mg), and nicotine lozenges (2 or 4 mg). Participants could order patches, gum, or lozenges alone as monotherapy or patches with either gum or lozenges as combination therapy. Participants were able to order NRT every 2 weeks for up to a 12-week course over the first 12 months of the study.

Biovalidation

Biovalidation was sought at 52 weeks in individuals who reported 7-day (or greater) PPA on the associated questionnaire. A video call with study staff and the participant was scheduled within 7 days following the participant's response to the associated questionnaire. At the beginning of the biovalidation visit, participants were asked their CPD, 7-day PPA status, and whether they had smoked any other noncigarette (eg, pipes, cigars, or hookah) or combustible materials (eg, cloves or marijuana) over the previous 24 hours.

Participants who indicated that they were not at least 7 days abstinent or that they had smoked ≥1 CPD were not eligible to undergo further biovalidation testing during the visit. Participants who indicated that they were at least 7 days abstinent and had not smoked cigarettes were eligible to proceed with the testing. Participants who indicated that they had smoked any other combustible materials over the previous 24 hours were eligible to undergo a biovalidation test at that same visit, with the possibility of scheduling a follow-up biovalidation test for the following day with instructions not to smoke the previously reported other combustible substance or substances over the intervening 24-hour period. If a participant was eligible for biovalidation and biovalidation was not achieved, the reason was noted (eg, did not schedule or attend a biovalidation study visit, reported change in smoking status at the outset of visit, or the participant's breath CO sample was ≥10 ppm).

Biovalidation was obtained through CO breath sampling. Participants in the intervention arm used their Pivot Breath Sensor for this test. Shortly before the visit, participants in the control arm were mailed a Pivot Breath Sensor that was limited to 10 breath samples. During the video call, participants held the breath sensor up to the screen immediately after completing the breath sample so that the study staff could see and record the CO ppm measurement on the sensor screen. A CO value of <10 ppm was considered consistent with abstinence [17,34,35].

Outcomes and Measures

Baseline

The following variables were collected at baseline: demographic information (age, gender, race, ethnicity, household income, education, employment status, and smartphone type); smoking status; smoking history; Heaviness of Smoking Index [36]; success to quit (scale of 1-10); DTQ (scale of 1-10) [37,38]; and Smoking Abstinence Self-Efficacy Questionnaire (SASEQ), a 6-item survey describing emotional or social situations for which smokers indicate on a 5-point Likert scale (0-4) whether they will be able to refrain from smoking, with higher scores representing higher self-efficacy [39].

52 Weeks

At the 52-week time point, the study outcomes focused on smoking behavior and self-efficacy. Assessments included the SASEQ; quit attempts; CPD; use of NRT; and smoking cessation via self-reported 7-day PPA, 30-day PPA, and continuous abstinence; biochemically confirmed abstinence; biochemically confirmed continuous abstinence; and self-reported abstinence from all tobacco products.

Self-efficacy was assessed using the SASEQ. Participants were considered to have made a quit attempt during the study if they answered ≥1 to the following question: "Since you began the study, how many times have you tried to quit smoking where you've gone at least 1 day without smoking a cigarette, even a single puff?" From this question, the mean quit attempts per participant were quantified as well. CPD were assessed through the mean percentage change and the proportion of participants who reduced their CPD by ≥50% compared with the baseline. NRT use included whether a participant ordered NRT (yes or no)—and, if so, the type of NRT they ordered—using participant-placed orders.

Participants were considered to have achieved self-reported 7-day (30-day) PPA if they answered "no" to the following question: "In the last 7 (30) days have you smoked any cigarettes, even a single puff?" Biochemically confirmed abstinence was defined as self-reporting 7-day abstinence and having a breath CO sample of <10 ppm at the 52-week biovalidation visit. Self-reported continuous abstinence was defined as self-reporting 7-day (or greater) PPA at 12 weeks, self-reporting 30-day PPA at 26 and 52 weeks, no more than 5 cigarettes between 12 and 26 weeks, and 0 cigarettes between 26 and 52 weeks. Biochemically confirmed continuous abstinence was defined as self-reported continuous abstinence, as detailed previously, with a breath CO sample of <10 ppm at each of the associated 12-, 26-, and 52-week biovalidation visits. Abstinence from all tobacco products was self-reported.

Sample Size

As this was a pilot RCT and the first assessment of Pivot compared with usual care, the sample size was powered to show differences in engagement, specifically, the number of times participants opened their assigned app over the first 12 weeks of the study. The methodology to power the study for engagement and the associated outcomes have been previously reported. Pivot participants self-reported a mean of 157.9 (SD 210.6) total app openings versus 86.5 (SD 66.3) in the QuitGuide group (incidence rate ratio [IRR]=1.8, 95% CI 1.4-2.3; P<.001) through week 12 [17].

Statistical Analyses

Comparisons

For results in which a change from baseline could be measured (CPD and SASEQ), each participant's baseline data served as their control to then calculate a difference at 52 weeks, and a paired 2-tailed t test was used to test for a significant difference from 0. The outcomes were evaluated using regression analyses adjusted for the 4 randomization stratification covariates to detect differences between the treatment and control arms. Linear regression was used for numerical data to obtain a point estimate of the mean difference. For count outcomes, the IRR was estimated using Poisson regression when the variance-to-mean ratio was close to 1 or using negative binomial regressions when the variance-to-mean ratio was >1. For binary outcomes, the OR was estimated using logistic regression, and the RR was estimated using either log-link binomial regression or log-link Poisson regression with robust estimators [40]. For binary outcomes where there was a very high-frequency response (eg, ≥95%), only the RR was presented. For multicategory outcomes of ≥3, multinomial logistic regression was used to test for proportion differences between the study arms.

In the assessment of quit rates (self-reported and biovalidated PPA, self-reported and biovalidated continuous abstinence, and self-reported abstinence from all tobacco products), 2 sets of analyses were performed. In the ITT analysis, individuals who did not respond to the PPA questions were assumed to be smoking. A study responder analysis was also performed, which only included individuals who completed the 52-week questionnaire. For the outcomes of quit attempts and the proportion of participants who reduced CPD by ≥50%, a study responder analysis was performed.

Predictors

We also performed exploratory post hoc analyses that may help inform the design of future studies. We used logistic regression to examine the associations between baseline characteristics and smoking behavior outcomes at 52 weeks. Each independent baseline variable was evaluated as a predictor in either 1 or 2 types of models using the ITT and responder data sets. The first model evaluated the interaction of the independent baseline variable with the randomized cohort. The interaction was tested for significance compared with a model without the interaction using the joint test. If the interaction was significant, the results were reported. If not, then the second model without the interaction was applied. In this model, the independent baseline variable was adjusted for the randomization cohort, and a type-3 test was used for detection of statistical significance. One *P* value is presented, either from the joint test if significant or from the type-3 test. For statistically significant models with baseline characteristics with ≥3 categories, additional permutations of the reference were completed to appropriately characterize the difference.

This evaluation was completed for the binary outcomes of self-reported 7-day and 30-day PPA, biovalidated 7-day PPA, self-reported continuous abstinence, and biovalidated continuous abstinence. In contrast to the analysis of outcomes, these models were not adjusted for the 4 randomization variables. Certain categorical baseline variables were collapsed either because of insufficient tallies for model convergence or to simplify the model evaluation. These baseline variables included ethnicity, education, income, health, use of noncigarette tobacco products, past quit methods used, and first cigarette smoked after waking. In addition, some of the interaction models produced quasi-complete separation, and the data were modeled again with logistic regression using a Firth bias correction [41]. Analyses were conducted using SAS (version 9.4; SAS Institute). Statistical significance was set at $P<.05$.

Data Collection

Data collection was performed using web-based questionnaires at baseline and at the 52-week follow-up. Study data were imported directly into a secure database (PostgreSQL; PostgreSQL Global Development Group).

Participants were compensated with US $50 for completing the 52-week web-based questionnaire and US $50 if they were eligible for and completed the associated biovalidation visit. Taking into consideration the 15 participant questionnaires conducted over the 52-week study period (with compensation of US $10-$50 per questionnaire) and the possibility of 2 previous biovalidation visits with compensation of US $50 each, participants could earn up to US $465 in total over the course of the 52-week study. Compensation was in the form of Visa or Mastercard gift cards that were emailed or mailed to their provided address approximately 2 to 3 weeks after completing the associated questionnaire or questionnaires or biovalidation visits. Remuneration was not tied to quitting smoking or use of one's study program.

Ethics Approval

The study was reviewed and approved by Solutions Institutional Review Board, LLC (protocol 2021/04/38) and registered with ClinicalTrials.gov (NCT04955639).

Results

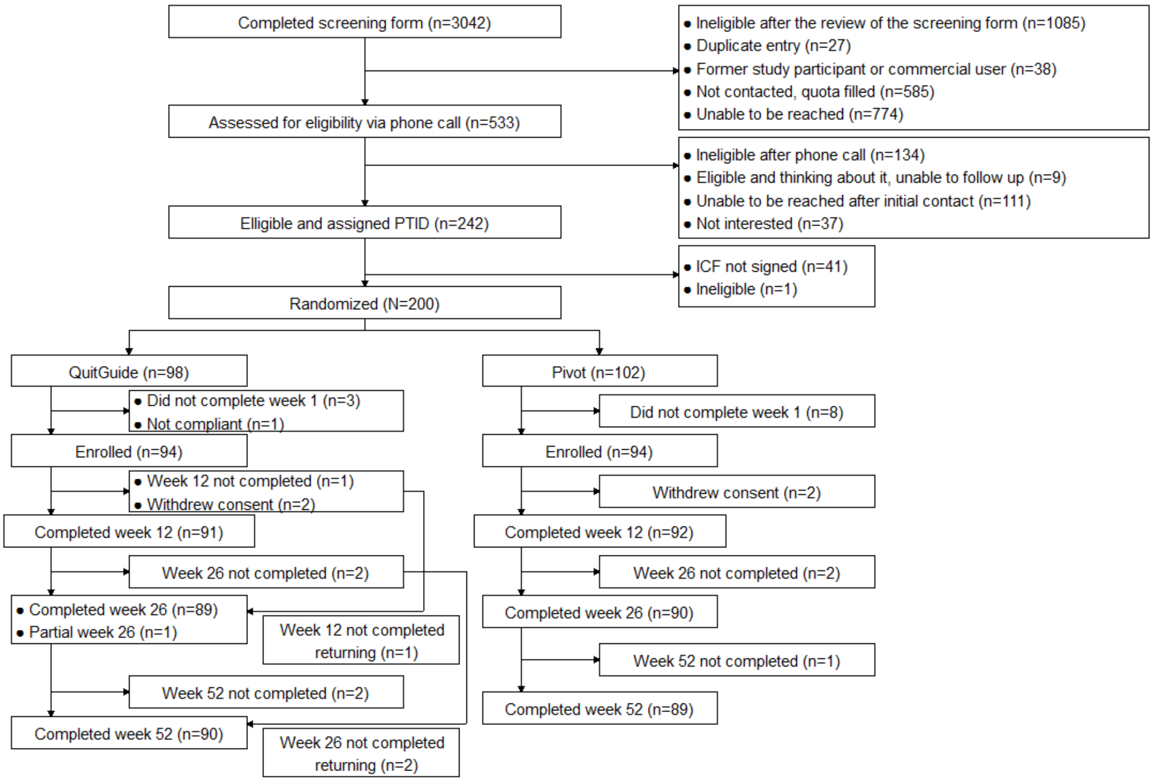
Enrollment and Questionnaire Completion

From June 2021 to October 2021, a total of 3042 web-based screening forms were received; 533 (17.52%) met the screening eligibility criteria and responded to an initial outbound phone call from the study staff. Of these 533 individuals, 188 (35.3%) were randomized and completed enrollment (n=94, 50% in each arm), comprising the ITT sample (188/3042, 6.18%).

Of the 3042 potential participants who completed the web-based screening form, 1085 (35.67%) were ineligible.Among these 1085 individuals, the reasons for ineligibility were readiness to quit (n=619, 57.1% not ready to quit within the next 30 days), quota filled for employment category (n=115, 10.6%), incompatible phone (n=72, 6.6%), form completed by someone other than the potential participant (n=44, 4.1%), currently smoking <7 days per week (n=44, 4.1%), currently smoking <5 CPD (n=38, 3.5%), being aged <21 years (n=5, 0.5%), and having ≥2 disqualifications (n=148, 13.6%).

In each arm, 2% (2/94) of the participants withdrew consent within the first 3 weeks of the study. The 52-week study questionnaire was completed by 95.2% (179/188) of the participants, specifically by 95% (89/94) in the Pivot arm and 96% (90/94) in the QuitGuide arm; these comprise the study responder samples. Study enrollment and attrition are depicted in the CONSORT (Consolidated Standards of Reporting Trials) flow diagram (Figure 1).

Figure 1. Study participant CONSORT (Consolidated Standards of Reporting Trials) flow diagram. ICF: informed consent form; PTID: participant identification number.



Baseline Characteristics

The study sample had a mean age of 46.4 (SD 9.2) years, comprised 55.3% (104/188) women, was predominantly White (128/188, 68.1%), smoked a mean of 17.6 (SD 9.0) CPD at baseline, and had been smoking for a mean of 26.8 (SD 10.3) years. The mean Heaviness of Smoking Index was 3.2 (SD 1.2). Participants represented 42 of the 50 states in the United States, along with the District of Columbia. The following states were not represented: Alaska, Delaware, Maine, Montana, North Dakota, New Hampshire, Vermont, and Wyoming. On average, participants had made 2.0 (SD 3.6) quit attempts over the 12 months before study entry. Baseline demographic characteristics and smoking behavior were balanced between the treatment groups at baseline. Participant baseline data are presented in Table 1.

Table 1. Participant baseline data (N=188).

Characteristic	All	Pivot (n=94)	QuitGuide (n=94)	*P* value
Demographics				
Age (years), mean (SD)	46.4 (9.2)	46.6 (10.1)	46.1 (8.2)	.70
Gender (women), n (%)	104 (55.3)	50 (53.2)	54 (57.4)	.56
Ethnicity and race, n (%)				.52
American Indian or Alaska Native	1 (0.5)	1 (1.1)	0 (0)	
Asian	1 (0.5)	0 (0)	1 (1.1)	
Black	36 (19.1)	15 (16)	21 (22.3)	
Hispanic, Latino, or Spanish origin	13 (6.9)	8 (8.5)	5 (5.3)	
Native Hawaiian	2 (1.1)	0 (0)	2 (2.1)	
White	128 (68.1)	66 (70.2)	62 (66)	
Some other race	3 (1.6)	2 (2.1)	1 (1.1)	
Prefer not to answer	4 (2.1)	2 (2.1)	2 (2.1)	
Education, n (%)				.64
Less than eighth grade	1 (0.5)	0 (0)	1 (1.1)	
Some high school	2 (1.1)	1 (1.1)	1 (1.1)	
High school or GED[a]	27 (14.4)	15 (16)	12 (12.8)	
Some college	80 (42.6)	35 (37.2)	45 (47.9)	
Associate’s (2-year) degree	28 (14.9)	13 (13.8)	15 (16)	
Bachelor’s (4-year) degree	31 (16.5)	18 (19.1)	13 (13.8)	
Master’s degree	15 (8)	10 (10.6)	5 (5.3)	
Professional or doctorate degree	4 (2.1)	2 (2.1)	2 (2.1)	
Income (US $), n (%)				.34
<25,000	32 (17)	14 (14.9)	18 (19.1)	
25,000-34,999	26 (13.8)	14 (14.9)	12 (12.8)	
35,000-49,999	42 (22.3)	19 (20.2)	23 (24.5)	
50,000-74,999	32 (17)	13 (13.8)	19 (20.2)	
75,000-99,999	23 (12.2)	12 (12.8)	11 (11.7)	
100,000-149,999	15 (8)	8 (8.5)	7 (7.4)	
≥150,000	10 (5.3)	8 (8.5)	2 (2.1)	
Prefer not to answer	8 (4.3)	6 (6.4)	2 (2.1)	
Employment, n (%)				.83
Yes, ≥20 h/wk	117 (62.2)	59 (62.8)	58 (61.7)	
Yes, <20 h/wk	37 (19.7)	17 (18.1)	20 (21.3)	
No	34 (18.1)	18 (19.1)	16 (17)	
Self-reported health, n (%)				.35
Excellent	5 (2.7)	4 (4.3)	1 (1.1)	
Very good	57 (30.3)	24 (25.5)	33 (35.1)	
Good	99 (52.7)	51 (54.3)	48 (51.1)	
Fair	26 (13.8)	14 (14.9)	12 (12.8)	
Poor	1 (0.5)	1 (1.1)	0 (0)	
Smartphone, n (%)				.30
iPhone	113 (60.1)	60 (63.8)	53 (56.4)	

Characteristic	All	Pivot (n=94)	QuitGuide (n=94)	*P* value
Android	75 (39.9)	34 (36.2)	41 (43.6)	
Smoking and quitting behavior				
Cigarettes smoked per day, mean (SD)	17.6 (9.0)	18.0 (9.6)	17.2 (8.5)	.55
Years smoking, mean (SD)	26.8 (10.3)	27.7 (10.4)	25.8 (10.1)	.21
First cigarette smoked after waking, n (%)				.52
Within 5 min	67 (35.6)	30 (31.9)	37 (39.4)	
6 to 30 min	92 (48.9)	47 (50)	45 (47.9)	
31 to 60 min	22 (11.7)	12 (12.8)	10 (10.6)	
After 60 min	7 (3.7)	5 (5.3)	2 (2.1)	
Tobacco products used, n (%)				.42
Cigarettes only	162 (86.2)	79 (84)	83 (88.3)	
Cigarettes+e-cigarettes or vaping	15 (8)	10 (10.6)	5 (5.3)	
Cigarettes+cigars	3 (1.6)	1 (1.1)	2 (2.1)	
Cigarettes+e-cigarettes or vaping+cigars	2 (1.1)	1 (1.1)	1 (1.1)	
Cigarettes+chew or snuff	2 (1.1)	2 (2.1)	0 (0)	
Cigarettes+e-cigarettes, vaping+chew, or snuff	1 (0.5)	0 (0)	1 (1.1)	
Cigarettes+e-cigarettes or vaping+pipe	1 (0.5)	0 (0)	1 (1.1)	
Cigarettes+hookah+cigars	1 (0.5)	0 (0)	1 (1.1)	
Cigarettes+hookah	1 (0.5)	1 (1.1)	0 (0)	
HSI[b], mean (SD)	3.2 (1.2)	3.2 (1.3)	3.2 (1.2)	.72
Quit attempts in the past 12 months, mean (SD)	2.0 (3.6)	1.9 (3.4)	2.2 (3.8)	.63
Methods used in past quit attempts[c], n (%)				—[d]
Cold turkey	140 (74.5)	67 (71.3)	73 (77.7)	
NRT[e]	92 (48.9)	53 (56.4)	39 (41.5)	
e-Cigarettes or vaping	65 (34.6)	33 (35.1)	32 (34)	
Varenicline or Chantix	50 (26.6)	23 (24.5)	27 (28.7)	
Bupropion, Zyban, or Wellbutrin	33 (17.6)	26 (27.7)	7 (7.4)	
None	16 (8.5)	5 (5.3)	11 (11.7)	
Hypnotherapy	11 (5.9)	7 (7.4)	4 (4.3)	
Classes for quitting smoking	10 (5.3)	7 (7.4)	3 (3.2)	
Acupuncture	10 (5.3)	7 (7.4)	3 (3.2)	
Smartphone app	9 (4.8)	6 (6.4)	3 (3.2)	
Counseling	4 (2.1)	2 (2.1)	2 (2.1)	
Other	4 (2.1)	1 (1.1)	3 (3.2)	
Attitudes toward quitting smoking, mean (SD)				—
DTQ[f]	3.5 (2.5)	3.5 (2.3)	3.6 (2.6)	.72
STQ[g]	4.5 (2.4)	4.6 (2.4)	4.3 (2.3)	.33

Characteristic	All	Pivot (n=94)	QuitGuide (n=94)	*P* value
SASEQ[h]	11.7 (4.8)	11.8 (4.7)	11.5 (4.9)	.69

[a]GED: general educational development.

[b]HSI: Heaviness of Smoking Index; low (0-1), medium (2-4), and high (5-6).

[c]Participants were asked to select all that applied.

[d]Not available.

[e]NRT: nicotine replacement therapy.

[f]DTQ: difficulty to stay quit; if you were to quit smoking right now, how difficult do you think it would be to stay smoke free? (1=really hard to stay quit; 10=really easy to stay quit).

[g]STQ: success to quit; if you were to quit smoking right now, how successful would you be? (1=not at all successful; 10=completely successful).

[h]SASEQ: Smoking Abstinence Self-Efficacy Questionnaire (score of 1-24).

Self-Efficacy

The SASEQ increased in both groups from baseline to week 52 (Table 2). The increase in the Pivot group of 3.2 (SD 7.8) points was significant (*P*<.001), whereas the increase in the QuitGuide group of 1.0 (SD 8.5) points was not (*P*=.26).

Table 2. Changes in the Smoking Abstinence Self-Efficacy Questionnaire (SASEQ) from baseline to 52 weeks (N=188).

SASEQ[a]	All		Pivot		QuitGuide		*P* value[b]	Point estimate[c] (95% CI)	*P* value[c]
	Participants, n (%)	Values, mean (SD)[d]	Participants, n (%)	Values, mean (SD)[e]	Participants, n (%)	Values, mean (SD)[f]			
Baseline	188 (100)	11.7 (4.8)	94 (50)	11.8 (4.7)	94 (50)	11.5 (4.9)	.69	N/A[g]	N/A
52 weeks	179 (95.2)	13.7 (8.1)	89 (47.3)	14.9 (7.9)	90 (47.9)	12.6 (8.1)	N/A	2.3 (0.01 to 4.6)	.049
Change	N/A	2.1 (8.2)	N/A	3.2 (7.8)	N/A	1.0 (8.5)	N/A	2.2 (−0.1 to 4.5)	.06

[a]SASEQ score of 1-24.

[b]2-tailed *t* test between Pivot and QuitGuide.

[c]Point estimate and corresponding *P* value obtained from linear regression adjusted with randomization covariates: daily smoking frequency (≤14 vs ≥15 cigarettes per day), employment status (full-time or part-time employment vs not employed), race and ethnicity (minority race and ethnicity vs non-Hispanic White), and expected difficulty staying quit (scale of 1-10; self-reported score of ≤5 vs ≥6).

[d]*P*<.001 (paired *t* test change, ie, difference from baseline to week 52).

[e]*P*<.001 (paired *t* test change, ie, difference from baseline to week 52).

[f]*P*=.26 (paired *t* test change, ie, difference from baseline to week 52).

[g]N/A: not applicable.

Smoking Behavior

Quit Attempts

Using an ITT analysis, at 52 weeks, 96.8% (182/188) of participants reported having made at least one quit attempt since the study started, with comparable proportions in each study group (90/94, 96% for Pivot and 92/94, 98% for QuitGuide; OR 0.5, 95% CI 0.1-2.8, *P*=.44; RR Poisson=1.0, 95% CI 0.9-1.0, *P*=.42). On average, QuitGuide participants reported more quit attempts (mean 7.0, SD 6.3 for Pivot vs 9.5, SD 7.5 for QuitGuide; IRR negative binomial=0.7, 95% CI 0.6-0.9; *P*=.01).

Change in CPD

Among participants who responded at 52 weeks (179/188, 95.2%), CPD were reduced by 63.4% (SD 39.3%) from baseline. Within each group, the reduction in CPD from baseline to week 52 was significant (*P*<.001 for both). The CPD reduction was similar between the 2 groups (mean −63.3%, SD 42% for Pivot vs −63.4%, SD 36.7% for QuitGuide; point estimate=0.5, 95% CI −10.8 to 11.9; *P*=.93).

Among the subset of participants who did not report 7-day (or greater) PPA at 52 weeks (111/188, 59%), CPD were reduced by 40.9% (SD 34.1%) from baseline. Within each group, the reduction in CPD from baseline to week 52 was significant (*P*<.001 for both). The reduction in CPD was greater in the QuitGuide group (mean −34.6%, SD 35.5% vs −46.1%, SD 32.3% for QuitGuide; point estimate=13.1, 95% CI 0.6-25.7; *P*=.04).

Among the participants who responded at 52 weeks (179/188, 95.2%), the proportion who reduced CPD by ≥50% was similar between the 2 groups (58/89, 65% for Pivot vs 63/90, 70% for QuitGuide; OR 0.8, 95% CI 0.4-1.5, *P*=.44; RR Poisson=0.9, 95% CI 0.8-1.1, *P*=.44).

Focusing on the subset of participants who did not report 7-day (or greater) PPA at 52 weeks (111/188, 59%), the proportion of those who reduced CPD by ≥50% was 38% (19/50) for Pivot versus 56% (34/61) for QuitGuide (OR 0.4, 95% CI 0.2-0.9, *P*=.04; RR=0.7, 95% CI 0.4-1.0, *P*=.07).

Cessation Rates

Cessation rates are detailed in Table 3, which includes previously published rates from 26 weeks for context [17]. At 52 weeks, differences between the 2 study groups in self-reported 7- and 30-day PPA rates and abstinence from all tobacco products were not statistically significant (all *P*>.05). In contrast, self-reported continuous abstinence, biochemically confirmed abstinence, and biochemically confirmed continuous abstinence rates were significantly higher in the Pivot group (all *P*<.05).

Self-reported continuous abstinence (ITT) was achieved in 20% (19/94) of the Pivot participants versus 9% (8/94) of the QuitGuide participants (OR 2.8, 95% CI 1.1-6.9, *P*=.03; RR=2.4, 95% CI 1.1-5.1, *P*=.03). Biochemically confirmed abstinence (ITT) was achieved in 31% (29/94) of the Pivot participants versus 18% (17/94) of the QuitGuide participants (OR 2.1, 95% CI 1.0-4.3, *P*=.04; RR=1.9, 95% CI 1.1-3.1, *P*=.02). Biochemically confirmed continuous abstinence (ITT) was achieved in 19% (18/94) of the Pivot participants versus 9% (8/94) of the QuitGuide participants (OR 2.6, 95% CI 1.1-6.5, *P*=.04; RR=2.2, 95% CI 1.0-4.9, *P*=.046).

Notably, the participation rate in the 52-week biovalidation visit was 75% (51/68) overall—79% (31/39) in the Pivot group and 69% (20/29) in the QuitGuide group (OR 1.5, 95% CI 0.5-4.8, *P*=.51; RR=Poisson 1.1, 95% CI 0.8-1.5, *P*=.63). The 5 study participants (n=2, 40% in the Pivot group and n=3, 60% in the QuitGuide group) who completed a biovalidation visit but did not achieve biovalidated abstinence all reported a change in smoking status at the outset of the visit and, therefore, did not provide a breath sample during the visit. At the 52-week visit, no study participants had a breath sample value that was discordant with their self-reported abstinence.

Table 3. Smoking cessation rates at 26 and 52 weeks (N=188).

Outcome	Overall, n (%)	Pivot (n=94), n (%)	QuitGuide (n=94), n (%)	Adjusted OR[a] (95% CI)	OR *P* value	RR[b] (95% CI)	RR *P* value
7-day PPA[c]							
52-week ITT[d]	68 (36.2)	39 (41.5)	29 (30.9)	1.6 (0.9-3.0)	.12	1.3 (0.9-2.0)	.14
26-week ITT	59 (31.4)	34 (36.2)	25 (26.6)	1.7 (0.9-3.2)	.12	1.4 (0.9-2.1)[e]	.12
52-week responder analysis[f]	68 (38)	39 (43.8)	29 (32.2)	1.6 (0.9-3.0)	.11	1.3 (0.9-2.0)	.15
26-week responder analysis[g]	59 (32.8)	34 (37.8)	25 (27.8)	1.7 (0.9-3.2)	.13	1.5 (1.0-2.3)	.06
30-day PPA							
52-week ITT	62 (33)	34 (36.2)	28 (29.8)	1.4 (0.7-2.5)	.33	1.2 (0.8-1.9)	.34
26-week ITT	51 (27.1)	30 (31.9)	21 (22.3)	1.7 (0.9-3.4)	.12	1.4 (0.9-2.2)	.18
52-week responder analysis	62 (34.6)	34 (38.2)	28 (31.1)	1.4 (0.7-2.6)	.32	1.2 (0.8-1.9)	.34
26-week responder analysis	51 (28.3)	30 (33.3)	21 (23.3)	1.7 (0.9-3.4)	.13	1.4 (0.9-2.22)	.19
Biovalidated abstinence							
52-week ITT	46 (24.5)	29 (30.9)	17 (18.1)	2.1 (1.0-4.3)	.04	1.9 (1.1-3.1)	.02
26-week ITT	40 (21.3)	26 (27.7)	14 (14.9)	2.3 (1.1-4.8)	.03	1.9 (1.1-3.5)	.02
52-week responder analysis	46 (25.7)	29 (32.6)	17 (18.9)	2.1 (1.1-4.3)	.04	1.9 (1.1-3.1)	.02
26-week responder analysis	40 (22.2)	26 (28.9)	14 (15.6)	2.3 (1.1-4.8)	.03	1.9 (1.1-3.4)	.02
Self-reported continuous abstinence							
52-week ITT	27 (14.4)	19 (20.2)	8 (8.5)	2.8 (1.1-6.9)	.03	2.4 (1.1-5.1)[e]	.03
26-week ITT	39 (20.7)	24 (25.5)	15 (16)	1.9 (0.9-3.8)	.10	1.6 (0.9-2.8)	.11
52-week responder analysis	27 (15.1)	19 (21.3)	8 (8.9)	2.8 (1.1-6.8)	.03	2.3 (1.1-5.0)	.03
26-week responder analysis	39 (21.7)	24 (26.7)	15 (16.7)	1.8 (0.9-3.9)	.11	1.6 (0.9-2.8)	.12
Biovalidated continuous abstinence							
52-week ITT	26 (13.8)	18 (19.1)	8 (8.5)	2.6 (1.1-6.5)	.04	2.2 (1.0-4.9)[e]	.046
26-week ITT	29 (15.4)	20 (21.3)	9 (9.6)	2.7 (1.1-6.4)	.03	2.2 (1.1-4.6)	.03
52-week responder analysis	26 (14.5)	18 (20.2)	8 (8.9)	2.6 (1.04-6.4)	.04	2.2 (1.0-4.8)	.047
26-week responder analysis	29 (16.1)	20 (22.2)	9 (10)	2.7 (1.1-6.3)	.03	2.3 (1.1-4.7)	.02
Self-reported abstinence from all tobacco products							
52-week ITT	60 (31.9)	33 (35.1)	27 (28.7)	1.4 (0.7-2.6)	.33	1.2 (0.8-1.8)	.41
26-week ITT	55 (29.3)	32 (34)	23 (24.5)	1.6 (0.9-3.1)	.13	1.5 (1.0-2.3)	.06
52-week responder analysis	60 (33.5)	33 (37.1)	27 (30)	1.4 (0.7-2.6)	.33	1.2 (0.8-1.8)	.44
26-week responder analysis	55 (30.6)	32 (35.6)	23 (25.6)	1.6 (0.8-3.1)	.16	1.5 (1.0-2.2)	.08

[a]OR: odds ratio.

[b]RR: adjusted relative risk using log-link binomial regression unless otherwise noted (eg, it does not converge and log-link Poisson regression is used).

[c]PPA: point-prevalence abstinence.

[d]ITT: intention-to-treat analysis at 52 and 26 weeks; N=188 (n=94, 50% in the Pivot group and n=94, 50% in the QuitGuide group).

[e]Log-link Poisson regression used.

[f]Responder analysis at 52 weeks; n=179 (n=89, 49.7% in the Pivot group and n=90, 50.3% in the QuitGuide group).

[g]Responder analysis at 26 weeks; n=180 (n=90, 50% in the Pivot group and n=90, 50% in the QuitGuide group).

Use of NRT

At 52 weeks, 99% (93/94) of the Pivot participants had ordered NRT compared with 85% (80/94) of the QuitGuide participants (RR Poisson=1.2, 95% CI 1.1-1.3; *P*<.001). The average number of NRT orders placed per participant was 3.3 (SD 2.0) in the Pivot group and 1.8 (SD 1.6) in the QuitGuide group (IRR=1.8, 95% CI 1.5-2.2; *P*<.001). Combination therapy (patch+gum or patch+lozenge) was the most common regimen among participants (Table 4).

Table 4. Nicotine replacement therapy (NRT) orders placed by participants through 52 weeks (*P*<.001; N=188)[a].

NRT order type	All, n (%)	Pivot (n=94), n (%)	QuitGuide (n=94), n (%)
≥1 NRT single therapy[b] order	31 (16.5)	22 (23.4)	9 (9.6)
≥1 NRT combination therapy[c] order	99 (52.7)	42 (44.7)	57 (60.6)
≥1 NRT single therapy+≥1 NRT combination therapy order	43 (22.9)	29 (30.9)	14 (14.9)
None	15 (8)	1 (1.1)	14 (14.9)

[a]Multinomial logistic regression adjusted for randomization covariates.

[b]Single therapy: nicotine patch alone, nicotine gum alone, or nicotine lozenge alone.

[c]Combination therapy: nicotine patch+nicotine gum or nicotine patch+nicotine lozenge.

Participant Feedback

Among participants who reported 7-day PPA at 52 weeks (68/188, 36.2%), most reported that their study program helped them quit smoking (true or false response; 35/39, 90% in the Pivot group vs 26/29, 90% in the QuitGuide group; RR Poisson=1.0, 95% CI 0.9-1.2; *P*=.73).

Predictors

We performed exploratory post hoc analyses using logistic regression to explore the qualitative associations between baseline characteristics and smoking behavior outcomes (Multimedia Appendix 1). We identified 3 variables with statistically significant correlation: age, education, and DTQ. None of these variables were considered predictive across all 5 smoking cessation outcomes.

Older age at baseline was associated with a decrease in self-reported 7- and 30-day PPA after adjusting for the randomization cohort. In addition, age had an interaction with cohort such that, in Pivot participants, older age was associated with achieving self-reported and biochemically confirmed continuous abstinence. In contrast, older age in QuitGuide participants was associated with a decreased likelihood of self-reported and biochemically confirmed continuous abstinence.

Regarding education, those who did not have any college experience were more likely to achieve biochemically confirmed abstinence compared with those with some college, a 2-year degree, or ≥4 years of college.

Finally, contrasting correlations were observed for DTQ in each cohort. In Pivot participants, a higher DTQ at entry was associated with a decreased likelihood of achieving self-reported 30-day PPA. In QuitGuide participants, a higher DTQ at entry was associated with an increased likelihood of self-reported 30-day PPA.

Discussion

Principal Findings

In this paper, we report the results at 52 weeks of this pilot RCT comparing the Pivot and QuitGuide mobile smoking cessation programs among 188 adult cigarette smokers in the United States. Pivot had higher smoking cessation rates for self-reported continuous abstinence (19/94, 20% for Pivot vs 8/94, 9% for QuitGuide; *P*=.03), biochemically confirmed abstinence (29/94, 31% for Pivot vs 17/94, 18% for QuitGuide; *P*=.04), and biochemically confirmed continuous abstinence (18/94, 19% for Pivot vs 8/94, 9% for QuitGuide; *P*=.046). Pivot participants also had a significant increase in self-efficacy via the SASEQ, whereas QuitGuide participants did not. In addition, more Pivot participants (93/94, 99% in the Pivot group vs 80/94, 85% in the QuitGuide group; *P*<.001) placed NRT orders (Pivot mean 3.3, SD 2.0 vs QuitGuide mean 1.8, SD 1.6; *P*<.001).

On average, QuitGuide participants made more quit attempts (7.0, SD 6.3 for Pivot vs 9.5, SD 7.5 for QuitGuide; *P*=.01). Among the subset of participants who did not report 7-day (or greater) PPA at 52 weeks, QuitGuide participants had a greater reduction in CPD (−34.6% for Pivot vs −46.1% for QuitGuide; *P*=.04). Among those who reported 7-day PPA at 52 weeks, most in each arm (35/39, 90% in the Pivot group and 26/29, 90% in the QuitGuide group) indicated that their study program helped them quit smoking.

Smoking Cessation Rates: Comparison With Prior Work

The availability of comparative long-term data from RCTs assessing mobile smoking cessation programs is limited, as shown in Table 5. At 1 year, 7-day PPA rates were similar between this study and the RCT by Webb et al [16] assessing Quit Genius (39/94, 42% in the Pivot group and 35% in the Quit Genius group) and slightly higher than the 29% reported by Bricker et al [14] for iCanQuit. A similar pattern was reported for durable abstinence rates at 1 year (20%-22% for this study and the RCT by Webb et al [16] and 10% for the RCT by Bricker et al [14]). Notably, this study and the RCT by Webb et al [16] provided participants with NRT; the RCT by Bricker et al [14] did not. This may have contributed to the differences in cessation rates.

Table 5. Comparative 1-year smoking cessation outcomes from randomized controlled trials assessing mobile smoking cessation programs.

Study	Outcome								
	7-day PPA[a], n/N (%)		Statistical analysis	30-day PPA, n/N (%)		Statistical analysis	Durable abstinence, n/N (%)		Statistical analysis
	Intervention	Control		Intervention	Control		Intervention	Control	
Bricker et al[b] [14]	356/1214 (29.3)	302/1201 (25.1)	OR[c] 1.26, 95% CI 1.05-1.52; *P*=.01	293/1214 (24.1)	225/1201 (18.7)	OR 1.40, 95% CI 1.14-1.71; *P*=.001	116/1214[d] (9.6)	65/1201 (5.4)	OR 1.84, 95% CI 1.34-2.53; *P*<.001
Webb et al[e] [16]	92/265 (34.7)	78/265 (29.4)	IRR[f] 1.20, 95% CI 0.94-1.54; *P*=.19	Not assessed	Not assessed	N/A[g]	58/265[h] (21.9)	30/265 (11.3)	RR[i] 1.93, 95% CI 1.29-2.90; *P*=.002
Marler et al[j] [17]	39/94 (41.5)	29/94 (30.9)	OR 1.6, 95% CI 0.9-3.0; *P*=.12	34/94 (36.2)	28/94 (30)	OR 1.4, 95% CI 0.7-2.5; *P*=.33	19/94[k] (20.2)	8/94 (8.5)	OR 2.8, 95% CI 1.1-6.9; *P*=.03

[a]PPA: point-prevalence abstinence. Participants were considered to have achieved self-reported 7-day (30-day) PPA if they answered "no" to the following question: "In the last 7 (30) days have you smoked any cigarettes, even a single puff?"

[b]Intervention: iCanQuit; control: QuitGuide.

[c]OR: odds ratio.

[d]Prolonged abstinence defined as the time since the date of the last cigarette. The dates of the last cigarette that occurred between 0 and 90 days after randomization were categorized as prolonged abstinence. Dates that occurred ≥91 days after randomization were categorized as not prolonged abstinence.

[e]Intervention: Quit Genius; control: very brief advice.

[f]IRR: incidence rate ratio.

[g]N/A: not applicable.

[h]Sustained abstinence was defined as smoking no more than 5 cigarettes from the quit date to the 52-week follow-up.

[i]RR: relative risk.

[j]Intervention: Pivot; control: QuitGuide.

[k]Continuous abstinence was defined as self-reporting 7-day (or greater) PPA at 12 weeks, self-reporting 30-day PPA at 26 and 52 weeks, no more than 5 cigarettes between 12 and 26 weeks, and 0 cigarettes between 26 and 52 weeks.

Differences in Other Smoking-Related Behaviors and Outcomes Between Pivot and QuitGuide

There were notable differences in other smoking-related outcomes and behaviors between the 2 study groups. Specifically, at 52 weeks, participants in the Pivot group reported a significant increase in self-efficacy via the SASEQ, whereas QuitGuide participants did not. Moreover, more Pivot participants ordered NRT over the course of the study, with higher mean NRT orders per person. Both higher self-efficacy and NRT use are associated with an increased likelihood of cessation [26,42-46], and these factors may have contributed to the higher cessation rates in the Pivot group.

QuitGuide participants made, on average, more quit attempts and achieved a greater reduction in CPD among individuals who did not report abstinence. The literature provides a mixed picture of quit attempts, acknowledging that it can take an average of 30 quit attempts before successfully quitting for 1 year or longer [47]. By this logic, more quit attempts bring one closer to quitting, and setting realistic expectations accordingly may be helpful. However, there have also been reports of psychological distress, anxiety, and frustration that come with failed quit attempts [48-54]. Together, these data suggest that both the quantity and quality of quit attempts are important factors in successful cessation. A reduction in CPD has been reported as a meaningful behavior change that increases the probability of future cessation [55]. As the study participants continue surveillance, it remains to be seen whether this outcome in the QuitGuide group will translate to improved quit rates in the future.

Predictors

Exploratory post hoc analyses using logistic regression identified age, education, and DTQ as baseline variables associated with smoking cessation outcomes. Considering the exploratory nature of these analyses, the relatively small number of participants in the different baseline categories, and that none of these baseline variables had statistically significant correlations for all 5 smoking cessation outcomes, caution should be used in the interpretation of these data. Accordingly, we have limited our interpretations to a qualitative assessment. However, a review of the literature also reveals previous identification of these variables as associated with smoking cessation.

Several studies have reported that older age is associated with smoking cessation [56-58], which was true in the Pivot group for self-reported and biochemically confirmed continuous abstinence. Interestingly, in the QuitGuide group, older age was associated with a lower likelihood of self-reported and biochemically confirmed continuous abstinence. This finding has been reported elsewhere; however, the age effect disappeared when controlling for heaviness of smoking [59].

The finding that a lower educational level (no college experience) was associated with a higher likelihood of biochemically confirmed abstinence has also been reported elsewhere [60,61]. However, reports on the association between higher educational levels and successful cessation are more common [19,57,58,62-64].

Finally, a higher DTQ at study entry was associated with a lower likelihood of self-reported 30-day PPA in the Pivot group and a higher likelihood in the QuitGuide group. Approaching the DTQ metric used in this study as a general proxy for self-efficacy, the literature predominantly supports the association between higher self-efficacy and a greater likelihood of cessation and between lower self-efficacy and a lower likelihood of cessation and a greater likelihood of relapse [19,42-46].

Overall, the data related to predictors of cessation are nuanced and likely heavily influenced by the presence, type, and use of a cessation program and by the cessation program characteristics. For example, some populations may respond better to specific program tools or delivery mechanisms than others. The value of these analyses is that they highlight variables of interest for future study designs and assist in teasing out optimal pairings between populations and cessation approaches.

Strengths and Limitations

This study has several strengths. First, nonproportional quota sampling achieved a diverse and balanced population. The smoking cessation programs compared in this study were also of the same modality, decreasing the likelihood of modality-related confounding. In addition, QuitGuide as a control program provides sufficient context for the study outcomes as a well-established and well-studied digital cessation program [14,15,17,33]. Another strength of this study is the inclusion of 3 separate biovalidation visits for those who reported at least 7-day abstinence at 12, 26, and 52 weeks, providing validation for self-reported claims of smoking abstinence. In addition, the 52-week follow-up provided additional insights into the effectiveness of each program for smoking cessation on a longer-term scale. Finally, the following measures of study participation remained robust for 52 weeks: retention (approximately 89/94, 95% in each arm), survey completion (≥92%; ≥87/94 for each survey), and biovalidation visit completion (151/186, 81.2% overall).

This study also has several limitations. First, as a pilot study, it was not powered for cessation outcomes. The self-reported PPA and abstinence from all tobacco products outcomes were not significantly different between the 2 study groups, whereas the biovalidated abstinence and self-reported continuous abstinence outcomes were. Whether a larger study powered for these outcomes would have resulted in significant differences in smoking abstinence outcomes is unknown.

Second, the Pivot program includes the following additional tools not included in the QuitGuide program: a CO breath sensor, SMS text messaging–based counseling with a tobacco cessation coach, and a moderated web-based community support forum. This study compares the 2 programs but cannot determine the specific effects of these additional tools. Accordingly, the appropriate focus of future assessments includes whether and to what extent these tools contribute to Pivot outcomes.

Third, we cannot rule out some influence on participant selection by the recruiting and enrollment process and how this might influence the generalizability of study outcomes to the population of smokers as a whole. We took steps to mitigate factors that might incentivize enrollment, such as the availability of free NRT and study compensation. Specifically, these study characteristics were not mentioned in the recruiting advertisements or web-based screening form or until later in the screening phone calls. Potential participants were deemed ineligible if they were already using NRT or other smoking cessation tools at study entry. Steps were also taken to minimize the possible influence of compensation, including not tying compensation to outcome or program use, delaying payments 2 to 3 weeks after the completion of compensated events, and keeping payment amounts conservative. Recruitment was conducted on social media throughout the United States, and ultimately, study participants represented 42 of the 50 states, along with the District of Columbia. We also used nonproportional quota sampling for several participant characteristics (age group, gender, race, and employment status), with potential participants called on a first-come, first-served basis accordingly. All nonproportional quota sampling goals were met. Nonetheless, it is always important to consider factors that may influence the generalizability of study outcomes to larger populations.

Fourth, after randomization, all researchers were unblinded to participant group allocation, which can result in unbalanced participant communication and data collection efforts. To mitigate this, the study design included scheduled, standardized, and scripted participant communications (written and verbal) reviewed by the institutional review board. High and comparable questionnaire and biovalidation visit completion rates (≥87/94, ≥92% for questionnaires and ≥69% for biovalidation visits at 12, 26, and 52 weeks in both study arms) reflect favorably on our attempt to minimize these possible effects.

Fifth, exhaled CO as a biovalidation test for smoking cessation is imperfect. The half-life of CO is, on average, 4 hours and is influenced by activity level (ie, shorter half-life when exercising and longer half-life when sleeping). Accordingly, smokers may be able to abstain from smoking for several hours before providing a breath sample and obtain a CO value consistent with “not smoking”; we cannot exclude this occurrence during the biovalidation visits. Moreover, secondhand smoke, use of other combustible substances such as marijuana, and environmental or occupational CO exposure can increase CO levels. That said, the limitations of other biovalidation methods made exhaled CO, which is noninvasive, less expensive, and easy for a lay user to perform, the preferred option. Specifically, although cotinine, a nicotine metabolite, has a longer half-life (≥8-30 hours) than CO and, therefore, requires longer abstinence periods (2-7 days) to reach “nonsmoking” levels, its collection from body fluids is more onerous and will yield positive results in individuals using NRT, which was problematic with our study design. Anabasine and anatabine are minor tobacco alkaloids that are specific for tobacco-derived products (eg, cigarettes, cigars, and smokeless tobacco). They are well suited for testing individuals who use NRT for tobacco use. However, these biomarkers require urine collection and chromatography–mass spectrometry measurement [65]. Altogether, when considering the remote nature of this study and the provision of NRT, we felt that exhaled CO, despite its imperfections, was the best option for biovalidation.

Finally, this study did not address the reported differential cessation outcomes in the context of cost-effectiveness. Although it is beyond the scope of this study, a cost-benefit analysis is an appropriate and logical next step. Resource use is an important consideration in the assessment of mobile tobacco cessation interventions. We have sought to establish foundational outcomes in Pivot; these inputs should be included in future cost-benefit assessments.

Conclusions

In this pilot RCT comparing the Pivot and QuitGuide mobile smoking cessation programs, Pivot had higher abstinence rates at 52 weeks. The context of this outcome in a study powered for engagement underscores the effectiveness of the intervention. Moreover, the treatment effect was durable, with stable biovalidated continuous abstinence rates from 26 to 52 weeks. This study adds to the small but growing body of RCT-derived evidence focused on longer-term outcomes among adult smokers using mobile smoking cessation programs. The data are encouraging and consistent, pointing to an important role for these types of programs in the larger effort to curb nicotine dependence.

Acknowledgments

The authors thank Alex Kolodychak for his direct role in the recruitment and collaborative tracking of study participants.

Data Availability

Deidentified data will be shared upon reasonable request stating the purpose to the corresponding author, which is subject to approval by the sponsor.

Authors' Contributions

JDM, CAF, DJB, and DSU designed the study. JDM, CAF, and MTU recruited participants. JDM and DJB oversaw the study. CAF managed the database. CAF and JAG determined the most appropriate analyses and performed them. MTU performed many of the functions of study conduct, including managing participant interactions, data collection, participant compliance with questionnaires and visits, scheduling and running of biovalidation visits, and study payments. CAF and JDM assisted with the aforementioned study conduct tasks as needed. JDM, MTU, and CAF prepared the original draft of the manuscript. JDM, CAF, MTU, DJB, DSU, and JAG reviewed and edited the manuscript before submission.

Conflicts of Interest

JDM, CAF, MTU, DJB, and DSU are current employees of Pivot Health Technologies Inc (Pivot), the developer of the Pivot smoking cessation program. They receive salary and stock options from Pivot. DSU is the president and chief executive officer of Pivot and an investor in the company. JAG is a paid statistical consultant.

Multimedia Appendix 1

Intention-to-treat (N=188) logistic regression analyses of baseline predictors of 7-day point-prevalence abstinence (PPA), 30-day PPA, biovalidated PPA, continuous abstinence, and biovalidated continuous abstinence at 52 weeks among all study participants.

[DOCX File , 34 KB - mhealth_v11i1e48157_app1.docx]

Multimedia Appendix 2

CONSORT-EHEALTH Checklist (V 1.6.1).

[PDF File (Adobe PDF File), 10854 KB - mhealth_v11i1e48157_app2.pdf]

References

1. Cornelius ME, Loretan CG, Wang TW, Jamal A, Homa DM. Tobacco product use among adults - United States, 2020. MMWR Morb Mortal Wkly Rep 2022 Mar 18;71(11):397-405 [FREE Full text] [doi: 10.15585/mmwr.mm7111a1] [Medline: 35298455]
2. Smoking and health: report of the advisory committee to the surgeon general of the public health service. U.S. Department of Health, Education, and Welfare. 1964. URL: https://www.govinfo.gov/content/pkg/GPO-SMOKINGANDHEALTH/pdf/GPO-SMOKINGANDHEALTH.pdf [accessed 2023-04-03]
3. Smoking cessation: a report of the surgeon general. U.S. Department of Health and Human Services. 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-full-report.pdf [accessed 2023-04-03]

4. Patnode CD, Henderson JT, Coppola EL, Melnikow J, Durbin S, Thomas RG. Interventions for tobacco cessation in adults, including pregnant persons: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2021 Jan 19;325(3):280-298. [doi: 10.1001/jama.2020.23541] [Medline: 33464342]
5. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3(3):CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
6. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
7. Mobile fact sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/#panel-305d9a4e-2c6c-4194-86c6-9ab21ac76df0 [accessed 2023-04-03]
8. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
9. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
10. Ubhi HK, Kotz D, Michie S, van Schayck OC, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
11. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
12. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]
13. Etter JF, Khazaal Y. The Stop-tabac smartphone application for smoking cessation: a randomized controlled trial. Addiction 2022 May;117(5):1406-1415 [FREE Full text] [doi: 10.1111/add.15738] [Medline: 34738687]
14. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
15. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
16. Webb J, Peerbux S, Ang A, Siddiqui S, Sherwani Y, Ahmed M, et al. Long-term effectiveness of a clinician-assisted digital cognitive behavioral therapy intervention for smoking cessation: secondary outcomes from a randomized controlled trial. Nicotine Tob Res 2022 Oct 26;24(11):1763-1772 [FREE Full text] [doi: 10.1093/ntr/ntac113] [Medline: 35470860]
17. Marler JD, Fujii CA, Utley MT, Balbierz DJ, Galanko JA, Utley DS. Outcomes of a comprehensive mobile smoking cessation program with nicotine replacement therapy in adult smokers: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Nov 24;10(11):e41658 [FREE Full text] [doi: 10.2196/41658] [Medline: 36257323]
18. Home page. Study Randomizer. URL: https://www.studyrandomizer.com/ [accessed 2021-03-24]
19. Marler JD, Fujii CA, Galanko JA, Balbierz DJ, Utley DS. Durability of abstinence after completing a comprehensive digital smoking cessation program incorporating a mobile app, breath sensor, and coaching: cohort study. J Med Internet Res 2021 Feb 15;23(2):e25578 [FREE Full text] [doi: 10.2196/25578] [Medline: 33482628]
20. Schuck K, Otten R, Kleinjan M, Bricker JB, Engels RC. Predictors of cessation treatment outcome and treatment moderators among smoking parents receiving quitline counselling or self-help material. Prev Med 2014 Dec;69:126-131. [doi: 10.1016/j.ypmed.2014.09.014] [Medline: 25278424]
21. Kendzor DE, Reitzel LR, Mazas CA, Cofta-Woerpel LM, Cao Y, Ji L, et al. Individual- and area-level unemployment influence smoking cessation among African Americans participating in a randomized clinical trial. Soc Sci Med 2012 May;74(9):1394-1401 [FREE Full text] [doi: 10.1016/j.socscimed.2012.01.013] [Medline: 22405506]
22. Falba T, Teng HM, Sindelar JL, Gallo WT. The effect of involuntary job loss on smoking intensity and relapse. Addiction 2005 Sep;100(9):1330-1339 [FREE Full text] [doi: 10.1111/j.1360-0443.2005.01150.x] [Medline: 16128722]
23. Weden MM, Astone NM, Bishai D. Racial, ethnic, and gender differences in smoking cessation associated with employment and joblessness through young adulthood in the US. Soc Sci Med 2006 Jan;62(2):303-316. [doi: 10.1016/j.socscimed.2005.06.009] [Medline: 16029919]
24. Arcaya M, Glymour MM, Christakis NA, Kawachi I, Subramanian SV. Individual and spousal unemployment as predictors of smoking and drinking behavior. Soc Sci Med 2014 Jun;110:89-95 [FREE Full text] [doi: 10.1016/j.socscimed.2014.03.034] [Medline: 24727666]
25. Kotz D, West R. Explaining the social gradient in smoking cessation: it's not in the trying, but in the succeeding. Tob Control 2009 Feb 01;18(1):43-46 [FREE Full text] [doi: 10.1136/tc.2008.025981] [Medline: 18936053]

26. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018 May 31;5(5):CD000146 [FREE Full text] [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
27. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Interventions for tobacco smoking cessation in adults, including pregnant persons: US preventive services task force recommendation statement. JAMA 2021 Jan 19;325(3):265-279. [doi: 10.1001/jama.2020.25019] [Medline: 33464343]
28. Vinci C. Cognitive behavioral and mindfulness-based interventions for smoking cessation: a review of the recent literature. Curr Oncol Rep 2020 May 16;22(6):58 [FREE Full text] [doi: 10.1007/s11912-020-00915-w] [Medline: 32415381]
29. Williams GC, Niemiec CP, Patrick H, Ryan RM, Deci EL. Outcomes of the smoker's health project: a pragmatic comparative effectiveness trial of tobacco-dependence interventions based on self-determination theory. Health Educ Res 2016 Dec;31(6):749-759 [FREE Full text] [doi: 10.1093/her/cyw046] [Medline: 27923864]
30. Hartmann-Boyce J, Livingstone-Banks J, Ordóñez-Mena JM, Fanshawe TR, Lindson N, Freeman SC, et al. Behavioural interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2021 Jan 04;1:CD013229. [doi: 10.1002/14651858.CD013229.pub2] [Medline: 33411338]
31. U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. Home page. Smokefree. URL: https://smokefree.gov/ [accessed 2022-11-05]
32. Smokefree apps. Smokefree. URL: https://smokefree.gov/tools-tips/apps [accessed 2022-08-01]
33. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
34. Shahab L. Why use CO-verified 4-week quit rates as the primary measure of stop smoking service success? NHS Centre for Smoking Cessation and Training Briefings. NHS Centre for Smoking Cessation and Training Briefings. 2014. URL: https://discovery.ucl.ac.uk/id/eprint/1450797/1/ [accessed 2022-07-12]
35. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary outcomes of a digital therapeutic intervention for smoking cessation in adult smokers: randomized controlled trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
36. Borland R, Yong HH, O'Connor RJ, Hyland A, Thompson ME. The reliability and predictive validity of the heaviness of smoking index and its two components: findings from the international tobacco control four country study. Nicotine Tob Res 2010 Oct;12 Suppl(Suppl 1):S45-S50 [FREE Full text] [doi: 10.1093/ntr/ntq038] [Medline: 20889480]
37. Hall SM, Havassy BE, Wasserman DA. Commitment to abstinence and acute stress in relapse to alcohol, opiates, and nicotine. J Consult Clin Psychol 1990 Apr;58(2):175-181. [doi: 10.1037//0022-006x.58.2.175] [Medline: 2335634]
38. Hall SM, Havassy BE, Wasserman DA. Effects of commitment to abstinence, positive moods, stress, and coping on relapse to cocaine use. J Consult Clin Psychol 1991 Aug;59(4):526-532. [doi: 10.1037//0022-006x.59.4.526] [Medline: 1918556]
39. Spek V, Lemmens F, Chatrou M, van Kempen S, Pouwer F, Pop V. Development of a smoking abstinence self-efficacy questionnaire. Int J Behav Med 2013 Sep;20(3):444-449 [FREE Full text] [doi: 10.1007/s12529-012-9229-2] [Medline: 22350635]
40. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005 Aug 01;162(3):199-200. [doi: 10.1093/aje/kwi188] [Medline: 15987728]
41. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002 Aug 30;21(16):2409-2419. [doi: 10.1002/sim.1047] [Medline: 12210625]
42. Herd N, Borland R, Hyland A. Predictors of smoking relapse by duration of abstinence: findings from the International Tobacco Control (ITC) four country survey. Addiction 2009 Dec;104(12):2088-2099 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02732.x] [Medline: 19922574]
43. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324 [FREE Full text] [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]
44. Palmer AM, Sutton SK, Correa JB, Simmons VN, Brandon TH. Abstinence-related motivational engagement for smoking cessation: longitudinal patterns and predictive validity. PLoS One 2021 Mar 04;16(3):e0247867 [FREE Full text] [doi: 10.1371/journal.pone.0247867] [Medline: 33661996]
45. Martínez-Vispo C, López-Durán A, Senra C, Becoña E. Specific relapse predictors: could cognitive-behavioral treatment for smoking cessation be improved? Int J Environ Res Public Health 2020 Jun 17;17(12):4317 [FREE Full text] [doi: 10.3390/ijerph17124317] [Medline: 32560325]
46. Blevins CE, Farris SG, Brown RA, Strong DR, Abrantes AM. The role of self-efficacy, adaptive coping, and smoking urges in long-term cessation outcomes. Addict Disord Their Treat 2016 Dec;15(4):183-189 [FREE Full text] [doi: 10.1097/ADT.0000000000000087] [Medline: 28133440]
47. Chaiton M, Diemert L, Cohen JE, Bondy SJ, Selby P, Philipneri A, et al. Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. BMJ Open 2016 Jun 09;6(6):e011045 [FREE Full text] [doi: 10.1136/bmjopen-2016-011045] [Medline: 27288378]

48. van der Deen FS, Carter KN, Wilson N, Collings S. The association between failed quit attempts and increased levels of psychological distress in smokers in a large New Zealand cohort. BMC Public Health 2011 Jul 28;11:598 [FREE Full text] [doi: 10.1186/1471-2458-11-598] [Medline: 21798059]
49. McClave AK, Dube SR, Strine TW, Kroenke K, Caraballo RS, Mokdad AH. Associations between smoking cessation and anxiety and depression among U.S. adults. Addict Behav 2009 Jun;34(6-7):491-497. [doi: 10.1016/j.addbeh.2009.01.005] [Medline: 19217720]
50. Dube SR, Caraballo RS, Dhingra SS, Pearson WS, McClave AK, Strine TW, et al. The relationship between smoking status and serious psychological distress: findings from the 2007 behavioral risk factor surveillance system. Int J Public Health 2009 Jun;54 Suppl 1:68-74. [doi: 10.1007/s00038-009-0009-y] [Medline: 19396580]
51. Kastaun S, Brose LS, Scholz E, Viechtbauer W, Kotz D. Mental health symptoms and associations with tobacco smoking, dependence, motivation, and attempts to quit: findings from a population survey in Germany (DEBRA study). Eur Addict Res 2022;28(4):287-296 [FREE Full text] [doi: 10.1159/000523973] [Medline: 35358964]
52. Berlin I, Chen H, Covey LS. Depressive mood, suicide ideation and anxiety in smokers who do and smokers who do not manage to stop smoking after a target quit day. Addiction 2010 Dec;105(12):2209-2216. [doi: 10.1111/j.1360-0443.2010.03109.x] [Medline: 20840207]
53. Capron DW, Allan NP, Norr AM, Zvolensky MJ, Schmidt NB. The effect of successful and unsuccessful smoking cessation on short-term anxiety, depression, and suicidality. Addict Behav 2014 Apr;39(4):782-788 [FREE Full text] [doi: 10.1016/j.addbeh.2013.12.014] [Medline: 24457901]
54. McDermott MS, Marteau TM, Hollands GJ, Hankins M, Aveyard P. Change in anxiety following successful and unsuccessful attempts at smoking cessation: cohort study. Br J Psychiatry 2013 Jan;202(1):62-67 [FREE Full text] [doi: 10.1192/bjp.bp.112.114389] [Medline: 23284151]
55. Hughes JR, Carpenter MJ. Does smoking reduction increase future cessation and decrease disease risk? A qualitative review. Nicotine Tob Res 2006 Dec;8(6):739-749. [doi: 10.1080/14622200600789726] [Medline: 17132521]
56. Hymowitz N, Cummings KM, Hyland A, Lynn WR, Pechacek TF, Hartwell TD. Predictors of smoking cessation in a cohort of adult smokers followed for five years. Tob Control 1997;6 Suppl 2(Suppl 2):S57-S62 [FREE Full text] [doi: 10.1136/tc.6.suppl_2.s57] [Medline: 9583654]
57. Stolz D, Scherr A, Seiffert B, Kuster M, Meyer A, Fagerström KO, et al. Predictors of success for smoking cessation at the workplace: a longitudinal study. Respiration 2014;87(1):18-25 [FREE Full text] [doi: 10.1159/000346646] [Medline: 23594795]
58. Holm M, Schiöler L, Andersson E, Forsberg B, Gislason T, Janson C, et al. Predictors of smoking cessation: a longitudinal study in a large cohort of smokers. Respir Med 2017 Nov;132:164-169 [FREE Full text] [doi: 10.1016/j.rmed.2017.10.013] [Medline: 29229092]
59. Arancini L, Borland R, Le Grande M, Mohebbi M, Dodd S, Dean OM, et al. Age as a predictor of quit attempts and quit success in smoking cessation: findings from the International Tobacco Control Four-Country survey (2002-14). Addiction 2021 Sep;116(9):2509-2520 [FREE Full text] [doi: 10.1111/add.15454] [Medline: 33651412]
60. Hall SM, Shi Y, Humfleet GL, Muñoz RF, Reus VI, Prochaska JJ. Smoking cessation abstinence goal in treatment-seeking smokers. Addict Behav 2015 Mar;42:148-153 [FREE Full text] [doi: 10.1016/j.addbeh.2014.11.012] [Medline: 25462664]
61. Wu L, He Y, Jiang B, Zuo F, Liu Q, Zhang L, et al. Relationship between education levels and booster counselling sessions on smoking cessation among Chinese smokers. BMJ Open 2015 Aug 05;5(8):e007885 [FREE Full text] [doi: 10.1136/bmjopen-2015-007885] [Medline: 26246076]
62. Breslau N, Peterson EL. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. Am J Public Health 1996 Feb;86(2):214-220. [doi: 10.2105/ajph.86.2.214] [Medline: 8633738]
63. Reid JL, Hammond D, Boudreau C, Fong GT, Siahpush M, ITC Collaboration. Socioeconomic disparities in quit intentions, quit attempts, and smoking abstinence among smokers in four western countries: findings from the International Tobacco Control Four Country Survey. Nicotine Tob Res 2010 Oct;12 Suppl(Suppl 1):S20-S33 [FREE Full text] [doi: 10.1093/ntr/ntq051] [Medline: 20889477]
64. Gram IT, Antypas K, Wangberg SC, Løchen ML, Larbi D. Factors associated with predictors of smoking cessation from a Norwegian internet-based smoking cessation intervention study. Tob Prev Cessat 2022 Oct 31;8:38 [FREE Full text] [doi: 10.18332/tpc/155287] [Medline: 36382026]
65. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2020 Jun 12;22(7):1086-1097 [FREE Full text] [doi: 10.1093/ntr/ntz132] [Medline: 31570931]

Abbreviations

CO: carbon monoxide
CONSORT: Consolidated Standards of Reporting Trials
CPD: cigarettes per day
DTQ: difficulty to stay quit

FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
IRR: incidence rate ratio
ITT: intention to treat
NCI: National Cancer Institute
NRT: nicotine replacement therapy
OR: odds ratio
PPA: point-prevalence abstinence
ppm: parts per million
RCT: randomized controlled trial
RR: relative risk
SASEQ: Smoking Abstinence Self-Efficacy Questionnaire
USCPG: United States Clinical Practice Guideline

Edited by L Buis; submitted 13.04.23; peer-reviewed by I Gram; comments to author 11.05.23; revised version received 07.06.23; accepted 15.08.23; published 18.09.23.

Please cite as:
Marler JD, Fujii CA, Utley MT, Balbierz DJ, Galanko JA, Utley DS
Long-Term Outcomes of a Comprehensive Mobile Smoking Cessation Program With Nicotine Replacement Therapy in Adult Smokers: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e48157
URL: https://mhealth.jmir.org/2023/1/e48157
doi:10.2196/48157
PMID:37585282

Original Paper

The Effectiveness of a Neurofeedback-Assisted Mindfulness Training Program Using a Mobile App on Stress Reduction in Employees: Randomized Controlled Trial

Beomjun Min[1], MD, PhD; Heyeon Park[2], PhD; Johanna Inhyang Kim[3], MD, PhD; Sungmin Lee[4], BSc; Soyoung Back[4], BSc; Eunhwa Lee[5], BSc; Sohee Oh[6], PhD; Je-Yeon Yun[5,7], MD, PhD; Bung-Nyun Kim[8], MD, PhD; Yonghoon Kim[9], MEng; JungHyun Hwang[9], BCS; Sanghyop Lee[9], BSc; Jeong-Hyun Kim[4,10,11], MD, PhD

[1]Department of Public Health Medical Services, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[2]Liberal Arts College, Dongduk Women's University, Seoul, Republic of Korea
[3]Department of Psychiatry, Hanyang University Medical Center, Seoul, Republic of Korea
[4]Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[5]Department of Psychiatry, Seoul National University Hospital, Seoul, Republic of Korea
[6]Medical Research Collaborating Center, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[7]Yeongeon Student Support Center, Seoul National University College of Medicine, Seoul, Republic of Korea
[8]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea
[9]OMNI C&S Inc, Seoul, Republic of Korea
[10]Department of Public Health Medical Services, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea
[11]Institute of Human Behavioral Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea

Corresponding Author:
Jeong-Hyun Kim, MD, PhD
Department of Public Health Medical Services
Seoul National University Bundang Hospital
82, Gumi-ro 173 beon-gil
Bundang-gu
Seongnam-si, 13620
Republic of Korea
Phone: 82 31 787 2025
Email: retrial3@hanmail.net

Abstract

Background: Mindfulness-based training programs have consistently shown efficacy in stress reduction. However, questions regarding the optimal duration and most effective delivery methods remain.

Objective: This research explores a 4-week neurofeedback-assisted mindfulness training for employees via a mobile app. The study's core query is whether incorporating neurofeedback can amplify the benefits on stress reduction and related metrics compared with conventional mindfulness training.

Methods: A total of 92 full-time employees were randomized into 3 groups: group 1 received mobile mindfulness training with neurofeedback assistance (n=29, mean age 39.72 years); group 2 received mobile mindfulness training without neurofeedback (n=32, mean age 37.66 years); and group 3 were given self-learning paper materials on stress management during their first visit (n=31, mean age 38.65 years). The primary outcomes were perceived stress and resilience scales. The secondary outcomes were mindfulness awareness, emotional labor, occupational stress, insomnia, and depression. Heart rate variability and electroencephalography were measured for physiological outcomes. These measurements were collected at 3 different times, namely, at baseline, immediately after training, and at a 4-week follow-up. The generalized estimating equation model was used for data analysis.

Results: The 4-week program showed significant stress reduction (Wald χ_2^2=107.167, *P*<.001) and improvements in psychological indices including resilience, emotional labor, insomnia, and depression. A significant interaction was observed in resilience (time

× group, Wald χ_4^2=10.846, *P*=.02). The post hoc analysis showed a statistically significant difference between groups 1 (least squares mean [LSM] 21.62, SE 0.55) and 3 (LSM 19.90, SE 0.61) at the posttraining assessment (*P*=.008). Group 1 showed a significant improvement (*P*<.001) at the posttraining assessment, with continued improvements through the 1-month follow-up assessment period (LSM 21.55, SE 0.61). Physiological indices were analyzed only for data of 67 participants (22 in group 1, 22 in group 2, and 23 in group 3) due to the data quality. The relaxation index (ratio of alpha to high beta power) from the right electroencephalography channel showed a significant interaction (time × group, Wald χ_2^2=6.947, *P*=.03), with group 1 revealing the highest improvement (LSM 0.43, SE 0.15) compared with groups 2 (LSM –0.11, SE 0.10) and 3 (LSM 0.12, SE 0.10) at the 1-month follow-up assessment.

Conclusions: The study demonstrated that the neurofeedback-assisted group achieved superior outcomes in resilience and relaxation during the 4-week mobile mindfulness program. Further research with larger samples and long-term follow-up is warranted.

Trial Registration: ClinicalTrials.gov NCT03787407; https://clinicaltrials.gov/ct2/show/NCT03787407

(JMIR Mhealth Uhealth 2023;11:e42851) doi:10.2196/42851

KEYWORDS

mindfulness; neurofeedback; stress; resilience; mobile app; employee

Introduction

Mental stress is a major public health concern and has a high prevalence. Stress is suggested to negatively impact both the physical and mental health of individuals [1,2], ranging from cardiovascular mortality [3], musculoskeletal pain [4], type 2 diabetes [5], depression, and anxiety [6-8]. Work-related stress is associated with presenteeism, absenteeism, diminished productivity, and high employee turnovers [9-11]. According to an annual UK survey, stress, depression, or anxiety accounts for 44% of all work-related health issues, and 54% of all working days were lost due to a recent increase in poor health of employees [12]. Thus, the importance of stress management in employees can be seen at both the industrial and individual levels.

Recently, there has been an increase in the interest, and thereby increased research effort, surrounding the mindfulness-based training programs to reduce stress and improve mental wellness for employees [13]. Many organizations and corporations have adopted mindfulness-based programs to enhance their employees' well-being and performance [14]. Several meta-analyses focusing on healthy individuals suggested that mindfulness-based training programs were able to reduce perceived stress, anxiety, and depression, as well as enhance the quality of life, spiritual values, and resilience [15,16]. Mindfulness-based training programs have also demonstrated benefits for clinicians, nursing staff, and health profession students who experience high stress levels and burnout rates [17-19]. The health benefits experienced by medical professionals could indirectly benefit patients by enhancing the quality of patient care and clinical practices [20]. However, despite these possible benefits, several issues, including training duration and delivery methods—and their varying effectiveness—have been addressed [17,21].

According to a qualitative study that reviewed 67 published studies [13], the most common method of delivery was in-person lectures (n=64, 96%), followed by audio recordings (n=39, 58%) and group discussions (n=34, 51%). Among these 67 studies, only 4 (6%) used online modules as a means of content delivery. However, with technological advancements and greater immersion of smartphones in daily life, there has recently been an increase in the interest in incorporating such technology into the field of mental health care [22]. With the advent of smartphone technology and mobile apps, health professionals are realizing their advantages and ways to enhance patient care. The advantages of mobile app–based approaches are increased accessibility, interactivity, and usability of various functions, such as treatment monitoring, appointment reminders, and recordkeeping [23,24]. These advantages could enhance patient's or employees' adherence to the mental health intervention program. A recent study suggested that mindfulness training using smartphone apps may provide immediate positive effects on mood and stress, as well as long-term benefits for attentional control [25].

In addition, in terms of training duration, a traditional 8-week mindfulness training program, such as mindfulness-based stress reduction (MBSR), may be a barrier to widespread organizational adoption and raises the question of whether a similar beneficial effect can be achieved with a shorter program [13]. A previous study reported that even a 4-day mindfulness training program was effective in reducing anxiety and fatigue, as well as increasing mindfulness [26]. In a systematic review, the correlation between the mean effect size and the number of in-class hours for MBSR training was not significant in clinical and nonclinical samples. This finding suggests that reduced—or shorter—class time may be worthwhile for those with time constraints or those with little to no motivation to participate [27].

Meanwhile, a recent meta-analysis concluded that the most consistent electroencephalography (EEG) findings associated with mindfulness training may be increased theta and alpha power [28]. Several previous studies have evaluated the relationship between mindfulness-based training and neurofeedback. Neurofeedback is a training method to control brain waves consciously, and an EEG is used to record these waves. During training, individuals are instructed to focus or relax, and EEG is presented as a type of visual stimuli in real time so that they can recognize their present state of brain activity. In a previous study, participants undergoing mindfulness meditation with alpha-neurofeedback demonstrated a higher alpha amplitude compared with their sham neurofeedback counterparts [29]. Other studies suggested that neurofeedback could mediate the effect of mindfulness meditation [30,31].

The purpose of this study is to verify the effects of a neurofeedback-assisted mindfulness training program delivered via a mobile app. We hypothesize that the neurofeedback-assisted mindfulness training might have augmenting effects on stress and related indices compared with mindfulness training alone.

Methods

Participants

Participants were recruited via advertisements at the Seoul National University Hospital and the Seoul National University Bundang Hospital between August 2018 and December 2018. The inclusion criteria were as follows: (1) age between 19 and 65 years; (2) total score of the Perceived Stress Scale (PSS) ≥14 at baseline, and (3) currently employed full-time. The exclusion criteria were (1) age <19 or >65 years; (2) having cognitive disorders, such as dementia or intellectual disability; (3) neurological disorders, such as epileptic disorders, stroke, brain tumors, or others; (4) current or previous history of psychosis, such as schizophrenia or bipolar I disorders; (5) current report of suicidal ideation; (6) other conditions that might influence the measurement of heart rate variability (HRV), such as cardiac or pulmonary disease; and (7) nonpharmacological psychiatric treatment, counseling, or meditation training within the past 6 months. As mental stress at the workplace is commonly associated with depression, anxiety, and insomnia [6-8], participants with these conditions, but in stable status, were not excluded, as long as the type and dosage of medication remained the same for the past 6 months. In the screening process, the Mini International Neuropsychiatric Interview (MINI) [32], which is a short, structured psychiatric interview designed to find a broad spectrum of psychiatric disorders as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, and International Classification of Diseases, Tenth Revision, was used to diagnose psychiatric disorders. The MINI was translated into Korean and has good validity and reliability [33]. The interviews were conducted by 2 psychologists with a master's degree who were familiar with the tool. The estimated total sample size using G-Power was 90 (30 participants per group) with a predicted effect size of Cohen d=0.3, α level of .05, and a desired power of 0.7. Considering a dropout rate of 10%, we targeted to recruit 100 participants. Multimedia Appendix 1 presents the CONSORT (Consolidated Standards of Reporting Trials) checklist completed for this study.

Assessments

Demographic information, including age, sex, marital status (unmarried or others), educational status (more than college education or less), and length of career (more than 3 years or less), was obtained using self-reported questionnaires.

As the primary psychological outcome measures, changes in the scores of the PSS and Brief Resilience Scale (BRS) were used. The secondary psychological outcomes were changes in the scores of scales assessing mindfulness (Korean version of the Mindfulness Attention Awareness Scale [K-MAAS]), emotional labor (Korean Emotional Labor Scale [KELS]), occupational stress (Korean Standard Occupational Stress Scale—short form [KOSS]), insomnia (Athens Insomnia Scale [AIS]), and depression (9-item Patient Health Questionnaire [PHQ-9]).

In its original form, as developed by Cohen et al [34], the PSS is a 14-item scale; however, a modified version with a 10-item scale is commonly used [35]. The questionnaire is rated on a 5-point Likert scale, with 0=never, 1=almost never, 2=sometimes, 3=fairly often, and 4=very often. Higher scores reflect higher levels of perceived stress.

The BRS, comprising 6 items and measured on a 5-point Likert Scale (1=strongly disagree to 5=strongly agree), was used to evaluate the resilience of each individual and assessed the ability to bounce back or recover from stress [36]. Higher scores indicated better resilience.

The K-MAAS was originally developed by Brown and Ryan [37] and validated by Kwon and Kim [38]. This 15-item scale focuses on the attention and awareness of mindful states, ranging from 1 (almost always) to 6 (almost never). Higher scores indicated higher levels of dispositional mindfulness.

The KELS was developed to assess the emotional labor of employees and was validated for its applicability to the Korean population by Lee et al [39]. Emotional labor was defined as the process by which employees have to control their feelings in accordance with the organizational demand and occupational role [40,41]. The KELS has 5 subscales: effort to control emotion (5 items), organizational monitoring system (4 items), demands of emotional labor (3 items), emotional damage (6 items), and organizational support system (7 items). Each item was rated on a 4-point Likert scale (1=not at all to 4=very much), with higher scores reflecting higher levels of emotional labor.

The KOSS-Short Form consists of 24 items, with each item rated on a 4-point Likert scale (1=never to 4=always). It was validated by Chang et al [42]. The scale is most commonly used and studied for the evaluation of job stress in South Korea. It consists of 7 subscales, including job demand (4 items), job control (4 items), interpersonal conflict (3 items), job insecurity (2 items), organizational system (4 items), lack of reward (3 items), and workplace environment (4 items). In this study, the sum of scores on each subscale was calculated and then converted to 100 points. Higher scores reflected higher levels of job stress.

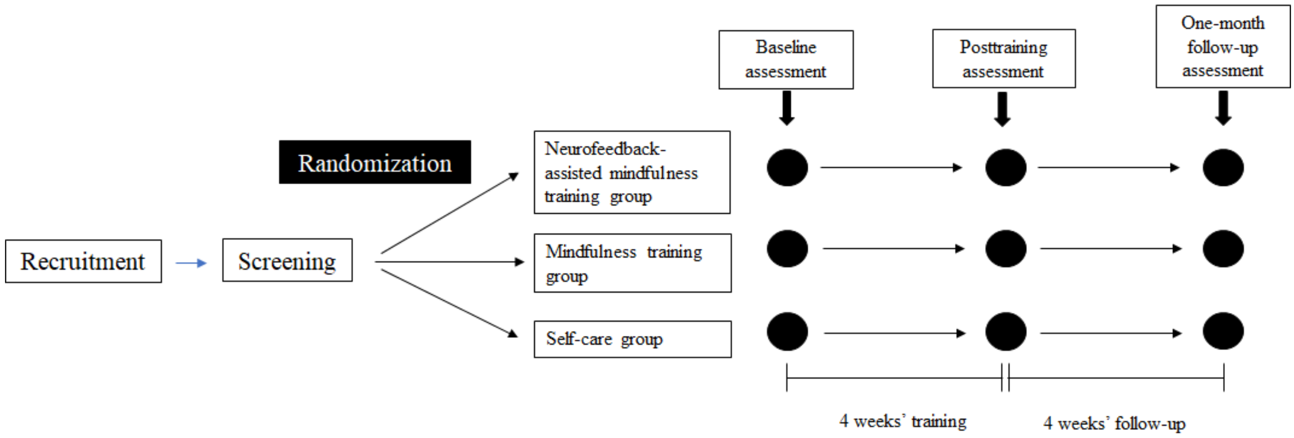
Insomnia was assessed by the AIS, which consisted of 8 items with a 4-point Likert scale, with higher scores reflecting greater severity of insomnia [43].

The PHQ-9 was used to evaluate depressive symptoms, with higher scores indicating a higher level of depressive symptoms [44].

In addition, we measured the physiological variables, including HRV and EEG, which are widely used as stress-related physiological biosignals [45,46]. These physiological variables were acquired using headsets (neuroNicle FX2; Laxtha) [47]. In a previous work that used the FX2 device, the authors established the reliability of a prefrontal EEG marker from the device and reported that no significant differences in the mean values of the tested variables were observed between the prefrontal and occipital regions [48]. In our study, EEG and HRV were recorded simultaneously for 5 minutes for all participants in a resting state with eyes closed. Ocular, muscular, and other types of artifacts were detected by an automatic removal algorithm, and contaminated periods longer than 10 seconds were removed. For EEG, the band-pass filter (infinite impulse response Butterworth filters, high-pass filter: first order with fc=2.6 Hz; low-pass filter: eighth order with fc=43 Hz) with 3-43-Hz range was applied to the data. All data were digitized in the continuous recording mode (5 minutes of recording; 250-Hz sampling rate; 15-bit resolution). For HRV, the plethysmograph (PPG) waveform and the heartbeat time interval in milliseconds for each heartbeat were detected by the pulse wave sensor. These EEG and PPG raw data were transmitted to the host device in real time. In the host device, the fast Fourier transform was applied to calculate the power spectrum of the acquired EEG and HRV data. As a result, the power spectra of theta (4-8 Hz), alpha (8-12 Hz), low beta (12-15 Hz), beta (15-20 Hz), high beta (20-30 Hz), and gamma (30-40 Hz) frequency bands were calculated from the EEG data. In addition, the relaxation index (ratio of alpha to high beta power) and concentration index (ratio of low beta to theta power) were calculated from the EEG data. For HRV parameters, the power of low-frequency (LF; 0.04-0.15 Hz) and high-frequency (HF; 0.15-0.4 Hz) bands was calculated. All these variables were normalized by sex and age.

Throughout the study period, all self-reported psychological and physiological indices were measured at 3 time points: during enrollment (baseline assessment), immediately after training (posttraining assessment), and 4 weeks after training (1-month follow-up assessment; Figure 1). All outcome measures were obtained when participants visited the Seoul National University Hospital and the Seoul National University Bundang Hospital.

Figure 1. Assessment process of the study.



Randomization and Intervention Conditions

Participants were assigned to 3 groups by 1:1:1 block randomization (stratified by organization and age) with randomly selected block sizes (3, 6, or 9) using REDCap (Research Electronic Data Capture; Vanderbilt University) tools hosted at the Medical Research Collaborating Center of Seoul National University Bundang Hospital. REDCap generates randomization codes with SAS (Statistical Analysis System; SAS Institute). The allocation sequence was concealed from the participants until they entered the trial, and from the investigators until the end of the study. Participants in the neurofeedback-assisted mindfulness training group (group 1) and mindfulness training-only group (group 2) were instructed to download a mobile app specifically developed for the research (developed by Omni C&S, Inc.). The app provides a mindfulness-based training program with an audio guide. In addition, for the neurofeedback-assisted mindfulness training group (group 1), headsets (OMNIFIT Brain; Omni C&S, Inc.) were provided. The neurofeedback function was embedded in the mobile app, with audio stimuli delivered to participants through the earphones of the headset. Group 1 participants received instructions on these processes during an educational session. Before initiating the training, the experimental groups (groups 1 and 2) received education to comprehend the concept of mindfulness and engaged in an hour-long practice session guided by a well-trained psychologist. All the participants in both groups attended 4 weekly meeting sessions with the psychologist, each lasting 30 minutes. In the meeting, the psychologist assessed the participants' adherence to the training from the previous week and provided encouragement to those who reported lower levels of accomplishment. They also discussed any issues related to the training and the way forward. The control group participants, who practiced self-care (group 3), were given self-learning paper materials on stress management during their first visit, without any additional weekly meetings. These materials covered the definition of stress, signs, and symptoms of stress, as well as strategies for managing stress.

Apparatus

EEG and HRV were measured using neuroNicle FX2 (Laxtha Inc.) [47]. The device is equipped with 2 EEG monopolar electrodes positioned over the left and right prefrontal areas

(FP1 and FP2), along with a ground and reference electrode attached to a clamp on the right earlobe. Additionally, there is 1 pulse wave sensor (PPG) for measuring HRV, which is also attached to the same clamp. Both EEG and PPG data were recorded simultaneously. The raw EEG and PPG data were transmitted to the host device via Bluetooth technology (version 4.2) and subsequently analyzed to derive quantitative data, including power spectral components and neurofeedback indices. These data were then transmitted to the server of Omni C&S Inc., our research collaborating company, using Wi-Fi (wireless fidelity) technology. The device received medical device approval from the South Korean Ministry of Food and Drug Safety (authorization number 16-4837).

The OMNIFIT Brain is a headset device provided to participants in group 1 to aid in their neurofeedback training during the intervention. It is compatible with both Android (Google LLC/Alphabet Inc.) and iOS (Apple Inc.) operating systems. This headset features 2 EEG sensors and 1 PPG sensor, all operating at a sampling rate of 250 Hz, and it transmits neurophysiological data in real-time to the mobile app via Bluetooth technology. The app generates various neurofeedback indices, which are then relayed back to the headset, and the headset's earphones provide auditory feedback to the participants.

Mobile App

The mindfulness-based training program comprised 3 components: awareness training (7 minutes), abdominal breathing (4 minutes), and body scan meditation (8 minutes), each accompanied by an audio guide (Figure 2). Following each session, participants had approximately 10 minutes for self-exercise. The order and frequency of these sessions were carefully structured as per the controlled research protocol, and participants were instructed to adhere to the provided schedule (refer to Table 1). The total duration of 1 training session, including self-exercise, was 20 minutes.

Figure 2. Images from the mobile app used in the mindfulness-based training program. (A) Main screen of the app; (B) introduction screen; (C) awareness program; and (D) self-exercise program.

Table 1. Schedule of the 4-week mindfulness training program.

Programs	Frequency	Duration (days)
First week		
Breathing with self-exercise	Twice a day	3
Awareness with self-exercise	Twice a day	2
Body scan with self-exercise	Twice a day	2
Second week		
Breathing with self-exercise then awareness with self-exercise	Both, twice a day	7
Third week		
Breathing with self-exercise then body scan with self-exercise	Both, twice a day	7
Fourth week		
Self-exercise	Twice a day	7

Neurofeedback Function

For group 1, the neurofeedback function was facilitated using the OMNIFIT Brain during the self-exercise sessions. The neurofeedback training incorporated an alpha protocol designed to augment the power of alpha frequency compared with other frequency bands. The alpha wave is known to be associated with a state of relaxation [49], and alpha protocols are widely used in neurofeedback techniques aimed at reducing anxiety and stress, while enhancing mental performance [50]. Participants in group 1 were informed that they would receive a positive feedback sound (a 1-second ringing of a bell) when they were in a focused and relaxed state, whereas a negative feedback sound (a 1-second chirping sound of a cricket) would be provided when they were in a distracted and unrelaxed state.

During the self-exercise session, the ratio of alpha power (8-12 Hz) to high beta power (20-30 Hz) was calculated at 2-second intervals. Participants received a positive auditory neurofeedback sound through the earphones when the ratio of alpha power to high beta power reached or exceeded a level of 2.775. This criterion was validated in a prior clinical study involving 1500 Korean healthy volunteers, using the same EEG system, and the value was determined based on the distribution of t-scores within the population [51].

Statistical Analysis

Demographic variables and baseline psychological measurements were analyzed using an ANOVA or Kruskal-Wallis test for continuous variables, while chi-square tests or Fisher exact tests were used to compare the 3 groups. Marital status, educational levels, and length of career were divided into 2 entities for simplicity. For all outcome measures, including psychological and physiological variables, a generalized estimating equation (GEE) was applied to examine the effects of interventions. A GEE is commonly used when the outcome variables are measured repeatedly from the same individual and probable dependencies are assumed between those variables [52]. In this study, we utilized GEE to examine the group, time, and their interaction term, as it allowed us to assess the differential group effect at each time point. In the absence of an interaction, we assessed the effect of the group. We applied a first-order autoregressive correlation structure and set the significance level at 2-tailed (paired) $P<.05$. All statistical analyses were conducted using IBM SPSS Statistics version 25 (IBM Corp.).

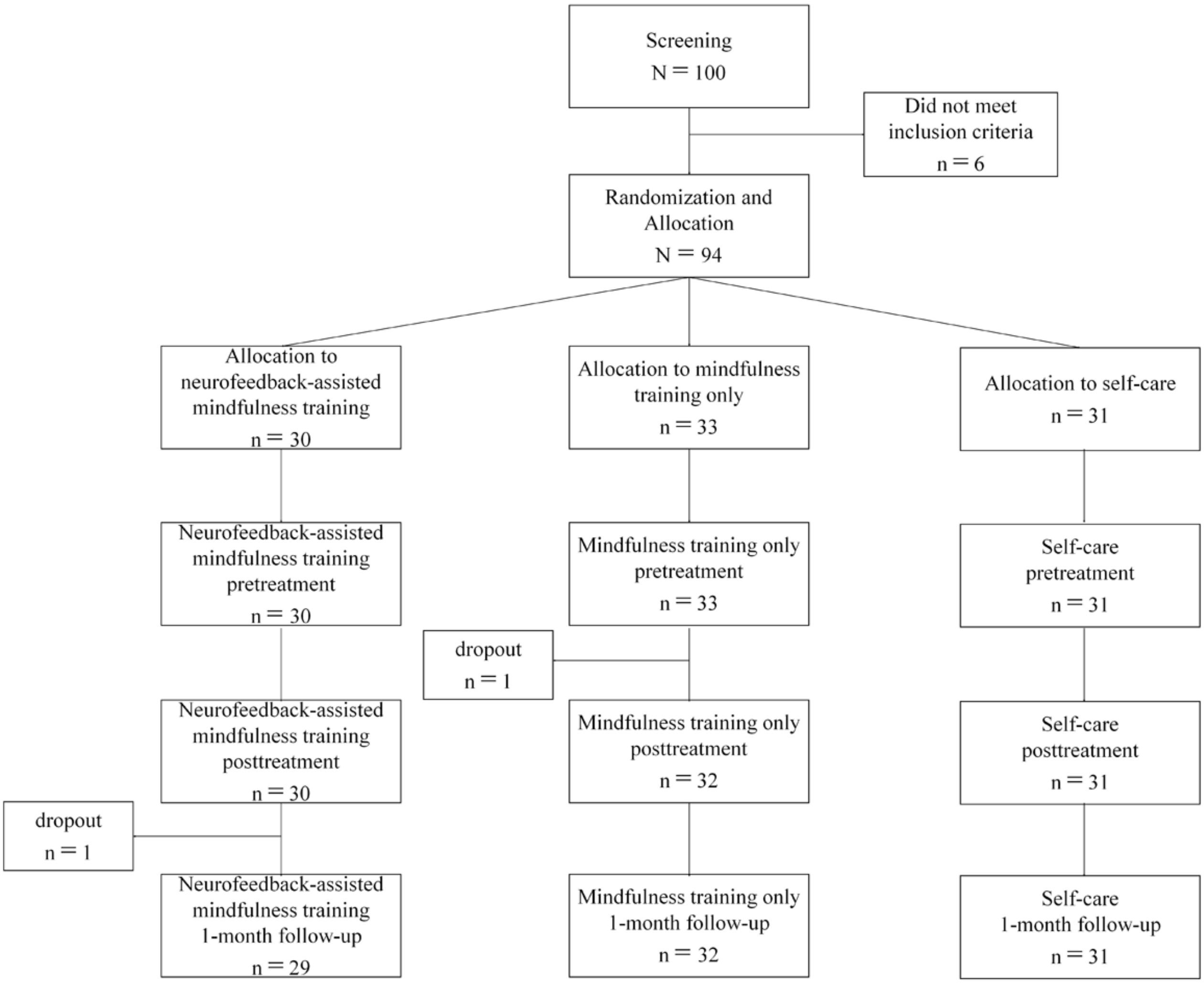
Ethics Statement

The study protocol was approved by the Institutional Review Board of Seoul National University Hospital in Seoul and Seoul National University Bundang Hospital in Seongnam, South Korea (approval number: B-1807-483-303). All participants were provided with information about the study and signed informed consent forms before participation. This study is registered in the ClinicalTrials.gov (NCT03787407).

Results

Of the 100 participants screened for this study, 6 did not meet the inclusion criteria. As a result, a total of 94 participants were initially included and randomly assigned to 1 of the 3 groups: 30 in group 1, 33 in group 2, and 31 in group 3. After the start of the intervention, 1 participant in group 2 dropped out due to compatibility issues between the mobile app and personal cellular phone. Thus, a total of 93 participants completed the training. One participant in group 1 dropped out before completing the 1-month follow-up assessment due to unemployment. Thus, 29 participants in group 1, 32 in group 2, and 31 in group 3 completed all assessments. Figure 3 presents a flow diagram of the study process. The dropout rates after training engagement were 3% (1/30), 3% (1/33), and 0% (0/31) for groups 1-3, respectively; this distribution was not statistically significant ($P=.76$) by Fisher exact test.

Figure 3. Flowchart of the study process. f/u: follow-up.



Demographic and clinical characteristics of the participants at baseline who completed all assessments are presented in Table 2. There were no significant differences in age (*P*=.76), sex (*P*=.67), marital status (*P*=.38), educational levels (*P*=.90), and length of career (*P*=.10) between the groups.

Table 2. Demographic and clinical characteristics of each group at baseline.

Characteristics	Neurofeedback-assisted mindfulness training (n=29)	Mindfulness-only training (n=32)	Self-care (n=31)	*F* or chi-square (*df*) statistic	*P* value
Age (years), mean (SD)	39.72 (10.01)	37.66 (10.59)	38.65 (11.88)	*F*=0.275 (2, 89)	.76
Sex (female), n (%)	27 (93.10)	29 (90.63)	27 (87.10)	χ^2=0.796 (2)	.67
Marital status: unmarried, n (%)	11 (37.93)	18 (56.25)	14 (45.16)	χ^2=1.913 (2)	.38
Education >college education, n (%)	27 (93.10)	29 (90.63)	27 (87.10)	χ^2=0.203 (2)	.90
Length of career >3 years, n (%)	28 (96.55)	25 (78.13)	25 (80.65)	χ^2=4.583 (2)	.10
Perceived Stress Scale score, mean (SD)	21.52 (4.38)	20.78 (4.82)	22.55 (4.95)	*F*=1.104 (2, 89)	.33
Brief Resilience Scale score, mean (SD)	19.72 (3.32)	19.97 (3.91)	19.81 (4.21)	*F*=0.033 (2, 89)	.96
Korean Version of the Mindfulness Attention Awareness Scale score, mean (SD)	65.28 (12.23)	60.72 (8.31)	57.81 (11.46)	*F*=3.664 (2, 89)	.03[a]
Korean Emotional Labor Scale score, mean (SD)	52.89 (7.20)	53.01 (11.89)	53.46 (15.09)	*F*=0.013 (2, 89)	.98
Korean Standard Occupational Stress Scale score, mean (SD)	53.33 (7.20)	51.69 (4.51)	48.90 (6.49)	*F*=4.016 (2, 89)	.02[b]
Athens Insomnia Scale score, mean (SD)	14.48 (3.12)	14.50 (2.23)	14.90 (3.35)	*F*=0.203 (2, 89)	.81
9-item Patient Health Questionnaire score, mean (SD)	5.41 (4.46)	6.34 (4.09)	6.77 (4.30)	*F*=0.761 (2, 89)	.47

[a]Post hoc test using the Tukey method was performed. The neurofeedback-assisted mindfulness training group exhibited higher scores on the Mindfulness Attention Awareness Scale compared with the self-care group.

[b]Post hoc test using the Tukey method was performed. The neurofeedback-assisted mindfulness training group exhibited higher scores on the Occupational Stress Scale compared with the mindfulness training group; the mindfulness training group exhibited higher scores compared with the self-care group.

In all groups, 5 psychological variables, namely, PSS, BRS, KELS, AIS, and PHQ-9, showed improvements. Figures 4-8 depict the changes in PSS, BRS, KELS, AIS, and PHQ-9 scores across time. In the GEE model with interaction terms, neither the interaction (time × group) nor a main effect of group difference was found in the PSS score. Only the main effect of time was statistically significant in PSS (Wald χ_2^2=40.63, *P*<.001). All 3 groups showed a lower PSS score after the intervention (Wald χ_2^2=107.167, *P*<.001). For BRS (Figure 5), a significant interaction was observed (time × group, Wald χ_4^2=10.846, *P*=.02). In the post hoc analysis, a significant difference was found between group 1 (least squares mean [LSM] 21.62, SE 0.55) and group 3 (LSM 19.90, SE 0.61) at the posttraining assessment (*P*=.008). In addition, a significant effect of time (Wald χ_2^2=31.238, *P*<.001) was found, indicating that the resilience scores increased for both groups 1 and 2. Moreover, BRS scores at posttraining (LSM 21.62, SE 0.55) and 1-month follow-up (LSM 21.55, SE 0.61) assessments in group 1 revealed significant differences compared with the baseline assessment (LSM 19.72, SE 0.59, *P*<.001). In group 2, the score of the posttraining assessment was the highest (LSM 21.19, SE 0.60) and differed significantly from the baseline assessment (LSM 19.97, SE 0.68, *P*=.01), but not at the 1-month follow-up assessment (LSM 20.69, SE 0.55, *P*=.18). By contrast, we found no interaction in KELS, AIS, and PHQ-9 scores. However, the main effects of time were statistically significant in KELS (Wald χ_2^2=17.061, *P*<.001; Figure 6), AIS (Wald χ_2^2=14.391, *P*<.001; Figure 7), and PHQ-9 (Wald χ_2^2=16.395, *P*<.001; Figure 8) scores. No improvements were found in K-MAAS and KOSS scores. These results are summarized in Table 3.

Figure 4. Perceived Stress Scale (PSS) scores across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

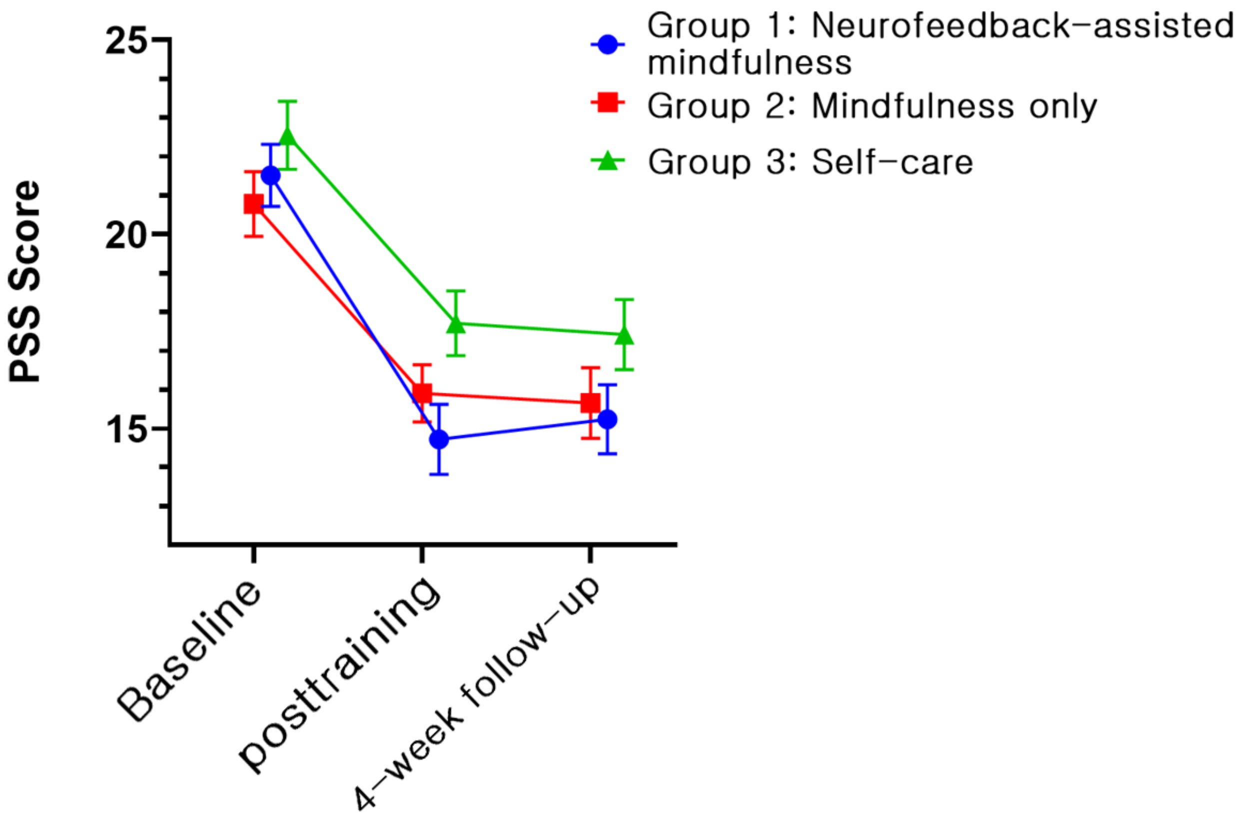


Figure 5. Brief Resilience Scale (BRS) scores across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

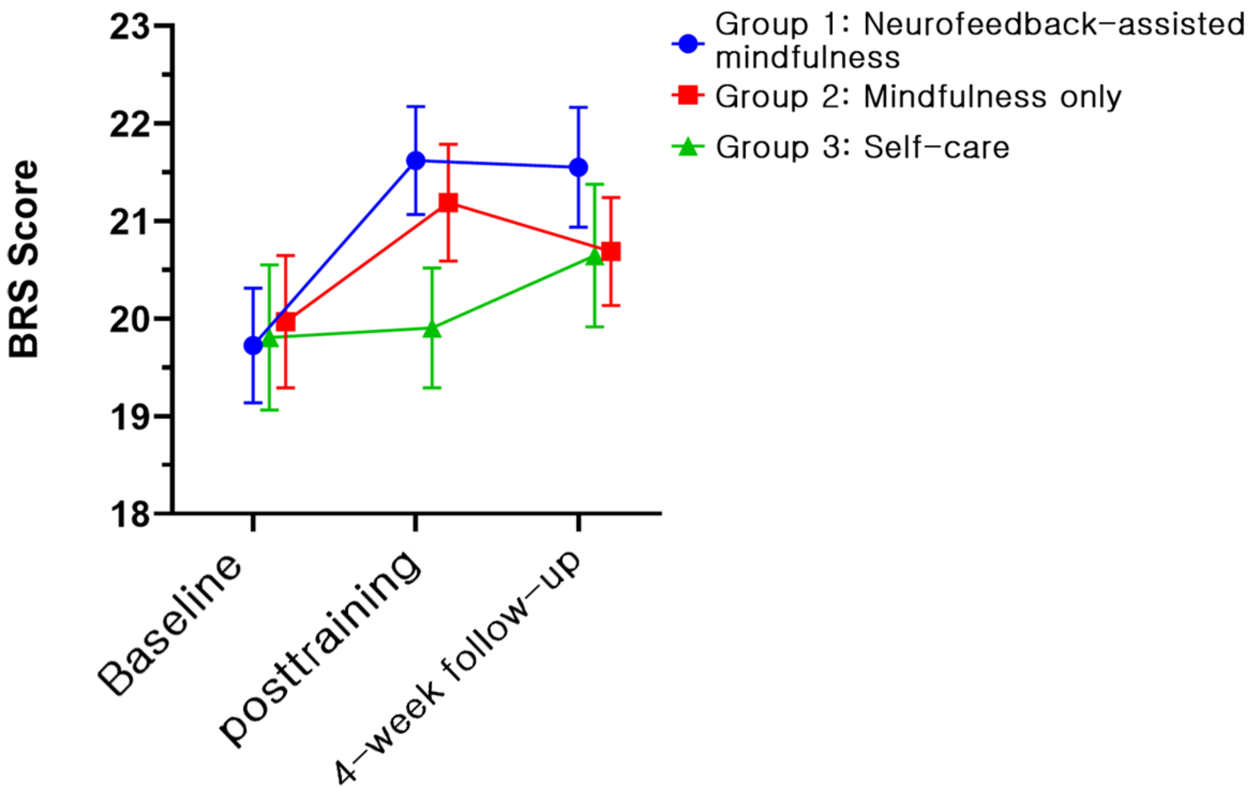


Figure 6. Korean Emotional Labor Scale (KELS) scores across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

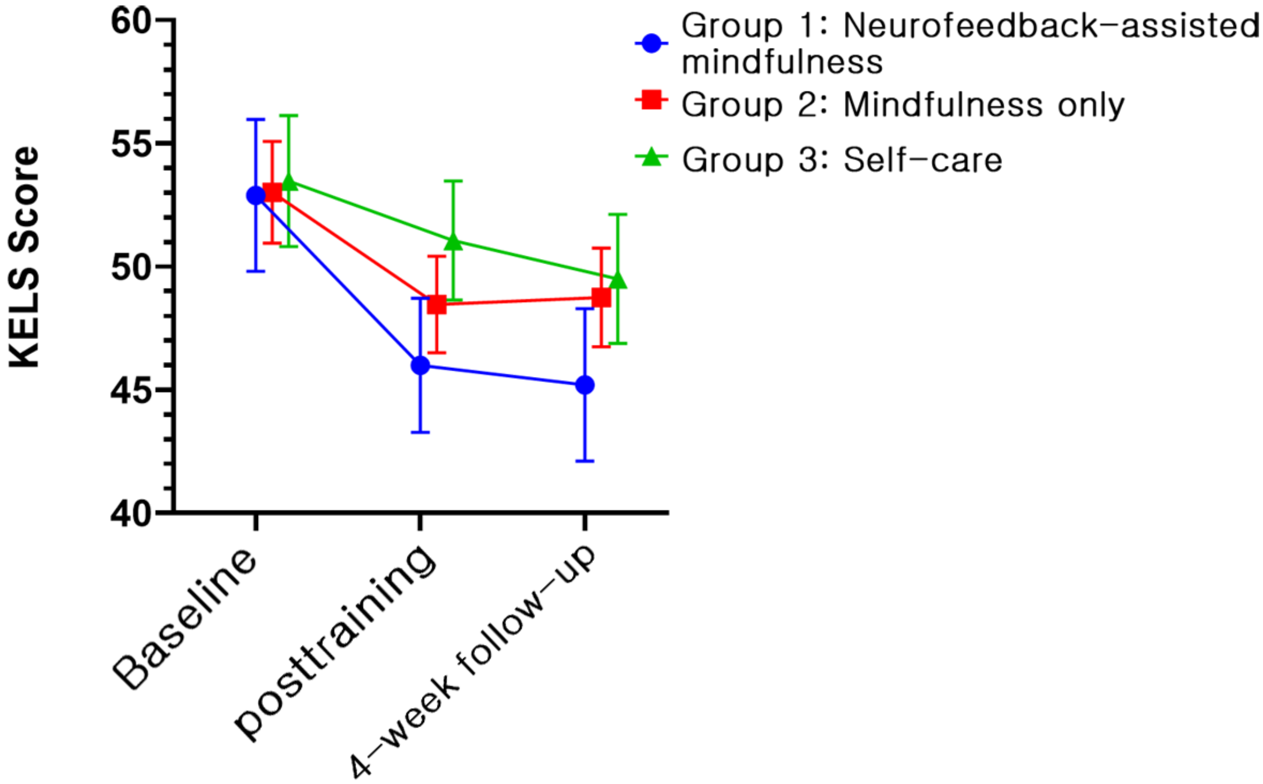


Figure 7. Athens Insomnia Scale (AIS) scores across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

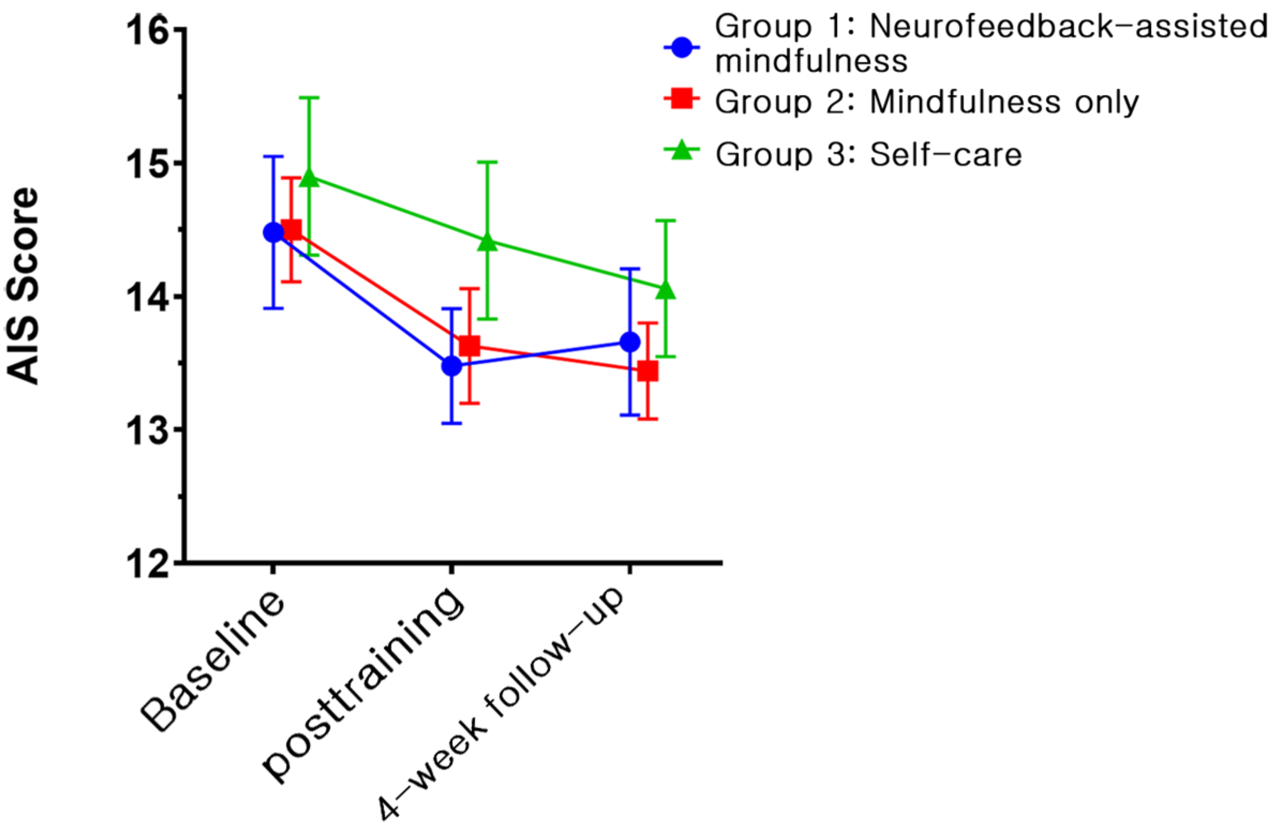


Figure 8. 9-item Patient Health Questionnaire (PHQ-9) scores across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

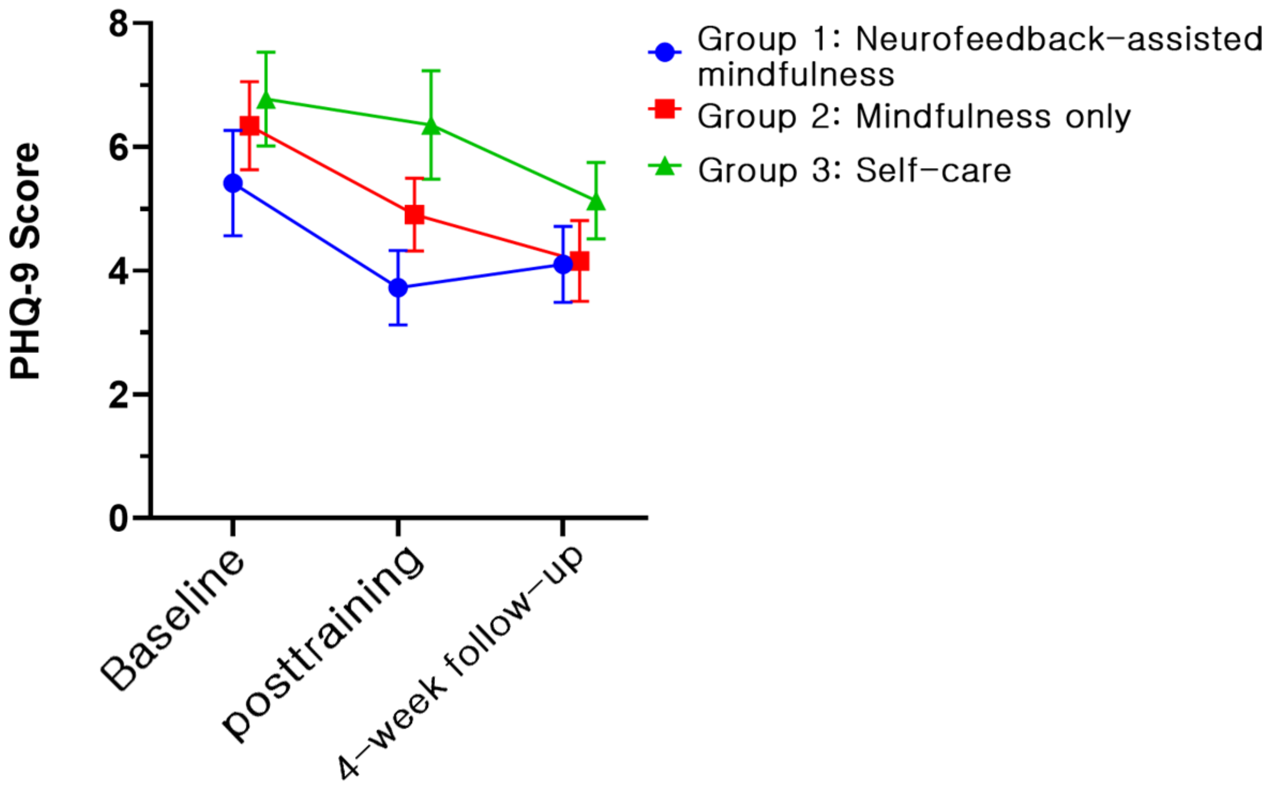


Table 3. Generalized estimating equation model results of psychological variables.

Variables	Effects	Wald chi-square (*df*) statistic	*P* value
PSS[a,b]			
	Group × time interaction	2.472 (4)	.64
	Group	4.293 (2)	.11
	Time	107.167 (2)	<.001
BRS[c,d]			
	Group × time interaction	10.846 (4)	.02
	Group	0.075 (2)	.96
	Time	31.238 (2)	<.001
KELS[e]			
	Group × time interaction	3.491 (4)	.47
	Group	0.591 (2)	.74
	Time	17.061 (2)	<.001
AIS[a,f]			
	Group × time interaction	1.947 (4)	.74
	Group	0.875 (2)	.64
	Time	14.391 (2)	<.001
PHQ-9[a,g]			
	Group × time interaction	8.555 (4)	.07
	Group	2.679 (2)	.26
	Time	16.395 (2)	<.001
K-MAAS[a,h]			
	Group × time interaction	4.640 (4)	.32
	Group	9.139 (2)	.01
	Time	2.861 (2)	.23
KOSS[a,i]			
	Group × time interaction	5.055 (4)	.28
	Group	4.817 (2)	.08
	Time	3.067 (2)	.21

[a]Results of group and time effects were derived from the model without interaction.

[b]PSS: Perceived Stress Scale.

[c]Results of group and time effects were derived from the model with interaction.

[d]BRS: Brief Resilience Scale.

[e]KELS: Korean Emotional Labor Scale.

[f]AIS: Athens Insomnia Scale.

[g]PHQ-9: 9-item Patient Health Questionnaire.

[h]K-MAAS: Korean version of the Mindfulness Attention Awareness Scale.

[i]KOSS: Korean Standard Occupational Stress Scale.

The EEG and HRV data for 25 participants were excluded from the final analysis; one of the 3 measurements (ie, baseline assessment, posttraining assessment, and 1-month follow-up assessment) was not available due to either inappropriate length or continuity of data after removal of artifacts for analysis. Thus, the physiologic data for 67 participants (22 in group 1, 22 in group 2, and 23 in group 3) were obtained. However, no significant results were obtained in terms of time × group interactions when both EEG and HRV data were subjected to GEE analysis in the model. Conversely, irrespective of interactions, the relaxation index of the right prefrontal channel showed significant effects by group (Wald χ_2^2=9.235, *P*=.009) and time (Wald χ_2^2=6.255, *P*=.04; Figure 9). In the post hoc analysis, group 1 (LSM 3.87, SE 0.11) showed a significant difference compared with groups 2 (LSM 3.47, SE 0.10, *P*=.009) and 3 (LSM 3.42, SE 0.12, *P*=.007). In terms of time, a significant difference (*P*=.017) was found between the baseline (LSM 3.48, SE 0.08) and posttraining (LSM 3.66, SE 0.08, *P*=.01) values. In an additional analysis, we examined differences between groups with relative changes compared with baseline. A significant time × group interaction (Wald χ_2^2=6.947, *P*=.03) was found. In the post hoc analysis, the relaxation index changed significantly (*P*=.03) at both posttraining (LSM 0.24, SE 0.12, *P*=.04) and 1-month follow-up assessments (LSM 0.43, SE 0.15 *P*=.003) compared with the baseline measurement only in group 1. Furthermore, we observed a significant group difference (Wald χ_2^2=9.609, *P*=.008) at the 1-month follow-up assessment. Group 1 demonstrated the highest increase in the index (LSM 0.43, SE 0.15), in contrast to group 2 (LSM –0.11, SE 0.10) and group 3 (LSM 0.12, SE 0.10). We illustrated the changes of alpha and high beta frequency power, which are components of the relaxation index, along with the serial measurements. Between baseline and the posttraining assessment, the alpha power of groups 1 and 2 increased, whereas it decreased in group 3 (Figure 10). In the high beta frequency band, the power of groups 1 and 2 remained similar, but that of group 3 decreased (Figure 11). In the next period, between the posttraining and 1-month follow-up assessments, the alpha frequency band of groups 1 and 2 moved downward together (Figure 10), but in the beta frequency band, they moved separately in opposite directions (Figure 11). LF of HRV parameters showed a significant effect by time (Wald χ_2^2=14.073, *P*=.001) in the model without an interaction term (Figure 12). However, no other significant results were revealed via the HRV parameters.

Figure 9. Relaxation Index across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

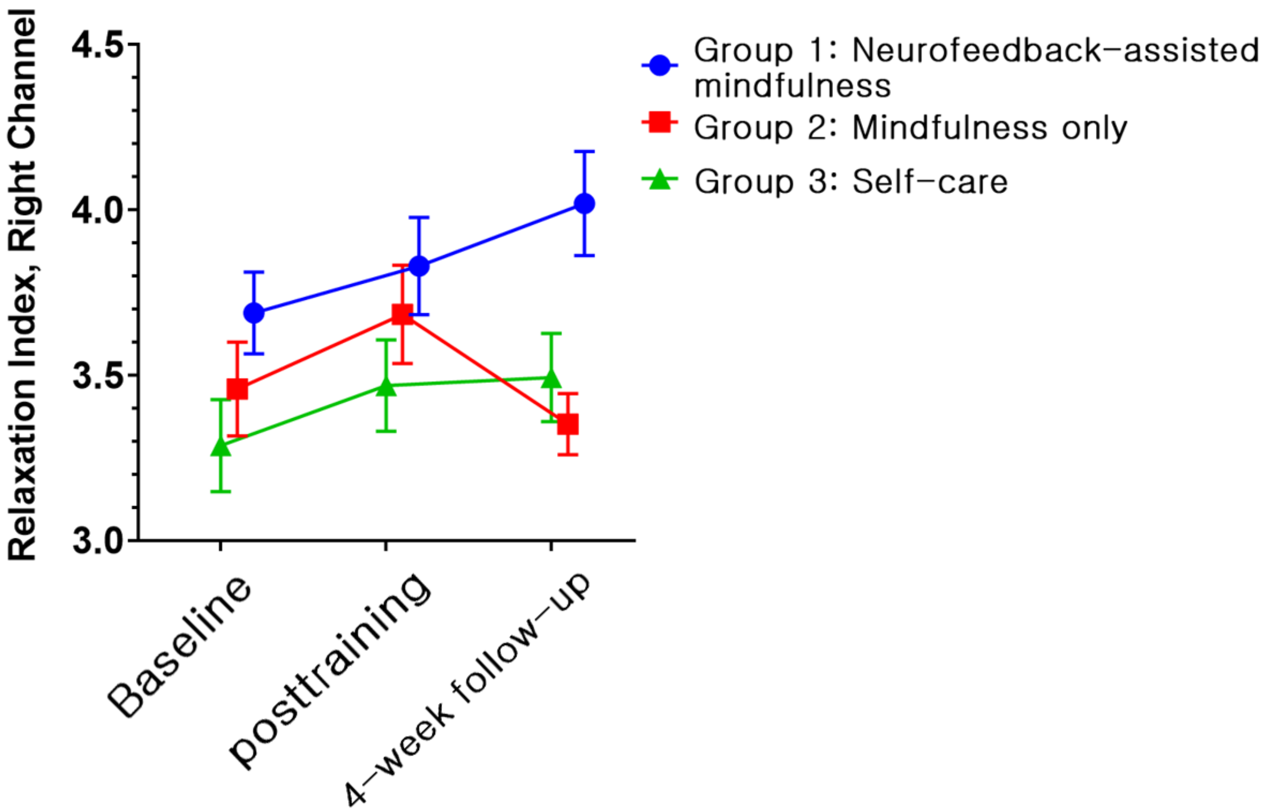


Figure 10. Power of the alpha frequency band of the right electroencephalogram channel (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

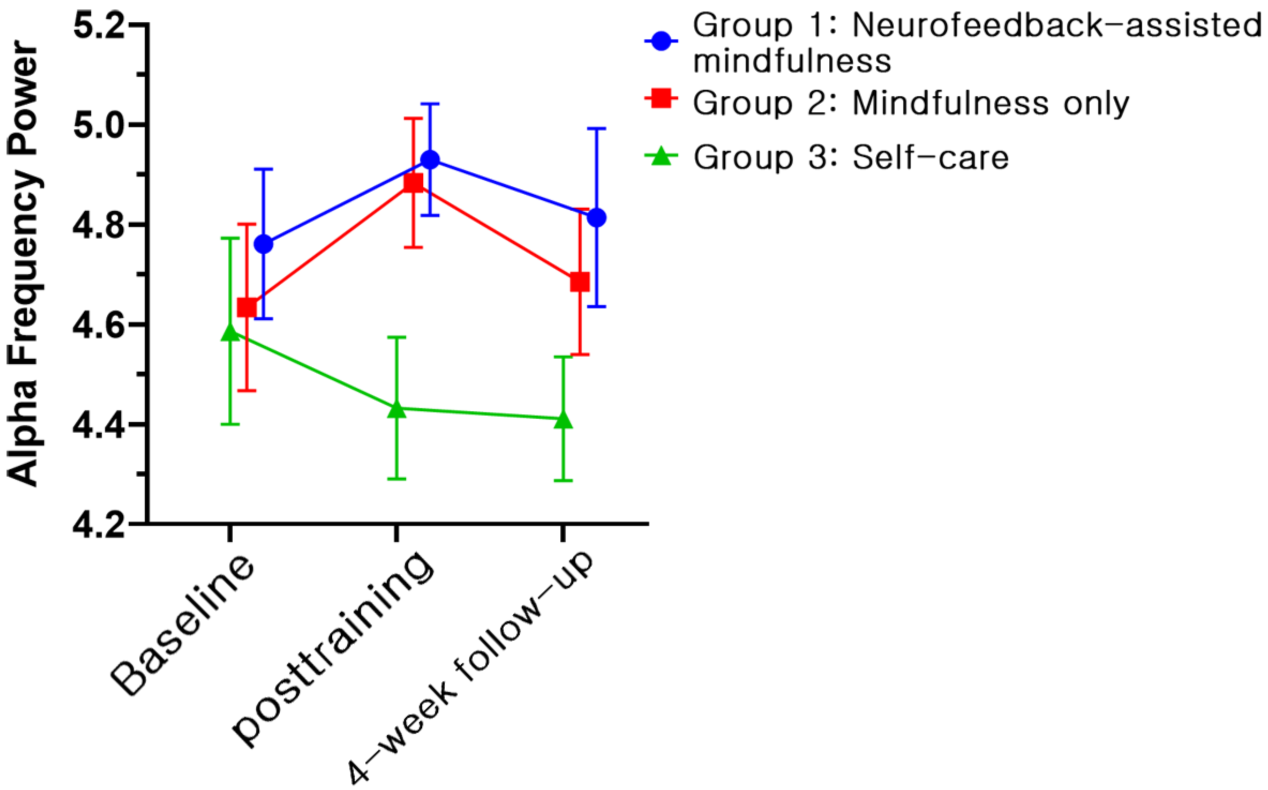


Figure 11. Power of the high beta frequency band of the right electroencephalogram channel (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.

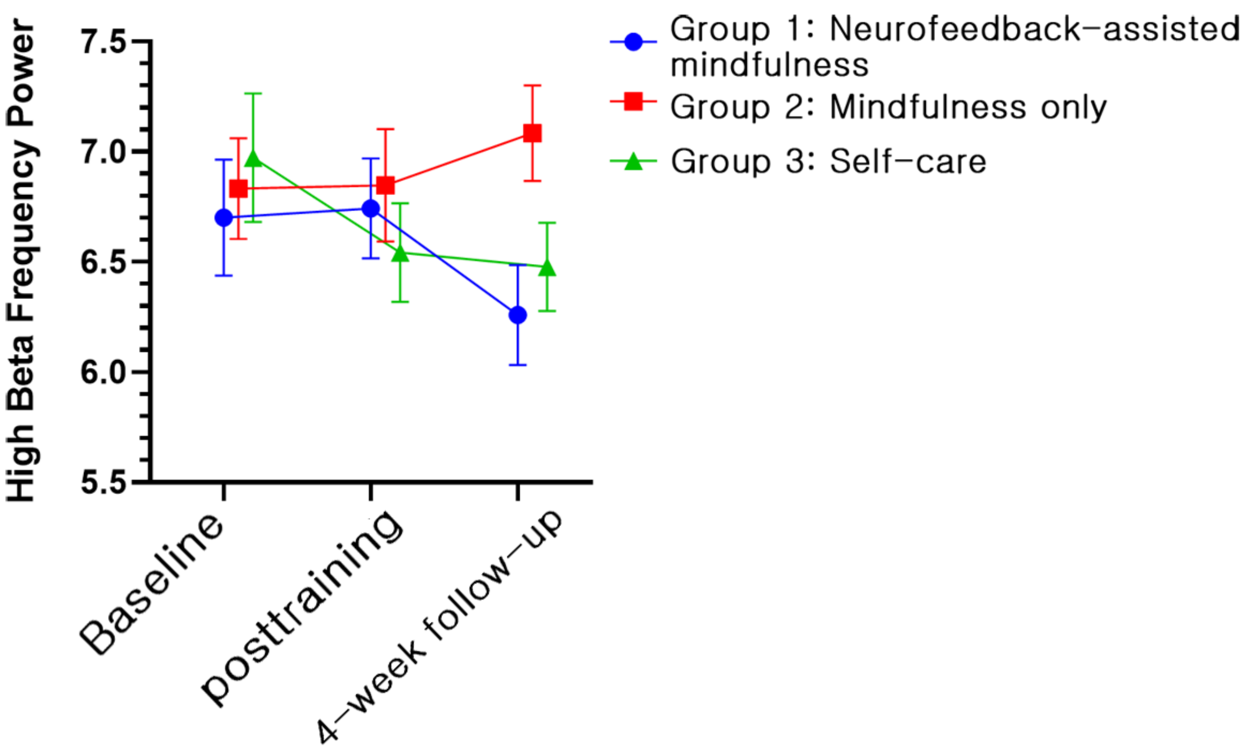
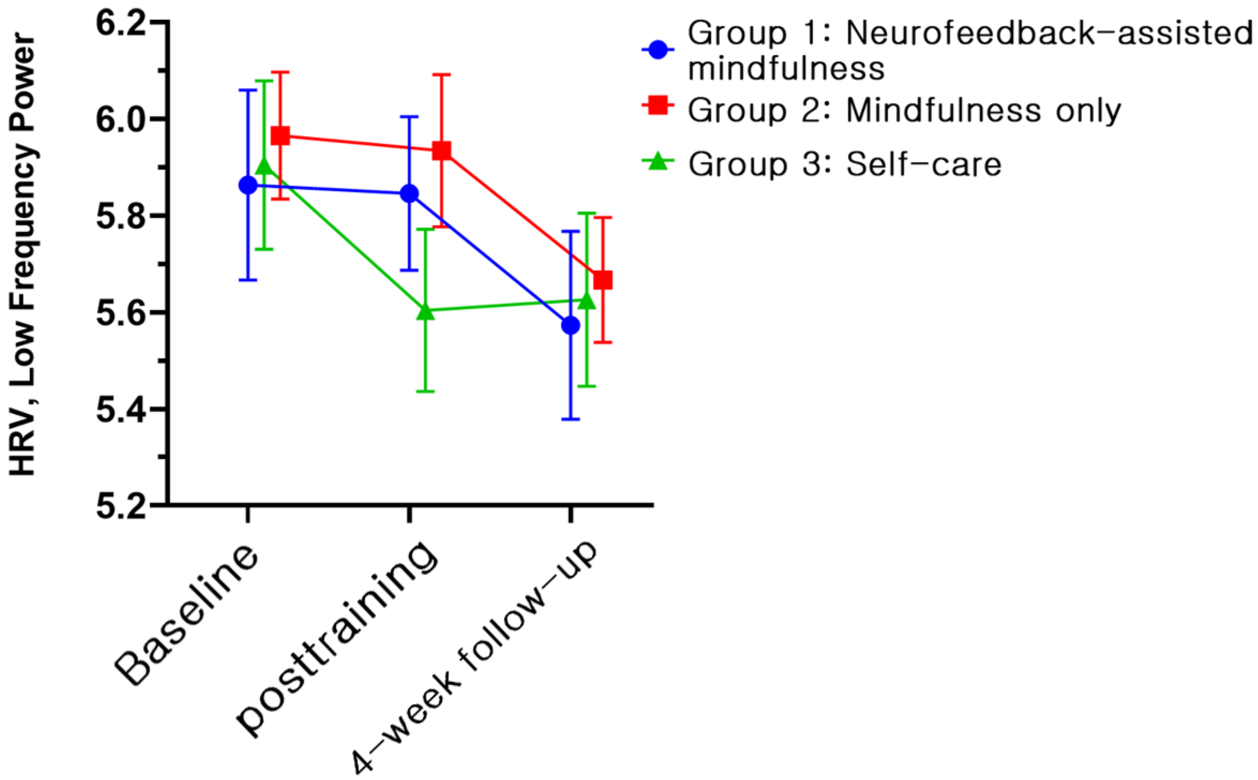


Figure 12. Low frequency power of heart rate variability across time according to condition (least squares means and SE). Group 1: neurofeedback-assisted mindfulness training; group 2: mindfulness training only; group 3: self-care.



Discussion

Principal Findings

In this study, we found that a 4-week mindfulness-based, self-training program delivered via a mobile app effectively reduced stress and improved other psychological indices among employees. Additionally, we hypothesized that the neurofeedback-assisted training might enhance the effects of mindfulness training. A significant interaction (times × group) was found in the BRS between the neurofeedback-assisted mindfulness training group and the control group, with the former showing a significant improvement at the posttraining assessment, with continued improvements observed during the 1-month follow-up assessment period. Group 2 (which received mindfulness training only) also showed significant improvement at the posttraining assessment, but at the 1-month follow-up, the improvement was not significant. In the relaxation index, the neurofeedback-assisted mindfulness training group showed ongoing improvements compared with the baseline assessment, with the improvement having significant differences compared with other groups at the 1-month follow-up assessment.

Comparison With Prior Studies

Originally, high-fidelity MBSR [53] required an 8-week training program with a 7.5-hour silent retreat during week 6. However, this type of program is difficult for general employees to participate in. To address this barrier, various mindfulness-based training programs have been developed, with shorter duration or incorporation of other kinds of programs. These modified programs have been proven to be effective [13]. In our study, the mobile app delivered 3 training modules (awareness training, abdominal breathing, and body scan) over 4 weeks. One of the primary outcome measures was resilience using BRS. Group 1 showed the largest improvement compared with other groups (Figure 5) at the second assessment. The group also showed a significant training effect compared with group 3, which was also the case at the 1-month follow-up assessment.

Previous studies have suggested that mindfulness training could enhance resilience [54-56]. In addition, one study has reported that increased mindfulness may be related to increased resilience and decreased burnout [54]. Another study has demonstrated that high resilience and higher mindfulness may be significant predictors of lower levels of burnout and psychological distress [57]. Moreover, resilience has been known to be correlated with less stress, better mental health, and more mindfulness [58].

The other outcome measures of our study were evaluated based on the effects of mindfulness intervention programs reported in previously published studies. Several studies have reported improvements in PSS scores after a 4-week mindfulness training program [59,60]. Regarding emotional labor, mindfulness-based intervention appeared to cause less emotional exhaustion [61,62]. One study reported the mediating role of emotional labor in the negative correlation between mindfulness and burnout [63]. Regarding the effect of mindfulness on sleep, while some evidence suggests improved sleep, a systematic review concluded that controlled studies have not consistently demonstrated positive effects of MBSR on sleep quality and duration [64]. There were no statistically significant differences between groups 2 and 3 in these psychological measures. This could be attributed to the self-learning effect observed in the control group, which received the self-learning materials. According to a recent qualitative review of 67 published studies, only 5 had an active control group (7%) [13]. There was only 1 study with a 3-arm design that included an inactive control group and an active control group. The design and total number of training hours in that study were similar to those of our study; however, it did not find significant effects of the mindfulness-based program, even in the experimental group [65]. Two other 2-armed studies with experimental and active control groups (education only) conducted over 8 weeks yielded mixed results. In one of these studies, no significant differences were found between the groups on the PSS [66], whereas in the other study, significant differences were observed between the 2 groups on work-related stress scales [67].

On the other hand, for the relaxation index of the right prefrontal channel, the neurofeedback-assisted mindfulness training group exhibited significant differences compared with the other groups at the 1-month follow-up assessment. We depict the changes in alpha and high beta frequency power in Figures 10 and 11, respectively. The alpha power of groups 1 and 2 moved in the same direction during the measurements, but the high beta power of groups 1 and 2 diverged after the posttraining assessment. Notably, the alpha and high beta power of the control group followed a completely different trajectory compared with groups 1 and 2. This transition resulted in a significant interaction, further emphasizing the differences between the groups. Group 1 demonstrated the highest relaxation index at the last assessment. In summary, our findings suggest that a 4-week mindfulness-based training program may increase the power of alpha frequency in groups 1 and 2. Several studies have indicated that mindfulness training may be associated with enhanced alpha and theta power, although changes in beta, delta, and gamma power may exhibit inconsistent patterns [28].

For the power of LF in HRV, all 3 groups exhibited a decreased pattern without significant differences between the groups or interactions (Figure 12). This suggests that neurofeedback-assisted mindfulness training did not demonstrate additional effects on HRV compared with the other groups. The power of the LF band in HRV is known to be influenced by parasympathetic and sympathetic nervous activity, as well as baroreflex function [68]. Interpretation of HRV should be approached with caution [69] due to various contextual factors, including artifacts [70], measurement techniques [71], and others. Several studies have reported decreased LF band power following mindfulness-based training programs [72,73], which aligns with our results. However, another study reported no significant changes in LF but observed an increase in the LF-to-HF ratio [74], which is not consistent with our findings.

Suggestions and Limitations

In this study, we found that the effects of neurofeedback-assisted mindfulness training persisted even during the follow-up period. Future research examining these effects over a more extended duration would be valuable. Additionally, we only recruited full-time employees, which could be considered a limitation. Therefore, future studies would benefit from including a broader range of participants across different conditions, encompassing a more diverse array of physiological variables, and employing larger sample sizes.

Conclusions

In this study, we developed a mindfulness training program embedded in a mobile app and verified its effectiveness over 4 weeks. The program delivered by the mobile app was effective on PSS, BRS, KELS, AIS, and PHQ-9 scores. The neurofeedback-assisted mindfulness training group showed a significant difference in BRS compared with the control group. In addition, the effect of neurofeedback-assisted mindfulness training on relaxation remained significant at the 1-month follow-up assessment. Therefore, we could suggest that a 4-week program via a mobile app may contribute to stress reduction of employees and could improve resilience and relaxation when the mindfulness training is supported by the neurofeedback function.

Acknowledgments

This research was supported by a fund from the Ministry of Trade, Industry and Energy of South Korea (No. 10069105 to J-HK).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 1208 KB - mhealth_v11i1e42851_app1.pdf]

References

1. European Agency for Safety and Health at Work, Milczarek M, Schneider E, González E. European Agency for Safety and Health at work. In: OSH in figures: stress at work--facts and figures. Luxembourg: Office for Official Publications of the European Communities; 2009.
2. Brun JP. Work-related stress : Scientific evidence-base of risk factors, prevention and costs. DocPlayer. 2007. URL: https://docplayer.net/20631009-Work-related-stress-scientific-evidence-base-of-risk-factors-prevention-and-costs.html [accessed 2023-09-05]

3. Kivimäki M, Leino-Arjas P, Luukkonen R, Riihimäki H, Vahtera J, Kirjonen J. Work stress and risk of cardiovascular mortality: prospective cohort study of industrial employees. BMJ 2002 Oct 19;325(7369):857 [FREE Full text] [doi: 10.1136/bmj.325.7369.857] [Medline: 12386034]
4. Lundberg U, Dohns IE, Melin B, Sandsjö L, Palmerud G, Kadefors R, et al. Psychophysiological stress responses, muscle tension, and neck and shoulder pain among supermarket cashiers. J Occup Health Psychol 1999;4(3):245-255. [doi: 10.1037/1076-8998.4.3.245]
5. Heraclides A, Chandola T, Witte D, Brunner E. Psychosocial stress at work doubles the risk of type 2 diabetes in middle-aged women: evidence from the Whitehall II study. Diabetes Care 2009 Dec;32(12):2230-2235 [FREE Full text] [doi: 10.2337/dc09-0132] [Medline: 19720842]
6. Siegrist J. Chronic psychosocial stress at work and risk of depression: evidence from prospective studies. Eur Arch Psychiatry Clin Neurosci 2008 Nov 5;258 Suppl 5(S5):115-119. [doi: 10.1007/s00406-008-5024-0] [Medline: 18985307]
7. Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE. Work stress precipitates depression and anxiety in young, working women and men. Psychol. Med 2007 Apr 04;37(8):1119-1129. [doi: 10.1017/s0033291707000414]
8. Hammen C. Stress and depression. Annu Rev Clin Psychol 2005;1:293-319. [doi: 10.1146/annurev.clinpsy.1.102803.143938] [Medline: 17716090]
9. Mor Barak M, Nissly J, Levin A. Antecedents to Retention and Turnover among Child Welfare, Social Work, and Other Human Service Employees: What Can We Learn from Past Research? A Review and Metanalysis. Social Service Review 2001 Dec;75(4):625-661. [doi: 10.1086/323166]
10. Cooper C, Dewe P. Well-being--absenteeism, presenteeism, costs and challenges. Occup Med (Lond) 2008 Dec 01;58(8):522-524. [doi: 10.1093/occmed/kqn124] [Medline: 19054749]
11. Dewa CS, Loong D, Bonato S, Thanh NX, Jacobs P. How does burnout affect physician productivity? A systematic literature review. BMC Health Serv Res 2014 Jul 28;14:325 [FREE Full text] [doi: 10.1186/1472-6963-14-325] [Medline: 25066375]
12. Work-related stress, anxiety or depression statistics in Great Britain. Health and Safety Executive. 2019. URL: https://www.hse.gov.uk/statistics/causdis/stress.pdf [accessed 2023-09-05]
13. Eby LT, Allen TD, Conley KM, Williamson RL, Henderson TG, Mancini VS. Mindfulness-based training interventions for employees: A qualitative review of the literature. Human Resource Management Review 2019 Jun;29(2):156-178. [doi: 10.1016/j.hrmr.2017.03.004]
14. Hyland PK, Lee RA, Mills MJ. Mindfulness at work: A new approach to improving individual and organizational performance. Ind Organ Psychol 2015 Jul 15;8(4):576-602. [doi: 10.1017/iop.2015.41]
15. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
16. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. J Psychosom Res 2015 Jun;78(6):519-528. [doi: 10.1016/j.jpsychores.2015.03.009] [Medline: 25818837]
17. McConville J, McAleer R, Hahne A. Mindfulness Training for Health Profession Students-The Effect of Mindfulness Training on Psychological Well-Being, Learning and Clinical Performance of Health Professional Students: A Systematic Review of Randomized and Non-randomized Controlled Trials. Explore (NY) 2017;13(1):26-45. [doi: 10.1016/j.explore.2016.10.002] [Medline: 27889445]
18. Ghawadra SF, Abdullah KL, Choo WY, Phang CK. Mindfulness-based stress reduction for psychological distress among nurses: A systematic review. J Clin Nurs 2019 Nov 22;28(21-22):3747-3758. [doi: 10.1111/jocn.14987] [Medline: 31267619]
19. Burton A, Burgess C, Dean S, Koutsopoulou GZ, Hugh-Jones S. How effective are mindfulness-based interventions for reducing stress among healthcare professionals? A systematic review and meta-analysis. Stress Health 2017 Feb 25;33(1):3-13 [FREE Full text] [doi: 10.1002/smi.2673] [Medline: 26916333]
20. Grepmair L, Mitterlehner F, Loew T, Bachler E, Rother W, Nickel M. Promoting mindfulness in psychotherapists in training influences the treatment results of their patients: a randomized, double-blind, controlled study. Psychother Psychosom 2007;76(6):332-338. [doi: 10.1159/000107560] [Medline: 17917468]
21. Irving J, Dobkin P, Park J. Cultivating mindfulness in health care professionals: a review of empirical studies of mindfulness-based stress reduction (MBSR). Complement Ther Clin Pract 2009 May;15(2):61-66. [doi: 10.1016/j.ctcp.2009.01.002] [Medline: 19341981]
22. Chan S, Torous J, Hinton L, Yellowlees P. Mobile Tele-Mental Health: Increasing Applications and a Move to Hybrid Models of Care. Healthcare (Basel) 2014 May 06;2(2):220-233 [FREE Full text] [doi: 10.3390/healthcare2020220] [Medline: 27429272]
23. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018 Mar;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
24. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
25. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active Randomized Controlled Trial and experience sampling study. JMIR Ment Health 2019 Jan 08;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]

26. Zeidan F, Johnson SK, Diamond BJ, David Z, Goolkasian P. Mindfulness meditation improves cognition: evidence of brief mental training. Conscious Cogn 2010 Jun;19(2):597-605. [doi: 10.1016/j.concog.2010.03.014] [Medline: 20363650]
27. Carmody J, Baer RA. How long does a mindfulness-based stress reduction program need to be? A review of class contact hours and effect sizes for psychological distress. J Clin Psychol 2009 Jun;65(6):627-638. [doi: 10.1002/jclp.20555] [Medline: 19309694]
28. Lomas T, Ivtzan I, Fu CH. A systematic review of the neurophysiology of mindfulness on EEG oscillations. Neurosci Biobehav Rev 2015 Oct;57:401-410. [doi: 10.1016/j.neubiorev.2015.09.018] [Medline: 26441373]
29. Chow T, Javan T, Ros T, Frewen P. EEG Dynamics of Mindfulness Meditation Versus Alpha Neurofeedback: a Sham-Controlled Study. Mindfulness 2016 Oct 19;8(3):572-584. [doi: 10.1007/s12671-016-0631-8]
30. Navarro Gil M, Escolano Marco C, Montero-Marín J, Minguez Zafra J, Shonin E, García Campayo J. Efficacy of Neurofeedback on the Increase of Mindfulness-Related Capacities in Healthy Individuals: a Controlled Trial. Mindfulness 2017 Jul 14;9(1):303-311. [doi: 10.1007/s12671-017-0775-1]
31. Sas C, Chopra R. MeditAid: a wearable adaptive neurofeedback-based system for training mindfulness state. Pers Ubiquit Comput 2015 Sep 4;19(7):1169-1182. [doi: 10.1007/s00779-015-0870-z]
32. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
33. Yoo S, Kim Y, Noh J, Oh K, Kim C, Namkoong K, et al. Validity of Korean Version of the Mini-International Neuropsychiatric Interview. Anxiety and Mood 2006;2(1):50-55 [FREE Full text]
34. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
35. Taylor JM. Psychometric analysis of the Ten-Item Perceived Stress Scale. Psychol Assess 2015 Mar;27(1):90-101. [doi: 10.1037/a0038100] [Medline: 25346996]
36. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
37. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
38. Sun Jung Kwon, Kim Kyo Heon. Validation of the Korean Version of Mindful Attention Awareness Scale. Korean Journal of Health Psychology 2007 Mar;12(1):269-287. [doi: 10.17315/kjhp.2007.12.1.014]
39. Lee JE, Han EM, Hong HG, Lee IJ. Validation of the Korean Version of the Emotional Labor Scale (ELS). The Korean Journal of Health Psychology 2016 Mar;21(1):243-256. [doi: 10.17315/kjhp.2016.21.1.012]
40. Grandey AA. Emotion regulation in the workplace: a new way to conceptualize emotional labor. J Occup Health Psychol 2000 Jan;5(1):95-110. [doi: 10.1037//1076-8998.5.1.95] [Medline: 10658889]
41. Grandey AA, Melloy RC. The state of the heart: Emotional labor as emotion regulation reviewed and revised. J Occup Health Psychol 2017 Jul;22(3):407-422. [doi: 10.1037/ocp0000067] [Medline: 28150996]
42. Chang S, Koh S, Kang D, Kim S, Kang M, Lee C, et al. Developing an Occupational Stress Scale for Korean Employees. Korean J Occup Environ Med 2005 Dec;17(4):297 [FREE Full text] [doi: 10.35371/kjoem.2005.17.4.297]
43. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [doi: 10.1016/s0022-3999(00)00095-7] [Medline: 11033374]
44. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
45. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
46. Hou X, Liu Y, Sourina O, Tan Y, Wang L, Mueller-Wittig W. EEG Based Stress Monitoring. New York, NY: IEEE; 2015 Presented at: The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC2015); October 9-12, 2015; Hong Kong, China p. 3110-3115 URL: https://www.researchgate.net/publication/283441378_EEG_based_Stress_Monitoring [doi: 10.1109/smc.2015.540]
47. Neuronicle FX2. Laxtha Inc. URL: http://laxtha.net/neuronicle-fx2/#specifications-neuronicle-fx2 [accessed 2023-09-05]
48. Choi J, Ku B, You YG, Jo M, Kwon M, Choi Y, et al. Resting-state prefrontal EEG biomarkers in correlation with MMSE scores in elderly individuals. Sci Rep 2019 Jul 18;9(1):10468 [FREE Full text] [doi: 10.1038/s41598-019-46789-2] [Medline: 31320666]
49. Evans JR, Abarbanel A. Introduction to Quantitative EEG and Neurofeedback. Amsterdam, The Netherlands: Elsevier; 1999.
50. Marzbani H, Marateb HR, Mansourian M. Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical Applications. Basic Clin Neurosci 2016 Apr;7(2):143-158 [FREE Full text] [doi: 10.15412/J.BCN.03070208] [Medline: 27303609]
51. Kim GS. Clinical Study for Statistical Standardization of Quantitative EEG and PTG Parameters based on Healthy Volunteers. In: Proceedings of the Korean Society for Emotion and Sensibility Conference. Daejeon, South Korea: Korean Society for

Emotion and Sensibility; 2008 Presented at: The Korean Society for Emotion and Sensibility Conference; May 22, 2009; Jeju, Korea p. 15.
52. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
53. Kabat-Zinn J, Hanh TN. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Random House Publishing Group; 2009.
54. Kaplan JB, Bergman AL, Christopher M, Bowen S, Hunsinger M. Role of Resilience in Mindfulness Training for First Responders. Mindfulness (N Y) 2017 Oct 19;8(5):1373-1380 [FREE Full text] [doi: 10.1007/s12671-017-0713-2] [Medline: 34484459]
55. Nila K, Holt DV, Ditzen B, Aguilar-Raab C. Mindfulness-based stress reduction (MBSR) enhances distress tolerance and resilience through changes in mindfulness. Mental Health & Prevention 2016 Mar;4(1):36-41. [doi: 10.1016/j.mhp.2016.01.001]
56. Joyce S, Shand F, Tighe J, Laurent SJ, Bryant RA, Harvey SB. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. BMJ Open 2018 Jun 14;8(6):e017858 [FREE Full text] [doi: 10.1136/bmjopen-2017-017858] [Medline: 29903782]
57. Harker R, Pidgeon AM, Klaassen F, King S. Exploring resilience and mindfulness as preventative factors for psychological distress burnout and secondary traumatic stress among human service professionals. Work 2016 Jun 08;54(3):631-637. [doi: 10.3233/WOR-162311] [Medline: 27286075]
58. Kemper KJ, Mo X, Khayat R. Are Mindfulness and Self-Compassion Associated with Sleep and Resilience in Health Professionals? J Altern Complement Med 2015 Aug;21(8):496-503 [FREE Full text] [doi: 10.1089/acm.2014.0281] [Medline: 26218885]
59. Baer RA, Carmody J, Hunsinger M. Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. J Clin Psychol 2012 Jul;68(7):755-765. [doi: 10.1002/jclp.21865] [Medline: 22623334]
60. Manotas M, Segura C, Eraso M, Oggins J, McGovern K. Association of brief mindfulness training with reductions in perceived stress and distress in Colombian health care professionals. Int J Stress Manag 2014 May;21(2):207-225. [doi: 10.1037/a0035150]
61. Hülsheger UR, Alberts HJEM, Feinholdt A, Lang JWB. Benefits of mindfulness at work: the role of mindfulness in emotion regulation, emotional exhaustion, and job satisfaction. J Appl Psychol 2013 Mar;98(2):310-325. [doi: 10.1037/a0031313] [Medline: 23276118]
62. Li J, Wong IA, Kim WG. Does mindfulness reduce emotional exhaustion? A multilevel analysis of emotional labor among casino employees. Int J Hosp Manage 2017 Jul;64:21-30. [doi: 10.1016/j.ijhm.2017.03.008]
63. Ma Y, Wang F, Cheng X. Kindergarten Teachers' Mindfulness in Teaching and Burnout: The Mediating Role of Emotional Labor. Mindfulness 2020 Nov 27;12(3):722-729. [doi: 10.1007/s12671-020-01538-9]
64. Winbush NY, Gross CR, Kreitzer MJ. The effects of mindfulness-based stress reduction on sleep disturbance: a systematic review. Explore (NY) 2007;3(6):585-591. [doi: 10.1016/j.explore.2007.08.003] [Medline: 18005910]
65. Burnett M, Pettijohn C. Investigating the efficacy of mind-body therapies and emotional intelligence on worker stress in an organizational setting: An experimental approach. J Organ Cult Commun Confl 2015;19(1):146-158 [FREE Full text]
66. Malarkey WB, Jarjoura D, Klatt M. Workplace based mindfulness practice and inflammation: a randomized trial. Brain Behav Immun 2013 Jan;27(1):145-154 [FREE Full text] [doi: 10.1016/j.bbi.2012.10.009] [Medline: 23078984]
67. Shonin E, Van Gordon W, Dunn TJ, Singh NN, Griffiths MD. Meditation Awareness Training (MAT) for Work-related Wellbeing and Job Performance: A Randomised Controlled Trial. Int J Ment Health Addiction 2014 Aug 7;12(6):806-823. [doi: 10.1007/s11469-014-9513-2]
68. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
69. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;5:1040 [FREE Full text] [doi: 10.3389/fpsyg.2014.01040] [Medline: 25324790]
70. Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. Front Physiol 2012;3:148 [FREE Full text] [doi: 10.3389/fphys.2012.00148] [Medline: 22654764]
71. Jeyhani V, Mahdiani S, Peltokangas M, Vehkaoja A. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5952-5955. [doi: 10.1109/EMBC.2015.7319747] [Medline: 26737647]
72. May R, Bamber M, Seibert G, Sanchez-Gonzalez M, Leonard J, Salsbury R, et al. Understanding the physiology of mindfulness: aortic hemodynamics and heart rate variability. Stress 2016;19(2):168-174. [doi: 10.3109/10253890.2016.1146669] [Medline: 26822542]
73. Krygier JR, Heathers JAJ, Shahrestani S, Abbott M, Gross JJ, Kemp AH. Mindfulness meditation, well-being, and heart rate variability: a preliminary investigation into the impact of intensive Vipassana meditation. Int J Psychophysiol 2013 Sep;89(3):305-313. [doi: 10.1016/j.ijpsycho.2013.06.017] [Medline: 23797150]

74. Delgado-Pastor LC, Perakakis P, Subramanya P, Telles S, Vila J. Mindfulness (Vipassana) meditation: effects on P3b event-related potential and heart rate variability. Int J Psychophysiol 2013 Dec;90(2):207-214. [doi: 10.1016/j.ijpsycho.2013.07.006] [Medline: 23892096]

Abbreviations

AIS: Athens Insomnia Scale
BRS: Brief Resilience Scale
EEG: electroencephalography
GEE: generalized estimating equation
HF: high frequency
HRV: heart rate variability
KELS: Korean Emotional Labor Scale
K-MAAS: Korean version of the Mindfulness Attention Awareness Scale
KOSS: Korean Standard Occupational Stress Scale
LF: low frequency
LSM: least squares mean
MBSR: mindfulness-based stress reduction
MINI: Mini International Neuropsychiatric Interview
PHQ-9: 9-item Patient Health Questionnaire
PPG: plethysmograph
PSS: Perceived Stress Scale
REDCap: Research Electronic Data Capture
SAS: Statistical Analysis System
Wi-Fi: wireless fidelity

Edited by L Buis; submitted 21.09.22; peer-reviewed by I Shubina, N Khalili-Mahani; comments to author 21.02.23; revised version received 14.06.23; accepted 18.08.23; published 03.10.23.

Please cite as:
Min B, Park H, Kim JI, Lee S, Back S, Lee E, Oh S, Yun JY, Kim BN, Kim Y, Hwang J, Lee S, Kim JH
The Effectiveness of a Neurofeedback-Assisted Mindfulness Training Program Using a Mobile App on Stress Reduction in Employees: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e42851
URL: https://mhealth.jmir.org/2023/1/e42851
doi:10.2196/42851
PMID:37788060

Original Paper

App-Based Mindfulness for Attenuation of Subjective and Physiological Stress Reactivity in a Population With Elevated Stress: Randomized Controlled Trial

Ulrich Kirk[1], PhD; Walter Staiano[1,2], PhD; Emily Hu[3], PhD; Christelle Ngnoumen[3,4], PhD; Sarah Kunkle[3], PhD; Emily Shih[3], PhD; Alicia Clausel[3], MSc; Clare Purvis[3,5], PhD; Lauren Lee[3], PhD

[1]Department of Psychology, University of Southern Denmark, Odense, Denmark

[2]Department of Physical Education and Sport, University of Valencia, Valencia, Spain

[3]Headspace Inc, Santa Monice, CA, United States

[4]Department of Psychology, Harvard University, Cambridge, MA, United States

[5]Clinical Excellence Research Center, School of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Ulrich Kirk, PhD
Department of Psychology
University of Southern Denmark
Campusvej 55
Odense, 5230
Denmark
Phone: 45 65502695
Email: ukirk@health.sdu.dk

Abstract

Background: Stress-related mental health disorders have steadily increased and contributed to a worldwide disease burden with up to 50% experiencing a stress-related mental health disorder worldwide. Data suggest that only approximately 20%-65% of individuals receive treatment. This gap in receiving treatment may be attributed to barriers such as limited treatment access, negative stigma surrounding mental health treatment, approachability (ie, not having a usual treatment plan or provider), affordability (ie, lack of insurance coverage and high treatment cost), and availability (ie, long waits for appointments) leaving those who need treatment without necessary care. To mitigate the limited access mental health treatment, there has been a rise in the application and study of digital mental health interventions. As such, there is an urgent need and opportunity for effective digital mental health interventions to alleviate stress symptoms, potentially reducing adverse outcomes of stress-related disorders.

Objective: This study examined if app-based guided mindfulness could improve subjective levels of stress and influence physiological markers of stress reactivity in a population with elevated symptoms of stress.

Methods: The study included 163 participants who had moderate to high perceived stress as assessed by the Perceived Stress Scale (PSS-10). Participants were randomly allocated to 1 of 5 groups: a digital guided program designed to alleviate stress (Managing Stress), a digital mindfulness fundamentals course (Basics), digitally delivered breathing exercises, an active control intervention (Audiobook), and a Waitlist Control group. The 3 formats of mindfulness interventions (Managing Stress, Basics, and Breathing) all had a total duration of 300 minutes spanning 20-30 days. Primary outcome measures were perceived stress using the PSS-10, self-reported sleep quality using the Pittsburgh Sleep Quality Index, and trait mindfulness using the Mindful Attention Awareness Scale. To probe the effects of physiological stress, an acute stress manipulation task was included, specifically the cold pressor task (CPT). Heart rate variability was collected before, during, and after exposure to the CPT and used as a measure of physiological stress.

Results: The results showed that PSS-10 and Pittsburgh Sleep Quality Index scores for the Managing Stress (all P<.001) and Basics (all P≤.002) groups were significantly reduced between preintervention and postintervention periods, while no significant differences were reported for the other groups. No significant differences among groups were reported for Mindful Attention Awareness Scale (P=.13). The physiological results revealed that the Managing Stress (P<.001) and Basics (P=.01) groups displayed reduced physiological stress reactivity between the preintervention and postintervention periods on the CPT. There were no significant differences reported for the other groups.

Conclusions: These results demonstrate efficacy of app-based mindfulness in a population with moderate to high stress on improving self-reported stress, sleep quality, and physiological measures of stress during an acute stress manipulation task.

Trial Registration: ClinicalTrials.gov NCT05832632; https://www.clinicaltrials.gov/ct2/show/NCT05832632

(JMIR Mhealth Uhealth 2023;11:e47371) doi:10.2196/47371

KEYWORDS

Mindfulness; mental health; stress; smartphone; technology; Headspace; mobile phone

Introduction

Background

Elevated stress is a state in which an individual experiences excessive or prolonged psychological and physiological strain [1,2]. The impact of elevated stress is wide-ranging and can affect various aspects of an individual's life. Sustained and elevated stress has been shown to be associated with adverse health effects such as obesity, type 2 diabetes, and cardiovascular disease [1-3]. Stress-related disorders are also a contributing factor to the onset of a range of mental health disorders, including depression and anxiety [4-9] and a heightened risk of cardiovascular disease [10]. As such, the connection between stress-related disorders, mental health disorders, and negative physical outcomes is well-established. Therefore, it is important to develop evidence-based mental health interventions to help reduce stress and therefore mitigate the adverse outcomes of stress-related disorders.

Mindfulness is defined as a state of being attentive to and bringing awareness to sensations that are taking place in the present moment without judgment [11]. Mindfulness is considered an evidence-based practice where one aims to reduce the effects of stress both psychologically and physiologically [11,12]. The most well-known mindfulness-based intervention is mindfulness-based stress reduction (MBSR) [11], which is delivered in group settings where participants meet on 8 weekly sessions and an experienced mindfulness teacher guides the sessions, provides instructions, and facilitates discussions. MBSR has been shown to be effective for managing stress and its detrimental effects on mental and physical health [13]. Research has shown that MBSR can lead to improved psychological well-being, reduced anxiety, and depressive symptoms, and enhanced overall resilience to stress [14-16].

However, the MBSR program is time consuming and costly. In recent years digital mental health interventions have been developed using self-guided mindfulness-based interventions [17,18]. Specifically, several research studies suggest that the Headspace mindfulness app decreases subjective levels of stress [19-26] and increases sleep quality [27]. Despite growing demands and usage of digital mental health apps, there is a lack of evidence on the efficacy of these interventions in clinical populations [28,29]. That is, research has primarily focused on effects of usage of digital mindfulness apps and interventions in healthy populations [20-26] and found reduced levels of stress in college students, nurses and in the workplace. However, there are a few studies that have looked at the effects of digital mindfulness apps in clinical populations [30,31]. Specifically, a meta-analysis included 15 randomized controlled trials to measure the effect of digital mindfulness interventions and reported significant effects on depression, anxiety, and stress [32]. Interestingly, the study reported higher effect sizes for digital mindfulness interventions with therapist guidance than for digital mindfulness without therapist guidance. Another study using a digital mindfulness intervention found reductions in psychiatric symptoms in adolescents undergoing mental health treatment [33].

Research has also found that mindfulness may improve cognitive functioning in healthy participants, which refers to the mental activities involved in maintaining and acquiring and using information. Specifically, studies using the Headspace mindfulness app have shown that 4 weeks of app-based mindfulness practice reduces behavioral indicators of mind wandering [19,34]. Mind wandering refers to the phenomenon in which the mind drifts away from the current task or focus of attention and becomes engaged in spontaneous thoughts unrelated to the present moment. Separate lines of research have discussed the component of *attention monitoring* embedded in mindfulness as a mechanism that may explain how mindfulness improves cognitive functioning outcomes by reducing mind wandering [35]. An additional component of mindfulness is *acceptance*, which according to the authors is necessary for reducing affective reactivity, such that attention monitoring and acceptance skills together act to improve negative affect and stress-related health outcomes and may explain the mechanism of action behind why mindfulness is thought to influence stress.

Effects of Mindfulness on Physiological Stress Reactivity

In addition to its psychological and cognitive effects, mindfulness has been found to have physiological effects that reduce stress. Regular mindfulness practice has been associated with decreased blood pressure, increased heart rate variability (HRV), and reduced cortisol levels [36-39], which together demonstrate the underlying physiological mechanism of why mindfulness is thought to influence stress-related health outcomes.

Elevated stress is widely regarded as healthy and functional when confronted with acute stressful situations, however permanent exposure to elevated stress—and in particular the failure to recover from stress—may lead to dysfunction of the underlying neurobiology [40]. The sympathetic nervous system (SNS) is one of the 2 main divisions of the autonomic nervous system (ANS), the other being the parasympathetic nervous system (PNS). The SNS plays a crucial role in mobilizing the body's response to perceived threats or stressful situations [41]. The PNS operates in opposition to the SNS and is responsible for facilitating bodily processes during periods of rest. The SNS increases the heart's cardiac output and decreases HRV, which is needed during acute stressful situations. Conversely, the PNS slows the heart rate and increases HRV to restore homeostasis. This natural interplay between these 2 systems allows the heart to quickly respond to different situations and needs based on the context [42]. It is generally assumed that HRV, a measure of beat-to-beat variability in heart rate, is mediated by the ANS [43,44]. Currently, there is not a universally recognized standard for stress evaluation [45]. However, studies on HRV and stress reactivity are increasing in frequency [45]. Evidence suggests that the physiological measure of HRV is impacted by stress and supports its use when assessing psychological health and stress [45]. Thus, we used HRV in this study to measure the effect of how digital mindfulness interventions impact physiological markers of stress.

Research studies have found that engaging in mindfulness can increase HRV [36-39,46,47]. Taken together, these studies suggest that mindfulness activates the PNS and promotes a state of relaxation. A higher HRV indicates a flexible ANS that can adapt to changing circumstances and shift between states of activation (SNS) and relaxation (PNS). By increasing HRV, mindfulness may thus improve emotional regulation, reduce stress reactivity, and enhance resilience to stress.

Effects of Mindfulness on Acute Stress

The cold pressor task (CPT) is a laboratory test commonly used to induce a physiological stress response in participants. The CPT involves immersing the participant's hand in an ice-cold water bath for 3 minutes, which causes vasoconstriction in the submerged hand. This triggers a physiological response, which activates the SNS, leading to an increase in heart rate, blood pressure, and peripheral vasoconstriction [48-54]. The CPT is widely used as a stress manipulation that elicits and models the effects of mild to moderate acute stress that participants might encounter in their everyday life [48,49,51,53,55,56]. Studies suggest that administration of the CPT disrupts executive functioning including working memory capacity [57]. Therefore, acute stress has a profound negative impact on cognitive functioning, and thus the CPT serves as an excellent probe for interventions such as digital mindfulness to test if digital mindfulness interventions influence physiological markers of stress reactivity during exposure to acute stress. A recent study demonstrated that administration of the Headspace mindfulness app for a period of 4 weeks mediated the relationship between cognitive performance and acute stress [24]. The results showed that the digital mindfulness intervention uncoupled the relationship between cognitive performance and acute stress, meaning that participants who underwent mindfulness training were less affected by stress during cognitive performance compared to the control group. The findings of this study suggest that mindfulness training may be a useful approach to mitigate the negative impact of acute stress.

Aims of the Study

This study examined if 3 different formats of digital mindfulness interventions demonstrated efficacy in terms of reducing self-reported levels of stress, sleep quality, and influencing physiological markers of stress reactivity in a population with elevated levels of stress.

To accomplish the experimental aim, participants with moderate and high stress according to the Perceived Stress Scale (PSS-10) [58] were randomized to 1 of 5 groups (3 formats of mindfulness, 1 active control, and 1 Waitlist Control). The study investigated if changes in subjective stress (PSS-10), sleep quality (PSQI) and trait mindfulness (Mindful Attention Awareness Scale [MAAS]) showed differences between the preintervention and postintervention periods.

The 3 types of mindfulness interventions were identical in total training dosage but varied in content and intervention length. That is, 2 of the 3 mindfulness interventions (Basics and Breathing) consisted of 30 sessions, whereas the third intervention (Managing Stress) consisted of 20 sessions. The type of content also varied whereby 2 interventions (Basics and Managing Stress) were programmatic that progressed from session to session, whereas the third interventions (Breathing) consisted of single succinct exercises. Furthermore, 1 intervention was designed specifically to reduce stress in people with elevated stress using mindfulness-based content (Managing Stress). This interventional setup allowed us to investigate if the unique characteristic of each intervention would result in differential efficacy in a population with moderate to high stress, while keeping the total training duration of each intervention identical. In other words, the rationale for employing 3 formats of mindfulness interventions was to explore if there were any differential effects between these distinct formats of mindfulness.

To probe psychological (self-reported) effects of stress, the study examined if changes in stress (PSS-10), sleep quality (PSQI), and mindfulness (MAAS) showed differences between the preintervention and postintervention periods.

We hypothesized (H1) that the app-based mindfulness interventions would yield significant improvements in self-reported stress (PSS-10) as compared to the active and Waitlist Control groups. We also hypothesized that trait mindfulness (MAAS) and sleep quality (PSQI) would improve in the mindfulness groups compared to control groups.

To probe physiological effects of stress, the study employed an acute stress manipulation task (ie, the CPT) at the preintervention and postintervention stage while measuring physiological activity (HRV) before, during, and after exposure to the acute stressor, and in addition self-reported stress perception immediately after stress exposure.

We hypothesized (H2) that the app-based mindfulness interventions would result in reduced physiological stress reactivity during exposure to the CPT, expressed as increased HRV activity compared to the active and Waitlist Control groups.

Methods

Participants

A total of 225 research participants who had moderate to high perceived stress based on PSS-10 total scores (14-40) were recruited for the study (see Figure 1). Additional inclusion criteria were men and women between 18 and 60 years of age.

Exclusion criteria were any medical diagnosis, for example, psychiatric or neurological conditions. The reason for these exclusion criteria was to limit comorbidity in our sample by recruiting participants with elevated PSS-10 scores. Other exclusion criteria were regular mindfulness practice for more than 1 month within the last year.

In total, 38 participants were excluded from the analysis either because they did not initiate the intervention practice (n=29), did not show up for the postintervention laboratory visit (n=6) or were unable to complete the CPT procedure (n=3). 24 participants were excluded because they did not meet the minimum engagement dosage of 100 minutes (total engagement duration was 300 minutes). Thus, the total number of participants in our analysis was 163 participants. The rationale for the cut off at a minimum engagement dosage of 100 minutes was to ensure that there were no significant differences in engagement duration across the 4 active intervention groups (Table 1).

Figure 1. CONSORT flow diagram showing the number of participants in each group and the phases of the study from enrollment to analysis. CPT: cold pressor task; CONSORT: consolidated standards of reporting trials.

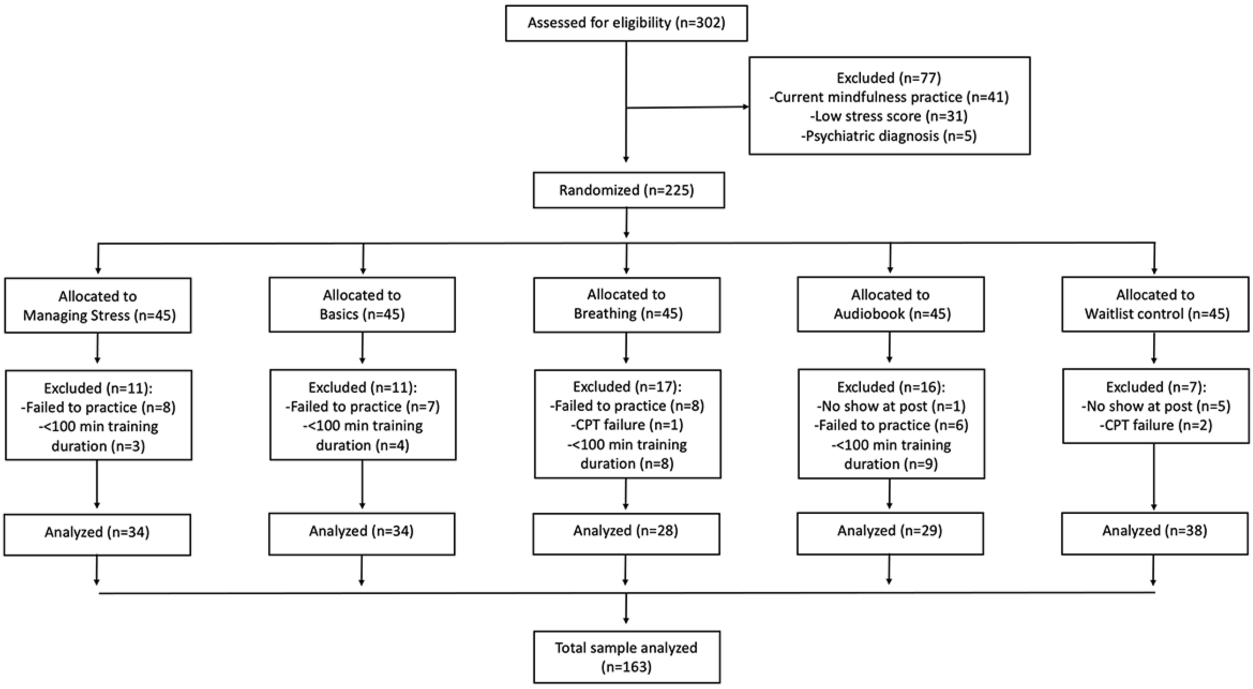


Table 1. Demographic data for the 5 groups, shown as mean and SD collected before and after the intervention.

	Managing Stress	Basics	Breathing	Audiobook	Waitlist Control
Number (females), mean (SD)	34 (17)	34 (18)	28 (13)	29 (14)	38 (18)
Age (years), mean (SD)	24.9 (2.2)	25.3 (2.1)	25.3 (2.3)	25.9 (2.0)	25.7 (2.3)
Race or ethnicity, n (%)					
White	23 (67.6)	21 (61.7)	16 (57.1)	19 (65.5)	25 (65.7)
Hispanic	0 (0)	1 (2.9)	2 (7.1)	1 (3.4)	2 (5.2)
Asian	5 (14.7)	4 (11.7)	4 (14.2)	4 (13.7)	4 (10.5)
Black	3 (8.8)	3 (8.8)	3 (10.7)	2 (6.8)	2 (5.2)
Middle Eastern	3 (8.8)	5 (14.7)	3 (10.7)	3 (10.3)	5 (13.1)
Engagement duration (minutes)[a], mean (SD)	187.5 (72.2)	172.3 (64.1)	163.2 (69.0)	164.8 (57.9)	N/A[b]

[a]The amount of engagement duration was calculated as the total minutes of engagement during the intervention period.

[b]N/A: not applicable.

Ethical Considerations

The participants were recruited through flyers and advertisements at a local university (University of Southern Denmark) and the Region of Southern Denmark. All procedures were conducted in accordance with the local ethical committee (*Videnskabsetisk Komité for Region Syddanmark - project ID: S-20170199*). The study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Participants received compensation for their participation in the study corresponding to a DKK 300 gift card (approximately US $45). All participants provided written consent prior to participation in the study.

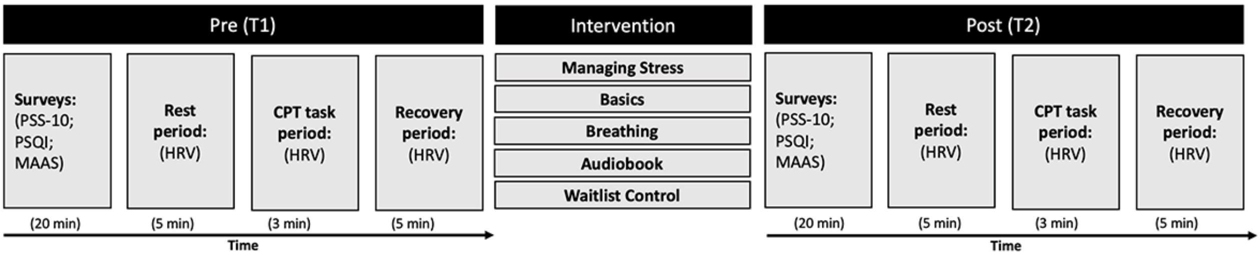
Experimental Procedures

The study included 5 experimental groups using a pre-post design.

Overall, 48 hours prior to the laboratory visit, eligible participants were emailed instructions to refrain from alcohol and nicotine before coming to the laboratory to avoid known influences of these factors on autonomic activity [59-61]. Participants were instructed not to engage in intense physical activity for the 48-hour period but were otherwise asked to maintain their daily and nightly routines. Both the pre- and postintervention laboratory visits were conducted on weekdays between 11 AM and 4 PM. This time window was kept constant across measurement periods to reduce variance of the circadian rhythm [62,63].

Upon arrival for the first laboratory visit (ie, preintervention or T1), the participants initially signed the consent form. Participants were subsequently instructed on how to complete the experimental procedures (Figure 2). The experimental procedures consisted of completing surveys (see *psychological measures*), and subsequently completing the CPT, that is, the stress manipulation whilst HRV was collected (see *physiological measures*). Specifically, participants were seated in a chair while the physiological HRV monitor was applied using 2 electrodes. After making sure that the HRV monitor was correctly applied, the participants were instructed about the CPT. The participants were asked to rest for 5 minutes by sitting in a chair in an upright position during the *rest period*. Subsequently, participants were instructed to initiate the CPT procedure (*CPT period*; see *acute stress manipulation*). Having completed the stress manipulation, that is, CPT, the participants were instructed to rest for 5 minutes by sitting in a chair in an upright position (*recovery period*).

Figure 2. Outline of the study procedures. CPT: cold pressor task; HRV: heart rate variability; MAAS: Mindful Attention Awareness Scale; PSQI: Pittsburgh Sleep Quality Index; PSS-10: Perceived Stress Scale.



The randomization procedure was determined after study recruitment, but before study launch. Specifically, participants were randomly allocated to 1 of 5 groups (Managing Stress, Basics, Breathing, Audiobook, or Waitlist Control) using sequence generation by the research team, who were not formally blinded to group allocation. Participants were not informed about group allocation until after completion of the preintervention experimental procedures. HRV was collected continuously during the CPT procedure. That is, the CPT procedure was broken up into 3 parts: *rest period*, *CPT period*, and *recovery period* (Figure 2). Following the preintervention laboratory visit, participants were instructed to initiate the 30-day intervention starting the following day. The second laboratory visit procedures (ie, postintervention or T2) were identical to the experimental procedures explained above.

Psychological Measures and Physiological Measures

The experimental methods and outcome measures are described in detail in the Multimedia Appendix 1 [64-68].

Interventions

The interventions were completed using a custom-built research smartphone app. Participants were given access to a smartphone app that was built for the purpose of conducting scientific research at the University of Southern Denmark that contained the training content for each specific intervention. The app has been applied in previous scientific research studies [27,36,69]. The app was set up to contain specific content related to each of the interventions. That is, participants could only access the content in the app belonging to the specific type of intervention they were completing at any given time. Compliance was provided by a backend in the app that tracked the timestamps to keep count of content length belonging to each intervention for each participant.

Participants did not receive an introductory session to the active intervention conditions but were provided with oral and written instructions about the app usage for the intervention period. Participants were instructed to follow the program in full.

Managing Stress (Mindfulness) Intervention

The content of the training was provided by Headspace [70] and based on well-established concepts and practices within the stress [71-74], stress management [75,76], and mindfulness training [11] literature. The program involved a daily combination of evidence-based stress management techniques including an advice video featuring stress psychoeducation, a mindfulness meditation, a relaxation method (ie, progressive muscle relaxation), and a reflective activity (ie, gratitude prompts). All were delivered through short, animated videos and sound files in the app. The Managing Stress program has been tailored for people with elevated stress levels (Figure S2 in Multimedia Appendix 2). Whereas the other 3 active interventions comprised 30 sessions, the Managing Stress intervention entailed 20 total sessions (5 sessions per week) with an average of 15 minutes duration per session. The Managing Stress intervention had the same total duration relative to the duration of the other 3 active interventions (approximately 300 min in total for each intervention), allowing us to investigate if the training frequency yielded a difference across the active interventions.

Basics (Mindfulness) Intervention

The content of the training was provided by Headspace [70] and based on well-established concepts and practices within the mindfulness literature [11]. The course entailed daily practice in guided mindfulness meditation, with instructions delivered through short, animated videos and sound files in the app (Figure S1 in Multimedia Appendix 3). The training course centered on mindfulness meditation, which included focusing on a selected object (ie, the body or the breath), monitoring the activity of the mind, noticing mind-wandering, and developing a nonjudgmental orientation toward one's experience (ie, equanimity). The Basics intervention entailed daily practice sessions of 10-minute duration for a total of 30 sessions.

Breathing (Mindfulness) Intervention

The content of the training was provided by Headspace [70] and based on well-established literature on the positive effects of deep breathing and diaphragmatic breathing on stress [77,78]. The Breathing intervention consisted of brief (around 1 min) guided deep breathing and diaphragmatic breathing exercises. There were 9 different deep breathing exercises that all contained identical instructions. The participants in this group were instructed to freely choose from the exercises and complete 10 exercises (corresponding to 10 min) per day for a total of 30 days. Importantly, this intervention was different from the 2 other mindfulness interventions in that it did not contain programmatic content that progressed over time, but rather consisted of single succinct exercises. However, it is to be noted that similar deep breathing exercises were part of all of the mindfulness interventions, including the Managing Stress, the Basics, and the Breathing interventions.

Audiobook Intervention

Mindfulness has been hypothesized to train attention and affect through interoceptive nonjudgmental awareness [79-81]. Therefore, to deliberately manipulate the active control intervention, this study aimed to de-emphasize these elements by isolating the mechanisms of action in mindfulness. The participants in the active control intervention therefore received instructions to listen to an audiobook. Audiobooks have been used previously as active control for mindfulness in several studies [82-86]. Specifically, in accordance with previous literature we used the following audiobook in this study: "The Natural History of Selborne" by White and Taylor [87]. The general topic of the audiobook contains observations of natural history of the area around Selborne organized more or less systematically by species and group. The intervention entailed listening to the audiobook for a daily duration of 10 minutes for a total of 30 sessions.

Waitlist Control

The Waitlist Control group required that participants did not follow an intervention. However, the Waitlist Control group was given the option to obtain access to 1 of the 3 active interventions after completion of the study.

Statistical Analysis

All data are presented as mean ± 1 SD unless otherwise stated. Assumptions of statistical tests for normal distribution and sphericity of data were checked. A series of mixed groups (Managing Stress, Basics, Breathing, Audiobook, Waitlist control) × time (pretest, posttest) ANOVAs were performed on PSS-10, PSQI, MAAS, and CPT self-reported stress. A series of mixed groups (Managing Stress, Basics, Breathing, Audiobook, Waitlist control) × time (pre, post) × task period (rest, during task, and recovery) ANOVAs were performed on HRV. Significant 3-way interactions were followed-up by Group by Time ANOVAs at each time point, and 2-way interactions were followed-up with relevant corrected pairwise comparisons using the Bonferroni method (post hoc analysis) for simple main effects within each group. Where no significant interactions were found, main effects of time, group, and task period were reported. Significance level was set at 0.05 (2-tailed) for all analyses. The effect sizes for the ANOVAs were calculated as partial eta squared (η^2p), with 0.02, 0.13, and 0.26 indicating small, medium, and large effects, respectively. Data analysis was conducted using SPSS (version 27; IBM Corp).

Results

Measures at Baseline (Pretest)

Age, gender, race, or ethnicity, and engagement duration (in minutes) is presented in Table 1, and behavioral measures (PSS-10, PSQI, MAAS, and self-reported stress during CPT) are presented in Table 2. No statistically significant differences were detected for any of these variables at baseline.

Table 2. Behavioral data for the 5 groups collected pre- and postintervention and percentage change in self-reported measures.

	Managing Stress	Basics	Breathing	Audiobook	Waitlist Control
PSS-10[a] (Pre), mean (SD)	21.6 (3.2)	21.6 (4.7)	22.0 (4.1)	22 (1.42)	21.7 (4.4)
PSS-10 (Post), mean (SD)	18.6 (3.0)	17.4 (4.3)	22.7 (3.5)	23.6 (5.0)	21.8 (4.0)
PSS-10 percentage change, %	–13.8	–14	3.1	7.2	0.4
PSQI[b] (Pre), mean (SD)	7.3 (1.4)	7.3 (1.3)	7.2 (1.6)	7.2 (1.5)	7.2 (1.7)
PSQI (Post), mean (SD)	6.1 (1.3)	6.4 (1.3)	7.0 (1.7)	7.1 (1.4)	7.3 (1.7)
PSQI percentage change, %	–16.4	–12.3	–2.7	–1.3	1.3
MAAS[c] (Pre), mean (SD)	3.0 (0.9)	2.9 (1.0)	2.9 (1.1)	3.0 (1.0)	2.8 (1.2)
MAAS (Post), mean (SD)	3.7 (1.3)	3.4 (1.1)	2.8 (1.0)	2.9 (1.2)	2.9 (1.2)
MAAS percentage change, %	23.3	17.2	–3.4	–3.3	3.5
CPT[d] self-reported stress (Pre), mean (SD)	7.5 (1.3)	7.3 (1.3)	7.9 (1.3)	7.4 (1.3)	7.7 (1.3)
CPT self-reported stress (Post), mean (SD)	6.9 (1.4)	6.9 (1.3)	7.2 (1.5)	7.2 (1.5)	7.3 (1.4)
CPT self-reported stress percentage change, %	–8	–5.4	–8.8	–2.7	–5.1

[a]PSS-10: Perceived Stress Scale.

[b]PSQI: Pittsburgh Sleep Quality Index.

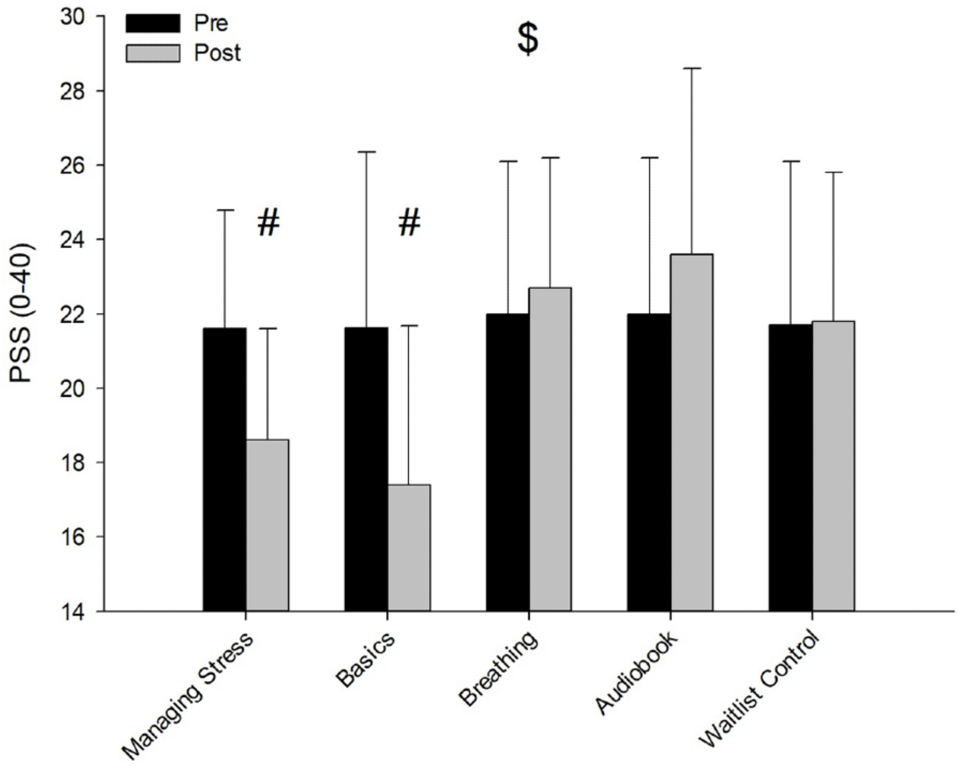
[c]MAAS: Mindful Attention Awareness Scale.

[d]CPT: cold pressor task.

Psychological Measures: PSS-10, PSQI, MAAS, and CPT Self-Reported Stress

There was an interaction for PSS-10 across the 5 groups and time ($F_{4,158}$=6.964; *P*<.001; η²p=0.188). Follow-up tests revealed that the PSS-10 score for the Managing Stress (*P*<.001; η²p=0.096) and the Basics (*P*=.01; η²p=0.060) groups were significantly reduced from the preintervention to the postintervention period, while no significant differences were reported for Breathing (*P*=.51; η²p=0.003), Audiobook (*P*=.08; η²p=0.007) and Control (*P*=.75; η²p=0.001; Figure 3; Table 2).

Figure 3. PSS-10 results pre and post across the 5 groups. $: interaction effect; #: simple main effects of time (follow-ups). PSS-10: Perceived Stress Scale.

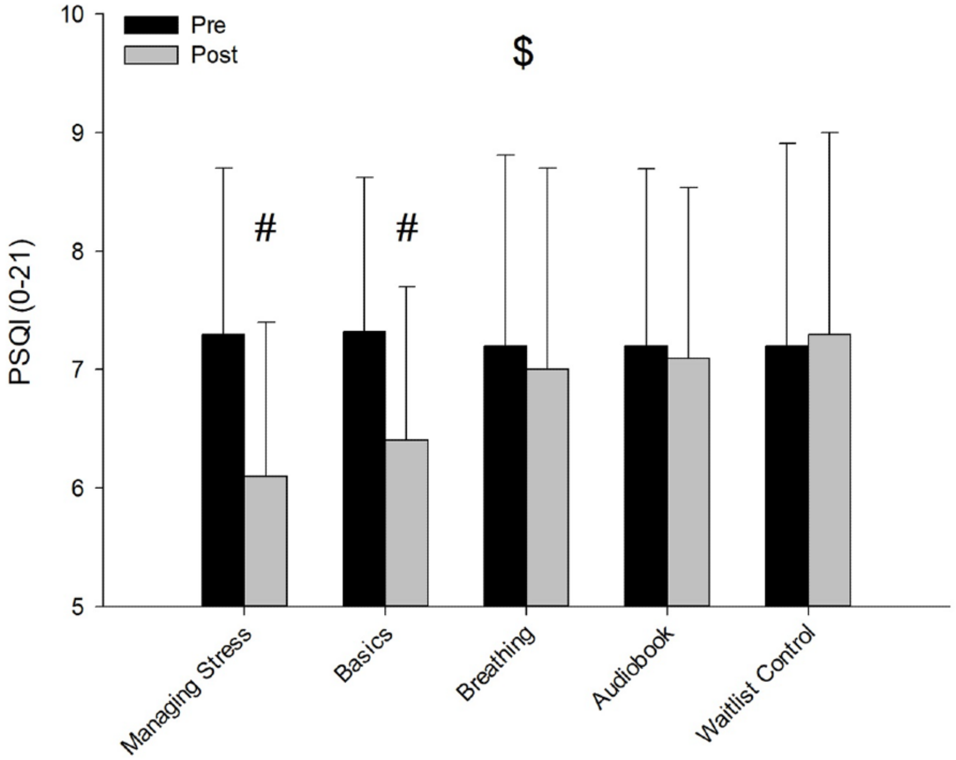


There was an interaction for PSQI across the 5 groups and time ($F_{4,158}$=4.941; *P*=.001; η²p=0.111). Follow-up tests revealed that the PSQI score for the Managing Stress (*P*<.001; η²p=0.131) and the Basics (*P*=.001; η²p=0.074) groups were significantly reduced from the preintervention to the postintervention period, while no significant differences were reported for Breathing (*P*=.50; η²p=0.003), Audiobook (*P*=.60; η²p=0.002), and Control (*P*=.73; η²p=0.001; Figure 4; Table 2).

There was no significant interaction ($F_{4,158}$=1.801; *P*=.13; η²p=0.044) or main effects of time ($F_{1,158}$=1.957; *P*=.16; η²p=0.012) and group ($F_{4,158}$=2.245; *P*=.07; η²p=0.054) for MAAS (Table 2).

No interaction ($F_{4,158}$=0.279; *P*=.90; η²p=0.007) or main effect of group ($F_{4,158}$=1.639; *P*=.17; η²p=0.040) were found for self-reported stress during the CPT (Table 2). A main effect of time was observed ($F_{1,158}$=10.960; *P*=.001; η²p=0.065).

Figure 4. PSQI results pre (A) and post (B) across the 5 groups. $: interaction effect; #: simple main effects of time (follow-ups). PSQI: Pittsburgh Sleep Quality Index.



Physiological Measures: HRV

No 3-way interaction (group × time × task period) was detected for HRV ($F_{8,316}$=0.030; *P*=.997; η²p=0.001). However, a significant 2-way interaction (group × time [$F_{4,158}$=2.637; *P*=.04; η²p=0.063]) was detected. Follow-up *t* tests revealed that HRV significantly increased from pretest to posttest for the Managing Stress (*P*<.001; η²p=0.089) and the Basics (*P*=.008; η²p=0.043) while no significant differences were reported for Breathing (*P*=.92; η²p=0.001), Audiobook (*P*=.53; η²p=0.002), and Control (*P*=.62; η²p=0.002). No group × task period ($F_{8,316}$=0.854; *P*=.56; η²p=0.021) and time × task period ($F_{2,316}$=1.670; *P*=.19; η²p=0.010) interactions were observed. A significant main effect of task period was observed ($F_{2,316}$=42.529; *P*<.001; η²p=0.212; Figure 5; Table 3).

Figure 5. HRV results pre and post across the 5 groups. $: interaction effect; #: simple main effects of time (follow-ups). HRV: heart rate variability; RMSSD: root-mean-square of successive differences between normal heartbeats.

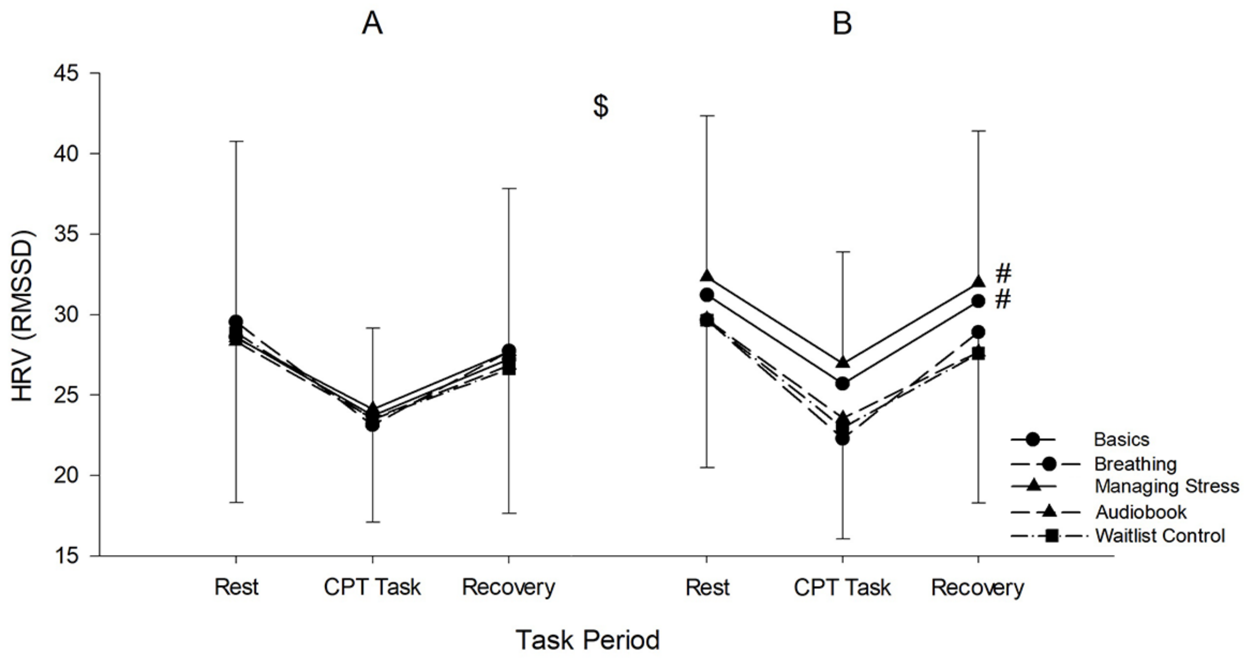


Table 3. Behavioral data for the 5 groups, shown as mean and SD collected pre- and postintervention.

	Managing Stress, mean (SD)	Basics, mean (SD)	Breathing, mean (SD)	Audiobook, mean (SD)	Waitlist Control, mean (SD)
HRV[a] (RMSSD[b]; Pre)					
Rest period	28.9 (8.0)	28.5 (9.5)	30.9 (11.3)	29.0 (8.4)	28.8 (8.0)
CPT[c] period	24.5 (6.0)	23.9 (6.6)	22.0 (5.1)	23.5 (5.9)	23.5 (6.4)
Recovery period	28.2 (9.1)	27.0 (9.6)	28.4 (10.0)	26.0 (8.3)	26.6 (9.3)
HRV (RMSSD; Post)					
Rest period	32.9 (10.0)	31.2 (8.8)	31.0 (9.0)	29.4 (7.9)	29.7 (8.9)
CPT period	27.5 (6.9)	25.7 (7.1)	21.1 (4.4)	23.5 (6.6)	23.0 (6.3)
Recovery period	32.8 (9.1)	30.3 (7.9)	29.6 (11.0)	27.6 (9.4)	27.6 (9.5)

[a]HRV: heart rate variability.

[b]RMSSD: root-mean-square of successive differences between normal heartbeats.

[c]CPT: cold pressor task.

Discussion

Principal Findings

This study sought to investigate if 3 formats of digital mindfulness interventions would show efficacy in reducing subjective and physiological levels of stress in a population with elevated symptoms of stress. The study found that self-reported stress as measured by PSS-10 was significantly reduced in the Managing Stress and Basics groups from the preintervention to the postintervention period compared to the other groups. Furthermore, the study found significant improvement in self-reported sleep quality as measured by the PSQI in the Managing Stress and Basics groups from the preintervention to the postintervention period compared to the other groups. Trait mindfulness, as measured by MAAS, did not yield significant differences across any of the groups. Finally, the results showed that only the Basics and Managing Stress groups displayed significantly reduced levels of physiological stress (ie, expressed as increased HRV activity) during exposure to the acute stress manipulation task (ie, CPT).

To our knowledge this study is the first study to demonstrate stress-reducing effects of a digital mindfulness app in a population with elevated baseline stress levels. Stress is prevalent in modern society and is accepted as a contributing factor to the onset of a range of mental health disorders, including depression and anxiety [6,7,88]. These findings highlight the promise of the Headspace mindfulness app in reducing both psychological and physiological stress in people with elevated stress.

Self-Reported Effects of Stress and Sleep Quality

The magnitude of change for the PSS-10 was larger in the Managing Stress group and the Basics groups compared to the Breathing and control groups, thus finding partial support for hypothesis 1 (H1). Our findings expand upon the results of previous studies that found digital mindfulness interventions improved stress among the general population [20-23,25,26]. This study showed efficacy among a sample composed of mostly university students with elevated stress. A meta-analysis showed that digital mindfulness interventions had significant effects on depression, anxiety, and stress, which was supported in this study in terms of stress reduction [32]. However, the meta-analysis also showed that digital mindfulness interventions with therapist guidance were more effective than for digital mindfulness without therapist guidance. Our study did not include therapist guidance, future studies could examine whether therapist-guided interventions lead to higher effect sizes. An implication from this study is that the Managing Stress and Basics interventions both improved stress in a population afflicted with elevated stress, with effect sizes being slightly higher for the Managing Stress relative to the Basics intervention for both physiological measures (ie, HRV) and psychological measures (ie, PSS-10 and PSQI). The results may point to the fact that the act of following a skill-building intervention centered around mindfulness is sufficient to show changes in these outcomes, which was present in both the Managing Stress and Basics groups. Both the Managing Stress and Basics interventions were effective despite the differences in frequency and duration of individual sessions (20 daily 15-minute sessions for Managing Stress and 30 daily 10-minute sessions for Basics). The results suggest that using mindfulness app-based interventions may be a practical approach to reducing stress in that it requires fewer resources (costs, mindfulness instructors, and brick-and-mortar clinics) where people can participate remotely with fewer practical constraints.

This study found that the PSQI indexing subjective sleep quality improved in the Managing Stress and Basics groups from the preintervention to the postintervention period. It is noteworthy that the participant's subjective sleep quality on the PSQI did not show statistical differences from the preintervention to the postintervention period for the Breathing group. However, the HRV results support this finding in that there was not a significant difference in HRV for the Breathing intervention. A recent meta-analysis reported a significant positive effect of mindfulness on sleep quality based on the results of 6 randomized controlled trials on people with insomnia [89]. In this study, we found a reduction in perceived stress and in a separate analysis an increase in sleep quality, however, only in 2 of the 3 aforementioned mindfulness interventions, which might be due to the skill-building program in Managing Stress and Basics that yielded an effect on both outcome measures.

In contrast, we found no effects of the MAAS for any of the interventions. This finding is surprising in that previous findings have revealed that web- or app-based mindfulness has a significant impact on mindfulness, albeit with small effect sizes [32]. However, we note that both the Managing Stress and Basics interventions showed a trend toward significance on the MAAS.

Physiological Effects of Stress

We did not find significant group differences in subjective levels of stress after administration of the CPT from the preintervention to the postintervention period. However, we did observe significant differences in HRV activity both in the expectation phase before the CPT, during the CPT, and in the recovery period after the CPT in both the Managing Stress group and the Basics group in support of hypothesis 2 (H2).

A recent study employed the CPT in the context of a 30-day app-based mindfulness intervention and showed that mindfulness training was less affected by acute stress on their cognitive performance compared to the control group suggesting that mindfulness may mitigate the negative impact of acute stress [24]. This previous finding may support the results from this study. Although this study cannot speak to the precise neurobiological mechanism underlying the elevated HRV activity observed in the Managing Stress and Basics groups, we can take advantage of our understanding of other systems involved in the stress response to interpret why these effects might occur. The stress response is driven by elevated noradrenaline levels during acute stress that reflects SNS activation [90,91]. Consequently, elevated noradrenaline levels during stress lead to the rapid decline in cognitive processing and impairment of the prefrontal cortex [92]. However, our finding of increased HRV activity in the Managing Stress and Basics group, reflects increased PNS activity [44] during acute stress which is in line with the abovementioned study showing that mindfulness in the context of the CPT does not impair cognitive functioning [24]. The results demonstrate that the Managing Stress and Basics mindfulness interventions show a stress-buffering effect compared to the other interventions, specifically by increasing PNS activity in a population afflicted with elevated levels of self-reported baseline stress. Furthermore, the results show that mindfulness can momentarily decrease stress and stress-related autonomic activity, which has also been observed in related studies, albeit in healthy populations and using salivary cortisol [93,94].

The results in this study found differential efficacy whereby only 2 of the 3 mindfulness groups, that is the Managing Stress and Basics group, but not the Breathing group, showed physiological and self-reported stress-buffering effects. This could be due to the relationship between the different ingredients of the mindfulness content across the different interventions. The content of the Breathing intervention was to engage in deep breathing and diaphragmatic breathing exercises for 10 min daily, which does not seem to translate into behavioral changes that could influence physiological reactivity as opposed to the Managing Stress and Basics interventions. This finding of differential effects between the 3 types of mindfulness is interesting. Reasons for such a difference might be found in the literature demonstrating mixed results in terms of the effect of deep breathing and diaphragmatic breathing interventions across various outcome measures. Further, 1 study found that breathing exercises did not yield pain reduction in a clinical population [95]. However, another study found effects in terms of reduced distress in the context of brief deep breathing and diaphragmatic breathing exercises, albeit this study only showed a limited immediate effect of a deep breathing exercise [96]. Thus, additional studies are required to tease apart the impact of deep breathing and diaphragmatic breathing exercises on stress outcomes and respiratory-physiological effects both in the immediate phase and more chronic effects. Further, 1 interpretation of the results from this study may be that essential components of the mindfulness interventions that were not present in the breathing exercises, such as *attention monitoring* and *acceptance* components, may be responsible for the nonsignificant finding in the breathing exercise condition. This may explain the mechanism of action behind why mindfulness is thought to influence stress in this study.

Strengths and Limitations

Although the results of this study are promising regarding the efficacy in 2 of the 3 app-based mindfulness interventions, several limitations must be noted. First, the sample primarily comprised young university students in their twenties, thus the generalizability of our findings may be limited. Second, this study did not investigate the efficacy of the app-based interventions beyond 30 days, nor whether any of the findings were maintained (regardless of app usage) beyond this period. Future studies should consider including follow-up measurements to evaluate sustained outcomes. Third, the study did not collect endocrine measures such as cortisol to assess the acute stress response during the CPT, although the HRV-data supported, it would have been interesting to also inspect correlations between endocrine measures and HRV to be completely certain that the CPT elicited reliable stress reactivity in participants. Fourth, participants were instructed to follow the program in full, however training adherence was not checked during the intervention, and participants were not reminded by the researchers to complete the daily training in the intervention period, which might have increased adherence.

Notably, a strength of this study was its relatively diverse population in terms of race, whereby 5 different ethnicities were represented in the study. In future interventions, it is important to explore the efficacy of diverse populations both in terms of race or ethnicity, age, socioeconomic status, and education level in order to generalize the findings to a broader population. A second strength of this study is that we employed multiple mindfulness interventions and were thus able to investigate if the unique characteristic of each intervention would result in differential efficacy. A third strength of this study was that both subjective and objective measures were employed which increased the inferences made in terms of reductions in both self-reported stress and objective measures of stress in the Managing Stress and Basics interventions.

Conclusions

In summary, our findings extend previous studies suggesting the efficacy of Headspace's app-based mindfulness interventions, specifically the Managing Stress and Basics content, to reduce stress in populations with elevated stress levels. Specifically, we found stress-buffering effects in a relatively diverse sample of participants afflicted with elevated stress. More research in this area is needed to establish efficacy and explore the degree to which effects are sustained in the long-term. The findings presented here provide important data that may be applied to the design of future studies or mental health interventions in people who experience elevated levels of stress.

Acknowledgments

UK and WS contributed equally to this study. This study was funded by Headspace.

Data Availability

The data sets generated or analyzed during the study are available from the corresponding author upon request.

Conflicts of Interest

The following authors are current or former employees of Headspace: EH, CN, SK, ES, AC, CP, and LL.

Multimedia Appendix 1 [DOCX File , 1237 KB - mhealth_v11i1e47371_app1.docx]

Multimedia Appendix 2
Managing Stress (mindfulness) guided program.
[PNG File , 276 KB - mhealth_v11i1e47371_app2.png]

Multimedia Appendix 3
Basics (mindfulness) course.
[PNG File , 180 KB - mhealth_v11i1e47371_app3.png]

Multimedia Appendix 4
CONSORT eHealth Checklist (V 1.6.2).
[PDF File (Adobe PDF File), 91 KB - mhealth_v11i1e47371_app4.pdf]

References

1. Ng DM, Jeffery RW. Relationships between perceived stress and health behaviors in a sample of working adults. Health Psychol 2003;22(6):638-642. [doi: 10.1037/0278-6133.22.6.638] [Medline: 14640862]
2. Steptoe A, Kivimäki M. Stress and cardiovascular disease. Nat Rev 2012;9:360-370 [FREE Full text] [doi: 10.1038/nrcardio.2012.45] [Medline: 22473079]
3. Mikolajczyk RT, El Ansari W, Maxwell AE. Food consumption frequency and perceived stress and depressive symptoms among students in three European countries. Nutr J 2009;8:31 [FREE Full text] [doi: 10.1186/1475-2891-8-31] [Medline: 19604384]
4. Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012;42:1320-1334 [FREE Full text] [doi: 10.1111/j.1559-1816.2012.00900.x]
5. Tian F, Shen Q, Hu Y, Ye W, Valdimarsdóttir UA, Song H, et al. Association of stress-related disorders with subsequent risk of all-cause and cause-specific mortality: a population-based and sibling-controlled cohort study. Lancet Reg Health Eur 2022;18:100402 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100402] [Medline: 35663363]
6. Iwata M, Ota KT, Duman RS. The inflammasome: pathways linking psychological stress, depression, and systemic illnesses. Brain Behav Immun 2013;31:105-114 [FREE Full text] [doi: 10.1016/j.bbi.2012.12.008] [Medline: 23261775]
7. Masi G, Brovedani P. The hippocampus, neurotrophic factors and depression: possible implications for the pharmacotherapy of depression. CNS Drugs 2011;25(11):913-931. [doi: 10.2165/11595900-000000000-00000] [Medline: 22054117]
8. Nochaiwong S, Ruengorn C, Thavorn K, Hutton B, Awiphan R, Phosuya C, et al. Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: a systematic review and meta-analysis. Sci Rep 2021;11(1):10173 [FREE Full text] [doi: 10.1038/s41598-021-89700-8] [Medline: 33986414]
9. Coombs NC, Meriwether WE, Caringi J, Newcomer SR. Barriers to healthcare access among U.S. adults with mental health challenges: a population-based study. SSM Popul Health 2021;15:100847 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100847] [Medline: 34179332]
10. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
11. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York: Delacorte; 1990.
12. Strauss C, Cavanagh K, Oliver A, Pettman D. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: a meta-analysis of randomised controlled trials. PLoS One 2014;9(4):e96110 [FREE Full text] [doi: 10.1371/journal.pone.0096110] [Medline: 24763812]
13. Rosenzweig S, Reibel DK, Greeson JM, Brainard GC, Hojat M. Mindfulness-based stress reduction lowers psychological distress in medical students. Teach Learn Med 2003;15(2):88-92 [FREE Full text] [doi: 10.1207/S15328015TLM1502_03] [Medline: 12708065]
14. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009;15(5):593-600 [FREE Full text] [doi: 10.1089/acm.2008.0495]
15. Jain S, Shapiro SL, Swanick S, Roesch SC, Mills PJ, Bell I, et al. A randomized controlled trial of mindfulness meditation versus relaxation training: effects on distress, positive states of mind, rumination, and distraction. Ann Behav Med 2007;33(1):11-21 [FREE Full text] [doi: 10.1207/s15324796abm3301_2] [Medline: 17291166]
16. Baer RA. Self-focused attention and mechanisms of change in mindfulness-based treatment. Cogn Behav Ther 2009;38(Suppl 1):15-20. [doi: 10.1080/16506070902980703] [Medline: 19697176]
17. Sorkin DH, Janio EA, Eikey EV, Schneider M, Davis K, Schueller SM, et al. Rise in use of digital mental health tools and technologies in the United States during the COVID-19 pandemic: survey study. J Med Internet Res 2021;23(4):e26994 [FREE Full text] [doi: 10.2196/26994] [Medline: 33822737]
18. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022;9(5):e35159 [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]
19. Bennike IH, Wieghorst A, Kirk U. Online-based mindfulness training reduces behavioral markers of mind wandering. J Cogn Enhanc 2017;1:172-181 [FREE Full text] [doi: 10.1007/s41465-017-0020-9]
20. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
21. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
22. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2019;10:863-876 [FREE Full text] [doi: 10.1007/s12671-018-1050-9]

23. Mascaro JS, Wehrmeyer K, Mahathre V, Darcher A. A longitudinal, randomized and controlled study of app-delivered mindfulness in the workplace. J Wellness 2020;2(1):4 [FREE Full text] [doi: 10.18297/jwellness/vol2/iss1/4]
24. Piil F, Axelsen JL, Staiano W, Kirk U. Mindfulness passes the stress test: attenuation of behavioral markers of mind wandering during acute stress. J Cogn Enhanc 2021;5:155-163 [FREE Full text] [doi: 10.1007/s41465-020-00193-1]
25. Wylde CM, Mahrer NE, Meyer RML, Gold JI. Mindfulness for novice pediatric nurses: smartphone application versus traditional intervention. J Pediatr Nurs 2017;36:205-212. [doi: 10.1016/j.pedn.2017.06.008] [Medline: 28888505]
26. Yang E, Schamber E, Meyer RML, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018;24(5):505-513 [FREE Full text] [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
27. Kirk U, Ngnoumen C, Clausel A, Purvis CK. Using actigraphy and Heart Rate Variability (HRV) to assess sleep quality and sleep arousal of three app-based interventions: sleep music, sleepcasts, and guided mindfulness. J Cogn Enhanc 2022;6:216-231 [FREE Full text] [doi: 10.1007/s41465-021-00233-4]
28. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
29. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol 2020;75(8):1093-1104 [FREE Full text] [doi: 10.1037/amp0000654] [Medline: 33252947]
30. Avalos LA, Aghaee S, Kurtovich E, Quesenberry C, Nkemere L, McGinnis MK, et al. A mobile health mindfulness intervention for women with moderate to moderately severe postpartum depressive symptoms: feasibility study. JMIR Ment Health 2020;7(11):e17405 [FREE Full text] [doi: 10.2196/17405] [Medline: 33180028]
31. Ainsworth B, Stanescu S, Stuart B, Russell D, Liddiard M, Djukanovic R, et al. A feasibility trial of a digital mindfulness-based intervention to improve asthma-related quality of life for primary care patients with asthma. J Behav Med 2022;45(1):133-147 [FREE Full text] [doi: 10.1007/s10865-021-00249-3] [Medline: 34448986]
32. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
33. Hutchison M, Russell BS, Starkweather AR, Gans KM. Outcomes from an online pilot mindfulness based intervention with adolescents: a comparison by categories of risk. J Child Fam Stud 2023;32(2):438-450 [FREE Full text] [doi: 10.1007/s10826-022-02448-1] [Medline: 36193193]
34. Axelsen JL, Meline JSJ, Staiano W, Kirk U. Mindfulness and music interventions in the workplace: assessment of sustained attention and working memory using a crowdsourcing approach. BMC Psychol 2022;10(1):108 [FREE Full text] [doi: 10.1186/s40359-022-00810-y] [Medline: 35478086]
35. Lindsay EK, Creswell JD. Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). Clin Psychol Rev 2017;51:48-59 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.011] [Medline: 27835764]
36. Kirk U, Axelsen JL. Heart rate variability is enhanced during mindfulness practice: a randomized controlled trial involving a 10-day online-based mindfulness intervention. PLoS One 2020;15(12):e0243488 [FREE Full text] [doi: 10.1371/journal.pone.0243488] [Medline: 33332403]
37. Crosswell AD, Moreno PI, Raposa EB, Motivala SJ, Stanton AL, Ganz PA, et al. Effects of mindfulness training on emotional and physiologic recovery from induced negative affect. Psychoneuroendocrinology 2017;86:78-86 [FREE Full text] [doi: 10.1016/j.psyneuen.2017.08.003] [Medline: 28923751]
38. Ditto B, Eclache M, Goldman N. Short-term autonomic and cardiovascular effects of mindfulness body scan meditation. Ann Behav Med 2006;32(3):227-234 [FREE Full text] [doi: 10.1207/s15324796abm3203_9] [Medline: 17107296]
39. Tang YY, Ma Y, Fan Y, Feng H, Wang J, Feng S, et al. Central and autonomic nervous system interaction is altered by short-term meditation. Proc Natl Acad Sci U S A 2009;106(22):8865-8870 [FREE Full text] [doi: 10.1073/pnas.0904031106] [Medline: 19451642]
40. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
41. Charmandari E, Tsigos C, Chrousos G. Endocrinology of the stress response. Annu Rev Physiol 2005;67:259-284. [doi: 10.1146/annurev.physiol.67.040403.120816] [Medline: 15709959]
42. Thayer JF, Ahs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]
43. Berntson GG, Bigger JT, Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 1997;34(6):623-648 [FREE Full text] [doi: 10.1111/j.1469-8986.1997.tb02140.x] [Medline: 9401419]
44. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014;5:1040 [FREE Full text] [doi: 10.3389/fpsyg.2014.01040] [Medline: 25324790]
45. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]

46. Garland SN, Carlson LE, Stephens AJ, Antle MC, Samuels C, Campbell TS. Mindfulness-based stress reduction compared with cognitive behavioral therapy for the treatment of insomnia comorbid with cancer: a randomized, partially blinded, noninferiority trial. J Clin Oncol 2014;32(5):449-457 [FREE Full text] [doi: 10.1200/JCO.2012.47.7265] [Medline: 24395850]
47. Adler-Neal AL, Waugh CE, Garland EL, Shaltout HA, Diz DI, Zeidan F. The role of heart rate variability in mindfulness-based pain relief. J Pain 2020;21(3-4):306-323 [FREE Full text] [doi: 10.1016/j.jpain.2019.07.003] [Medline: 31377215]
48. Lovallo W. The cold pressor test and autonomic function: a review and integration. Psychophysiology 1975;12(3):268-282. [doi: 10.1111/j.1469-8986.1975.tb01289.x] [Medline: 1153632]
49. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. Brain Cogn 2007;65(3):209-237 [FREE Full text] [doi: 10.1016/j.bandc.2007.02.007] [Medline: 17466428]
50. Forgas JP. Affective influences on self-disclosure: mood effects on the intimacy and reciprocity of disclosing personal information. J Pers Soc Psychol 2011;100(3):449-461. [doi: 10.1037/a0021129] [Medline: 21171788]
51. Raio CM, Hartley CA, Orederu TA, Li J, Phelps EA. Stress attenuates the flexible updating of aversive value. Proc Natl Acad Sci U S A 2017;114(42):11241-11246 [FREE Full text] [doi: 10.1073/pnas.1702565114] [Medline: 28973957]
52. Brown CC, Raio CM, Neta M. Cortisol responses enhance negative valence perception for ambiguous facial expressions. Sci Rep 2017;7(1):15107 [FREE Full text] [doi: 10.1038/s41598-017-14846-3] [Medline: 29118319]
53. Velasco M, Gómez J, Blanco M, Rodriguez I. The cold pressor test: pharmacological and therapeutic aspects. Am J Ther 1997;4(1):34-38. [doi: 10.1097/00045391-199701000-00008] [Medline: 10423589]
54. Victor RG, Leimbach WN, Seals DR, Wallin BG, Mark AL. Effects of the cold pressor test on muscle sympathetic nerve activity in humans. Hypertension 1987;9(5):429-436. [doi: 10.1161/01.hyp.9.5.429] [Medline: 3570420]
55. Raio CM, Orederu TA, Palazzolo L, Shurick AA, Phelps EA. Cognitive emotion regulation fails the stress test. Proc Natl Acad Sci U S A 2013;110(37):15139-15144 [FREE Full text] [doi: 10.1073/pnas.1305706110] [Medline: 23980142]
56. McRae AL, Saladin ME, Brady KT, Upadhyaya H, Back SE, Timmerman MA. Stress reactivity: biological and subjective responses to the cold pressor and Trier social stressors. Hum Psychopharmacol 2006;21(6):377-385. [doi: 10.1002/hup.778] [Medline: 16915579]
57. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND. Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci USA 2013;110(52):20941-20946 [FREE Full text] [doi: 10.1073/pnas.1312011110] [Medline: 24324166]
58. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]
59. Heart Rate Variability. Standards of measurement, physiological interpretation, and clinical use. Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Eur Heart J 1996;17(3):354-381. [Medline: 8737210]
60. Sjoberg N, Saint DA. A single 4 mg dose of nicotine decreases heart rate variability in healthy nonsmokers: implications for smoking cessation programs. Nicotine Tob Res 2011;13(5):369-372. [doi: 10.1093/ntr/ntr004] [Medline: 21350044]
61. Ralevski E, Petrakis I, Altemus M. Heart rate variability in alcohol use: a review. Pharmacol Biochem Behav 2019;176:83-92 [FREE Full text] [doi: 10.1016/j.pbb.2018.12.003] [Medline: 30529588]
62. Weitzman ED, Fukushima D, Nogeire C, Roffwarg H, Gallagher TF, Hellman L. Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. J Clin Endocrinol Metab 1971;33(1):14-22. [doi: 10.1210/jcem-33-1-14] [Medline: 4326799]
63. Krieger DT, Allen W, Rizzo F, Krieger HP. Characterization of the normal temporal pattern of plasma corticosteroid levels. J Clin Endocrinol Metab 1971;32(2):266-284. [doi: 10.1210/jcem-32-2-266] [Medline: 4321505]
64. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
65. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
66. Ottaviani C, Shahabi L, Tarvainen M, Cook I, Abrams M, Shapiro D. Cognitive, behavioral, and autonomic correlates of mind wandering and perseverative cognition in major depression. Front Neurosci 2015;8:433 [FREE Full text] [doi: 10.3389/fnins.2014.00433] [Medline: 25601824]
67. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:8099-8102. [doi: 10.1109/EMBC.2015.7320273] [Medline: 26738173]
68. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV—heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]
69. Kirk U, Ngnoumen C, Clausel A, Kennedy C. Effects of three genres of focus music on heart rate variability and sustained attention. J Cogn Enhanc 2022;6:143-158 [FREE Full text] [doi: 10.1007/s41465-021-00226-3]
70. Make it your year with Headspace. URL: https://www.headspace.com/ [accessed 2023-08-29]

71. Cohen S, Kessler RC, Gordon LU. Strategies for Measuring Stress in Psychiatric and Physical Disorders. In: Cohen S, Kessler RC, Gordon LU, editors. Measuring Stress. New York: Oxford University Press; 1995:3-28.
72. Lazarus RS. From psychological stress to the emotions: a history of changing outlooks. Annu Rev Psychol 1993;44:1-21 [FREE Full text] [doi: 10.1146/annurev.ps.44.020193.000245] [Medline: 8434890]
73. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. United States: Springer publishing company; 1984.
74. Baum A. Behavioral, Biological, and Environmental Interactions in Disease Processes. In: Blumenthal S, Matthews K, Weiss S, editors. New Research Frontiers in Behavioral Medicine. Proceedings of the National Conference. Washington, DC: NIH Publications; 1993:61-70.
75. Kirby ED, Williams VP, Hocking MC, Lane JD, Williams RB. Psychosocial benefits of three formats of a standardized behavioral stress management program. Psychosom Med 2006;68(6):816-823. [doi: 10.1097/01.psy.0000238452.81926.d3] [Medline: 17132834]
76. Shiralkar MT, Harris TB, Eddins-Folensbee FF, Coverdale JH. A systematic review of stress-management programs for medical students. Acad Psychiatry 2013;37(3):158-164. [doi: 10.1176/appi.ap.12010003] [Medline: 23446664]
77. Ma X, Yue ZQ, Gong ZQ, Zhang H, Duan NY, Shi YT, et al. The effect of diaphragmatic breathing on attention, negative affect and stress in healthy adults. Front Psychol 2017;8:874 [FREE Full text] [doi: 10.3389/fpsyg.2017.00874] [Medline: 28626434]
78. Benson H, Klipper MZ. The Relaxation Response. New York: Morrow; 1975:240.
79. Kirk U, Gu X, Sharp C, Hula A, Fonagy P, Montague PR. Mindfulness training increases cooperative decision making in economic exchanges: evidence from fMRI. Neuroimage 2016;138:274-283 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.05.075] [Medline: 27266443]
80. Kirk U, Gu X, Harvey AH, Fonagy P, Montague PR. Mindfulness training modulates value signals in ventromedial prefrontal cortex through input from insular cortex. Neuroimage 2014;100:254-262 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.06.035] [Medline: 24956066]
81. Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci 2008;12(4):163-169 [FREE Full text] [doi: 10.1016/j.tics.2008.01.005] [Medline: 18329323]
82. Allen M, Dietz M, Blair KS, van Beek M, Rees G, Vestergaard-Poulsen P, et al. Cognitive-affective neural plasticity following active-controlled mindfulness intervention. J Neurosci 2012;32(44):15601-15610 [FREE Full text] [doi: 10.1523/JNEUROSCI.2957-12.2012] [Medline: 23115195]
83. Zeidan F, Johnson SK, Diamond BJ, David Z, Goolkasian P. Mindfulness meditation improves cognition: evidence of brief mental training. Conscious Cogn 2010;19(2):597-605 [FREE Full text] [doi: 10.1016/j.concog.2010.03.014] [Medline: 20363650]
84. Zeidan F, Emerson NM, Farris SR, Ray JN, Jung Y, McHaffie JG, et al. Mindfulness meditation-based pain relief employs different neural mechanisms than placebo and sham mindfulness meditation-induced analgesia. J Neurosci 2015;35(46):15307-15325 [FREE Full text] [doi: 10.1523/JNEUROSCI.2542-15.2015] [Medline: 26586819]
85. Zeidan F, Adler-Neal AL, Wells RE, Stagnaro E, May LM, Eisenach JC, et al. Mindfulness-meditation-based pain relief is not mediated by endogenous opioids. J Neurosci 2016;36(11):3391-3397 [FREE Full text] [doi: 10.1523/JNEUROSCI.4328-15.2016] [Medline: 26985045]
86. Quaglia JT, Zeidan F, Grossenbacher PG, Freeman SP, Braun SE, Martelli A, et al. Brief mindfulness training enhances cognitive control in socioemotional contexts: behavioral and neural evidence. PLoS One 2019;14(7):e0219862 [FREE Full text] [doi: 10.1371/journal.pone.0219862] [Medline: 31323050]
87. White G. In: Taylor J, editor. The Natural History of Selborne. Willesden Green City: London State: Greater London: The Copyright Group; 2011.
88. Shi J, Huang A, Jia Y, Yang X. Perceived stress and social support influence anxiety symptoms of Chinese family caregivers of community-dwelling older adults: a cross-sectional study. Psychogeriatrics 2020;20(4):377-384. [doi: 10.1111/psyg.12510] [Medline: 31971319]
89. Gong H, Ni CX, Liu YZ, Zhang Y, Su WJ, Lian YJ, et al. Mindfulness meditation for insomnia: a meta-analysis of randomized controlled trials. J Psychosom Res 2016;89:1-6 [FREE Full text] [doi: 10.1016/j.jpsychores.2016.07.016] [Medline: 27663102]
90. Thoma MV, Kirschbaum C, Wolf JM, Rohleder N. Acute stress responses in salivary alpha-amylase predict increases of plasma norepinephrine. Biol Psychol 2012;91(3):342-348. [doi: 10.1016/j.biopsycho.2012.07.008] [Medline: 22954623]
91. Ehlert U, Erni K, Hebisch G, Nater U. Salivary alpha-amylase levels after yohimbine challenge in healthy men. J Clin Endocrinol Metab 2006;91(12):5130-5133 [FREE Full text] [doi: 10.1210/jc.2006-0461] [Medline: 16968802]
92. Arnsten AFT. Stress signalling pathways that impair prefrontal cortex structure and function. Nat Rev Neurosci 2009;10(6):410-422 [FREE Full text] [doi: 10.1038/nrn2648] [Medline: 19455173]
93. Pascoe MC, Thompson DR, Jenkins ZM, Ski CF. Mindfulness mediates the physiological markers of stress: systematic review and meta-analysis. J Psychiatr Res 2017;95:156-178. [doi: 10.1016/j.jpsychires.2017.08.004] [Medline: 28863392]
94. Sanada K, Montero-Marin J, Díez MA, Salas-Valero M, Pérez-Yus MC, Morillo H, et al. Effects of mindfulness-based interventions on salivary cortisol in healthy adults: a meta-analytical review. Front Physiol 2016;7:471 [FREE Full text] [doi: 10.3389/fphys.2016.00471] [Medline: 27807420]

95. Guan NC, Beng TS, Sue-Yin L, Kanagasundram S. The effect of 5-min mindful breathing on pain in palliative care cancer patients: a randomized controlled study. Indian J Palliat Care 2021;27(1):83-88 [FREE Full text] [doi: 10.4103/IJPC.IJPC_122_20] [Medline: 34035622]
96. Ng CG, Lai KT, Tan SB, Sulaiman AH, Zainal NZ. The effect of 5 minutes of mindful breathing to the perception of distress and physiological responses in palliative care cancer patients: a randomized controlled study. J Palliat Med 2016;19(9):917-924. [doi: 10.1089/jpm.2016.0046] [Medline: 27110900]

Abbreviations

ANS: autonomic nervous system
CPT: cold pressor task
HRV: heart rate variability
MAAS: Mindful Attention Awareness Scale
MBSR: mindfulness-based stress reduction
PNS: parasympathetic nervous system
PSQI: Pittsburgh Sleep Quality Index
PSS-10: Perceived Stress Scale
SNS: sympathetic nervous system

Edited by L Buis; submitted 17.03.23; peer-reviewed by A Werntz, P Watson; comments to author 24.05.23; revised version received 19.07.23; accepted 08.08.23; published 13.10.23.

Please cite as:
Kirk U, Staiano W, Hu E, Ngnoumen C, Kunkle S, Shih E, Clausel A, Purvis C, Lee L
App-Based Mindfulness for Attenuation of Subjective and Physiological Stress Reactivity in a Population With Elevated Stress: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e47371
URL: https://mhealth.jmir.org/2023/1/e47371
doi:10.2196/47371
PMID:37831493

Original Paper

Use of e-Cigarettes in Cigarette Smoking Cessation: Secondary Analysis of a Randomized Controlled Trial

Margarita Santiago-Torres[1*], PhD; Kristin E Mull[1*], MS; Brianna M Sullivan[1*], MS; Jonathan B Bricker[1,2*], PhD

[1]Fred Hutchinson Cancer Center, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Margarita Santiago-Torres, PhD
Fred Hutchinson Cancer Center
1100 Fairview Avenue North
Seattle, WA, 98109
United States
Phone: 1 2066674780
Email: msantiag@fredhutch.org

Abstract

Background: Many adults use e-cigarettes to help them quit cigarette smoking. However, the impact of self-selected use of e-cigarettes on cigarette smoking cessation, particularly when concurrently receiving app-based behavioral interventions, remains unexplored.

Objective: This study used data from a randomized trial of 2 smartphone apps to compare 12-month cigarette smoking cessation rates between participants who used e-cigarettes on their own (ie, adopters: n=465) versus those who did not (ie, nonadopters: n=1097).

Methods: The study population included all participants who did not use e-cigarettes at baseline. “Adopters” were those who self-reported the use of e-cigarettes at either 3- or 6-month follow-ups. “Nonadopters” were those who self-reported no use of e-cigarettes at either follow-up time point. The primary cessation outcome was self-reported, complete-case, 30-day point prevalence abstinence from cigarette smoking at 12 months. Secondary outcomes were missing-as-smoking and multiple imputation analyses of the primary outcome, prolonged abstinence, and cessation of all nicotine and tobacco products at 12 months. In logistic regression models, we first examined the potential interaction between e-cigarette use and treatment arm (iCanQuit vs QuitGuide) on the primary cessation outcome. Subsequently, we compared 12-month cigarette smoking cessation rates between adopters and nonadopters separately for each app.

Results: There was suggestive evidence for an interaction between e-cigarette use and treatment arm on cessation (P=.05). In the iCanQuit arm, 12-month cigarette smoking cessation rates were significantly lower among e-cigarette adopters compared with nonadopters (41/193, 21.2% vs 184/527, 34.9%; P=.003; odds ratio 0.55, 95% CI 0.37-0.81). In contrast, in the QuitGuide arm, 12-month cigarette smoking cessation rates did not differ between adopters and nonadopters (46/246, 18.7% vs 104/522, 19.9%; P=.64; odds ratio 0.91, 95% CI 0.62-1.35).

Conclusions: The use of e-cigarettes while concurrently receiving an app-based smoking cessation intervention was associated with either a lower or an unimproved likelihood of quitting cigarette smoking compared to no use. Future behavioral treatments for cigarette smoking cessation should consider including information on the potential consequences of e-cigarette use.

Trial Registration: ClinicalTrials.gov NCT02724462; https://clinicaltrials.gov/study/NCT02724462

(JMIR Mhealth Uhealth 2023;11:e48896) doi:10.2196/48896

KEYWORDS

acceptance and commitment therapy; cigarette smoking; digital behavioral interventions; e-cigarettes; smoking cessation; smartphone apps; vaping; mobile phone

Introduction

Background

Electronic nicotine delivery systems or vaping (referred to hereafter as "e-cigarettes") are battery-powered devices that simulate the experience of smoking by heating a liquid solution into an aerosol, which is then inhaled by the user [1]. This makes them a popular alternative for individuals seeking a less harmful nicotine delivery method. One of the key distinctions is that e-cigarettes operate without combustion, resulting in significantly lower levels of carcinogens and harmful compounds compared to the smoke produced by traditional cigarettes [2-4]. While e-cigarettes offer the potential to substantially reduce exposure to the harmful substances found in cigarette smoke [5,6], it is also important to acknowledge that e-cigarettes are not risk free. The long-term effects of vaping are currently under investigation, and concerns have been raised particularly regarding nonsmokers and young individuals who may take up vaping [7-9].

According to a survey conducted in the United States (US) by the Centers for Disease Control and Prevention in 2020, among adult e-cigarette users, 34.7% reported using e-cigarettes to quit smoking cigarettes, and 29.8% reported using e-cigarettes to reduce their cigarette smoking [1]. While there is significant interest among health providers, regulators, and treatment-seeking individuals alike regarding the impact of e-cigarettes on cigarette smoking cessation [10], the current evidence on the benefits and harms of using e-cigarettes for cigarette smoking cessation is mixed. For example, observational studies using data from the US Population Assessment of Tobacco and Health cohort suggest that e-cigarette use facilitates cigarette smoking cessation in real-world settings [11-13]. Moreover, a recent Cochrane review of e-cigarette randomized controlled trials (RCTs) for smoking cessation found evidence, albeit of very low certainty, of higher quit rates in participants using e-cigarettes compared with behavioral support alone or no support [14]. However, it is important to note that the authors of the review highlighted the very low certainty of these results due to imprecision and potential bias. In contrast, results from 2 prospective studies of RCTs among treatment-seeking individuals who used e-cigarettes while concurrently receiving behavioral cessation treatments suggest that e-cigarette use impedes cigarette smoking cessation [15,16]. Thus, it remains unclear whether e-cigarette use aids or hinders smoking cessation among treatment-seeking individuals while concurrently receiving behavioral support.

In-person traditional behavioral interventions such as counseling, support groups, and pharmacotherapy (eg, nicotine replacement therapy or medications to aid cessation) have been used to aid in cigarette smoking cessation [17]. However, the widespread use of smartphones and other digital devices has opened new avenues for delivering behavioral interventions through digital therapeutics [18], potentially enhancing the efficacy and accessibility of cessation treatments. The impact of e-cigarettes on the efficacy of digital interventions for smoking cessation is completely unexplored. With hundreds of smartphone apps designed for smoking cessation currently available, and given that 85% of US adults own smartphones, these apps are highly accessible [19,20]. In the US alone, smoking cessation apps are downloaded around 2 million times per year, highlighting their popularity and potential impact (R Nelson; SensorTower.com; personal communication; April 15, 2020). App-based interventions are particularly relevant because they can provide access to evidence-based cessation treatments that may otherwise be limited. Due to their low cost and potential for high population reach, app-delivered interventions could be an effective solution for many people seeking to quit smoking.

Among smoking cessation apps, iCanQuit is the only app that has demonstrated efficacy for cigarette smoking cessation in a full-scale RCT with long-term follow-up [21]. The iCanQuit app is based on acceptance and commitment therapy (ACT) principles [22] for behavior change, which teach users skills to accept their cravings to smoke and allow them to pass without smoking [23-25]. Our group developed and tested the iCanQuit app against the QuitGuide app. QuitGuide was selected because it follows standard behavioral approaches recommended by the US Clinical Practice Guidelines (USCPG) [21]; it is a smartphone app and thus avoids confounding treatment content with delivery modality; and it is nonproprietary and freely available to the public, providing maximal transparency and replicability.

The results from the iCanQuit parent trial showed (1) potential for high reach nationwide, (2) higher efficacy for cigarette smoking cessation of iCanQuit relative to QuitGuide (12-month quit smoking rates: 28% in iCanQuit vs 21% in QuitGuide; $P<.001$), and (3) higher engagement with iCanQuit compared to QuitGuide. Although the iCanQuit app has shown promising results in helping individuals quit smoking, it is unknown to what extent concurrent use of e-cigarettes might impact their cigarette smoking cessation rates. Understanding the impact of e-cigarettes on the efficacy of app-based cessation interventions is crucial as it can inform the development of content for users who intend to combine e-cigarettes with app-based interventions.

Objectives

To help address this unexplored question, this study aimed to compare cigarette smoking cessation rates between individuals who adopted e-cigarettes on their own and those who did not, within the context of app-based smoking cessation interventions. We hypothesized that the use of e-cigarettes would be associated with a greater likelihood of quitting cigarette smoking compared with no use. This hypothesis is based on the role of nicotine-containing e-cigarettes in alleviating withdrawal symptoms associated with dependence, thereby enhancing abstinence. We further hypothesized that e-cigarette adopters within the iCanQuit arm would have higher smoking cessation rates than those in the QuitGuide arm. This hypothesis was grounded in the evidence demonstrating a greater efficacy of iCanQuit for smoking cessation compared to QuitGuide [21].

To test these hypotheses, the potential interaction between the use of e-cigarette and treatment arm was first investigated (iCanQuit vs QuitGuide) on the primary cessation outcome. Subsequently, 12-month rates of cigarette smoking cessation were compared between adopters and nonadopters separately for each app. Finally, to gain a deeper understanding of the potential mechanisms underlying the impact of e-cigarettes on combustible cigarette smoking cessation, we conducted post hoc exploratory analyses to compare the following aspects between adopters and nonadopters: (1) self-selected use of cessation pharmacotherapy, (2) rates of cigarette smoking cessation for participants indicating they adopted e-cigarettes specifically as a cessation aid versus those who did not [26,27], (3) app engagement, and (4) number of quit attempts [28,29]. Moreover, we examined the associations between (5) the number of quit attempts and (6) levels of nicotine dependence with rates of cigarette smoking cessation at 12 months.

Methods

Overview

Data from this secondary analysis were from participants (aged 18 years or older) in the iCanQuit parent RCT who reported no e-cigarette use at baseline. Details of the iCanQuit RCT have been previously published [21]. Briefly, the iCanQuit RCT enrolled a racially or ethnically diverse sample of 2415 adults who smoke combustible cigarettes to participate in a 12-month app-delivered cigarette smoking cessation intervention. The primary aim of the trial was to test the efficacy of the ACT-based iCanQuit app against the USCPG-based QuitGuide app for cigarette smoking cessation at the 12-month follow-up.

Ethical Considerations

Study procedures were approved by the Fred Hutchinson Cancer Center Institutional Review Board (reference number: 8317). All participants were provided with information about the study and signed informed consent forms before participation. This study is registered in the ClinicalTrials.gov (NCT02724462).

Eligibility Criteria

Eligibility criteria included daily cigarette smoking, smartphone access, wanting to quit combustible cigarette smoking within the next 30 days, and wanting to quit within the next 30 days any other concurrently used nicotine or tobacco products. Exclusion criteria included being unable to read English, receiving smoking cessation treatment, having used either iCanQuit or QuitGuide in the past, or having a household member already enrolled in the study. Any use of e-cigarettes or other noncigarette nicotine or tobacco products was not an exclusion criterion, and data on their use were collected at each data collection time point. The approved study protocol was published alongside the main outcomes paper and is readily accessible for review [21].

Recruitment and Enrollment

Facebook ads were the primary source of recruitment for all trial participants (1281/1562, 82%), followed by a survey sampling company (203/1562, 13%), friends or family (47/1562, 3%), and search engine results (16/1562, 1%). The period of recruitment of all trial participants was May 2017 through September 2018. Participants completed a web-based screening survey and were notified of their eligibility via email. They then clicked on a secured emailed link to the study website, where they provided consent and completed the baseline survey. Consent to participate included collecting app use data via Google Analytics (number of logins, time spent using the app, and specific features used). Activation of the assigned app on the participant's smartphone was conducted by entering the login code sent to them in their trial enrollment email.

At each enrollment step, the study was presented as a comparison of 2 smartphone apps for cigarette smoking cessation. After completing the baseline survey, participants were then randomized 1:1 to receive iCanQuit or QuitGuide, both of which were accessible for 12 months. Randomization by permuted blocks of size 2, 4, and 6 was stratified by positive screening for depression (Center for Epidemiological Studies Depression Scale-20 score ≤15 vs ≥16), daily smoking frequency (≤20 vs ≥21 cigarettes per day), minoritized race or ethnicity backgrounds, and education level (less than high school vs some college or higher education). Neither research staff nor study participants had access to upcoming randomized study group assignments. For blinding, each app was branded as "iCanQuit" and did not mention either ACT or QuitGuide. Contamination between apps was avoided with a unique username and password provided only to the individual user and by having an eligibility criterion of not having other household members participating in the study.

Data collection occurred between August 2017 through December 2019 via web-based study surveys at the 3-, 6-, and 12-month follow-ups. For trial integrity, the follow-up data were collected by the survey research unit that was blinded to random assignments, and cessation outcomes data were collected outside of the intervention apps. Participants were compensated for completed data collection. For each follow-up time point, participants received US $25 for completing the follow-up survey. Participants would receive an additional US $10 bonus if the encrypted web-based survey was completed within 24 hours of the initial email invitation to complete the survey, with up to US $105 in total compensation per participant.

Study Population

The data for this secondary analysis included a subsample of trial participants from the iCanQuit parent RCT who had available data on e-cigarette use after randomization. This subset comprised 1562 participants, which accounted for 64.7% (1562/2415) of the total sample. Within this subsample, we categorized participants into two distinct groups based on their self-reported use of e-cigarettes after randomization: (1) "adopters," defined as individuals who reported no e-cigarette use in the past 30 days at baseline and reported using e-cigarettes within the past 30 days at either the 3- or 6-month follow-ups (465 adopters); and (2) "nonadopters," defined as individuals who reported no e-cigarette use in the past 30 days at baseline, 3-month, and 6-month follow-ups (1097 nonadopters). Participants who reported using e-cigarettes at baseline (n=575) were excluded from the analysis, as were those with missing e-cigarette use data (n=278). It is important to emphasize that the use of e-cigarettes during the 12-month intervention period was entirely self-selected by the participants.

Smartphone App–Based Interventions

iCanQuit

Details of the iCanQuit app have been previously published [21]. Briefly, participants who had access to the ACT-based iCanQuit app for 12 months received 8 levels of intervention content based on 2 key processes of ACT: acceptance of cravings to smoke and enactment of core life values that motivate living a smoke-free life. After setting up a personalized quit plan, users are taken to the home screen, where they can progress through 8 levels of the intervention content. The program is self-paced, and content is unlocked in a sequential manner. In the "Preparing to Quit" phase, iCanQuit focuses on helping the user develop acceptance of physical sensations, emotions, and thoughts that trigger smoking and allowing these triggers to pass without smoking via mindfulness and perspective-taking. There is an "Urge Help" feature that is tailored to the type of trigger experienced by the user as well as a tracking feature that encourages participants to track the number of cigarettes smoked and urges passed. In the "After You Quit" phase, iCanQuit focuses on helping the user stay motivated and preventing relapse.

QuitGuide

Details of the QuitGuide app have been previously published [21]. Briefly, the QuitGuide app developed by the National Cancer Institute is based on the USCPG for smoking cessation [30]. QuitGuide contained 4 sections of content. "Thinking About Quitting" focuses on motivations to quit by encouraging users to think of reasons for quitting and providing information on the general health consequences of smoking and quitting. "Preparing to Quit" helps users develop a quit plan; identify smoking behaviors, triggers, and reasons for being smoke-free; and identify social support for quitting. "Quitting" teaches skills for avoiding cravings to smoke, such as finding replacement behaviors (eg, chewing on carrot sticks) and staying busy. "Staying Quit" presents tips, motivations, and actions to stay smoke-free and skills for coping with cravings and trying to be positive.

Major Similarities Between iCanQuit and QuitGuide

Both apps provide education and skills for preparing to quit cigarette smoking and for preventing relapse after quitting; actionable plans for quitting cigarette smoking; skills for coping with cravings to smoke; and education on common triggers to smoke, barriers to cessation, and how to seek support for smoking cessation. Both apps also provided education on US Food and Drug Administration–approved medications for smoking cessation. However, neither app provided e-cigarettes as part of the behavioral interventions, encouraged their use, or provided information on the potential risks or benefits of using e-cigarettes to quit combustible cigarette smoking. The details on the features and functionalities of the apps as well as a comparison have been previously published [21].

Assessments

Baseline Survey

The baseline survey collected information on sociodemographic factors such as age, gender, race, education level, and household income. In addition, participants were screened for depression using the Center for Epidemiological Studies Depression Scale-20 and panic disorder and posttraumatic stress disorder (PTSD) using self-report measures [31-33]. Questions related to smoking behavior were also included, such as the number of combustible cigarettes smoked per day, smoking history (years of smoking and past quit attempts), use of e-cigarettes in the past month, and whether family and friends also smoked. The Fagerström Test for Cigarette Dependence (FTCD) [34] was used to assess the level of nicotine dependence. Participants were also asked about their confidence in quitting cigarette smoking and alcohol use.

Smoking Cessation

Study questionnaires at the 3-, 6-, and 12-month follow-ups asked participants whether they abstained from combustible cigarette smoking for the past 30 days and the date of their last cigarette. Consistent with the parent trial, the primary smoking cessation outcome was self-reported, complete-case, and 30-day point prevalence abstinence (PPA) from cigarette smoking at 12 months. Secondary outcomes at 12 months included 30-day PPA missing-as-smoking and multiple imputation; 30-day PPA from all nicotine and tobacco products (ie, e-cigarettes, chewing tobacco, snus, hookahs, cigars, cigarillos, tobacco pipes, and kreteks); prolonged abstinence (between 3 and 12 months); and harm reduction, defined as a substantial reduction (≥50%) in average daily cigarette consumption between baseline and 12 months [35,36].

Self-Selected Use of e-Cigarettes and Cessation Pharmacotherapy

Participants were asked via web-based study questionnaires at baseline and 3-, 6-, and 12-month follow-ups: "In the last 30 days, how often did you use any kind of e-cigarette or vaping?" Response options included (1) not at all, (2) less than once a month, (3) once a month or more but less than once a week, (4) once a week or more but not daily, and (5) at least daily. To assess self-selected use of cessation pharmacotherapy, participants were asked: "Since the date you joined the study, did you ever use any of the following nicotine replacement therapies or medications to help you quit smoking?" Response options included (1) nicotine gum, (2) nicotine patches, (3) e-cigarettes, (4) Chantix, (5) Zyban or Wellbutrin, or (6) "free text." The response to this question was also used to assess self-selected use of cessation pharmacotherapy and whether e-cigarettes were intentionally used to aid quitting.

Engagement With the App-Based Interventions

The engagement was measured via Google Analytics and included the number of times users interacted with their assigned app (ie, number of logins) by 6 months after randomization.

Quit Attempts

A number of quit attempts were assessed at each time point using the survey question, "About how many times since the start of the study did you quit smoking for at least 24 hours?

Statistical Analysis

Sociodemographic characteristics and smoking behaviors of participants were compared between adopters and nonadopters using 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables. To test whether there was an interaction between adopting e-cigarettes and app assignment (iCanQuit vs QuitGuide) on the primary cessation outcome, we evaluated the interaction term in a logistic regression model. To compare 12-month cigarette smoking cessation rates between adopters and nonadopters, logistic regression models were used. For the multiple imputation of missing 30-day PPA from cigarette smoking, multivariate imputation by chained equations in the R package *mice* [37] was used to create 10 complete data sets and pool logistic regression model results [38]. Differences in app engagement and the number of quit attempts between adopters and nonadopters were explored using negative binomial models in the R package *MASS* [39]. We used logistic regression models to explore whether the adoption of e-cigarettes as an aid to cessation impacted smoking cessation. All cessation models were adjusted for factors used in stratified randomization to avoid losing power and obtaining an incorrect 95% CI [40]. These baseline factors included positive screening for depression, daily smoking frequency, education, and minority race or ethnicity backgrounds. We also adjusted for baseline characteristics that both differed between e-cigarette adopters and nonadopters and were associated with the cessation outcome to reduce the potential for confounding [41,42]. The R software (version 4.2.3; R Foundation for Statistical Computing) was used for all statistical analyses, and all statistical tests were 2-sided with α=.05 [43].

Results

Participants Characteristics

Participants included in this analysis (ie, adopters and nonadopters combined) were on average 38.9 years old, 72.9% (1139/1562) female, and 32.8% (512/1562) were from minoritized racial groups (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, or multiple races; Table 1). A substantial proportion (657/1562, 42.1%) reported having a high school diploma or less education; 47.3% (739/1562) were unemployed, disabled, or out of the labor force; and 37.3% (583/1562) reported household incomes of US $20,000 per year or less. Almost half (711/1553, 45.8%) screened positive for depression, 26.1% (401/1537) screened positive for panic disorder, and 41.7% (647/1550) screened positive for PTSD. The FTCD measure of nicotine dependence was mean 5.9 (SD 2.0), indicating moderate to high nicotine dependence, and an overwhelming majority (1323/1562, 84.7%) reported smoking for ≥10 years.

Table 1. Baseline characteristics between nonadopters and adopters.

Characteristic	All participants (N=1562)	Nonadopters (n=1097)	Adopters (n=465)	*P* value
Age (years; N=1562), mean (SD)	38.9 (10.9)	39.3 (10.9)	37.7 (10.8)	.008
Female (N=1562), n (%)	1139 (72.9)	809 (73.7)	330 (71)	.26
Race (N=1562), n (%)				.003
American Indian or Alaska Native	39 (2.5)	31 (2.8)	8 (1.7)	
Asian	3 (0.2)	2 (0.2)	1 (0.2)	
Black or African American	341 (21.8)	266 (24.2)	75 (16.1)	
Native Hawaiian or Pacific Islander	2 (0.1)	0 (0)	2 (0.4)	
White	1050 (67.2)	713 (65)	337 (72.5)	
Multiracial	103 (6.6)	70 (6.4)	33 (7.1)	
Hispanic or Latino ethnicity (N=1562), n (%)	130 (8.3)	83 (7.6)	47 (10.1)	.10
Education (N=1562), n (%)				.11
Less than GED[a] or high school	657 (42.1)	452 (41.2)	205 (44.1)	
Some college, no degree	581 (37.2)	402 (36.6)	179 (38.5)	
College degree or higher	324 (20.7)	243 (22.2)	81 (17.4)	
Employment (N=1562), n (%)				.30
Employed	823 (52.7)	573 (52.2)	250 (53.8)	
Unemployed	192 (12.3)	141 (12.9)	51 (11)	
Disabled	242 (15.5)	178 (16.2)	64 (13.8)	
Out of labor force	305 (19.5)	205 (18.7)	100 (21.5)	
Income (US $; N=1562), n (%)				.94
<$20,000 per year	583 (37.3)	412 (37.6)	171 (36.8)	
$20,000-$54,999 per year	715 (45.8)	499 (45.5)	216 (46.5)	
≥$55,000 per year	264 (16.9)	186 (17)	78 (16.8)	
Married (N=1562), n (%)	503 (32.2)	338 (30.8)	165 (35.5)	.07
LGBT[b] (N=1562), n (%)	251 (16.1)	166 (15.1)	85 (18.3)	.12
Rural residence (N=1562), n (%)	361 (23.1)	262 (23.9)	99 (21.3)	.27
Mental health positive screening results, n (%)				
Depression (n=1553)[c]	711 (45.8)	478 (43.9)	233 (50.2)	.02
Panic disorder (n=1537)[d]	401 (26.1)	278 (25.8)	123 (26.7)	.71
PTSD (n=1550)[e,f]	647 (41.7)	435 (39.9)	212 (46)	.03
Alcohol use (n=1518)				
Drinks per day, mean (SD)	1.8 (3.6)	1.8 (3.5)	1.8 (3.9)	.96
Heavy drinker, n (%)[g]	216 (14.2)	157 (14.7)	59 (13.1)	.43
Smoking behaviors				
Cigarettes smoked per day (n=1518), mean (SD)	19.2 (14.4)	18.9 (13.9)	19.9 (15.3)	.17
FTCD[h] score (n=1518), mean (SD)	5.9 (2.0)	5.8 (2.1)	6.0 (1.9)	.02
Cigarette within 5 minutes of waking (n=1518), n (%)	849 (54.4)	574 (52.3)	275 (59.1)	.01
Smokes greater than one-half pack per day (n=1518), n (%)	1174 (75.2)	816 (74.4)	358 (77)	.28
Smokes >1 pack per day (n=1518), n (%)	313 (20)	211 (19.2)	102 (21.9)	.22

Characteristic	All participants (N=1562)	Nonadopters (n=1097)	Adopters (n=465)	*P* value
Smoked for ≥10 years (n=1518), n (%)	1323 (84.7)	929 (84.7)	394 (84.7)	.98
Past 12-month quit attempts (n=1496), mean (SD)	1.4 (6.6)	1.3 (3.7)	1.7 (10.7)	.26
Confidence to quit (n=1518), mean (SD)[i]	64.2 (26.8)	64.5 (26.9)	63.4 (26.5)	.45
Close friends who smoke (n=1518), mean (SD)	2.6 (1.7)	2.6 (1.8)	2.7 (1.7)	.10
No. housemates who smoke (n=1518), mean (SD)	1.5 (0.9)	1.4 (0.8)	1.5 (1.0)	.03
Living with partner who smokes (n=1518), n (%)	567 (36.3)	384 (35)	183 (39.4)	.10

[a]GED: General Education Development.

[b]LGBT: lesbian, gay, bisexual, or transgender.

[c]Positive screening results for depression via the Center for Epidemiological Studies Depression Scale (cutoff ≥16).

[d]Positive screening results for panic disorder via the 5-item Autonomic Nervous System Questionnaire (reporting ≥1 panic attack or worry about a recurrence within the past month indicates a positive screen).

[e]PTSD: posttraumatic stress disorder.

[f]Positive screening results for PTSD via the PTSD Checklist (scores of ≥14 indicate a positive screen).

[g]Heavy drinking is defined as 4 or more drinks on a typical drinking day for women and 5 or more drinks on a typical drinking day for men within the past 30 days.

[h]FTCD: Fagerström Test for Cigarette Dependence.

[i]Range 0-100, where 0 indicates not at all confident and 100 indicates extremely confident.

Overall, 29.8% (465/1562) of participants adopted e-cigarette use between baseline and the 6-month follow-up, while 70.2% (1097/1562) reported no use of e-cigarettes between baseline and the 6-month follow-up. A descriptively higher, but nonsignificantly different, proportion of QuitGuide participants adopted e-cigarettes compared with iCanQuit participants (QuitGuide: 254/797, 31.9% and iCanQuit: 211/765, 27.6%; *P*=.17).

Adopters were younger, more likely to be White, and screened positive for depression and PTSD compared with nonadopters. Adopters were also more dependent on nicotine than nonadopters (FTCD score: mean 6.0, SD 1.9 for adopters vs mean 5.8, SD 2.1 for nonadopters) and more likely to report having their first cigarette within 5 minutes of waking and to live with other people who smoke. Outcome data retention rates were high and did not differ between adopters and nonadopters (439/465, 94.4% vs 1049/1097, 95.6%; *P*=.32).

Interaction Effect of Adopting e-Cigarettes and Treatment Arm on Smoking Cessation

There was suggestive evidence (*P*=.05) that the association between adopting e-cigarettes and 12-month combustible cigarette smoking cessation was moderated by treatment arm assignment (iCanQuit vs QuitGuide). Therefore, the comparison between cigarette smoking cessation rates between adopters and nonadopters was conducted separately for each arm.

Smoking Cessation Rates Between Adopters and Nonadopters and by Treatment Arm

Specifically, results showed that 12-month smoking cessation rates in the iCanQuit arm were lower among adopters as compared to nonadopters (41/193, 21.2% vs 184/527, 34.9%; *P*=.003; odds ratio [OR] 0.55, 95% CI 0.37-0.81; Table 2). Results were similar and statistically significant for the secondary outcomes: missing-as-smoking imputation (41/211, 19.4% vs 184/554, 33.2%; *P*=.002), multiple imputation (461/2110, 21.8% vs 1909/5540, 34.5%; *P*=.01), and 30-day PPA from all nicotine and tobacco products (30/193, 15.5% vs 167/526, 31.7%; *P*<.001). Furthermore, prolonged abstinence from cigarette smoking between 3- and 12-month follow-ups was about 3 times lower among adopters as compared with nonadopters (7/148, 4.7% vs 75/434, 17.3%; *P*=.001; OR 0.26, 95% CI 0.11-0.57). Finally, there was no difference in daily cigarette reduction rates (ie, reduced by at least half) between adopters and nonadopters (*P*=.99).

Table 2. Smoking cessation rates at 12 months between adopters and nonadopters and by arm[a].

Cessation outcomes and treatment arm	Nonadopters, n/N (%)	Adopters, n/N (%)	OR[b] (95% CI)	*P* value
30-day PPA[c] from cigarette smoking				
Complete case[d]				
iCanQuit	184/527 (34.9)	41/193 (21.2)	0.55 (0.37-0.81)	.003
QuitGuide	104/522 (19.9)	46/246 (18.7)	0.91 (0.62-1.35)	.64
Missing-as-smoking[d]				
iCanQuit	184/554 (33.2)	41/211 (19.4)	0.53 (0.36-0.79)	.002
QuitGuide	104/543 (19.2)	46/254 (18.1)	0.93 (0.63-1.37)	.70
Multiple imputation[d]				
iCanQuit	1909/5540 (34.5)	461/2110 (21.8)	0.57 (0.38-0.85)	.01
QuitGuide	1078/5430 (19.9)	471/2540 (18.5)	0.91 (0.62-1.34)	.63
30-day PPA from nicotine and tobacco products				
iCanQuit	167/526 (31.7)	30/193 (15.5)	0.42 (0.27-0.65)	<.001
QuitGuide	91/522 (17.4)	31/246 (12.6)	0.68 (0.44-1.06)	.09
Prolonged cigarette abstinence[d,e]				
iCanQuit	75/434 (17.3)	7/148 (4.7)	0.26 (0.11-0.57)	.001
QuitGuide	34/435 (7.8)	6/187 (3.2)	0.37 (0.15-0.91)	.03
Reduction in the number of cigarettes smoked by ≥50%[f,g]				
iCanQuit	65/207 (31.4)	29/90 (32.2)	1.00 (0.58-1.72)	.99
QuitGuide	84/268 (31.3)	32/112 (28.6)	0.85 (0.51-1.44)	.55

[a]All models were adjusted for factors used in stratified randomization including positive screening for depression, daily smoking frequency, education, and minority race or ethnicity backgrounds.

[b]OR: odds ratio.

[c]PPA: point prevalence abstinence.

[d]Additional covariate is posttraumatic stress disorder positive screening.

[e]Prolonged abstinence is defined as no smoking since 3-month after randomization, using self-reported date of last cigarette.

[f]Reduction in the number of cigarettes smoked from baseline to 12 months defined as 50% reduction or more.

[g]Additional covariates are age, posttraumatic stress disorder positive screening, and smoking first cigarette within 5 minutes of waking.

In the QuitGuide arm, adopting e-cigarettes was not associated with 12-month combustible cigarette smoking cessation (46/246, 18.7% adopters vs 104/522, 19.9% nonadopters; *P*=.64; OR 0.91, 95% CI 0.62-1.35) with similar results for the secondary outcomes of missing-as-smoking and multiple imputation. There were no significant differences between adopters and nonadopters in 30-day PPA from all nicotine or tobacco products or in daily cigarette reduction. However, the odds of prolonged abstinence from cigarette smoking were roughly 3 times lower among adopters as compared with nonadopters (6/187, 3.2% vs 34/435, 7.8%; *P*=.03; OR 0.37, 95% CI 0.15-0.91).

Post Hoc Analyses

Self-Selected Use of Cessation Pharmacotherapy and e-Cigarettes as an Aid to Cessation

Over a third (188/444, 42.3% of adopters vs 438/1097, 39.9% of nonadopters; *P*=.43) of all participants reported using another form of nicotine replacement therapy (ie, nicotine patch, gum, or lozenge) or medications (ie, varenicline or bupropion) to help them quit smoking at 3 and 6 months. We found no evidence of an interaction between the use of pharmacotherapy and treatment arm assignment on the primary cessation outcome (*P*=.91). We also observed no difference in the rate of pharmacotherapy usage between adopters and nonadopters in either treatment arm (92/211, 43.6% adopters vs 227/554, 41% nonadopters; *P*=.66 in the iCanQuit arm and 107/254, 42.1% adopters vs 212/543, 39% nonadopters in the QuitGuide arm; *P*=.53).

Nearly half (219/465, 47.1%) of adopters reported using e-cigarettes as an aid to cessation. However, there was no difference in cigarette smoking cessation rates between those reporting they adopted e-cigarettes as an aid to cessation of cigarette smoking versus those who did not—in either the iCanQuit arm (19/85, 22% vs 22/105, 21%; *P*=.97; OR 0.99, 95% CI 0.47-2.06) or the QuitGuide arm (19/110, 17.3% vs 25/129, 19.4%; *P*=.69; OR 0.87, 95% CI 0.45-1.71).

Engagement With the App-Based Interventions

Table 3 provides information on participants' engagement with their assigned app. In the iCanQuit arm, adopters had fewer logins to their assigned app than nonadopters, but the difference was not statistically significant (*P*=.19). Participants engaged less with the QuitGuide app overall, and in this treatment arm, there was no difference in app engagement between adopters compared with nonadopters (*P*=.27).

Table 3. Engagement with the app-based interventions between adopters and nonadopters for each treatment arm by 6 monthsa.

Treatment arm	Participants, n/N (%)	Nonadopters		Adopters		Incidence rate ratio (95% CI)	*P* value
		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
iCanQuit	749/765 (97.9)	35.7 (68.2)	9 (3-39)	24.8 (51.4)	8 (2.5-20.5)	0.86 (0.68-1.08)	.20
QuitGuide	790/797 (99.1)	9.1 (18.4)	5 (2-11)	10.0 (45.3)	5 (2-10)	1.10 (0.93-1.31)	.27

[a]Negative binomial models include the following covariates: age, posttraumatic stress disorder positive screening, smoking first cigarette within 5 minutes of waking, and number of housemates who smoke.

Number of Quit Attempts

Adopters in both arms reported a significantly higher number of quit attempts compared with nonadopters (Table 4). In the iCanQuit arm, adopters reported a significantly higher number of quit attempts than nonadopters at 6 months (*P*=.02) but not at 12 months (*P*=.44). In the QuitGuide arm, adopters reported a significantly higher number of quit attempts than nonadopters at 6 months (*P*=.001) and 12 months (*P*<.001). Given the higher number of quit attempts among adopters in both treatment arms, we further evaluated whether the number of quit attempts at the 3- and 6-month follow-ups was associated with cigarette smoking cessation rates at 12 months. We found a very small negative association between 12-month cigarette smoking cessation and number of quit attempts by 3 months (*P*<.001; OR 0.999, 95% CI 0.999-0.999) and 6 months (*P*<.001; OR 0.999, 95% CI 0.999-0.9996).

Table 4. Association between e-cigarette use and number of quit attempts by treatment arm at each follow-up time point[a].

Treatment arm	Participants, n/N (%)	Nonadopters		Adopters		Incidence rate ratio of point estimate (95% CI)	*P* value
		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
At 3 months[b]							
iCanQuit	750/765 (98)	5.6 (12.1)	2 (1-5)	6.4 (12.3)	3 (1-6)	1.15 (0.91-1.45)	.26
QuitGuide	782/797 (98.1)	4.2 (8.0)	2 (0-4)	4.2 (6.9)	2 (1-4)	1.03 (0.83-1.27)	.80
At 6 months[c]							
iCanQuit	747/765 (97.6)	8.8 (22.8)	3 (1-6)	10.8 (22.4)	3 (2-8)	1.32 (1.04-1.68)	.02
QuitGuide	788/797 (98.9)	6.2 (14.2)	2 (1-5)	8.7 (19.9)	3 (2-6)	1.46 (1.17-1.81)	.001
At 12 months[d]							
iCanQuit	711/765 (92.9)	13.2 (46.3)	3 (1-6)	12.0 (39.5)	3 (2-8)	1.11 (0.85-1.45)	.44
QuitGuide	764/797 (95.6)	8.0 (23.6)	3 (1-6)	13.9 (40.6)	4 (2-9)	1.56 (1.25-1.94)	<.001

[a]All models were adjusted for factors used in stratified randomization including positive screening for depression, daily smoking frequency, education, and minority race or ethnicity backgrounds.

[b]Additional covariates are posttraumatic stress disorder positive screening and number of housemates who smoke.

[c]Additional covariates are posttraumatic stress disorder positive screening, smoking first cigarette within 5 minutes of waking, and number of housemates who smoke.

[d]Additional covariates are number of housemates who smoke.

Levels of Nicotine Dependence and Cigarette Smoking Cessation

Regarding the baseline difference in FTCD scores between adopters and nonadopters (6.0 vs 5.8; *P*=.01; Table 1) in our study, we did not find any evidence indicating that baseline levels of nicotine dependence, as measured by FTCD, were associated with combustible cigarette smoking cessation rates at 12 months (*P*=.69; OR 0.99, 95% CI 0.92-1.06). Similar observations emerged when assessing the cessation rates of cigarette smoking at 12 months, comparing individuals with low to moderate nicotine dependence (FTCD scores below 6.0) and those with high nicotine dependence (FTCD scores of 6.0 or higher) at baseline (147/594, 24.7% vs 228/894, 25.5%; *P*=.82; OR 1.03, 95% CI 0.80-1.34).

Discussion

Principal Findings

This study used data from an RCT to compare combustible cigarette smoking cessation rates between participants who used e-cigarettes on their own (ie, "adopters") versus those who did not (ie, "nonadopters"). For the first time, this study provides empirical evidence that adults who adopted e-cigarettes while concurrently receiving an app-delivered cigarette smoking cessation intervention had either a lower or an unimproved likelihood of quitting combustible cigarette smoking—despite their strong motivation to quit. In the iCanQuit arm, 12-month combustible cigarette smoking cessation rates were significantly lower among adopters compared with nonadopters (41/193, 21.2% vs 184/527, 34.9%; *P*=.003). In contrast, in the QuitGuide arm, 12-month combustible cigarette smoking cessation rates did not differ between adopters and nonadopters (46/246, 18.7% vs 104/522, 19.9%; *P*=.64).

Our results are consistent with 2 previous prospective RCT studies for smoking cessation. Both studies suggested that e-cigarette use impeded cigarette smoking cessation among individuals concurrently receiving a behavioral intervention for smoking cessation [15,16]. The first study used data from an RCT (n=1040) among hospitalized patients who wanted to quit combustible cigarette smoking and received cessation behavioral counseling while at the hospital [15]. Patients were followed after discharge and assessed for cigarette smoking abstinence at the 6-month follow-up. The study found that those who started using e-cigarettes after discharge were significantly less likely than non–e-cigarette users to abstain from combustible cigarette smoking (quit rates, 10.1% for e-cigarette users vs 26.6% for non–e-cigarette users; *P*<.001). The second study used data from a large RCT (n=2637) among adults who smoked combustible cigarettes daily and were assigned to receive 12-month web-based behavioral interventions for cigarette smoking cessation [16]. The study found that those who used e-cigarettes were significantly less likely to quit combustible cigarettes than non–e-cigarette users at the 12-month follow-up (quit rates, 21.4% e-cigarette users vs 29.7% non–e-cigarette users; *P*=.006). In contrast to these studies, observational data suggest that e-cigarette use may facilitate smoking cessation under "free living conditions" [11-13]. Moreover, data from RCTs testing e-cigarettes as a potential treatment for smoking cessation suggest that e-cigarette use may help adults quit cigarette smoking [14,44,45]. The disparate results between observational studies, prospective studies, and RCTs of e-cigarettes may relate to discrepancies in how e-cigarettes are used in the real world versus in controlled research studies.

There are unclear reasons why e-cigarette adopters in this study had either a lower or an unimproved likelihood of quitting combustible cigarette smoking. Notably, we found no evidence that the self-selected use of cessation pharmacotherapy differed between adopters and nonadopters. We also examined whether other factors such as adopting e-cigarettes specifically to aid cessation had any impact on the 12-month cessation outcomes. Our findings revealed no significant associations in these regards. Regarding app engagement, however, we found that adopters in the iCanQuit arm used their assigned app less than nonadopters, albeit nonsignificantly, suggesting that adopting e-cigarettes may have been perceived as a method for quitting smoking—which could have thereby replaced app use. Given that engagement with an app-based intervention is a strong predictor of smoking cessation [46-51], it is possible that less engagement with the iCanQuit app contributed, at least in part, to the reduced odds of quitting smoking among adopters compared with nonadopters. However, in the QuitGuide arm, we found no difference in app use between adopters and nonadopters. Thus, further research is needed to understand why adopting e-cigarettes while concurrently receiving an app-based behavioral intervention to quit smoking may lead to reduced use of QuitGuide.

Finally, the number of quit attempts at 3-, 6-, and 12-month follow-ups between adopters and nonadopters was compared. Interestingly, we discovered that adopters had notably higher instances of quit attempts in comparison to nonadopters at the 6-month follow-up in the iCanQuit arm and at 6- and 12-month follow-ups in the QuitGuide arm. These results suggest that, in contrast to nonadopters, adopters encountered more difficulty with quitting cigarette smoking. Notably, these results could not be attributed to differences in levels of nicotine dependence between adopters and nonadopters in this study, which agrees with previous studies with mixed results [16,52-55]. These studies collectively underscore the complexities and uncertainties surrounding the direction of causality between nicotine dependence and e-cigarette use. Therefore, further randomized controlled research on the use of e-cigarettes for smoking cessation is necessary to draw more definitive conclusions.

Strengths

This study has several strengths. First, this is the only known study to examine the impact of e-cigarette use among adults receiving an app-delivered behavioral intervention to help them quit smoking. Second, the original iCanQuit RCT enrolled a racially and geographically diverse large sample (n=2415) of adults from all 50 US states [21], increasing the generalizability of the findings to a broad range of adults who smoke across the country. Third, outcome data retention rates at the 12-month follow-up were high and did not differ between groups. Finally, the use of app-delivered interventions is also a notable strength, as it provides a convenient and accessible way for adults to access evidence-based cessation support.

Limitations

There are also limitations of this study. As with any secondary analysis, the study was not specifically designed to examine the causal relationship between adopting e-cigarettes and cigarette smoking cessation; as such, a causal relationship may not be inferred from the associations found. Additionally, participants self-selected into using or not using e-cigarettes (as allocation was not randomized). This means that both measured and unmeasured confounding likely impact any associations seen. Nonetheless, we adjusted for any baseline characteristics that were significantly different between adopters and nonadopters and were associated with the cessation outcomes. Second, the study did not assess for type of e-cigarette used or its nicotine content. Third, self-reported data are subject to bias and inaccuracies, including in reporting smoking behavior and use of e-cigarettes. In this trial, the self-reported outcome was prespecified due to methodological issues with remote biochemical verification [56,57]. Although previous studies have demonstrated strong agreement between self-reported and biochemically verified smoking status [58,59], some have also shown significant discordance [60,61]. Therefore, the external validity of the self-reported smoking status in this trial is uncertain. However, because the trial was double-blinded, there is no compelling reason that the false reporting rate would be higher in one arm than the other.

Conclusions

Adults who adopted e-cigarettes while simultaneously receiving an app-delivered smoking cessation intervention had either a lower or an unimproved likelihood of quitting cigarette smoking. This is concerning, given the increasing popularity of e-cigarette use. The study can inform future app-based cessation interventions and best practices to mitigate harm from cigarette smoking. Future studies and interventions may include education on the potential negative or unhelpful impact of using e-cigarettes alongside app-delivered behavioral interventions for cigarette smoking cessation.

Acknowledgments

The authors appreciate the tireless contributions of the entire study staff, most notably, Eric Meier, Eric Strand, Carolyn Ehret, Alanna Boynton; the design services of Ayogo, Inc; and the development services of Moby, Inc. The authors are very appreciative of the study participants. This paper was funded by the National Cancer Institute (R01 CA192849).

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Smoking and tobacco use. about electronic cigarettes (e-cigarettes). Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion. 2022. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/about-e-cigarettes.html#e-cigarettes-help-adults-quit-cigarettes [accessed 2023-10-13]
2. Pankow JF, Kim K, Luo W, McWhirter KJ. Gas/particle partitioning constants of nicotine, selected toxicants, and flavor chemicals in solutions of 50/50 propylene glycol/glycerol as used in electronic cigarettes. Chem Res Toxicol 2018;31(9):985-990 [FREE Full text] [doi: 10.1021/acs.chemrestox.8b00178] [Medline: 30113826]
3. Goniewicz ML, Gawron M, Smith DM, Peng M, Jacob P, Benowitz NL. Exposure to nicotine and selected toxicants in cigarette smokers who switched to electronic cigarettes: a longitudinal within-subjects observational study. Nicotine Tob Res 2017;19(2):160-167 [FREE Full text] [doi: 10.1093/ntr/ntw160] [Medline: 27613896]
4. Cobb CO, Foulds J, Yen MS, Veldheer S, Lopez AA, Yingst JM, et al. Effect of an electronic nicotine delivery system with 0, 8, or 36 mg/mL liquid nicotine versus a cigarette substitute on tobacco-related toxicant exposure: a four-arm, parallel-group, randomised, controlled trial. Lancet Respir Med 2021;9(8):840-850 [FREE Full text] [doi: 10.1016/S2213-2600(21)00022-9] [Medline: 33857436]
5. Hartmann-Boyce J, Butler AR, Theodoulou A, Onakpoya IJ, Hajek P, Bullen C, et al. Biomarkers of potential harm in people switching from smoking tobacco to exclusive e-cigarette use, dual use or abstinence: secondary analysis of cochrane systematic review of trials of e-cigarettes for smoking cessation. Addiction 2023;118(3):539-545 [FREE Full text] [doi: 10.1111/add.16063] [Medline: 36208090]
6. Rigotti NA. Balancing the benefits and harms of e-cigarettes: a national academies of science, engineering, and medicine report. Ann Intern Med 2018;168(9):666-667. [doi: 10.7326/M18-0251] [Medline: 29435573]
7. Wise J. e-Cigarettes can help people quit smoking but more evidence needed on long term harms, review concludes. BMJ 2020;371:m3982. [doi: 10.1136/bmj.m3982] [Medline: 33055098]
8. Allen JG, Flanigan SS, LeBlanc M, Vallarino J, MacNaughton P, Stewart JH, et al. Flavoring chemicals in e-cigarettes: diacetyl, 2,3-pentanedione, and acetoin in a sample of 51 products, including fruit-, candy-, and cocktail-flavored e-cigarettes. Environ Health Perspect 2016;124(6):733-739 [FREE Full text] [doi: 10.1289/ehp.1510185] [Medline: 26642857]
9. Singh S, Windle SB, Filion KB, Thombs BD, O'Loughlin JL, Grad R, et al. e-Cigarettes and youth: patterns of use, potential harms, and recommendations. Prev Med 2020;133:106009 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106009] [Medline: 32027913]
10. Leone FT, Douglas IS. The emergence of e-cigarettes: a triumph of wishful thinking over science. Ann Am Thorac Soc 2014;11(2):216-219 [FREE Full text] [doi: 10.1513/AnnalsATS.201312-428ED] [Medline: 24575986]
11. Glasser AM, Vojjala M, Cantrell J, Levy DT, Giovenco DP, Abrams D, et al. Patterns of e-cigarette use and subsequent cigarette smoking cessation over 2 years (2013/2014-2015/2016) in the population assessment of tobacco and health study. Nicotine Tob Res 2021;23(4):669-677 [FREE Full text] [doi: 10.1093/ntr/ntaa182] [Medline: 32939555]
12. Gravely S, Meng G, Hammond D, Hyland A, Cummings KM, Borland R, et al. Differences in cigarette smoking quit attempts and cessation between adults who did and did not take up nicotine vaping: findings from the ITC four country

smoking and vaping surveys. Addict Behav 2022;132:107339 [FREE Full text] [doi: 10.1016/j.addbeh.2022.107339] [Medline: 35605409]
13. Harlow AF, Stokes AC, Brooks DR, Benjamin EJ, Leventhal AM, McConnell RS, et al. Prospective association between e-cigarette use frequency patterns and cigarette smoking abstinence among adult cigarette smokers in the United States. Addiction 2022;117(12):3129-3139 [FREE Full text] [doi: 10.1111/add.16009] [Medline: 35913015]
14. Hartmann-Boyce J, Lindson N, Butler AR, McRobbie H, Bullen C, Begh R, et al. Electronic cigarettes for smoking cessation. Cochrane Database Syst Rev 2022;11(11):CD010216 [FREE Full text] [doi: 10.1002/14651858.CD010216.pub7] [Medline: 36384212]
15. Rigotti NA, Chang Y, Tindle HA, Kalkhoran SM, Levy DE, Regan S, et al. Association of e-cigarette use with smoking cessation among smokers who plan to quit after a hospitalization: a prospective study. Ann Intern Med 2018;168(9):613-620 [FREE Full text] [doi: 10.7326/M17-2048] [Medline: 29582077]
16. Watson NL, Mull KE, Bricker JB. The association between frequency of e-cigarette use and long-term smoking cessation outcomes among treatment-seeking smokers receiving a behavioral intervention. Drug Alcohol Depend 2021;218:108394 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108394] [Medline: 33203525]
17. Adams JM. Smoking cessation-progress, barriers, and new opportunities: the surgeon general's report on smoking cessation. JAMA 2020;323(24):2470-2471. [doi: 10.1001/jama.2020.6647] [Medline: 32432659]
18. Lutz J, Offidani E, Taraboanta L, Lakhan SE, Campellone TR. Appropriate controls for digital therapeutic clinical trials: a narrative review of control conditions in clinical trials of digital therapeutics (DTx) deploying psychosocial, cognitive, or behavioral content. Front Digit Health 2022;4:823977 [FREE Full text] [doi: 10.3389/fdgth.2022.823977] [Medline: 36060538]
19. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019;10(10):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
20. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-01-11]
21. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
22. Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. Behav Ther 2013;44(2):180-198 [FREE Full text] [doi: 10.1016/j.beth.2009.08.002] [Medline: 23611068]
23. McCallion EA, Zvolensky MJ. Acceptance and Commitment Therapy (ACT) for smoking cessation: a synthesis. Curr Opin Psychol 2015;2:47-51. [doi: 10.1016/j.copsyc.2015.02.005]
24. Karekla M, Savvides SN. Smoking cessation avatar-led Acceptance and Commitment Therapy digital intervention: feasibility and acceptability in young adults. Transl Behav Med 2021;11(1):198-205 [FREE Full text] [doi: 10.1093/tbm/ibz128] [Medline: 31628476]
25. Mak YW, Leung DYP, Loke AY. Effectiveness of an individual acceptance and commitment therapy for smoking cessation, delivered face-to-face and by telephone to adults recruited in primary health care settings: a randomized controlled trial. BMC Public Health 2020;20(1):1719 [FREE Full text] [doi: 10.1186/s12889-020-09820-0] [Medline: 33198700]
26. Caraballo RS, Shafer PR, Patel D, Davis KC, McAfee TA. Quit methods used by US adult cigarette smokers, 2014-2016. Prev Chronic Dis 2017;14:E32 [FREE Full text] [doi: 10.5888/pcd14.160600] [Medline: 28409740]
27. Suttiratana SC, Morean ME, Krishnan-Sarin S, Bold KW. Qualitative exploration of longer versus shorter quit attempts among adults using e-cigarettes for combustible cigarette cessation. Addict Behav 2023;143:107710 [FREE Full text] [doi: 10.1016/j.addbeh.2023.107710] [Medline: 37015170]
28. Everard CD, Silveira ML, Kimmel HL, Marshall D, Blanco C, Compton WM. Association of electronic nicotine delivery system use with cigarette smoking relapse among former smokers in the United States. JAMA Netw Open 2020;3(6):e204813 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.4813] [Medline: 32501492]
29. Pierce JP, Chen R, Kealey S, Leas EC, White MM, Stone MD, et al. Incidence of cigarette smoking relapse among individuals who switched to e-cigarettes or other tobacco products. JAMA Netw Open 2021;4(10):e2128810 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.28810] [Medline: 34665239]
30. Fiore MC. US public health service clinical practice guideline: treating tobacco use and dependence. Respir Care 2000;45(10):1200-1262. [Medline: 11054899]
31. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
32. Stein MB, Roy-Byrne PP, McQuaid JR, Laffaye C, Russo J, McCahill ME, et al. Development of a brief diagnostic screen for panic disorder in primary care. Psychosom Med 1999;61(3):359-364. [doi: 10.1097/00006842-199905000-00016] [Medline: 10367617]

33. Lang AJ, Wilkins K, Roy-Byrne PP, Golinelli D, Chavira D, Sherbourne C, et al. Abbreviated PTSD Checklist (PCL) as a guide to clinical response. Gen Hosp Psychiatry 2012;34(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.02.003] [Medline: 22460001]
34. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
35. McRobbie H, Bullen C, Hartmann-Boyce J, Hajek P. Electronic cigarettes for smoking cessation and reduction. Cochrane Database Syst Rev 2014(12):CD010216 [FREE Full text] [doi: 10.1002/14651858.CD010216.pub2] [Medline: 25515689]
36. Pisinger C, Godtfredsen NS. Is there a health benefit of reduced tobacco consumption? a systematic review. Nicotine Tob Res 2007;9(6):631-646. [doi: 10.1080/14622200701365327] [Medline: 17558820]
37. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
38. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons; 1987.
39. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th Edition. New York: Springer; 2023.
40. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol 1999;52(1):19-26 [FREE Full text] [doi: 10.1016/s0895-4356(98)00138-3] [Medline: 9973070]
41. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;21(19):2917-2930. [doi: 10.1002/sim.1296] [Medline: 12325108]
42. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355(9209):1064-1069. [doi: 10.1016/S0140-6736(00)02039-0] [Medline: 10744093]
43. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2020. URL: https://www.R-project.org/ [accessed 2023-10-13]
44. Eisenberg MJ, Hébert-Losier A, Windle SB, Greenspoon T, Brandys T, Fülöp T, et al. Effect of e-cigarettes plus counseling vs counseling alone on smoking cessation: a randomized clinical trial. JAMA 2020;324(18):1844-1854 [FREE Full text] [doi: 10.1001/jama.2020.18889] [Medline: 33170240]
45. Hajek P, Phillips-Waller A, Przulj D, Pesola F, Smith KM, Bisal N, et al. A randomized trial of e-cigarettes versus nicotine-replacement therapy. N Engl J Med 2019;380(7):629-637 [FREE Full text] [doi: 10.1056/NEJMoa1808779] [Medline: 30699054]
46. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
47. Graham AL, Cobb NK, Papandonatos GD, Moreno JL, Kang H, Tinkelman DG, et al. A randomized trial of internet and telephone treatment for smoking cessation. Arch Intern Med 2011;171(1):46-53 [FREE Full text] [doi: 10.1001/archinternmed.2010.451] [Medline: 21220660]
48. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011;13(4):227-238 [FREE Full text] [doi: 10.1093/ntr/ntq252] [Medline: 21350042]
49. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009;104(11):1792-1804 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
50. Webb TL. Commentary on Shahab and McEwen (2009): understanding and preventing attrition in online smoking cessation interventions: a self-regulatory perspective. Addiction 2009;104(11):1805-1806. [doi: 10.1111/j.1360-0443.2009.02751.x] [Medline: 19832784]
51. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
52. Martinez U, Simmons VN, Sutton SK, Drobes DJ, Meltzer LR, Brandon KO, et al. Targeted smoking cessation for dual users of combustible and electronic cigarettes: a randomised controlled trial. Lancet Public Health 2021;6(7):e500-e509 [FREE Full text] [doi: 10.1016/S2468-2667(20)30307-8] [Medline: 34175001]
53. González-Roz A, MacKillop J. No evidence of differences in smoking levels, nicotine dependence, carbon monoxide or motivational indices between cigarette smokers and cigarette + e-cigarette dual users in two samples. Addict Behav 2021;112:106543. [doi: 10.1016/j.addbeh.2020.106543] [Medline: 32911351]
54. Selya AS, Dierker L, Rose JS, Hedeker D, Mermelstein RJ. The role of nicotine dependence in e-cigarettes' potential for smoking reduction. Nicotine Tob Res 2018;20(10):1272-1277 [FREE Full text] [doi: 10.1093/ntr/ntx160] [Medline: 29065204]
55. Martínez Ú, Martínez-Loredo V, Simmons VN, Meltzer LR, Drobes DJ, Brandon KO, et al. How does smoking and nicotine dependence change after onset of vaping? a retrospective analysis of dual users. Nicotine Tob Res 2020;22(5):764-770 [FREE Full text] [doi: 10.1093/ntr/ntz043] [Medline: 30883640]
56. Herbec A, Brown J, Shahab L, West R. Lessons learned from unsuccessful use of personal carbon monoxide monitors to remotely assess abstinence in a pragmatic trial of a smartphone stop smoking app—a secondary analysis. Addict Behav Rep 2019;9:100122 [FREE Full text] [doi: 10.1016/j.abrep.2018.07.003] [Medline: 31193683]

57. Thrul J, Meacham MC, Ramo DE. A novel and remote biochemical verification method of smoking abstinence: predictors of participant compliance. Tob Prev Cessat 2018;4:20 [FREE Full text] [doi: 10.18332/tpc/90649] [Medline: 30417159]
58. van der Aalst CM, de Koning HJ. Biochemical verification of the self-reported smoking status of screened male smokers of the Dutch-Belgian randomized controlled lung cancer screening trial. Lung Cancer 2016;94:96-101. [doi: 10.1016/j.lungcan.2016.02.001] [Medline: 26973213]
59. Wong SL, Shields M, Leatherdale S, Malaison E, Hammond D. Assessment of validity of self-reported smoking status. Health Rep 2012;23(1):47-53 [FREE Full text] [Medline: 22590805]
60. Piñeiro B, Vidrine DJ, Wetter DW, Hoover DS, Frank-Pearce SG, Nguyen N, et al. Implementation of ask-advise-connect in a safety net healthcare system: quitline treatment engagement and smoking cessation outcomes. Transl Behav Med 2020;10(1):163-167 [FREE Full text] [doi: 10.1093/tbm/iby108] [Medline: 30476236]
61. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994;84(7):1086-1093. [doi: 10.2105/ajph.84.7.1086] [Medline: 8017530]

Abbreviations

ACT: acceptance and commitment therapy
FTCD: Fagerström Test for Cigarette Dependence
OR: odds ratio
PPA: point prevalence abstinence
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
USCPG: US Clinical Practice Guidelines

Edited by L Buis; submitted 10.05.23; peer-reviewed by M Metcalf, N Machado, S Shiffman; comments to author 30.06.23; revised version received 18.08.23; accepted 06.10.23; published 09.11.23.

Please cite as:
Santiago-Torres M, Mull KE, Sullivan BM, Bricker JB
Use of e-Cigarettes in Cigarette Smoking Cessation: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e48896
URL: https://mhealth.jmir.org/2023/1/e48896
doi:10.2196/48896
PMID:37943594

Original Paper

The Effects of a Digital, Transdiagnostic, Clinically and Peer-Moderated Treatment Platform for Young People With Emerging Mental Health Complaints: Repeated Measures Within-Subjects Study

Marilon van Doorn[1,2], MSc; Anne Monsanto[1,3], MSc; Chen Lu Wang[1], BSc; Sander C J Verfaillie[1,4], PhD; Thérèse A M J van Amelsvoort[5], Prof Dr; Arne Popma[1], Prof Dr; Monique W M Jaspers[6], Prof Dr; Ferko Öry[7], Dr med; Mario Alvarez-Jimenez[8,9], Prof Dr; John F Gleeson[10], Prof Dr; Dorien H Nieman[1], Dr med

[1]Amsterdam University Medical Centers (Location AMC), Amsterdam, Netherlands

[2]Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[3]Antes, Rotterdam, Netherlands

[4]GGZ inGeest Specialized Mental Health Care, Amsterdam, Netherlands

[5]Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, Netherlands

[6]Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers (Location AMC), Amsterdam, Netherlands

[7]Buurtzorg Jong, Almelo, Netherlands

[8]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia

[9]Orygen, Melbourne, Australia

[10]Healthy Brain and Mind Research Centre and School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Australia

Corresponding Author:
Marilon van Doorn, MSc
Amsterdam University Medical Centers (Location AMC)
5 Meibergdreef 9
Amsterdam, 1105AZ
Netherlands
Phone: 31 208913747
Email: m.vandoorn@amsterdamumc.nl

Abstract

Background: To address the growing prevalence of youth mental health problems, early intervention is crucial to minimize individual, societal, and economic impacts. Indicative prevention aims to target emerging mental health complaints before the onset of a full-blown disorder. When intervening at this early stage, individuals are more responsive to treatment, resulting in cost-effective outcomes. The Moderated Online Social Therapy platform, which was successfully implemented and proven effective in Australia, is a digital, peer- and clinically moderated treatment platform designed for young people. The Netherlands was the first country outside Australia to implement this platform, under the name Engage Young People Early (ENYOY). It has the potential to reduce the likelihood of young people developing serious mental health disorders.

Objective: This study aims to investigate the effects on young people using the ENYOY-platform in relation to psychological distress, psychosocial functioning, and positive health parameters.

Methods: Dutch-speaking young people with emerging mental health complaints (N=131) participated in the ENYOY-platform for 6 months in a repeated measures within-subjects study. Psychological distress, psychosocial functioning, and positive health parameters were assessed at baseline and 3, 6, and 12 months. Repeated measures ANOVA was conducted and adjusted for age, sex, therapy, and community activity. The Reliable Change Index and Clinically Significant Index were computed to compare the baseline with the 6- and 12-month measurements. The missing data rate was 22.54% and the dropout rate 62.6% (82/131).

Results: The primary analysis (77/131, 58.8%) showed that psychological distress decreased and psychosocial functioning improved over time with large effect sizes (*P*<.001 in both cases; η_p^2=0.239 and 0.318, respectively) independent of age (*P*=.76 for psychological distress and *P*=.48 for psychosocial functioning), sex (*P*=.24 and *P*=.88, respectively), therapy activity (*P*=.49

and *P*=.80, respectively), or community activity (*P*=.59 and *P*=.48, respectively). Similarly, secondary analyses (51/131, 38.9%) showed significant effects of time on the quality of life, well-being, and meaningfulness positive health parameters ($P<.05$; η_p^2=0.062, 0.140, and 0.121, respectively). Improvements in all outcome measures were found between baseline and 3 and 6 months ($P\leq.001$-.01; *d*=0.23-0.62) and sustained at follow-up (*P*=.18-.97; *d*=0.01-0.16). The Reliable Change Index indicated psychological distress improvements in 38% (39/102) of cases, no change in 54.9% (56/102) of cases, and worsening in 5.9% (6/102) of cases. Regarding psychosocial functioning, the percentages were 50% (51/102), 43.1% (44/102), and 6.9% (7/102), respectively. The Clinically Significant Index demonstrated clinically significant changes in 75.5% (77/102) of cases for distress and 89.2% (91/102) for functioning.

Conclusions: This trial demonstrated that the ENYOY-platform holds promise as a transdiagnostic intervention for addressing emerging mental health complaints among young people in the Netherlands and laid the groundwork for further clinical research. It would be of great relevance to expand the population on and service delivery of the platform.

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-021-03315-x

(JMIR Mhealth Uhealth 2023;11:e50636) doi:10.2196/50636

KEYWORDS

indicative prevention; youth mental health; Moderated Online Social Therapy; MOST+; eHealth; well-being; early detection and intervention; Engage Young People Early; ENYOY

Introduction

Background

Mental health problems in young people have never been more prevalent in the Netherlands [1]. In 2021, approximately 1 in 4 young adults (aged 18-25 y) were labeled “psychologically unhealthy” (scored <60 on a screening instrument for mental health complaints) [1,2]. Worldwide, the COVID-19 pandemic has been associated with further increases in psychological complaints in the past few years [3].

Of all mental health problems, 75% start before the age of 25 years, and mental health problems are the leading contributors to disease burden among young people [4,5]. The onset of mental health problems in early adulthood poses a significant threat to subsequent personal development [6-8]. Youth mental health problems are precursors to physical health problems (eg, respiratory, cardiovascular, and infectious diseases [9]) and more severe mental health problems later in life [10]. Owing to the disruptiveness of youth mental health problems, it is estimated that they are associated with 10-fold higher societal costs during adulthood compared with mental health problems emerging in adulthood [11]. Because of the growing waitlist for mental health care [12-15] and the need for a formal classification to receive treatment in the Netherlands [16], care is often unavailable; presenting problems tend to worsen during the waiting time; and, subsequently, more intensive treatment is needed. As mental health problems are the single highest source of global economic burden (with the Netherlands estimated to have spent €4.7 billion [US $5.02 billion] in 2019 [17]), prevention and treatment of mental health complaints in youth should be the top priority to help young people fulfill their potential.

To reduce the individual, societal, and economic impacts of mental health problems among youth, it is of importance to intervene as early as possible. Indicative prevention—intervening at an early stage, namely, in people with *emerging* mental health problems and *before* the onset of a full-blown disorder—is more effective or cost-effective as, at this stage, fewer psychosocial and neurobiological consequences occur and individuals need fewer treatment sessions as they are still more responsive to treatment [18]. This approach has been adopted in many other fields of medicine, such as cardiology and oncology, and has been shown to improve prognosis [19,20]. With the purpose of adopting this approach for mental health problems, the clinical staging model was developed [21-23] (Textbox 1 [21,24]). The stages in the model differentiate emerging mental health problems with mild symptoms and functional impairment (stage 1a) or moderate symptoms and functional impairment (stage 1b) [25-27] from more discrete (stage 2), recurrent (stage 3), and treatment-resistant (stage 4) disorders. The model has been found to be usable in clinical practice and has excellent interrater reliability [21,28,29].

Textbox 1. Clinical staging model (adapted from Hickie et al [21]; see also the study by van Doorn et al [24]). Italics indicate the stages we focus on in this study.

Stage 0

- Asymptomatic individuals at risk of a disorder who have not yet presented for care

Stage 1a

- *Help-seeking individuals with mild symptoms and mild functional impairment*

Stage 1b

- *People with attenuated syndromes with partial specificity, often with mixed or ambiguous symptoms and moderate functional impairment*

Stage 2

- People with discrete disorders: clear episodes of psychotic, manic, or severe depressive symptoms

Stage 3

- People with recurrent or persistent disorders

Stage 4

- People with severe, treatment-resistant, or unremitting disorders

As emerging mental health complaints (stages 1a and 1b) are often diffuse, have various trajectories, and are underpinned by overlapping mechanisms [21,24], a transdiagnostic approach could potentially add to the effectiveness of preventive interventions [30]. For example, transdiagnostic cognitive behavioral therapy (CBT) aims to pinpoint fundamental cognitive behavioral processes that are believed to be significant across various disorders and formulate a treatment strategy aimed at these shared factors [31]. Transdiagnostic treatment protocols for youth based on CBT and third-wave behavior therapies such as acceptance and commitment therapy, dialectical behavior therapy, meta-cognitive therapy, and mindfulness-based interventions are available and promote change through a common lens that applies across emotional disorders, including anxiety, depression, obsessive-compulsive disorders, and others [32,33]. Transdiagnostic models are more personalized, implementable, and scalable and, therefore, are ideal for the large-scale implementation of complex interventions [30,34]. Several meta-analyses indicate that, relative to diagnostic-specific interventions, transdiagnostic protocols are at least as or more effective at addressing the primary clinical diagnosis, more effective at alleviating comorbid diagnoses, easier to implement, and more engaging [35-41]. Transdiagnostic CBT has proven its efficacy, extending even into preventive approaches [33,42,43] and with promising preliminary results when delivered digitally [44,45]. Similarly, Păsărelu et al [46] provided evidence on the feasibility and efficacy of an internet-delivered intervention for anxiety and depressive disorders in adolescents based on rational motive behavior therapy.

Following the introduction of the clinical staging model, early intervention has been successfully implemented in Australia. There are >150 Headspace centers where young people with emerging mental health complaints can receive stepped and evidence-based care [47-49]. Moreover, usually when a young person turns 18 years, the division between child, adolescent, and adult mental health services hampers the continuity of care and often leaves them without professional help [50-52]. Headspace has successfully bridged this gap by providing services to young people between the ages of 12 and 25 years [52,53].

Given the barriers to access, the initial step toward seeking help is a difficult one. Only 30% of young people in Western countries seek help for their mental health problems [54,55], and most receive mental health care at a much later stage, when mental health symptoms have worsened or become chronic [55-58]. Barriers that contribute to the low rates of help seeking [9,59,60] include mental health stigma or self-stigma [61,62], mental health services that are not aligned with young people's needs [50,51], lack of accessible help, the preference for solving one's own problems, downgrading one's problems ("others need it more than me"), and lack of knowledge about mental health (services) [55,63].

Offering digital, nonstigmatizing, and accessible care could help overcome some of these barriers. A recent scoping review found that digital indicated preventive interventions hold promising potential in (1) reducing a range of mental health problems, (2) enhancing aspects of positive health such as mental well-being and resilience, and (3) having high levels of usability and acceptability. An important observation was that web-based interventions that combined both clinical and peer moderation tended to yield the most consistent and impactful results. Nevertheless, significant gaps have been identified in the literature. For instance, there is an absence of comprehensive transdiagnostic approaches, a lack of clear definitions of emerging mental health complaints, and a lack of suitable assessment tools for emerging mental health symptoms. In addition, studies have mostly focused on either children aged <18 years or adults aged >18 years, thereby accentuating the existing gap between child and adult psychiatry. In addition, there was a scarcity of studies with follow-up data [64].

These gaps have largely been addressed in the Moderated Online Social Therapy (MOST) platform, which is a digital treatment platform moderated by both peers and clinical professionals for young people with emerging mental health issues and has been implemented in Australia. On the MOST platform, young people can (1) undertake evidence-based therapy exercises (when and where they want), (2) have one-on-one sessions with psychologists and youth with lived experience with mental health problems (peer workers), and (3) have peer-to-peer support on the platform's community page [65,66]. On the basis of a transdiagnostic approach, MOST's guided therapy journeys are evidence-based therapy pathways rooted in CBT, acceptance and commitment therapy, compassion-focused therapy, mindfulness-based cognitive therapy, meta-cognitive therapy, and social cognition strategies and cater to specific mechanisms underlying common complaints such as anxiety, social anxiety, social functioning, and depression. Using a process-focused approach, the intervention addresses key cognitive and behavioral processes (ie, repetitive negative thinking, cognitive affective biases, avoidance, and emotional regulation) common across mental health presentations [66]. These journeys are individually and personally tailored to individual strengths and complaints through the use of an embedded algorithm. The constructed therapy journey comprises psychoeducation and therapeutic exercises proven to be effective for specific shared mechanisms that underlie conditions such as anxiety. Furthermore, these journeys remain adaptable under the guidance of clinical moderators, ensuring a closer alignment with the unique needs of each individual [66,67]. In >13 years of research, the platform has been associated with improvements in psychological distress, perceived stress, psychological well-being, depression, (social) anxiety, loneliness, and suicidal ideation in young people with complaints across the diagnostic spectrum [66,68-70]. Moreover, there have been significant increases in vocational and educational recovery and reduced rates of hospital admissions and visits to emergency services, as well as high levels of feasibility, acceptability, engagement, and safety [71,72].

Inspired by MOST, the Netherlands was the first country outside Australia to implement the MOST platform under the name Engage Young People Early (ENYOY). ENYOY aims to support young people (aged 16-25 y) with emerging mental health complaints (Textbox 1, stages 1a and 1b) and reduce their chance of developing a serious mental health disorder [24]. Research has already shown adequate to high usability rates, and young people have reported that they considered ENYOY to be a user-friendly, safe, accessible, and inclusive initiative that helped them reduce their mental health complaints and improve their quality of life [67]. Moreover, providing users with a smartwatch to transform physiological parameters into an observable signal ("biocueing") and offering stress-regulating exercises on the platform has been found to help them become more emotionally aware [73].

Objectives

In this paper, we report outcomes from the ENYOY-platform using a participatory design in a within-group approach, focusing on the following parameters: psychological distress, psychosocial functioning, and positive health parameters (such as well-being and quality of life) [24]. Our hypothesis was that ENYOY would attain similar results of improvement in psychological distress [74] and psychosocial functioning [75] as have been found for the young people visiting the Headspace centers in Australia [49], as well as a positive change in positive health parameters [76]. It was expected that spending more time on the platform and being more active in the digital community would be associated with increased improvements. In addition, individual changes (using the Reliable Change Index [RCI] and Clinically Significant Index [CSI]) were calculated to gain a nuanced understanding of intervention impact at the individual level [77,78]. These indexes could help gain insights into the specific individuals who experienced significant changes and those who did not, thus enabling a more nuanced understanding of the intervention's impact, which could contribute to a more personalized and effective treatment approach in the future. Finally, potential differences in effects between individuals in clinical stages 1a and 1b were investigated for explorative purposes. This project has the potential to realize the goal of specialized, precise, and digitally integrated treatment by providing cost-effective, nonstigmatizing, constantly available support to young people with emerging mental health complaints in the Netherlands.

Methods

Study Context and Design

This study took place within the context of the ENYOY project [24]. The study setting was a safe digital treatment environment, and all measurements were conducted via videocall. This study had a repeated measures within-subjects participatory design. Participants engaged with the platform for a duration of 6 months, during which measurements occurred at baseline, 3, and 6 months. Furthermore, a follow-up assessment was conducted at 12 months. The inclusion of the 3-month assessment time point was intended to explore potential dose- and time-dependent effects, namely, to ascertain whether prolonging the time on the platform would result in more substantial improvements.

Participants

The study included Dutch-speaking young people aged 16 to 25 years with emerging mental health complaints (stages 1a and 1b; Textbox 1) who were able to provide informed consent. Those without mental health problems (stage 0), with more severe mental health problems (stage ≥2), or with an acute risk of self-harm or suicide were excluded [24]. An operationalization of the clinical staging model by Hickie et al [21] was used. Hickie et al [21] provide clear descriptions and cutoffs that were used to ascertain the clinical stage. Subsequent to the initial assessment, in which a clinical interview was conducted, the research assistant determined the clinical stage, followed by a definitive consensus reached during a weekly collaborative meeting involving a team comprising researchers, psychologists, registered mental health care and clinical psychologists, and youth with lived experience. At the time of writing this paper, an assessment tool with a digital algorithm that holds promise for use in practice had been developed and is currently under review [79].

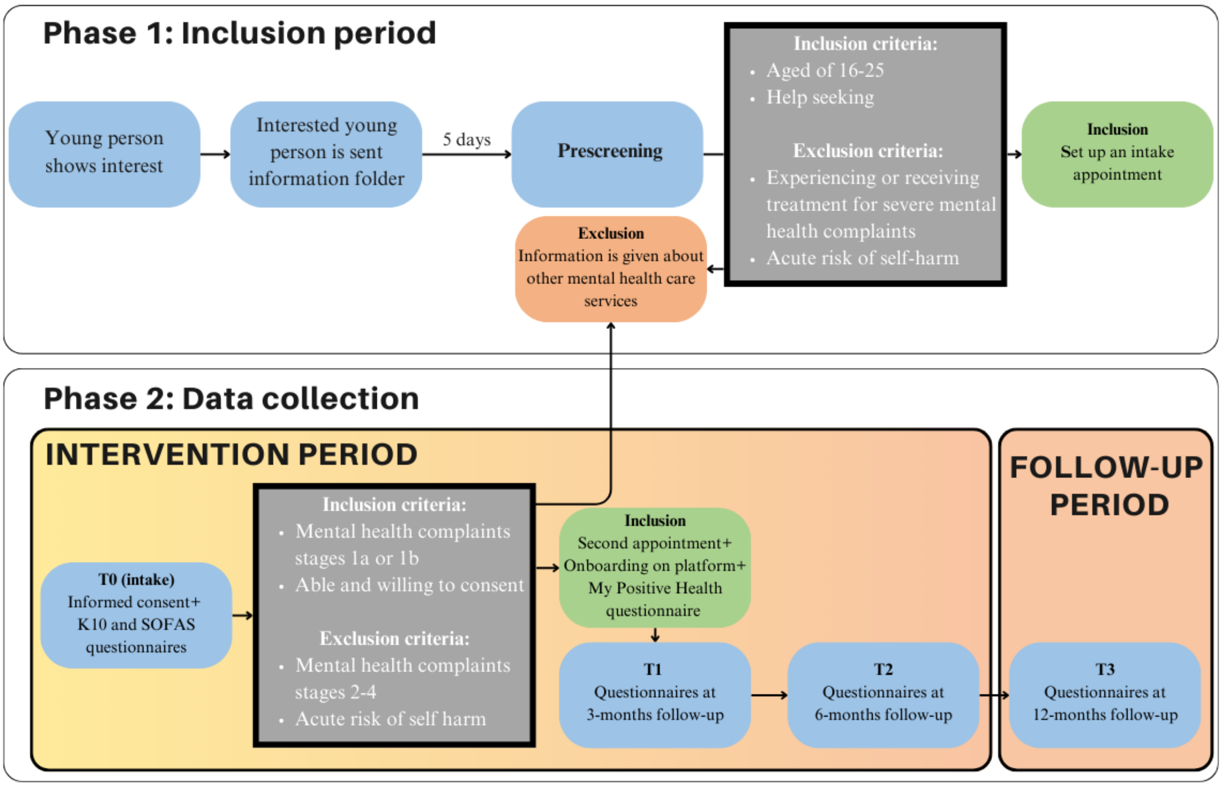
The sample size of 43 was based on a comparable study in Australia [49] (a priori analysis with 80% power and an α of .01). Dropout rates for digital interventions for young people with emerging mental health complaints have ranged from 2% to 73% [64]. A sample size of 125 at baseline was estimated to give us ample power to investigate our hypotheses considering an estimated 50% dropout.

Procedure

Participants approached the team via the ENYOY website or email address, university platforms, social media, advertisements, or a recruitment platform (Link2Trials) to show interest. As a response, contact details were collected, and the study information and consent forms were sent by post. Contact was made with the participant 5 days after the package was sent to conduct a short prescreening that was done digitally because of the COVID-19 pandemic (ie, via phone or Teams [Microsoft Corporation]; participants who were experiencing or receiving treatment for severe mental health issues were excluded and given information about other mental health care services) and to set up an intake appointment.

During a 1-hour baseline assessment interview (intake), participants provided informed consent and were screened for eligibility using a semistructured interview and questionnaires (see the *Questionnaires* section) and by means of the operationalization of the clinical model by Hickie et al [21] (see the *Participants* section). Participants were informed that participation was voluntary and that they could stop participation at any moment. If this happened, reason or reasons for dropout were noted. The intake information was reviewed in a weekly consensus meeting with a (clinical or mental health) psychologist and peer workers to determine the clinical stage [21]. Noneligible participants were contacted to offer further guidance for other mental health care services, and eligible participants were contacted to schedule the second appointment. During this appointment, participants were "onboarded" on the ENYOY-platform and filled out the My Positive Health (MPH) questionnaire (see the *Questionnaires* section [76]), and biweekly coaching appointments were made (Figure 1). Participants were allowed to use the platform as much as they liked over the course of 6 months. Follow-up measurements were taken at 3, 6, and 12 months. All data were stored using the Castor electronic data capture software [80].

Figure 1. Participant timeline. K10: Kessler Psychological Distress Scale; SOFAS: Social and Occupational Functioning Assessment Scale.



Materials

Intervention

The platform contained the modules (building blocks) outlined in Textbox 2.

Textbox 2. Modules of the platform.

Guided therapy journeys

- Each participant completed a questionnaire (based on the positive psychology framework [81,82]) and an algorithm tailored to a treatment journey specific to an individual's strengths and needs. On the basis of the preferences of an individual, together with a therapist, the therapy journey could be tailored even further (eg, changing the order of exercises, and adding or deleting exercise types or content), thereby enabling participants to modify the intervention to their personal needs and shape their own therapy journey. The activities entailed the following:
  - Reflective actions, which are behavioral or cognitive experiments used to improve the ability to notice thought processes in order to build insight and self-awareness of these processes;
  - Regular actions, which are behavioral experiments in real-world contexts used to generalize adaptive coping strategies and behaviors by increasing self-efficacy and by challenging cognitions;
  - Therapy comics, which are engaging, powerful, and accessible means to understand and negotiate mental ill health challenges;
  - Talking points, which are topics of discussion that provide an opportunity to share effective coping strategies to encourage social problem-solving and peer modeling and learning.

Personalized therapy toolkit

- A library of therapy work and favorite strategies.

Safe digital social network

- A moderated digital support network of all the included young people, there to support each other if and when they need it on their recovery journey.

Professional digital support

- Digital support from peer workers and clinical moderators.

The ENYOY-platform follows safety protocols based on the MOST intervention [83] that have been approved by 3 ethics committees and have been successfully implemented in 4 pilot studies and 2 randomized controlled trials (RCTs) [84,85]. The platform has an automated alert system for identifying increased risk of suicide or self-harm, and therapists screen for clinical risk information twice daily. High-risk words activate the safety protocol, following which the clinical moderator conducts a telephone risk assessment and, where necessary, implements one or more of the following procedures: (1) informing the general practitioner, (2) informing the nominated emergency contact, and (3) liaising with suitable emergency services. The platform also includes safety features such as a reporting function for users and visible 24/7 emergency numbers. Information and communications technology safety is in compliance with European laws and safety regulations developed in cooperation with a privacy officer from Amsterdam University Medical Centers.

Questionnaires

Questionnaires were administered at baseline and at the 3-, 6-, and 12-month follow-ups to assess the platform intervention (3- and 6-month measurements) and evaluate whether the initial effects were sustained over time (12-month measurement). The overall missing data rate was 22.54%. The overall level of functioning was measured using the Social and Occupational Functioning Assessment Scale (SOFAS) [75], which has a validated translation to Dutch [86]. The SOFAS consists of 15 open-ended questions (eg, “How is your contact with your family and/or partner?”). A 1-item rating ranges from 1 to 100 (1=inability to function; 100=superior functioning). It has excellent interrater reliability [87] (intraclass correlation coefficient=0.83).

Psychological distress was measured using the validated Dutch version [88] of the Kessler Psychological Distress Scale (K10) [74] self-report questionnaire. The K10 consists of 10 questions (eg, “during the past month, how often did you feel restless?”) that are scored on a 5-point Likert scale (1=*always*; 5=*never*). Higher total scores indicate more severe psychological distress. Strong psychometric qualities have been found, that is, good internal consistency (Cronbach α=.91), strong interitem correlation (0.350-0.659) [89], and high reliability (Cronbach α=.94) [88].

The Dutch self-report questionnaire MPH [76] assessed positive health. The MPH subscales measure daily functioning, physical functioning, mental well-being, participation, quality of life, and meaningfulness. For the purpose of this study, the mental well-being (eg, “I feel happy”), quality of life (eg, “I enjoy my life”), and meaningfulness (eg, “I know what things I would like to do in my life”) subscales were used. A total of 44 items were scored on a 10-point Likert scale (1=*not at all*; 10=*extremely*). The reliability was found to be good to very good (Cronbach α=.820-.933), and the discriminant validity was acceptable but with some overlap between domains [90]. An updated version of the instrument with improved validity is now available [76].

Data Analysis

User Statistics

Data management and confidentiality measures have been described previously [24]. Age (into 2 groups based on age, with a median split at 21 y: 16-21 y and 22-25 y), educational level (lower, middle, and higher education [91]), and sex differences were assessed using chi-square analyses [49].

Mean Changes in Outcomes Over Time

A repeated measures ANOVA was used to assess changes over time (baseline and 3, 6, and 12 mo) for the K10 and SOFAS data. For the secondary analysis, the MPH data were analyzed using a repeated measures ANOVA to explore the changes over time. As the parameters of daily and physical functioning and participation are already included in the SOFAS measure [75], the mental well-being, quality of life, and meaningfulness subscales were used as outcome variables. All analyses were adjusted for age, sex, therapy activity, and community activity to determine whether the effects over time on the K10 and SOFAS scores were independent of educational level, sex, and participant involvement as common confounders. Post hoc pairwise comparisons with Bonferroni correction were performed to investigate the effects between each temporal condition. Post hoc analyses indicated sufficient statistical power (>0.9) [24]. Effect sizes were reported for the primary and secondary analyses (η_p^2) and post hoc pairwise comparisons (Cohen *d*). To investigate whether dropout and missing data (22.54%) influenced the results, we performed a sensitivity analysis using a repeated measures ANOVA on the complete data set (baseline; N=131). Multiple imputations (50 [92,93]) were used to fill in the missing values using both the dependent and time variables.

Reliable and Clinically Significant Change

To assess change indicators, reliable improvement (RCI) and belonging to a clinical versus nonclinical population (CSI) were computed individually on the sample [77] comparing the baseline versus 6-month measurements and the baseline versus 12-month measurements. For the RCI, a distinction was made between groups 1a and 1b to compare the course of mental health complaints. A moderate effect size of ≥0.5 was considered a significant change [77]. For the K10, the RCI was estimated at a 7-point change, and the CSI was estimated at 23 points (following the studies by Rickwood et al [49] and Slade et al [94]). For the SOFAS, the RCI was estimated at a 10-point change, and the CSI was estimated at 69 points [49]. Furthermore, potential variations in individual outcomes between clinical stages 1a and 1b were subject to exploratory investigation as the staging model indicates that these groups are at different risk stages [21,23].

Predictors for Improvement

Finally, we conducted a logistic regression analysis to investigate the potential protective effects (of independent variables [quality of life and meaningfulness]) at 6 months (after the intervention) in relation to progression (on dependent variables [psychological distress and psychosocial functioning]) at 12 months.

Ethical Considerations

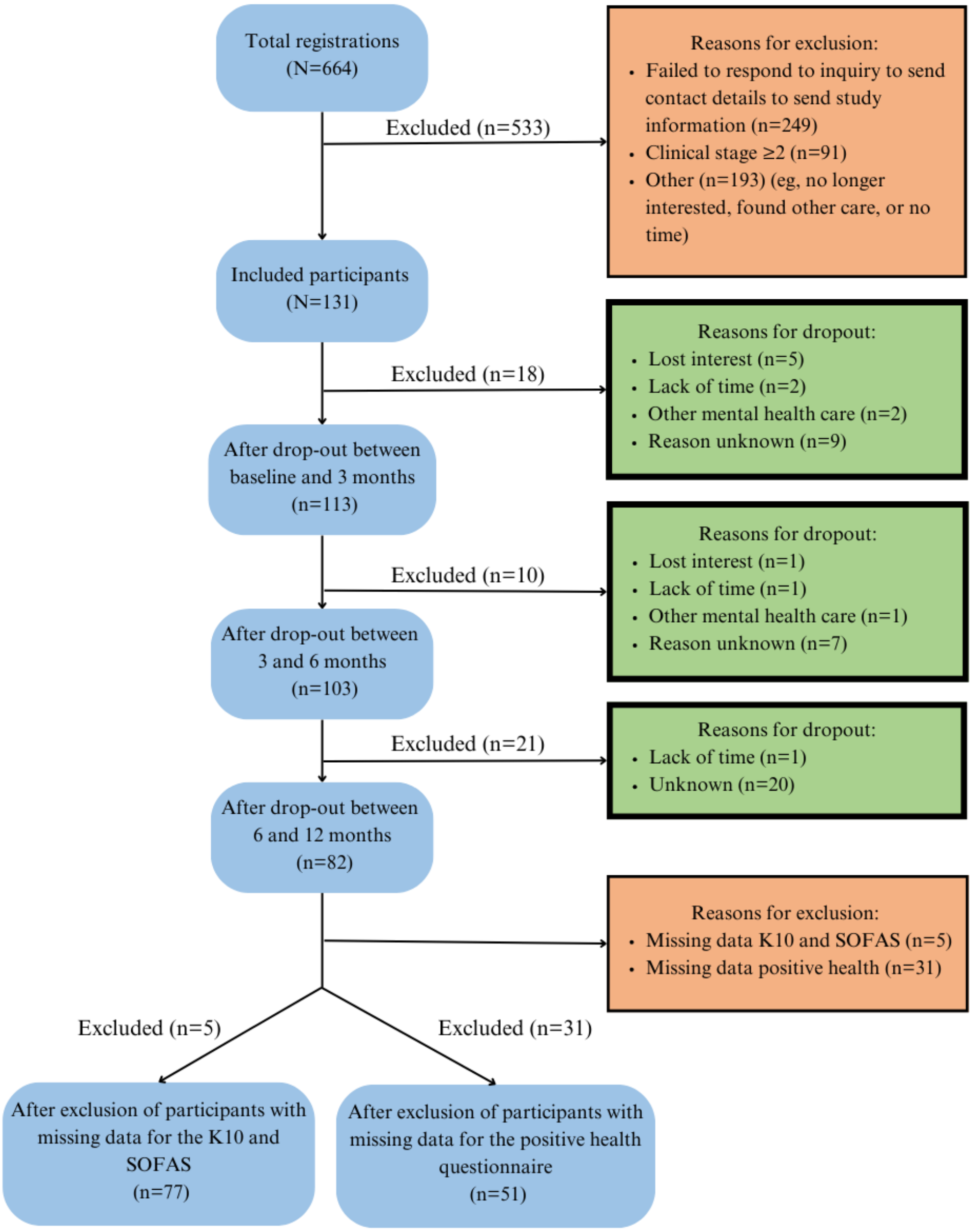
The study was reviewed and approved by the Medical Research Ethics Committee at Amsterdam University Medical Centers (NL66345.018.18). The participants provided written informed consent to take part in this study.

Results

Participants

Of the interested young people, 37.5% (249/664) failed to respond or had personal reasons for not joining, and 42.8% (284/664) were excluded. A sample of 131 remained. Excluding participants with missing data resulted in a final sample of 77 for the primary measurements and 51 for the secondary measurements (Figure 2).

Figure 2. Distribution of participants (dropout flow). K10: Kessler Psychological Distress Scale; SOFAS: Social and Occupational Functioning Assessment Scale.



Of the participants, 88.5% (116/131) were female with an average age of 21.61 (SD 2.2) years, and 85.5% (112/131) had completed higher education. A total of 43.5% (57/131) were categorized as stage 1a and 56.5% (74/131) were categorized as stage 1b. Participants' experiences of psychological distress ranged from "low" to "very high," and their psychosocial functioning ranged from "serious impairment" to "good functioning." Participants lived in different provinces in the Netherlands, with 89.3% (117/131) being Dutch or of Dutch and other ethnic and racial backgrounds (Table 1). Most found ENYOY via a recruitment platform (Link2Trials; 57/131, 43.5%), followed by social media (29/131, 22.1%), their general practitioner (6/131, 4.6%), family or friends (2/131, 1.5%), @ease center (1/131, 0.8%), or unknown (36/131, 27.5%). There were more female and relatively more highly educated participants (P<.001 in both cases). No differences in clinical stage 1a versus 1b were found (P=.16). A chi-square test indicated that there were no significant differences between dropouts and nondropouts on age (P=.62), educational level (P=.76), sex (P=.22), or clinical stage (P=.46). A 2-tailed independent-sample t test showed that there were no significant differences between dropouts and nondropouts in baseline K10 scores (P=.31), SOFAS scores (P=.96), or MPH scores (meaningfulness: P=.69; quality of life: P=.77; mental well-being: P=.34).

Table 1. Demographics of participants (N=131).

Variable	Values
Age (years), mean (SD)	21.63 (2.25)
Clinical stage, n (%)	
1a	57 (43.5)
1b	74 (56.5)
Psychological distress[a], n (%)	
Low[b]	4 (3.1)
Moderate	31 (23.7)
High	68 (51.9)
Very high	28 (21.4)
Psychosocial functioning[c], n (%)	
Serious impairment[d]	3 (2.3)
Moderate impairment	16 (12.2)
Some impairment	66 (50.4)
Slight impairment	43 (32.8)
Good functioning	3 (2.3)
Superior functioning	0 (0)
Sex, n (%)	
Female	116 (88.5)
Male	14 (10.7)
Intersex	1 (0.8)
Education[e], n (%)	
Primary education	2 (1.5)
Intermediate vocational education	17 (13)
Higher vocational education	69 (52.7)
University	43 (32.8)
Ethnicity, n (%)	
Dutch or Dutch and other[f]	117 (89.3)
Surinamese	6 (4.6)
Other[g]	3 (2.3)
Unknown	5 (3.8)

[a]Measured using the Kessler Psychological Distress Scale [74].
[b]Used cutoffs from the study by Slade et al [94].
[c]Measured using the Social and Occupational Functioning Assessment Scale [75].
[d]Used cutoffs from the study by Goldman et al [75].
[e]Highest level of education completed.
[f]Dutch and other ethnic and racial backgrounds.
[g]Ecuador and Colombia, Indonesia, and Iraq.

User Statistics

At baseline, participants (N=131) were "onboarded" on treatment journeys with a main focus (young people can complete several different pathways; only the primary ones they started with are listed in this paper) on anxiety (84/131, 64.1%), social anxiety (34/131, 26%), or depressive complaints (10/131, 7.6%). A self-compassion pathway was developed and implemented after 1 year when clinical moderators noticed that young people struggled with negative self-image, with 2.3% (3/131) following this path. On average, youth visited the platform for 18.75 (SD 22.62; range 1-107) days, opened 42.29

(SD 46.78; range 0-278) therapy exercises, and were active on the community with 4.06 (SD 9.23; range 0-86) messages (posting and commenting on posts).

Mean Changes in Outcomes Over Time

In the first analyses, we investigated whether platform use had an impact on psychological distress and daily functioning. This main analysis demonstrated a significant main effect of time on K10 scores (*P*<.001; η_p^2=0.239) and SOFAS scores (*P*<.001; η_p^2=0.318). These effects were not influenced by sex (*P*=.24 for psychological distress and *P*=.88 for psychosocial functioning), age (*P*=.76 and *P*=.48, respectively), community activity (*P*=.59 and *P*=.48, respectively), or therapy activity (*P*=.49 and *P*=.80, respectively).

Subsequently, we analyzed and compared the effects of time at different intervals (baseline and 3, 6, and 12 mo) post hoc (Table 2 and Figure 3). Pairwise comparisons with Bonferroni correction showed significant improvements in K10 scores between baseline and 3 months (*P*<.001; *d*=0.62) and between 3 and 6 months (*P*<.001; *d*=0.37). No significant differences were found between 6 and 12 months (*P*=.54; *d*=0.09). Pairwise comparisons showed significant improvements in SOFAS scores between baseline and 3 months (*P*<.001; *d*=0.50) and between 3 and 6 months (*P*<.001; *d*=0.50). No significant differences were found between 6 and 12 months (*P*=.21; *d*=0.15).

Table 2. Mean, SD, mean difference, *P* value, and effect sizes (Cohen d) of the outcome measures psychosocial functioning (Social and Occupational Functioning Assessment Scale [SOFAS]), psychological distress (Kessler Psychological Distress Scale [K10]), and positive health parameters (quality of life, meaningfulness, and mental well-being subscales).

Outcome measure and time point	Values, mean (SD)	Time point comparison	Mean difference	*P* value	Cohen *d*
SOFAS					
1[a]	68.68 (7.89)	N/A[b]	N/A	N/A	N/A
2[c]	72.97 (8.44)	2-1	4.30	<.001	0.50
3[d]	77.30 (8.23)	3-2	4.33	<.001	0.50
4[e]	78.58 (9.91)	4-3	1.29	.21	0.15
K10					
1	24.90 (6.18)	N/A	N/A	N/A	N/A
2	21.42 (4.97)	2-1	−3.48	<.001	0.62
3	19.32 (5.06)	3-2	−2.10	<.001	0.37
4	19.82 (6.35)	4-3	−0.49	.54	0.09
Quality of life					
1	5.83 (1.24)	N/A	N/A	N/A	N/A
2	6.26 (1.31)	2-1	0.43	.01	0.38
3	6.63 (1.32)	3-2	0.37	.02	0.28
4	6.81 (1.38)	4-3	0.18	.28	0.14
Meaningfulness					
1	6.39 (1.40)	N/A	N/A	N/A	N/A
2	6.80 (1.14)	2-1	0.41	.01	0.32
3	7.09 (1.39)	3-2	0.30	.02	0.23
4	7.31 (1.28)	4-3	0.21	.18	0.16
Mental well-being					
1	6.25 (1.16)	N/A	N/A	N/A	N/A
2	6.72 (1.10)	2-1	0.47	<.001	0.43
3	7.18 (1.04)	3-2	0.46	.002	0.42
4	7.17 (1.10)	4-3	0.01	.97	0.01

[a]Baseline.

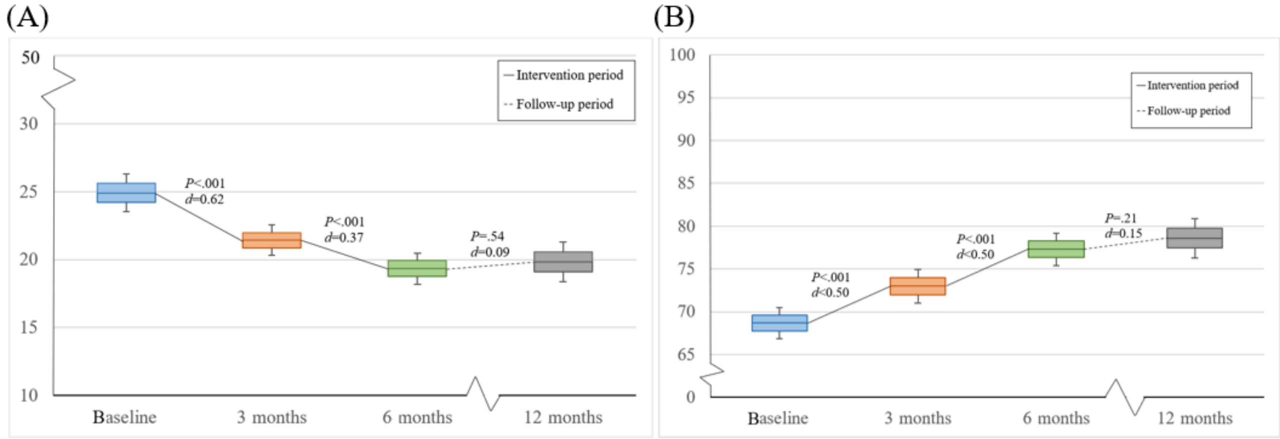
[b]N/A: not applicable.

[c]3-month measurement.

[d]6-month measurement.

[e]12-month follow-up.

Figure 3. Changes over time for (A) psychological distress (Kessler Psychological Distress Scale [K10]) and (B) psychosocial functioning (Social and Occupational Functioning Assessment Scale [SOFAS]). Ranges for the scales' total scores (minimum to maximum): 10 to 50 for the K10 and 0 to 100 for the SOFAS.

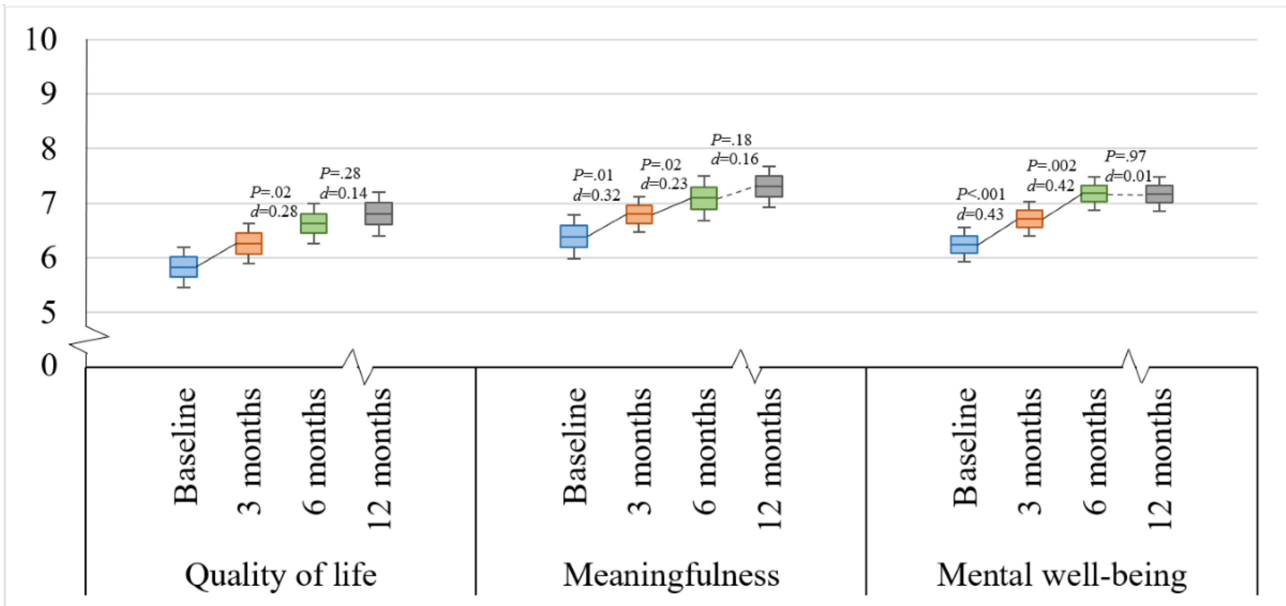


To investigate whether dropout influenced our results, we performed additional sensitivity analyses on SOFAS and K10 data using linear mixed models (50 imputations), and the results were comparable for all outcome measures (*P*<.001 in all cases; effect size ranges for K10 scores: η_p^2=0.217-0.355; effect size ranges for SOFAS scores: η_p^2=0.110 and 0.211).

For the secondary analysis of MPH data, we examined whether the platform had an impact on positive health parameters. Repeated measures ANOVAs demonstrated a significant main effect of time on mental well-being (*P*<.001; η_p^2=0.140), quality of life (*P*=.03; η_p^2=0.062), and meaningfulness (*P*<.001; η_p^2=0.121). When adding age (*P*=.44 for mental well-being, *P*=.97 for quality of life, and *P*=.84 for meaningfulness), sex (*P*=.74, *P*=.78, and *P*=.56, respectively), therapy activity (*P*=.80, *P*=.36, and *P*=.46, respectively), and community activity (*P*=.58, *P*=.76, and *P*=.97, respectively) to the model, the effects remained essentially unchanged.

Subsequently, we conducted a post hoc pairwise comparison analysis to examine and compare the effects of time at different intervals (baseline and 3-, 6-, and 12-month measurements; Table 2 and Figure 4). The analysis showed significant improvements in mental well-being, quality of life, and meaningfulness between baseline and 3 months (*P*<.001 and *d*=0.43; *P*=.01 and *d*=0.38; and *P*=.01 and *d*=0.32, respectively) and between 3 and 6 months (*P*=.002 and *d*=0.42; *P*=.02 and *d*=0.28; and *P*=.02 and *d*=0.23, respectively). No differences were found in mental well-being, quality of life, or meaningfulness between 6 and 12 months (*P*=.97 and *d*=0.01; *P*=.28 and *d*=0.14; and *P*=.18 and *d*=0.16, respectively).

Figure 4. My Positive Health scores over time. Ranges for the subscales' total scores (minimum to maximum): 0 to 10.



Reliable and Clinically Significant Change

For the percentages of young people showing reliable and clinically significant changes, refer to Table 3.

Table 3. Proportion of young people showing reliable and clinically significant changes in psychological distress and psychosocial functioning between the baseline and 6-month measurements and between the baseline and 12-month measurements (N=131).

Measure, method, and clinical stage	Change category at 6 months (n=102), n (%)				Change category at 12 months (n=82), n (%)			
	Participants	Improvement	No change	Worsening	Participants	Improvement	No change	Worsening
K10[a]								
RCI[b]								
1a	50 (49)	19 (18.6)	28 (27.5)	3 (3)	36 (43.9)	19 (23.2)	17 (20.7)	0 (0)
1b	52 (51)	20 (19.6)	29 (28.4)	3 (3)	46 (56.1)	18 (22)	20 (24.4)	8 (9.8)
CSI[c]								
1a and 1b	102 (100)	77 (75.5)	25 (24.5)	N/A[d]	82 (100)	63 (76.8)	19 (23.2)	N/A
SOFAS[e]								
RCI								
1a	48 (47.1)	26 (25.5)	19 (18.6)	3 (3)	36 (43.9)	22 (26.8)	12 (14.6)	2 (2.4)
1b	54 (52.9)	25 (24.5)	25 (24.5)	4 (3.9)	46 (56.1)	17 (20.7)	26 (31.7)	3 (3.7)
CSI								
1a and 1b	102 (100)	91 (89.2)	11 (10.8)	N/A	82 (100)	75 (91.5)	7 (8.5)	N/A

[a]K10: Kessler Psychological Distress Scale.

[b]RCI: Reliable Change Index.

[c]CSI: Clinically Significant Index.

[d]N/A: not applicable; participants in the clinical population stay in the clinical category.

[e]SOFAS: Social and Occupational Functioning Assessment Scale.

Comparing the 6-month measurement to baseline with regard to psychological distress, participants in stage 1a reliably improved in 38% (19/50) of cases, did not change in 56% (28/50) of cases, and reliably worsened in 6% (3/50) of cases. Participants in stage 1b reliably improved in 38% (20/52) of cases, showed no change in 56% (29/52) of cases, and reliably worsened in 6% (3/52) of cases. Overall, 75.5% (77/102) of cases showed clinically significant changes, and 24.5% (25/102) of cases showed no change. Regarding psychosocial functioning, participants in stage 1a reliably improved in 54% (26/48) of cases, did not change in 40% (19/48) of cases, and reliably worsened in 6% (3/48) of cases. Participants in stage 1b improved in 46% (25/54) of cases, showed no change in 46% (25/54) of cases, and reliably worsened in 7% (4/54) of cases. Overall, 89.2% (91/102) of cases showed clinically significant changes, and 10.8% (11/102) of cases showed no change.

Comparing the 12-month measurement to baseline with regard to psychological distress, participants in stage 1a reliably improved in 53% (19/36) of cases and did not change in 47% (17/36) of cases, and no cases reliably worsened. Participants in stage 1b reliably improved in 39% (18/46) of cases, showed no change in 43% (20/46) of cases, and reliably worsened in 17% (8/46) of cases. Overall, 77% (63/82) of cases showed clinically significant changes, and 23% (19/82) of cases showed no change. Regarding psychosocial functioning, participants in stage 1a reliably improved in 61% (22/36) of cases, did not change in 33% (12/36) of cases, and reliably worsened in 6% (2/36) of cases. Participants in stage 1b improved in 37% (17/46) of cases, showed no change in 57% (26/46) of cases, and reliably worsened in 7% (3/46) of cases. Overall, 91% (75/82) of cases showed clinically significant changes, and 9% (7/82) of cases showed no change.

Predictors for Improvement

Improvement in both psychological distress and social functioning was predicted at the 6-month measurement by higher psychological distress ($P<.001$) and lower social functioning ($P=.04$) at baseline, respectively. This means that both severity of mental health complaints and impairment of social functioning predicted greater improvement in young people. In addition, age ($P=.24$), sex ($P=.89$), educational level ($P=.19$), and clinical stage ($P>.99$) did not predict improvement.

In an additional set of exploratory analyses, we investigated the potential protective effects in relation to the progression of symptoms between 6 and 12 months. Logistic regression analyses showed that higher levels of meaningfulness (odds ratio 3.59, 95% CI 1.22-10.53) but not quality of life at 6 months (after the intervention) were associated with a lower risk of complaints worsening at the 12-month follow-up.

Discussion

Principal Findings

The aim of this study was to investigate the outcomes of the ENYOY-platform in relation to the following parameters: psychological distress, psychosocial functioning, and positive health parameters (such as well-being and quality of life [24]). Our hypothesis was that ENYOY would attain similar results of improvement in psychological distress and psychosocial functioning as have been found for the young people visiting the Headspace centers in Australia [49], as well as a positive change in positive health parameters. It was expected that spending more time on the platform and being more active in the digital community would be associated with increased improvements.

The main analysis revealed significant and substantial improvements in psychological distress and psychosocial functioning over time (large effect sizes) independent of sex, age, therapy activity, or community activity (negligible to small effect sizes). This contradicts the initial hypothesis that spending more time on the platform or being more active in the digital community would lead to greater improvements. Regardless of use frequency, the platform consistently had large effects among young people. In addition, the secondary analysis showed significant improvements in well-being, quality of life, and meaningfulness (large, medium, and large effect sizes, respectively) independent of sex, age, therapy activity, or community activity (negligible to small effect sizes). Notably, significant improvements in all outcome measures occurred within the first 3 months (moderate effect sizes for psychological distress and psychosocial functioning, small to moderate effect sizes for mental well-being and quality of life, and small effect sizes for meaningfulness), with further significant increases in the subsequent 3 months (moderate effect sizes for psychosocial functioning, small to moderate effect sizes for psychological distress and mental well-being, and small effect sizes for quality of life and meaningfulness), and remained stable after active platform use (ie, 12-month follow-up; very small to negligible effect sizes). These findings highlight the platform's effects in decreasing psychological distress, improving psychosocial functioning, and improving positive health parameters independent of individual characteristics or engagement levels. Moreover, there were no age effects, which implies that a broad age range (ages of 16-25 y) could benefit from this intervention. The sustained positive outcomes over time underscore the potential long-term benefits of using this intervention. This is also in line with a previous qualitative study that showed that young people perceived the ENYOY-platform to be a user-friendly, safe, accessible, and inclusive initiative that helped alleviate their mental health complaints and enhanced their overall quality of life [67].

Interestingly, in line with previous research [68-70], we initially anticipated similar medium effect sizes. However, the observed effect sizes in our study turned out to be even higher. A possible explanation for this difference is the longer engagement period of Dutch young people on the platform compared with the Australian studies (6 vs 3 mo). The extended duration likely contributed to the enhanced effects, as evidenced by the significant improvements over additional months. Another factor could be the homogeneity of our participant group, with a majority of female and higher-educated individuals. In contrast, user motivation—measured through platform visits—was not notably higher than that in previous studies [66], and the amount of community and therapy activity exhibited a wide variation among participants. Discrepancies with Australian studies [49,66] could also stem from differences in concern profiles or severity. The Australian studies did not select participants based on concern severity, possibly including individuals with a higher severity of complaints (clinical stage ≥2). Furthermore, the methodology of the study by Rickwood et al [49], which analyzed K10 and SOFAS scores at specific time points, makes direct comparisons challenging as it involved physical center visits rather than a web-based platform.

Of interest was the unexpected finding that the effects of the platform were independent of platform use. This could potentially be attributed to the platform's unique composition, consisting of 4 main modules: therapy exercises, conversations with peer workers, sessions with psychologists, and a community for peer-to-peer support. Guided by the principles of self-determination theory [95,96], we empowered young individuals to determine their own use patterns for each module, tailoring their engagement based on their specific needs. This individualized approach enabled young people to create their own optimal use of the platform's modules, rendering the measurement of therapy or community activity alone redundant for capturing the platform's overall impact.

Individual Change

In line with the main analysis, clinical and reliable changes were observed in a large portion of the sample at the 6-month follow-up. Regarding psychological distress, 38% (19/50) of young people in stage 1a and 39% (18/46) of young people in stage 1b reliably improved. Similarly, for psychosocial functioning, 54% (26/48) of young people in stage 1a and 46% (25/54) of young people in stage 1b showed reliable improvement. The percentage of individuals who experienced a worsening of complaints (3/50, 6%) and functioning (4/54, 7%) was minimal. A substantial majority of cases (77/102, 75.5% for psychological distress and 91/102, 89.2% for psychosocial functioning) no longer met the criteria for clinical levels of complaints. An interesting finding was that approximately one-sixth of individuals in clinical stage 1b (8/46, 17%) experienced increased psychological distress at the follow-up, whereas no one in stage 1a deteriorated. This aligns with previous research indicating a clear distinction between the 2 stages [25-27]. Individuals in the latter stage may require continued care. In Australia, young people are initially provided with an "active" intervention period of 3 months during which clinicians reach out to them weekly. After this phase, the intervention shifts to a more passive approach, allowing young people to use the platform and seek help as needed. This approach shows promise for the 1b group, which is at a higher risk of deterioration, by ensuring ongoing support while promoting independence. Despite increased psychological distress, individuals in stage 1b maintained their social and occupational functioning, which may indicate that they are more accepting of their mental health symptoms.

Explorative Analysis

It is revealed that a greater severity of mental health complaints and impairment of social functioning predicted greater reliable change in young individuals. This finding may be attributed to the presence of ceiling effects for social functioning and floor effects for mental health complaints, which are more likely to occur among young people with milder mental health complaints

[97,98]. In other words, young individuals with higher initial levels of mental health complaints have more room for improvement, leading to greater positive changes over time. This finding challenges the previous assertion that the lower effect size of studies in Australia could be due to the inclusion of young people with higher-severity complaints. Although we can only speculate, this might suggest that the platform may also have positive effects on young people with more severe mental health complaints (beyond clinical stage 1b). Explorative analyses further revealed that, for individuals in stage 1b, higher levels of meaningfulness at 6 months (after the intervention) predicted a lower change in worsening at 12 months (follow-up), which could indicate that meaningfulness is a protective factor. For future interventions, this could be taken into account, for example, by providing the subgroup in stage 1b with low levels of meaningfulness through more care or continuity of care or offering additional interventions aimed at increasing meaningfulness.

Despite exploring numerous variables, no specific indications were found regarding any particular variable that may have contributed to an increased likelihood of dropout. Previous studies have identified several predictors of dropout, such as psychosocial difficulties, symptom severity [99], being young, being male, the involvement of more than one therapist in treatment, having no history of psychiatric disorders [100], lower social support [101], and lower change motivation [102]. A comprehensive review of multiple studies on dropout characteristics concluded that predictors may vary across studies and samples and over time [103], making it challenging to draw conclusions in general.

Strengths and Limitations

A notable limitation of this study is the identified sample bias, which favored White female individuals with higher education. Young people with a different ethnic background represented only 6.9% (9/131) compared with the 30% among young people in the Netherlands [104], an explanation being the platform's sole availability in Dutch, thereby excluding individuals who speak other languages. Similarly, lower education was underrepresented compared with the equal distribution in the Netherlands [105]. Although different recruitment measures were implemented to obtain a diverse sample [24], certain young people were not reached. It is worth noting that similar studies have also found a bias toward female individuals with higher education [64]. Consequently, caution should be exercised when interpreting the results as they may only reflect an effect among this specific group of young people. No significant interaction effects of sex were observed, suggesting that the effects may extend to male individuals as well. However, to draw conclusive statements about male participants, a larger proportion of male individuals in the sample would be required. To comprehensively assess the intervention's effects, future research should incorporate more diverse samples [106]. In addition, questions might arise about the platform's suitability for a wider age range than 16 to 25 years. Notably, young people significantly influenced the ENYOY design across all phases, ensuring its relevance to the target population. Although our initial intention was to encompass young individuals aged 12 to 25 years [24], a decision was made to pilot the platform with the age range of 16 to 25 years based on recommendations from the institutional review board. Furthermore, the absence of a statistically significant age effect on the outcome measures suggests that age did not significantly influence the results. Therefore, it is expected that the platform accommodated a wide range of developmental ages, including 16 to 25 years, to a significant extent. It is important to note that accommodating participants aged <16 years in future studies would require platform adjustment [24,106].

Finally, it is important to acknowledge that this study was not an RCT, resulting in a lower level of evidence [107]. The extent to which the observed effects are attributable to the intervention or other factors remains unclear. It is essential to consider that these effects may be a result of other interventions. We adopted a participatory within-group design for ongoing platform enhancement based on user feedback, recognizing that starting with an RCT in implementation research could oversimplify real-world patient diversity, become time-consuming, and have strong internal but limited external validity [108]. Rapid technological advancements in web-based interventions could also risk making RCT results obsolete upon publication [109]. Although we acknowledge the importance of a more robust RCT design in the future, including a larger number of participants and a control group after the broader implementation of the ENYOY, we opted for flexibility to adapt quickly to user feedback and capture the complexities of real-world use. This choice involved a trade-off, potentially affecting the sample's representativeness, but allowed us to balance scientific rigor with accommodating the diverse and evolving needs of our user base [109]. Nonetheless, the observed improvements during platform use and their sustained stability after 6 months, combined with the consistent effectiveness found in >13 years of research (including RCTs) in Australia [65,66,68-72], strongly suggest the potential benefits of using the platform. This study has established a foundation, and the subsequent step is to validate these findings through an RCT to ensure a higher level of evidence.

A strength of this study was its high retention rate and statistical power. Over the course of a year, the study had a dropout rate of only 62.6% (82/131), and the missing data accounted for only 22.54%. These figures surpass those found in most comparable research studies, which typically report dropout rates of up to 73% [64]. Group-level analysis enabled the examination of overall trends and patterns, whereas individual-level analysis provided valuable insights into each participant's response [110]. This comprehensive approach could enhance decision-making and facilitate the future customization of the intervention to address the diverse needs of young people. Finally, the repeated measures design offered valuable insights into the trajectory of improvements over time and the sustainability of the effects following the intervention.

Recommendations

This trial has demonstrated that the ENYOY-platform holds promise as an intervention for addressing emerging mental health complaints among young people in the Netherlands. By addressing these complaints, the platform has the potential to reduce the burden of mental illness and associated costs among youth [111-113], potentially alleviating long waiting lists for mental health care, and offers a solution to bridge the gap between child, adolescent, and adult psychiatry [15,114,115]. To reach a wider audience, the platform could expand its language options and incorporate features for individuals with low literacy levels. For future projects, the inclusion of external informants (such as family and friends) should be considered to explore the potential benefits of involving them in the system, which could contribute to treatment customization, offer a holistic view of individual complaints, provide context, and possibly enhance therapy outcomes [106]. This trial represents a significant milestone in the implementation of a transdiagnostic, digital, and clinically and peer-moderated indicative prevention treatment platform for youth with emerging mental health complaints in the Netherlands [24]. Furthermore, the success of this trial opens up possibilities for the platform's implementation in other countries as well.

Future Research

This study has laid the groundwork for enhanced clinical research in the Netherlands using the ENYOY-platform. The results provide support for progressing to an RCT. In addition, it would be relevant to expand access to the platform to other populations. For example, in Australia, young people with mild to more severe problems have been granted access to the platform via youth mental health services to enhance face-to-face care during waiting periods, supplement face-to-face treatment, and support follow-up care (eg, relapse prevention) [66,71]. To inform decision-making regarding the criteria for offering ongoing care, it is important to calculate cutoff scores for meaningfulness for stage 1b individuals. This could help identify those who would benefit from ongoing support, thus enabling targeted interventions. In addition, future research could further investigate which factors predict dropout to develop more targeted interventions and support systems that effectively address the underlying causes and mitigate the risk of dropout. A recent study in Australia demonstrated the cost-effectiveness of the platform, which could be repeated for the Dutch situation [116]. Finally, as the most vulnerable period for developing mental health complaints is between the ages of 12 and 25 years [117,118], there is a need to include individuals aged 12 to 15 years in the trials (on the same platform or a customization of the current platform that better matches this age group).

Acknowledgments

Funding was provided by the Netherlands Organisation for Health Research and Development for personnel, materials, and implementation (file 60–63600–98–319). The authors would like to thank Mees Egeler, Laurens Nijhuis, Ria Vranken Lorenzo, Chantal Angeneind, Linde Oussoren, Annika Meuche, Lisa Cappendijk, Pien Verhoeven, Senna Dik, Bram Dijkstra, Sarah Dierckx, Anne Monsanto, Cato Boeschoten, Rosalie Horst, Maud Goddeeris, Zoë Zuilhof, Saartu Musa, Emma Boers, Helder van der Hoop, Esmée Vesseur, Alanur Kursun, Melissa Dorrestein, Hannah Stewart, Raven Barendregt, Chen Lu Wang, Gabrielle Oosterhoff, Tim Altorf, Paul Gelissen, Noortje Rotterdam, Willemijn Leenheer, Lauren van Stratum, Frederiek Brummelkamp, Benjamin Skalonjic, Jona Cohen, Wesly van Dijen, Tamara Schouten, Wouter Moolhuizen, Mayra Kapteijn, Flore van Doornik-Joskin, Arianne Westhuis, Margit Mulder, Nelly Blessinga, Claire Veldkamp, Jane Haskings, Tamsyn Gilbertson, Stefan von Cavallar, Simon Baker, Carla McEnery, Christopher Miles, Skey Mclean, and Simon d'Alfonso for their help with the execution of this study.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MvD, AP, TAMJvA, DHN, and MA-J completed the initial study design. MWMJ, DHN, TAMJvA, AP, JFG, FÖ, and MA-J provided expert assessment and feedback. MvD, AM, SCJV, and CLW performed the data analysis and wrote the manuscript. All authors contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Mentale gezondheid in eerste helft 2021 op dieptepunt. Centraal Bureau voor de Statistiek. 2021 Sep 3. URL: https://www.cbs.nl/nl-nl/nieuws/2021/35/mentale-gezondheid-in-eerste-helft-2021-op-dieptepunt [accessed 2021-09-10]
2. Rivera-Riquelme M, Piqueras JA, Cuijpers P. The Revised Mental Health Inventory-5 (MHI-5) as an ultra-brief screening measure of bidimensional mental health in children and adolescents. Psychiatry Res 2019 Apr;274:247-253. [doi: 10.1016/j.psychres.2019.02.045] [Medline: 30818147]
3. COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. World Health Organization. 2022 Mar 2. URL: https://www.who.int/news/item/

02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide [accessed 2023-09-09]

4. Doran CM, Kinchin I. A review of the economic impact of mental illness. Aust Health Rev 2019 Feb;43(1):43-48. [doi: 10.1071/AH16115] [Medline: 29129189]
5. Ormel J, Raven D, van Oort F, Hartman CA, Reijneveld SA, Veenstra R, et al. Mental health in Dutch adolescents: a TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. Psychol Med 2015 Jan;45(2):345-360. [doi: 10.1017/S0033291714001469] [Medline: 25066533]
6. Mojtabai R, Stuart EA, Hwang I, Eaton WW, Sampson N, Kessler RC. Long-term effects of mental disorders on educational attainment in the national comorbidity survey ten-year follow-up. Soc Psychiatry Psychiatr Epidemiol 2015 Oct 17;50(10):1577-1591 [FREE Full text] [doi: 10.1007/s00127-015-1083-5] [Medline: 26082040]
7. Curtis AC. Defining adolescence. J Adolesc Fam Health 2015;7(2) [FREE Full text]
8. Seabury SA, Axeen S, Pauley G, Tysinger B, Schlosser D, Hernandez JB, et al. Measuring the lifetime costs of serious mental illness and the mitigating effects of educational attainment. Health Aff (Millwood) 2019 Apr;38(4):652-659 [FREE Full text] [doi: 10.1377/hlthaff.2018.05246] [Medline: 30933598]
9. Janssen EM, McGinty EE, Azrin ST, Juliano-Bult D, Daumit GL. Review of the evidence: prevalence of medical conditions in the United States population with serious mental illness. Gen Hosp Psychiatry 2015 May;37(3):199-222 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.03.004] [Medline: 25881768]
10. Gustavson K, Knudsen AK, Nesvåg R, Knudsen GP, Vollset SE, Reichborn-Kjennerud T. Prevalence and stability of mental disorders among young adults: findings from a longitudinal study. BMC Psychiatry 2018 Mar 12;18(1):65 [FREE Full text] [doi: 10.1186/s12888-018-1647-5] [Medline: 29530018]
11. Suhrcke M, Pillas D, Selai C. Economic aspects of mental health in children and adolescents. In: Social Cohesion for Mental Wellbeing Among Adolescents. Geneva, Switzerland: World Health Organization; 2008:43-64.
12. Magnée T, de Beurs D, Verhaak P. Wachttijden voor de poh-ggz. Huisarts Wet 2017 May 13;60:205. [doi: 10.1007/s12445-017-0131-7]
13. Huisman C. In de jeugdzorg moeten juist de jongeren met de grootste problemen te lang wachten. Wat gaat er mis? de Volkskrant. 2021 Apr 26. URL: https://www.volkskrant.nl/nieuws-achtergrond/in-de-jeugdzorg-moeten-juist-de-jongeren-met-de-grootste-problemen-te-lang-wachten-wat-gaat-er-mis~b66a7681/ [accessed 2022-06-03]
14. Onvoldoende tijdige en juiste hulp voor jongeren met ernstige psychische problemen. Inspectie Gezondheidszorg en Jeugd. URL: https://www.igj.nl/publicaties/publicaties/2021/03/15/factsheet-onvoldoende-tijdige-en-juiste-hulp-voor-jongeren-met-ernstige-psychische-problemen [accessed 2020-03-19]
15. Oosterom R. Ggz-instellingen: lange wachtlijsten in de geestelijke gezondheidszorg zijn onoplosbaar. Trouw. 2021 Nov 6. URL: https://www.trouw.nl/zorg/ggz-instellingen-lange-wachtlijsten-in-de-geestelijke-gezondheidszorg-zijn-onoplosbaar~b27989a3/?referrer=https://mhealth.jmir.org/ [accessed 2022-04-19]
16. Nieman DH, McGorry PD. Detection and treatment of at-risk mental state for developing a first psychosis: making up the balance. Lancet Psychiatry 2015 Sep;2(9):825-834. [doi: 10.1016/S2215-0366(15)00221-7] [Medline: 26360901]
17. Kosten van ziekten 2019. Rijksinstituut voor Volksgezondheid en Milieu (RIVM). 2022 Aug 10. URL: https://statline.rivm.nl/#/RIVM/nl/dataset/50091NED/table?ts=1679053604984 [accessed 2023-03-17]
18. Fox S, Southwell A, Stafford N, Goodhue R, Jackson D, Smith C. Better systems, better chances: a review of research and practice for prevention and early intervention. Australian Research Alliance for Children and Youth (ARACY). 2015. URL: https://www.aracy.org.au/publications-resources/command/download_file/id/274/filename/Better-systems-better-chances.pdf [accessed 2023-11-28]
19. Shapiro MD, Maron DJ, Morris PB, Kosiborod M, Sandesara PB, Virani SS, et al. Preventive cardiology as a subspecialty of cardiovascular medicine: JACC council perspectives. J Am Coll Cardiol 2019 Oct 15;74(15):1926-1942 [FREE Full text] [doi: 10.1016/j.jacc.2019.08.1016] [Medline: 31601373]
20. Broberg AM, Geisler J, Tuohinen S, Skytta T, Hrafnkelsdóttir Þ, Nielsen KM, et al. Prevention, detection, and management of heart failure in patients treated for breast cancer. Curr Heart Fail Rep 2020 Dec;17(6):397-408 [FREE Full text] [doi: 10.1007/s11897-020-00486-8] [Medline: 32979150]
21. Hickie IB, Scott EM, Hermens DF, Naismith SL, Guastella AJ, Kaur M, et al. Applying clinical staging to young people who present for mental health care. Early Interv Psychiatry 2013 Feb;7(1):31-43. [doi: 10.1111/j.1751-7893.2012.00366.x] [Medline: 22672533]
22. McGorry PD, Hickie IB, Yung AR, Pantelis C, Jackson HJ. Clinical staging of psychiatric disorders: a heuristic framework for choosing earlier, safer and more effective interventions. Aust N Z J Psychiatry 2006 Aug;40(8):616-622. [doi: 10.1080/j.1440-1614.2006.01860.x] [Medline: 16866756]
23. McGorry PD, Hickie IB. Clinical Staging in Psychiatry Making Diagnosis Work for Research and Treatment. Cambridge, UK: Cambridge University Press; 2019.
24. van Doorn M, Popma A, van Amelsvoort T, McEnery C, Gleeson JF, Ory FG, et al. ENgage YOung people earlY (ENYOY): a mixed-method study design for a digital transdiagnostic clinical - and peer- moderated treatment platform for youth with

beginning mental health complaints in the Netherlands. BMC Psychiatry 2021 Jul 23;21(1):368 [FREE Full text] [doi: 10.1186/s12888-021-03315-x] [Medline: 34301213]
25. Cross SP, Hermens DF, Hickie IB. Treatment patterns and short-term outcomes in an early intervention youth mental health service. Early Interv Psychiatry 2016 Feb;10(1):88-97. [doi: 10.1111/eip.12191] [Medline: 25263607]
26. Cross SP, Hermens DF, Scott EM, Ottavio A, McGorry PD, Hickie IB. A clinical staging model for early intervention youth mental health services. Psychiatr Serv 2014 Jul;65(7):939-943. [doi: 10.1176/appi.ps.201300221] [Medline: 24828746]
27. Hickie IB, Scott EM, Cross SP, Iorfino F, Davenport TA, Guastella AJ, et al. Right care, first time: a highly personalised and measurement-based care model to manage youth mental health. Med J Aust 2019 Nov;211 Suppl 9(S9):S3-46. [doi: 10.5694/mja2.50383] [Medline: 31679171]
28. Shah JL, Scott J, McGorry PD, Cross SP, Keshavan MS, Nelson B, et al. Transdiagnostic clinical staging in youth mental health: a first international consensus statement. World Psychiatry 2020 Jun;19(2):233-242 [FREE Full text] [doi: 10.1002/wps.20745] [Medline: 32394576]
29. Iorfino F, Scott EM, Carpenter JS, Cross SP, Hermens DF, Killedar M, et al. Clinical stage transitions in persons aged 12 to 25 years presenting to early intervention mental health services with anxiety, mood, and psychotic disorders. JAMA Psychiatry 2019 Nov 01;76(11):1167-1175 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.2360] [Medline: 31461129]
30. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to mental health problems: current status and future directions. J Consult Clin Psychol 2020 Mar;88(3):179-195 [FREE Full text] [doi: 10.1037/ccp0000482] [Medline: 32068421]
31. Roefs A, Fried EI, Kindt M, Martijn C, Elzinga B, Evers AW, et al. A new science of mental disorders: using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. Behav Res Ther 2022 Jun;153:104096 [FREE Full text] [doi: 10.1016/j.brat.2022.104096] [Medline: 35500541]
32. Ehrenreich-May J, Kennedy SM, Sherman JA, Bilek EL, Buzzella BA, Bennett SM, et al. Unified Protocols for Transdiagnostic Treatment of Emotional Disorders in Children and Adolescents: Therapist Guide. Oxford, UK: Oxford University Press; Feb 22, 2018.
33. Sherman JA, Ehrenreich-May J. Changes in risk factors during the unified protocol for transdiagnostic treatment of emotional disorders in adolescents. Behav Ther 2020 Nov;51(6):869-881. [doi: 10.1016/j.beth.2019.12.002] [Medline: 33051030]
34. Black M, Hitchcock C, Bevan A, O Leary C, Clarke J, Elliott R, et al. The HARMONIC trial: study protocol for a randomised controlled feasibility trial of shaping healthy minds-a modular transdiagnostic intervention for mood, stressor-related and anxiety disorders in adults. BMJ Open 2018 Aug 05;8(8):e024546 [FREE Full text] [doi: 10.1136/bmjopen-2018-024546] [Medline: 30082367]
35. Brown TA, Barlow DH. A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: implications for assessment and treatment. Psychol Assess 2009 Sep;21(3):256-271 [FREE Full text] [doi: 10.1037/a0016608] [Medline: 19719339]
36. Harvey AG, Watkins E, Mansell W. Cognitive Behavioural Processes Across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment. Oxford, UK: Oxford University Press; 2004.
37. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017 Sep 1;74(9):875-884. [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
38. Barlow DH, Farchione TJ, Fairholme CP, Ellard KK, Boisseau CL, Allen LB, et al. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide. Oxford, UK: Oxford University Press; Dec 2010.
39. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
40. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
41. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
42. Andersen P, Toner P, Bland M, McMillan D. Effectiveness of transdiagnostic cognitive behaviour therapy for anxiety and depression in adults: a systematic review and meta-analysis. Behav Cogn Psychother 2016 Nov;44(6):673-690. [doi: 10.1017/S1352465816000229] [Medline: 27301967]
43. García-Escalera J, Valiente RM, Sandín B, Ehrenreich-May J, Prieto A, Chorot P. The unified protocol for transdiagnostic treatment of emotional disorders in adolescents (UP-A) adapted as a school-based anxiety and depression prevention program: an initial cluster randomized wait-list-controlled trial. Behav Ther 2020 May;51(3):461-473. [doi: 10.1016/j.beth.2019.08.003] [Medline: 32402261]

44. Sandín B, García-Escalera J, Valiente RM, Espinosa V, Chorot P. Clinical utility of an internet-delivered version of the unified protocol for transdiagnostic treatment of emotional disorders in adolescents (iUP-A): a pilot open trial. Int J Environ Res Public Health 2020 Nov 10;17(22):8306 [FREE Full text] [doi: 10.3390/ijerph17228306] [Medline: 33182711]
45. Schmitt JC, Valiente RM, García-Escalera J, Arnáez S, Espinosa V, Sandín B, et al. Prevention of depression and anxiety in subclinical adolescents: effects of a transdiagnostic internet-delivered CBT program. Int J Environ Res Public Health 2022 Apr 28;19(9):5365 [FREE Full text] [doi: 10.3390/ijerph19095365] [Medline: 35564759]
46. Păsărelu CR, Dobrean A, Andersson G, Zaharie GC. Feasibility and clinical utility of a transdiagnostic Internet-delivered rational emotive and behavioral intervention for adolescents with anxiety and depressive disorders. Internet Interv 2021 Nov 2;26:100479 [FREE Full text] [doi: 10.1016/j.invent.2021.100479] [Medline: 34804812]
47. McGorry P, Trethowan J, Rickwood D. Creating headspace for integrated youth mental health care. World Psychiatry 2019 Jun;18(2):140-141 [FREE Full text] [doi: 10.1002/wps.20619] [Medline: 31059618]
48. Rickwood DJ, Telford NR, Mazzer KR, Parker AG, Tanti CJ, McGorry PD. The services provided to young people through the headspace centres across Australia. Med J Aust 2015 Jun 01;202(10):533-536. [doi: 10.5694/mja14.01695] [Medline: 26021365]
49. Rickwood DJ, Mazzer KR, Telford NR, Parker AG, Tanti CJ, McGorry PD. Changes in psychological distress and psychosocial functioning in young people visiting headspace centres for mental health problems. Med J Aust 2015 Jun 01;202(10):537-542. [doi: 10.5694/mja14.01696] [Medline: 26021366]
50. Singh SP, Tuomainen H. Transition from child to adult mental health services: needs, barriers, experiences and new models of care. World Psychiatry 2015 Oct;14(3):358-361 [FREE Full text] [doi: 10.1002/wps.20266] [Medline: 26407794]
51. Amelsvoort van TA. Transitiepsychiatrie: bridging the gap. Maastricht University. 2013 Jan 18. URL: https://doi.org/10.26481/spe.20130118ta [accessed 2023-11-28]
52. McGorry PD, Mei C, Chanen A, Hodges C, Alvarez-Jimenez M, Killackey E. Designing and scaling up integrated youth mental health care. World Psychiatry 2022 Feb;21(1):61-76 [FREE Full text] [doi: 10.1002/wps.20938] [Medline: 35015367]
53. Alvarez-Jimenez M, Gleeson JF, Rice S, Gonzalez-Blanch C, Bendall S. Online peer-to-peer support in youth mental health: seizing the opportunity. Epidemiol Psychiatr Sci 2016 Apr;25(2):123-126 [FREE Full text] [doi: 10.1017/S2045796015001092] [Medline: 26880239]
54. Mitchell C, McMillan B, Hagan T. Mental health help-seeking behaviours in young adults. Br J Gen Pract 2017 Jan;67(654):8-9 [FREE Full text] [doi: 10.3399/bjgp17X688453] [Medline: 28034926]
55. Aguirre Velasco A, Cruz IS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
56. Bandelow B, Schüller K. Mean age and gender distribution of patients with major mental disorders participating in clinical trials. Eur Arch Psychiatry Clin Neurosci 2020 Sep;270(6):655-659. [doi: 10.1007/s00406-018-0972-5] [Medline: 30600352]
57. Seidler ZE, Rice SM, Ogrodniczuk JS, Oliffe JL, Shaw JM, Dhillon HM. Men, masculinities, depression: implications for mental health services from a Delphi expert consensus study. Prof Psychol Res Pract 2019;50(1):51-61. [doi: 10.1037/pro0000220]
58. Stagnaro JC, Cia AH, Vommaro H, Sustas S, Vázquez N, Serfaty E, et al. Delays in making initial treatment contact after the first onset of mental health disorders in the Argentinean study of mental health epidemiology. Epidemiol Psychiatr Sci 2019 Apr;28(2):240-250 [FREE Full text] [doi: 10.1017/S2045796018000094] [Medline: 29540248]
59. Kazdin AE. Annual research review: expanding mental health services through novel models of intervention delivery. J Child Psychol Psychiatry 2019 Apr;60(4):455-472. [doi: 10.1111/jcpp.12937] [Medline: 29900543]
60. Leijdesdorff SM, Rosema S, Klaassen RM, Popma A, van Amelsvoort T. Who is @ease? Visitors' characteristics and working method of professionally supported peer-to-peer youth walk-in centres, anonymous and free of charge. Early Interv Psychiatry 2022 Dec;16(12):1391-1397 [FREE Full text] [doi: 10.1111/eip.13294] [Medline: 35343056]
61. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. Br J Gen Pract 2016 Oct;66(651):e686-e692 [FREE Full text] [doi: 10.3399/bjgp16X687313] [Medline: 27688518]
62. Yap MB, Reavley N, Jorm AF. Where would young people seek help for mental disorders and what stops them? Findings from an Australian national survey. J Affect Disord 2013 May;147(1-3):255-261. [doi: 10.1016/j.jad.2012.11.014] [Medline: 23228570]
63. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
64. van Doorn M, Nijhuis LA, Egeler MD, Daams JG, Popma A, van Amelsvoort T, et al. Online indicated preventive mental health interventions for youth: a scoping review. Front Psychiatry 2021 Apr 29;12:580843 [FREE Full text] [doi: 10.3389/fpsyt.2021.580843] [Medline: 33995136]
65. Alvarez-Jimenez M, Gleeson JF, Bendall S, Penn DL, Yung AR, Ryan RM, et al. Enhancing social functioning in young people at ultra high risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]

66. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (moderated online social therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-mental health service. J Med Internet Res 2020 Aug 13;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
67. van Doorn M, Monsanto A, Boeschoten CM, van Amelsvoort T, Popma A, Öry FG, et al. Moderated digital social therapy for young people with emerging mental health problems: a user-centered mixed-method design and usability study. Front Digit Health 2022 Jan 9;4:1020753 [FREE Full text] [doi: 10.3389/fdgth.2022.1020753] [Medline: 36698649]
68. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey CG, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020 Apr 03;17(7):2435 [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
69. Rice S, O'Bree B, Wilson M, McEnery C, Lim MH, Hamilton M, et al. Leveraging the social network for treatment of social anxiety: pilot study of a youth-specific digital intervention with a focus on engagement of young men. Internet Interv 2020 May 8;20:100323 [FREE Full text] [doi: 10.1016/j.invent.2020.100323] [Medline: 32435600]
70. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
71. Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. World Psychiatry 2021 Jun;20(2):233-243 [FREE Full text] [doi: 10.1002/wps.20858] [Medline: 34002511]
72. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
73. van Doorn M, Nijhuis LA, Monsanto A, van Amelsvoort T, Popma A, Jaspers MW, et al. Usability, feasibility, and effect of a biocueing intervention in addition to a moderated digital social therapy-platform in young people with emerging mental health problems: a mixed-method approach. Front Psychiatry 2022 May 25;13:871813. [doi: 10.3389/fpsyt.2022.871813] [Medline: 35693972]
74. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
75. Goldman HH, Skodol AE, Lave TR. Revising axis V for DSM-IV: a review of measures of social functioning. Am J Psychiatry 1992 Sep;149(9):1148-1156. [doi: 10.1176/ajp.149.9.1148] [Medline: 1386964]
76. Doornenbal BM, Vos RC, Van Vliet M, Kiefte-De Jong JC, van den Akker-van Marle ME. Measuring positive health: concurrent and factorial validity based on a representative Dutch sample. Health Soc Care Community 2022 Sep 18;30(5):e2109-e2117 [FREE Full text] [doi: 10.1111/hsc.13649] [Medline: 34791738]
77. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
78. Rickwood D, McEachran J, Saw A, Telford N, Trethowan J, McGorry P. Sixteen years of innovation in youth mental healthcare: outcomes for young people attending Australia's headspace centre services. PLoS One 2023 Jun 30;18(6):e0282040 [FREE Full text] [doi: 10.1371/journal.pone.0282040] [Medline: 37390108]
79. Chong MK, Hickie IB, Cross SP, McKenna S, Varidel M, Capon W, et al. Validation of a digital algorithm for clinical staging to support stratification in youth mental health services: proof-of-concept study. ResearchGate Preprint posted online December 2022 [FREE Full text] [doi: 10.2196/preprints.44738]
80. Castor. URL: https://www.castoredc.com/ [accessed 2023-03-13]
81. Duckworth AL, Steen TA, Seligman ME. Positive psychology in clinical practice. Annu Rev Clin Psychol 2005;1:629-651. [doi: 10.1146/annurev.clinpsy.1.102803.144154] [Medline: 17716102]
82. Seligman ME, Rashid T, Parks AC. Positive psychotherapy. Am Psychol 2006 Nov;61(8):774-788. [doi: 10.1037/0003-066X.61.8.774] [Medline: 17115810]
83. Gleeson JF, Lederman R, Wadley G, Bendall S, McGorry PD, Alvarez-Jimenez M. Safety and privacy outcomes from a moderated online social therapy for young people with first-episode psychosis. Psychiatr Serv 2014 Apr 01;65(4):546-550. [doi: 10.1176/appi.ps.201300078] [Medline: 24687106]
84. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
85. Gleeson JF, Koval P, Zyphur M, Lederman R, Herrman H, Eleftheriadis D, et al. A randomized controlled trial of moderated online social therapy for family carers of first-episode psychosis patients in a specialist treatment setting. Schizophr Res 2023 May;255:203-212. [doi: 10.1016/j.schres.2023.03.019] [Medline: 37004332]
86. Yung L, Phillips M, Ward J, Thompson P, French PP. Comprehensive assessment of at risk mental states. In: van der Gaag M, van der Werf J, Wunderink L, Malda A, Klaassen R, editors. CAARMS Authorized Dutch Translation. Melbourne, Australia: The PACE Clinic, University of Melbourne Department of Psychiatry; 2008.

87. Jovanović AA, Gasić MJ, Ivković M, Milovanović S, Damjanović A. Reliability and validity of DSM-IV Axis V scales in a clinical sample of veterans with posttraumatic stress disorder. Psychiatr Danub 2008 Sep;20(3):286-300. [Medline: 18827754]
88. Donker T, Comijs H, Cuijpers P, Terluin B, Nolen W, Zitman F, et al. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. Psychiatry Res 2010 Mar 30;176(1):45-50. [doi: 10.1016/j.psychres.2009.01.012] [Medline: 20071036]
89. Pereira A, Oliveira CA, Bártolo A, Monteiro S, Vagos P, Jardim J. Reliability and factor structure of the 10-item Kessler Psychological Distress Scale (K10) among Portuguese adults. Cien Saude Colet 2019 Mar;24(3):729-736 [FREE Full text] [doi: 10.1590/1413-81232018243.06322017] [Medline: 30892495]
90. Van Vliet M, Doornenbal BM, Boerema S, van den Akker-van Marle EM. Development and psychometric evaluation of a positive health measurement scale: a factor analysis study based on a Dutch population. BMJ Open 2021 Feb 05;11(2):e040816 [FREE Full text] [doi: 10.1136/bmjopen-2020-040816] [Medline: 33550237]
91. Kessels RP. LLT location learning test – Herziene uitgave. Hogrefe. 2014 Apr. URL: https://www.hogrefe.com/nl/index.php?eID=dumpFile&t=f&f=4562&token=11ef0b805d2593f4c7ffe33c585a233abd1f4492 [accessed 2023-06-20]
92. Kenward MG, Carpenter J. Multiple imputation: current perspectives. Stat Methods Med Res 2007 Jun;16(3):199-218. [doi: 10.1177/0962280206075304] [Medline: 17621468]
93. Vallejo G, Fernández MP, Livacic-Rojas PE. Multivariate analysis of covariance for heterogeneous and incomplete data. Psychol Methods 2023 Feb 16 (forthcoming). [doi: 10.1037/met0000558] [Medline: 36795436]
94. Slade T, Grove R, Burgess P. Kessler psychological distress scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2011 Apr;45(4):308-316. [doi: 10.3109/00048674.2010.543653] [Medline: 21332432]
95. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: The Guilford Press; Feb 14, 2017.
96. Fotiadis A, Abdulrahman K, Spyridou A. The mediating roles of psychological autonomy, competence and relatedness on work-life balance and well-being. Front Psychol 2019 May 29;10:1267 [FREE Full text] [doi: 10.3389/fpsyg.2019.01267] [Medline: 31191420]
97. Jensen PS, Weersing R, Hoagwood KE, Goldman E. What is the evidence for evidence-based treatments? A hard look at our soft underbelly. Ment Health Serv Res 2005 Mar;7(1):53-74. [doi: 10.1007/s11020-005-1965-3] [Medline: 15832693]
98. Weisz JR, Kuppens S, Ng MY, Vaughn-Coaxum RA, Ugueto AM, Eckshtain D, et al. Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. Perspect Psychol Sci 2019 Mar;14(2):216-237. [doi: 10.1177/1745691618805436] [Medline: 30571478]
99. Cross SP, Karin E, Staples LG, Bisby MA, Ryan K, Duke G, et al. Factors associated with treatment uptake, completion, and subsequent symptom improvement in a national digital mental health service. Internet Interv 2022 Mar;27:100506 [FREE Full text] [doi: 10.1016/j.invent.2022.100506] [Medline: 35242587]
100. Reneses B, Muñoz E, López-Ibor JJ. Factors predicting drop-out in community mental health centres. World Psychiatry 2009 Oct;8(3):173-177 [FREE Full text] [doi: 10.1002/j.2051-5545.2009.tb00246.x] [Medline: 19812755]
101. Moshe I, Terhorst Y, Paganini S, Schlicker S, Pulkki-Råback L, Baumeister H, et al. Predictors of dropout in a digital intervention for the prevention and treatment of depression in patients with chronic back pain: secondary analysis of two randomized controlled trials. J Med Internet Res 2022 Aug 30;24(8):e38261 [FREE Full text] [doi: 10.2196/38261] [Medline: 36040780]
102. Lawler K, Earley C, Timulak L, Enrique A, Richards D. Dropout from an internet-delivered cognitive behavioral therapy intervention for adults with depression and anxiety: qualitative study. JMIR Form Res 2021 Nov 12;5(11):e26221 [FREE Full text] [doi: 10.2196/26221] [Medline: 34766909]
103. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
104. Cijfers over jeugd met een migratieachtergrond. Nederlands Jeugdinstituut. 2023 Feb 13. URL: https://www.nji.nl/cijfers/jeugd-met-een-migratieachtergrond#:~:text=In%202021%20was%2015%2C4,voor%20de%20Statistiek%2C%202022) [accessed 2023-05-31]
105. Bevolking; hoogst behaald onderwijsniveau en herkomst. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/nl-nl/cijfers/detail/85453NED?q=jongeren%20opleidingsniveau [accessed 2023-05-31]
106. Youth-centred digital health interventions: a framework for planning, developing and implementing solutions with and for young people. World Health Organization. 2021 Jul 19. URL: https://www.who.int/publications/i/item/9789240011717 [accessed 2023-11-28]
107. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011 Jul;128(1):305-310 [FREE Full text] [doi: 10.1097/PRS.0b013e318219c171] [Medline: 21701348]
108. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T. Randomized controlled trials - a matter of design. Neuropsychiatr Dis Treat 2016 Jun 10;12:1341-1349 [FREE Full text] [doi: 10.2147/NDT.S101938] [Medline: 27354804]

109. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018 Dec;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
110. Leavy P. Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches. New York, NY: Guilford Publications; 2017.
111. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
112. de Ruiter G, van Greuningen M, Luijk R. Inzicht in de geestelijke gezondheidszorg. Vektis. 2017. URL: https://www.vektis.nl/uploads/Publicaties/Zorgthermometer/Zorgthermometer%20Inzicht%20in%20de%20ggz%202017.pdf [accessed 2023-11-17]
113. Nieman D. Prevention in Mental Health Care: Time for a New Approach. Oxford, UK: Routledge; 2016.
114. Signorini G, Singh SP, Marsanic VB, Dieleman G, Dodig-Ćurković K, Franic T, et al. Correction to: the interface between child/adolescent and adult mental health services: results from a European 28-country survey. Eur Child Adolesc Psychiatry 2018 Nov 13;27(11):1511-1512. [doi: 10.1007/s00787-018-1132-1] [Medline: 29536191]
115. Achtergronddocumentatie bij de wachttijdgegevens. Vektis. URL: https://www.vektis.nl/uploads/Zorginzichten/FS%20GGZ%20Wachttijden/Achtergronddocumentatie%20bij%20de%20wachttijdgegevens.pdf [accessed 2023-06-20]
116. Engel L, Alvarez-Jimenez M, Cagliarini D, D'Alfonso S, Faller J, Valentine L, et al. The cost-effectiveness of a novel online social therapy to maintain treatment effects from first-episode psychosis services: results from the Horyzons randomized controlled trial. Schizophr Bull 2023 Jun 01:sbad071 (forthcoming). [doi: 10.1093/schbul/sbad071] [Medline: 37261464]
117. Fusar-Poli P. Integrated mental health services for the developmental period (0 to 25 years): a critical review of the evidence. Front Psychiatry 2019 Jun 7;10:355 [FREE Full text] [doi: 10.3389/fpsyt.2019.00355] [Medline: 31231250]
118. Jones PB. Adult mental health disorders and their age at onset. Br J Psychiatry Suppl 2013 Jan;54:s5-10. [doi: 10.1192/bjp.bp.112.119164] [Medline: 23288502]

Abbreviations

CBT: cognitive behavioral therapy
CSI: Clinically Significant Index
ENYOY: Engage Young People Early
K10: Kessler Psychological Distress Scale
MOST: Moderated Online Social Therapy
MPH: My Positive Health
RCI: Reliable Change Index
RCT: randomized controlled trial
SOFAS: Social and Occupational Functioning Assessment Scale

Edited by L Buis; submitted 07.07.23; peer-reviewed by B Sandín, G Donohoe; comments to author 18.08.23; revised version received 19.09.23; accepted 18.10.23; published 13.12.23.

Please cite as:
van Doorn M, Monsanto A, Wang CL, Verfaillie SCJ, van Amelsvoort TAMJ, Popma A, Jaspers MWM, Öry F, Alvarez-Jimenez M, Gleeson JF, Nieman DH
The Effects of a Digital, Transdiagnostic, Clinically and Peer-Moderated Treatment Platform for Young People With Emerging Mental Health Complaints: Repeated Measures Within-Subjects Study
JMIR Mhealth Uhealth 2023;11:e50636
URL: https://mhealth.jmir.org/2023/1/e50636
doi:10.2196/50636
PMID:38090802

A Novel Smartphone App for Self-Monitoring of Neonatal Jaundice Among Postpartum Mothers: Qualitative Research Study

Aminath Shiwaza Moosa[1,2,*], MMed; Alvin Jia Hao Ngeow[3,4,5,*], MMed; Yuhan Yang[6]; Zhimin Poon[1,2], MMed; Ding Xuan Ng[1], BEng; Eileen Koh Yi Ling[1], BSc; Ngiap Chuan Tan[1,2], MMed

1

2

3

4

5

6

*these authors contributed equally

Corresponding Author:
Aminath Shiwaza Moosa, MMed

Abstract

Background: Neonatal jaundice (NNJ) or hyperbilirubinemia is a ubiquitous condition in newborn infants. Currently, the transcutaneous bilirubinometer is used to screen for NNJ in health care facilities, where neonates need to be physically present (ie, a centralized model of care for NNJ screening). Mobile health (mHealth) apps present a low-cost, home-based, and noninvasive system that could facilitate self-monitoring of NNJ and could allow mothers the convenience of screening for NNJ remotely. However, end users' acceptability of such mHealth apps is of fundamental importance before the incorporation of such apps into clinical practice.

Objective: The study aimed to explore the perception of postpartum mothers toward self-monitoring of NNJ using a novel mHealth app.

Methods: Mothers attending video consultations for early postpartum care at 2 Singapore primary care clinics watched an instructional video for a hyperbilirubinemia-screening mHealth app (HSMA). An independent researcher used a semistructured topic guide to conduct in-depth interviews with 25 mothers, assessing their views on HSMAs. All interviews were audio recorded, transcribed verbatim, and checked for accuracy before data analysis. Two researchers independently analyzed the transcripts via thematic analysis. Data were managed using NVivo qualitative data management software.

Results: The identified themes were grouped under perceived usability and utility. Mothers valued the convenience and utility of HSMAs for remote monitoring of NNJ. They appreciated the objectivity the app readings provided compared to visual inspection. However, they perceived that the app's applicability would be restricted to severe jaundice, were concerned about its accuracy and restriction to the English language, and lacked confidence in using it. Nevertheless, they were willing to use it once its accuracy was proven and when they received adequate guidance from health care professionals. They also suggested including an action plan for the measured readings and clinical signs within the app. Mothers proposed pairing teleconsultations with HSMAs to boost their confidence and enhance adoption.

Conclusions: Mothers were receptive to using HSMAs but had concerns. Multiple languages, proof of accuracy, and resources to guide users should be incorporated into the app in the next phase to increase its successful adoption. Complementing such apps with a teleconsultation service presents a plausible and pragmatic NNJ care delivery model in general practice.

(JMIR Mhealth Uhealth 2023;11:e53291) doi:10.2196/53291

KEYWORDS

neonatal jaundice screening; mobile health; usability; primary care; neonatology; neonatal; care; self-monitoring; monitoring; smartphone application; application; app; mHealth app; users; screening; postpartum; postpartum mother; usability; utility; data management; applicability; teleconsultation; neonatal jaundice; mHealth

Introduction

Neonatal jaundice (NNJ) is common in newborn infants. It is secondary to elevated total serum bilirubin or hyperbilirubinemia [1]. The prevalence of NNJ is 60% and 80% in term and preterm neonates, respectively [2]. Physiological jaundice accounts for most cases of neonatal hyperbilirubinemia [1]. It is usually mild and self-limiting [1]. However, in severe hyperbilirubinemia, unconjugated bilirubin may cross the blood-brain barrier and lead to neurological deficits [1]. The incidences of acute and chronic bilirubin encephalopathy are approximately 1 in 10,000 live births and 1 in 50,000 to 100,000 live births, respectively [1]. Bilirubin-induced encephalopathy can result in cerebral palsy, with an incidence of 0.57 in 100,000 live births [3,4]. As such, infants with NNJ require close surveillance, and immediate medical attention is warranted in severe hyperbilirubinemia.

Total serum bilirubin measurement remains the gold standard for ascertaining NNJ [5]. However, this method involves invasive blood sampling of the neonate, whereas the existing local standard of care requires NNJ monitoring using the transcutaneous bilirubinometer. The latter is a specialized equipment only available in health care facilities. This creates a necessity for parents to bring their babies to clinics for jaundice checks [5], resulting in the incurrence of travel costs and time away from work by parents, as well as the potential risk of exposure to infectious pathogens at health care facilities [6,7].

Monitoring for NNJ through smartphone apps presents a low-cost, home-based, and noninvasive system that could facilitate self-monitoring of NNJ and reduce unnecessary exposure of the child to health care facilities. Such smartphone apps have been developed in China, the Unites States, and Norway [8-10]. Currently, these apps are in the research phase with promising results. Rong et al [8], Taylor et al [9], and Aune et al [10] reported good overall associations between the app-based bilirubin estimates and total serum bilirubin, with Pearson correlation coefficients (r) of 0.788, 0.91, and 0.84, respectively. Thus, the use of such smartphone apps offers a convenient and viable alternative for laypersons such as postpartum mothers for the purpose of NNJ screening.

At present, such mobile health (mHealth)–based, NNJ-monitoring tools are novel to the local multiethnic parents in cosmopolitan Singapore. Little is known about parental levels of awareness and acceptability. Scaling up the use of such apps is only possible if the end users are receptive to their use [11]. Engaging end users in the development and design of mHealth technologies eases the adoption and accessibility of such technologies [12]. Thus, this study aimed to explore the views of mothers toward a novel mHealth app for their self-monitoring of NNJ based on their perceived usefulness and concerns. Findings and feedback from these end users can be used to enhance the features of the mHealth app, which will improve its usability and utility to monitor for NNJ remotely.

Methods

Study Design

Qualitative research methodology was used to garner the perceptions of mothers on the usefulness of NNJ-monitoring or hyperbilirubinemia-screening mHealth apps (HSMAs). Biliscan (Shenzhen Beishen Healthcare Technology Co) is one such app being validated at a local tertiary hospital [13]. The usefulness of an app is defined as a combination of its usability and utility [14]; usability is the ease of use of the user interface, and utility refers to the functionality of the app and its value to users [14].

This study is reported according to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1).

Study Setting

The study sites were 2 public primary care clinics (polyclinics) in the Bukit Merah (BM) and Sengkang (SK) estates located in southern and eastern Singapore, respectively. The latter is populated with young families and manages over 1000 patients daily during office hours. The primary care physicians and nurses in these 2 polyclinics provide comprehensive primary care to mothers and children, including screening and monitoring neonates for NNJ.

Research Team

The research team included 2 female primary care physicians (ASM and ZP), 1 male primary care physician (NCT), 1 male neonatologist (AJHN), and 1 female student (YY). ASM, AJHN, and ZP are trained in qualitative and quantitative research. NCT is a trainer and experienced researcher in both qualitative and quantitative research.

Study Population and Recruitment

Mothers aged 21 years and older, within 4 weeks of their postpartum period, and who spoke English were eligible for the study. Primary care physicians and nurses at the study site helped identify these mothers when they attended the clinic for their child’s NNJ follow-up. Mothers were recruited from December 2021 to August 2022.

Ethical Considerations

The study was approved by the SingHealth Centralised Institutional Review Board (CIRB 2021/2732). The research was conducted according to International Conference on Harmonization guidelines for Good Clinical Practice. Clinical research coordinators approached potential participants face-to-face and informed them of the study purpose before they signed informed written consent. Each mother was assigned a unique study identification number to deidentify them and ensure confidentiality. They were remunerated with an SGD $20 (~US $15) grocery voucher for their time and contribution to the study.

Study Instruments and Data Collection

Demographic Characteristics Survey

A web-based questionnaire was designed to collect the mothers’ demographic characteristics. The electronic survey form was hosted on FormSG, a secure, officially approved web-based platform available to the public in Singapore [15]. The questionnaire collected demographic and selected clinical formation from the mothers on their age, ethnicity, marital status, educational level, housing type, financial status, delivery date, and parity at recruitment.

Instructional Video on NNJ Screening Using an mHealth App

The instructional video *Jaundice Screening Using a Smartphone* [16] was created by AJHN for the purpose of patient education and consent taking as part of an earlier validation study on

Biliscan [13,17]; it shows an HSMA being used for remote monitoring of NNJ. The use of HSMAs is described in detail by Alvin et al [17]. In brief, it involves placing a color calibration card with a central aperture on the baby's sternum, aligning the app's photograph frame with the color card, and taking a picture. After processing the image and applying a machine learning regression algorithm, bilirubin levels are estimated and displayed on the phone screen. The video was uploaded on the institute's official YouTube channel [16]. Mothers were invited to watch this video using a QR code before their in-depth interviews.

Topic Guide and In-Depth Interviews

A semistructured topic guide was developed based on a literature review and discussion with team members. The questions inquired about the mother's perceived utility and usability of HSMAs to monitor their child for NNJ. The topic guide was pilot-tested and iteratively modified based on data that emerged during both pilot and subsequent interviews.

An independent researcher (ASM, ZP, or NCT) used the topic guide to interview individual mothers over the phone. The interviewer was in the primary care clinic while the mother was at home. All mothers were unaccompanied during the interview. All interviews were audio recorded and each interview lasted 25-45 minutes. The study team recorded field notes during each interview.

Data Analysis

Professionals transcribed the interviews, and the content of the audio recordings was audited for accuracy. Next, 2 researchers (ASM and ZP) independently coded the initial 3 transcripts to generate an initial coding frame. Based on the initial coding frame, 2 researchers (ASM and YY) coded and indexed subsequent transcripts using NVivo qualitative data management software (Lumivero). The researchers regularly discussed the qualitative data and their coding.

The researchers analyzed and grouped the codes to identify emerging themes. Any disagreements were resolved with discussion and consensus. Subsequently, the research team reviewed and revised the results to finalize the themes. Representative quotes were selected to illustrate the results. Regular iterations after each interview and referral to field notes allowed ASM to reflect on the findings and to assess for idea saturation [18]. No new idea emerged after the 23rd interview. Two more interviews were conducted before a mutually agreed decision was reached to stop further qualitative data collection.

Results

Participants

A total of 56 postpartum mothers were approached; 30 consented to participate in the study (a response rate of 54%), and 25 mothers completed the in-depth interviews. A total of 5 mothers were lost to follow-up: 2 after giving consent and 3 after watching the video. Mothers who declined to participate cited time constraints, a lack of interest, or other commitments that prevented them from participating in the study. All mothers completed the preconsultation questionnaire.

Participant Characteristics

Table 1 reports the sociodemographic characteristics of the 30 mothers in the main study. Most were aged 30-34 years (15/30, 50%), were of Chinese ethnicity (21/30, 70%), obtained university or posttertiary education (26/30, 87%), lived in Housing and Development Board (Singapore's public housing) 4- to 5-room flats (20/30, 67%), and did not receive medical subsidy (29/30, 97%). More than half (17/30, 57%) of the mothers underwent normal vaginal delivery. Around two-thirds (19/30, 63%) were first-time mothers.

Table . Demographic characteristics of the mothers.

Characteristics		Mothers (n=30), n (%)
Age group (y)		
	20-24	1 (3)
	25-29	5 (17)
	30-34	15 (50)
	35-39	9 (30)
Ethnicity		
	Chinese	21 (70)
	Malay	1 (3)
	Indian	5 (17)
	Others	3 (10)
Highest education qualification		
	Secondary	1 (3)
	A-level or diploma (ITE[a], polytechnic, or private school)	3 (10)
	University or posttertiary	26 (87)
Housing type		
	Rental room or apartment	3 (10)
	HDB[b] 1- to 3-room flat	5 (17)
	HDB 4- to 5-room flat	20 (67)
	Masonite or executive condominium	2 (7)
Type of delivery		
	Assisted delivery	2 (7)
	Caesarean section	11 (37)
	Normal vaginal delivery	17 (57)
Number of children		
	1	19 (63)
	2	10 (33)
	3 or more	1 (3)
Current child feeding method		
	Breastfeeding only	4 (13)
	Formula feeding only	1 (3)
	Mixed feeding (breastfeeding and formula feeding)	25 (83)
Breastfeeding difficulty		
	No	12 (40)
	Yes	18 (60)
Medical subsidy status		
	No	29 (97)
	Yes	1 (3)

[a]ITE: Institute of Technical Education.

[b]Housing and Development Board (Singapore’s public housing).

Main Findings

Findings regarding the mothers' perceived usefulness were grouped into usability and utility. The themes, their descriptions, and subthemes are listed in Table 2.

Table . Themes, their descriptions, and subthemes on mother's perceived usefulness of hyperbilirubinemia-screening mobile health apps.

Themes	Descriptions	Subthemes
Utility	Functionality of the app and the value to users	• Preference and convenience to check jaundice level at home • Provides objectivity compared to visual inspection • Limited applicability for severe jaundice • Accuracy and reliability of the app
Usability	Ease of use of the app's user interface and improving the ease of use	• Easy to use • Instructions restricted to the English language • Higher trust in health care professionals • Guidance for app use and the management of jaundice • Incorporation into teleconsultations

Utility

All mothers were receptive to using HSMAs because of their convenience and quantitative measurement without the need to access health care facilities. However, they were concerned about the app's accuracy, leading to their lack of confidence in using it.

Preference and Convenience to Check Jaundice Levels at Home

Most mothers perceived the benefits of using HSMAs, which allowed them to monitor their children for NNJ daily in the comfort of their homes instead of needing multiple visits to a clinic or hospital. They were concerned about the risk of infection outdoors. They felt reassured to know their child's NNJ status before the in-person review by a health care professional.

> *I'm sure from time to time the mother will be curious if the numbers are going down. For paranoid mothers like me, I think it would be good if I can monitor it on my own on day-to-day basis instead of going to the polyclinic. [SK05; 36-year-old Chinese mother of 2]*

> *As a small kid, you know it's very troublesome to visit a clinic…[due to] the viruses and all outside because he's still so young. He is a baby, so it's not that good to make him travel around. Usually, we don't let the child travel until he's two months or three months [and] completed [the vaccination]. [BM06; 39-year-old Indian mother of 2]*

Mothers also preferred not undergoing blood sampling for NNJ assessment, as they perceived it to be painful for their child. They recognized the use of HSMAs as a less traumatic alternative to the invasive blood test.

> *I went to poly four times to do the jaundice check. And every time we went through the blood test, it pains me to see my child being pricked by the needle. I think this will be something that I will definitely use. [SK03; 38-year-old Javanese mother of 2]*

Provides Objective Assessment Compared to Visual Inspection

Mothers appreciated the quantitative evaluation of NNJ by HSMAs. They were more assured of the numerical measurement of jaundice than subjective visual inspection of the skin color.

> *…when it comes to jaundice, we always try to see baby yellow or not, eyes yellow or not. With such a device, at least it is more certain for us to do the first cut of assessment. [SK03; 38-year-old Javanese mother of 2]*

> *I think it's quite useful, at least we have some idea instead of just pressing his skin to know his level. I am quite a statistic person. If I have numbers, I will be more reassured that the jaundice is coming down. [SK10; 38-year-old Chinese mother of 1]*

Limited Applicability for Severe Jaundice

Mothers were unsure of the use of HSMAs for severe NNJ, as they considered that it would warrant a clinic visit for a laboratory-based serum bilirubin measurement.

> *…[in severe jaundice] we know already that we will have an appointment and then this appointment is set every 2 days because the level was still quite high up on that day [because of this] I don't think I will use the app anyway. I will be seeing a doctor, so there's no point for me to install the app. [BM10; 38-year-old Chinese mother of 1]*

Mothers would select in-person clinic visits to check bilirubin levels for severe NNJ. They were worried about undue delay for their child to receive the required treatment, which might result in harm.

> *…if the jaundice is not that serious, I guess I am okay to check at home with this mobile app but if baby has very serious jaundice, I would still prefer to come*

down to polyclinic. [SK01; 38-year-old Chinese mother of 1]

…maybe it [jaundice level] increasing and we are not able to notify the help of doctors, it will be very painful for my baby. [SK08; 31-year-old Indian mother of 1]

Accuracy and Reliability of the App

Many mothers were uncertain about the accuracy and reliability of the HSMA. They raised concerns that the impact of different camera quality and lighting on its measurements.

If there's a way to cite like how accurate the app is…because your phone is different right? the colour is different, then the camera is different; those are things that of [affects] the confidence level of the application. [SK04; 33-year-old Chinese mother of 1]

Hmm, because I think it depends on the lighting, so that's why I don't use [it]. [BM10; 38-year-old Chinese mother of 1]

They knew that bilirubin levels measured using a transcutaneous bilirubinometer and blood test differed. Thus, they anticipated a similar difference in the readings from the HSMA and the actual jaundice level. If there was a difference, they wanted to know the difference in the accuracy of the app reading compared to the actual jaundice level, as a "small difference" could change how the doctor will manage the child.

I don't think it's comparable because even when we do the first round of checks, using that meter [transcutaneous bilirubinometer] it can also differ from the blood test results. I don't think it will be totally comparable. [BM15; 31-year-old Chinese mother of 2]

If let's say it's high, then I still need phototherapy [and] I will go down [to the clinic]. if it [threshold for phototherapy] is supposed to be below 200, then my levels were like 190, then to me, it seems ok, but is it accurate? Is the app accurate? What if it's 200…what if there's a 10% discrepancy. [SK04; 33-year-old Chinese mother of 1]

Nevertheless, mothers trusted that the app would have been trialed and tested before being rolled out. Their trust in the service provider influenced their intention to use HSMAs in the future.

I assume that if you have rolled it out you have done a lot of pilot tests and ascertain that the app is accurate. [BM04; 33-year-old Chinese mother of 1]

Mothers suggested that health care professionals demonstrate and calibrate the HSMA on their devices in comparison with the current transcutaneous and serum bilirubin measurement. This would inform and reassure them that the app is accurate and reliable in assessing the jaundice level.

If there's a level of comparison, the phone app and the actual scanning, then you can see the level of discrepancy, let's say if it's 1% then it's like ok, it's quite accurate, not so bad. [SK04; 33-yeary-old Chinese mother of 1]

Usability

The simple interface of the app and the clear demonstration shown in the video facilitated the perceived ease of use. However, mothers were concerned about the procedural steps in using HSMAs and the app's instructions being restricted to the English language, which may hinder its use. Mothers alluded to improving its ease of use by receiving adequate guidance from health care professionals and pairing teleconsultations with it.

Easy to Use

Most mothers perceived HSMAs as easy to use, as the video demonstrated the use of the app in simple, clear steps.

Yes,it is quite easy. The steps are all very clear. [BM13; 30-year-old Chinese mother of 1]

Instructions Restricted to the English Language

A mother worried that if the video instructions were only available in English, it might restrict the app's use to a specific population. She suggested providing the instructions in multiple languages.

I guess for most people, should be quite straight-forward but maybe because the video is only in English, I don't know if it's easy for everybody to understand the instructions…if it can be provided in different languages it would be helpful. [SK01; 38-year-old Chinese mother of 1]

Higher Trust in Health Care Professionals

A few mothers doubted that they could use the HSMA correctly. They preferred having a health care professional evaluate their child for NNJ; they dreaded a delay in their child's treatment if they misused the app.

I think we would feel better, we will feel secure if we could actually come down [to the clinic] and get the nurse and the doctor to check properly. At the same time, you don't know whether you're doing it right and you don't want to do, not knowing that's the wrong way of doing and if there's any appropriate action that you need to take and you are unable to…you go to the hospital/clinic also, I don't know if I'm doing it right or wrong you know. I think it's safer for a mother, anything to do with the child…go to the clinic. Let the doctors do whatever they need to do. [BM01; 28-year-old Sikh mother of 1]

Guidance for App Use and the Management of Jaundice

Instruction by health care professionals on the use of the HSMA at the initial visit was favored by the mothers, as this provides an opportunity for them to clarify their doubts. Mothers suggested including a training manual in the app, a list of frequently asked questions with answers, and a hotline to assist them to resolve any problems.

…they [the nurse or the doctor] can demonstrate to me and then if I'm not too sure about anything they

can show me. I think a first demonstration by them will be fine. [BM01; 28-year-old Sikh mother of 1]

…if there's a standard setting to take the photo then that can be communicated to mums...that could be helpful. [SK04; 33-year-old Chinese mother of 1]

As long as I've been taught how to do it, like troubleshoot with the FAQ, or hotline I could call, then I think that would be good. [SK05; 36-year-old Chinese mother of 2]

Mothers wanted the app to include a guide to provide an action plan associated with the readings and clinical signs, as they perceived a paucity of web-based references to the measurements.

After you know the result, you need to [inform] parents what to do next. If it is this colour, can we still monitor for the next two days? or need to call doctor or need to go to hospital? [BM07; 34-year-old Chinese mother of 1]

Actually, there was no online gauge to what's the level that it should be…only when we went to the hospital, then we found out like…physical signs to look at. I think that could be part of the app as well. [SK04; 33-yeary-old Chinese mother of 1]

Incorporation Into Teleconsultations

Mothers proposed pairing the use of HSMAs with teleconsultations. They felt that using an app during teleconsultations would allow them to avoid bringing their child to the clinic and help demonstrate an objective NNJ assessment to the attending health care professional in real time.

I will use [teleconsultation]. I am very scared to bring the baby out. [BM07; 34-yeary-old Chinese mother of 1; when asked about using teleconsultations to monitor for NNJ]

Discussion

Principal Findings

This qualitative research study highlights the mothers' perceived usefulness of HSMAs. Mothers valued HSMAs for the convenience and objectivity when compared to visual inspection but had reservations about the accuracy. The app's user-friendly interface as shown in the demonstration video put parents at ease, although concerns were raised with regard to the instructions being restricted to the English language. Mothers suggested that guidance from health care professionals, validation of its accuracy, and integration with teleconsultations could boost the usability of the app.

Mothers viewed self-monitoring of NNJ using an mHealth app at home as a convenient alternative to multiple primary care clinic check-ups. Similarly, in an earlier study by Yan et al [19], more than half of mothers rated a smartphone app as convenient for monitoring their child for NNJ remotely. Remote access to an NNJ-monitoring service can overcome multiple inconveniences such as rushed clinic visits, long waits in the clinic, the unavailability of transportation, long travel distances, and the cost of traveling [6]. An earlier study also demonstrated that remote monitoring of NNJ reduced the number of outpatient visits without increasing the risk of readmission and severe neonatal hyperbilirubinemia [20].

In addition, pairing HSMAs with telemedicine or video consultations, as the mothers suggested, would be favored by those who prefer remote consultations in the convenience of their homes. Telemedicine can also reduce nosocomial infection or other adverse health effects in neonates and improves mothers' mental health [19,21]. The use of teleconsultation in a multisite neonatal follow-up program improved the consultation rates with a show rate of 95% [22]. This pairing will also improve mothers' confidence in using the app when NNJ is monitored objectively using HSMAs along with guidance from a health care professional via video consultation.

Mothers indicated that the app was easy to use based on the video demonstration. In an earlier study, Jin et al [23] highlighted the ease of use as a significant public consideration in the adoption of mHealth apps. Most mothers were familiar with phone-based app because of Singapore's high smartphone penetration; the current smartphone penetration in 2023 is 95.4%, which is expected to increase to 99.4% by 2028 [24]. Furthermore, the Singapore government encourages harnessing technology to monitor and manage health conditions [25]. These enablers support the uptake of validated mHealth apps in health care delivery.

The mothers' concerns about the accuracy of HSMAs and the inherent reliance of such apps on good image quality and standardized lighting are valid. Earlier studies on NNJ-monitoring mHealth apps demonstrated discrepancies and false positives, which could potentially be contributed to heterogeneous lighting conditions and photo quality [10,26]. Thus, for imaged-based mHealth apps, optimal image quality and standardized lighting conditions are crucial factors to be addressed to achieve reliable results. These concerns about image quality and lighting can be addressed through the app's built-in capability to reject low-quality images [17,26]. Additionally, the use of a color calibration card compensates for variations in light intensity [17,26]. A novel mHealth app that overcomes these limitations has been developed by the lead authors and will undergo clinical validation; this validation study will include a quantitative component to complement the qualitative findings.

Mothers in this study emphasized the need to conduct training before using the app; this would improve the app's usability. Lewis and Wyatt [27] reported inadequate user training of mHealth apps, resulting in inappropriate app use and users' failure to detect app malfunction or accurately diagnose the health condition. As the mothers in this study perceived, a live demonstration of the app by health care professionals is more valuable than an instructional video. Prompt user guidance on monitoring and advice on individualized actionable management of NNJ would encourage the use of mHealth apps appropriately [28]. Providing the relevant information for specific health conditions and corresponding feedback for tests enhances user satisfaction [28]. Thus, user training is pivotal to ensure the app's correct use and reduce possible risks to patient safety.

Another barrier to the HSMA's usability was its sole dependency on the English language to provide instructions to users. Communicating with users in languages they can understand is one of the most pertinent factors influencing an app's utility [29]. Nevertheless, English is the official language for education and work in Singapore. The country has a high literacy rate of 97.1%, and almost half of the population (48.3%) use English as their main language at home [30]. However, a minor group of residents or emigrants from overseas countries such as China, South Asia, Asia-Pacific, and other regions may not be as proficient in English as native Singaporeans and could deter them from using an English-based app. As such, incorporating more languages in the app could extend the reach to more users and will be developed at a later stage.

Strength and Limitations

This is likely the first study exploring Asian mothers' attitudes toward using a smartphone app for self-monitoring of NNJ in a technologically advanced community. Using qualitative research to explore the attitude among mothers provided in-depth knowledge of their needs, barriers, and intent to use such an mHealth app.

The study has several limitations. First, the study describes the perceived usefulness of the NNJ-monitoring mHealth app based on a video demonstration instead of the actual use of the app; thus, the usefulness of the app in the real world may differ. Nevertheless, understanding the end user's perspective is essential in designing and developing the app. The data gathered on the mother's needs of such an mHealth app from this study are sufficient to guide the developers and health care professionals to design and test the prototype on screening for NNJ remotely. Second, the mothers in this study were highly educated, restricting the study's generalizability. A more diverse range of participants in terms of education levels and socioeconomic backgrounds could enhance the generalizability of the findings. Nevertheless, the Singapore census shows that 80% to 90% of people aged 25-50 years (the usual age group of mothers) have completed postsecondary or higher education [30]. This is consistent with most of the tertiary-educated mothers in this study.

Conclusion

Mothers valued the convenience and objectivity of HSMAs and perceived them as easy to use. The app's exclusive use of the English language was raised as potential barriers to its usability and utility. To increase its successful adoption, multiple languages, proof of accuracy, and resources to guide users would be incorporated in the next phase of the study. Combining mHealth apps and teleconsultations to monitor for NNJ remotely represents an accessible and pragmatic care delivery model.

Acknowledgments

The authors would like to thank the staff of SingHealth Polyclinics–Bukit Merah, SingHealth Polyclinics–Sengkang, and the research department in the institution who assisted with the study implementation.

The study was supported by National Medical Research Council (NMRC) Centre Grant Programme (SingHealth Polyclinics): Integrated Platform for Research in Advancing Metabolic Health Outcomes in Women & Children in Asia (I-PRAMHO) Seed Fund (NMRC/CG/C008/2017). The publication cost is supported by SingHealth Polyclinics Centre Grant (CG21APR3006; NMRC/CG3/001/2022-SHP).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

ASM contributed to conceptualization, methodology, funding acquisition, project administration, investigation, formal analysis, writing–original draft, and writing–review and editing. YY contributed to formal analysis, writing–original draft, and writing–review and editing. ZP contributed to conceptualization, methodology, investigation, and formal analysis. AJHN contributed to conceptualization, methodology, the provision of content expertise, and the creation of the demonstration video. DXN contributed to formal analysis and visualization. EKYL contributed to formal analysis. NCT contributed to conceptualization, methodology, formal analysis, writing–review and editing, and supervision. All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Checklist 1
Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.
[PDF File, 187 KB - mhealth_v11i1e53291_app1.pdf]

References

1. Ansong-Assoku B, Shah SD, Adnan M, Ankola PA. Neonatal jaundice. In: StatPearls: StatPearls Publishing; 2023. [Medline: 30422525]

2. Mitra S, Rennie J. Neonatal jaundice: aetiology, diagnosis and treatment. Br J Hosp Med (Lond) 2017 Dec 2;78(12):699-704. [doi: 10.12968/hmed.2017.78.12.699] [Medline: 29240507]
3. Karimzadeh P, Fallahi M, Kazemian M, Taslimi Taleghani N, Nouripour S, Radfar M. Bilirubin induced encephalopathy. Iran J Child Neurol 2020;14(1):7-19. [Medline: 32021624]
4. Wu YW, Kuzniewicz MW, Wickremasinghe AC, et al. Risk for cerebral palsy in infants with total serum bilirubin levels at or above the exchange transfusion threshold: a population-based study. JAMA Pediatr 2015 Mar;169(3):239-246. [doi: 10.1001/jamapediatrics.2014.3036] [Medline: 25560920]
5. Riskin A, Tamir A, Kugelman A, Hemo M, Bader D. Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia? J Pediatr 2008 Jun;152(6):782-787.E2. [doi: 10.1016/j.jpeds.2007.11.003] [Medline: 18492516]
6. Brady JM, Pouppirt N, Bernbaum J, et al. Why do children with severe bronchopulmonary dysplasia not attend neonatal follow-up care? parental views of barriers. Acta Paediatr 2018 Jun;107(6):996-1002. [doi: 10.1111/apa.14265] [Medline: 29420848]
7. Moore D, Canadian Paediatric Society (CPS), Infectious Diseases and Immunization Committee. Infection control in paediatric office settings. Paediatr Child Health 2008 May 1;13(5):408-419. [doi: 10.1093/pch/13.5.408] [Medline: 30894793]
8. Rong ZH, Luo F, Ma LY, et al. Evaluation of an automatic image-based screening technique for neonatal hyperbilirubinemia. Article in Chinese. Zhonghua Er Ke Za Zhi 2016 Aug;54(8):597-600. [doi: 10.3760/cma.j.issn.0578-1310.2016.08.008] [Medline: 27510872]
9. Taylor JA, Stout JW, de Greef L, et al. Use of a smartphone app to assess neonatal jaundice. Pediatrics 2017 Sep;140(3):e20170312. [doi: 10.1542/peds.2017-0312] [Medline: 28842403]
10. Aune A, Vartdal G, Bergseng H, Randeberg LL, Darj E. Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels. Acta Paediatr 2020 Dec;109(12):2532-2538. [doi: 10.1111/apa.15287] [Medline: 32267569]
11. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414. [doi: 10.2196/10414] [Medline: 30061091]
12. Gonzales A, Custodio R, Lapitan MC, Ladia MA. End users' perspectives on the quality and design of mHealth technologies during the COVID-19 pandemic in the Philippines: qualitative study. JMIR Form Res 2023 Apr 20;7:e41838. [doi: 10.2196/41838] [Medline: 36943932]
13. Huang D, Yang B, Gao X, et al. Influences on accuracy of automated image-based estimation of neonatal serum bilirubin level using smartphone application under different circumstances. Article in Chinese. Chinese Journal of Perinatal Medicine 2019 Apr 16;22:269-277. [doi: 10.3760/cma.j.issn.1007-9408.2019.04.012]
14. Nielsen J. Usability 101: introduction to usability. Nielsen Norman Group. 2012 Jan 3. URL: www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2023-09-30]
15. FormSG. GovTech Singapore. URL: www.tech.gov.sg/products-and-services/formsg/ [accessed 2023-09-30]
16. Jaundice screening using a smartphone. SingHealth Polyclinics YouTube page. 2022 Jan 28. URL: https://m.youtube.com/watch?v=wB5opHp9_YY [accessed 2023-09-30]
17. Ngeow AJH, Tan MG, Dong X, et al. Validation of a smartphone-based screening tool (Biliscan) for neonatal jaundice in a multi-ethnic neonatal population. J Paediatr Child Health 2023 Feb;59(2):288-297. [doi: 10.1111/jpc.16287] [Medline: 36440650]
18. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
19. Yan Q, Gong Y, Luo Q, et al. Effects of a smartphone-based out-of-hospital screening app for neonatal hyperbilirubinemia on neonatal readmission rates and maternal anxiety. J Med Internet Res 2022 Nov 23;24(11):e37843. [doi: 10.2196/37843] [Medline: 36416882]
20. Xu C, Bao Y, Zhu J, et al. Remote monitoring of neonatal jaundice in newborns with ABO hemolytic disease. Article in Chinese. Zhejiang Da Xue Xue Bao Yi Xue Ban 2020 Oct 25;49(5):651-655. [doi: 10.3785/j.issn.1008-9292.2020.08.08] [Medline: 33210495]
21. Joshi SS, Benroy BR, Lawrence IN, Suresh TJ. Telemedicine as progressive treatment approach for neonatal jaundice due to the coronavirus disease 2019 pandemic. Clin Exp Pediatr 2022 May;65(5):269-271. [doi: 10.3345/cep.2021.01235] [Medline: 35130430]
22. DeMauro SB, Duncan AF, Hurt H. Telemedicine use in neonatal follow-up programs - what can we do and what we can't - lessons learned from COVID-19. Semin Perinatol 2021 Aug;45(5):151430. [doi: 10.1016/j.semperi.2021.151430] [Medline: 33892961]
23. Jin X, Yuan Z, Zhou Z. Understanding the antecedents and effects of mHealth app use in pandemics: a sequential mixed-method investigation. Int J Environ Res Public Health 2023 Jan 2;20(1):834. [doi: 10.3390/ijerph20010834] [Medline: 36613156]
24. Statista Research Department. Smartphone penetration rate as share of the population in Singapore from 2019 to 2028. Statista. 2023 Sep 12. URL: www.statista.com/statistics/625441/smartphone-user-penetration-in-singapore/ [accessed 2023-12-11]

25. Smart nation: the way forward executive summary. Smart Nation Singapore. 2018 Nov. URL: www.smartnation.gov.sg/files/publications/smart-nation-strategy-nov2018.pdf [accessed 2023-08-12]
26. Abebe MA, Hardeberg JY, Vartdal G. Smartphones' skin colour reproduction analysis for neonatal jaundice detection. Journal of Imaging Science and Technology 2021 Nov 1;65(6):060407-1-060407-15. [doi: 10.2352/J.ImagingSci.Technol.2021.65.6.060407]
27. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
28. Birkmeyer S, Wirtz BW, Langer PF. Determinants of mHealth success: an empirical investigation of the user perspective. Int J Inf Manage 2021 Aug;59:102351. [doi: 10.1016/j.ijinfomgt.2021.102351] [Medline: 36540093]
29. Alzahrani A, Gay V, Alturki R. Exploring Saudi individuals' perspectives and needs to design a hypertension management mobile technology solution: qualitative study. Int J Environ Res Public Health 2022 Oct 10;19(19):12956. [doi: 10.3390/ijerph191912956] [Medline: 36232254]
30. Singapore Census of Population 2020, statistical release 1: demographic characteristics, education, language and religion. Department of Statistics Singapore. URL: www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20_stat_release1 [accessed 2023-12-14]

Abbreviations

BM: Bukit Merah
COREQ: Consolidated Criteria for Reporting Qualitative Research
HSMA: hyperbilirubinemia-screening mobile health app
mHealth: mobile health
NNJ: neonatal jaundice
SK: Sengkang

Edited by L Buis; submitted 04.10.23; peer-reviewed by M Aydin, V Kain; revised version received 22.11.23; accepted 27.11.23; published 22.12.23.

Please cite as:
Moosa AS, Ngeow AJH, Yang Y, Poon Z, Ng DX, Yi Ling EK, Tan NC
A Novel Smartphone App for Self-Monitoring of Neonatal Jaundice Among Postpartum Mothers: Qualitative Research Study
JMIR Mhealth Uhealth 2023;11:e53291
URL: https://mhealth.jmir.org/2023/1/e53291
doi:10.2196/53291

App-Based Hearing Screenings in Preschool Children With Different Types of Headphones: Diagnostic Study

Pornsek Tananuchittikul[1,2], MD; Kwanchanok Yimtae[2], MD; Nichtima Chayaopas[2], MD; Panida Thanawirattananit[2], BSc, MSc; Pornthep Kasemsiri[2], MD; Patorn Piromchai[2], MD, MSc, PhD

1

2

Corresponding Author:
Patorn Piromchai, MD, MSc, PhD

Abstract

Background: Hearing disability in preschool children can delay or impact oral communication and social skills. Provision of hearing screening tests by standard audiometry in low- to middle-income countries is problematic due to a lack of pediatric audiologists, standard hearing equipment, and standard soundproof rooms. Therefore, an innovative hearing screening tool that is easily accessible and inexpensive such as a mobile app should be considered. Headphones have been a crucial part of hearing screenings. Audiometric headphones, which serve as the reference standard, have been used in most studies. However, since audiometric headphones are not accessible in rural areas, we hypothesized that generic headphones can also be used in hearing screenings.

Objective: This study aimed to determine the sensitivity, specificity, κ coefficiency, and time consumption of the PASS-Pro (Preschool Audiometry Screening System–Pro) app when using TDH39 headphones, Beyerdynamic DT 770 PRO headphones, and generic earmuff headphones compared to standard conditioned play audiometry.

Methods: We recruited preschool children aged 4 to 5 years to participate in this study. The children received 3 PASS-Pro screening tests using different types of headphones in a quiet room and 1 standard conditioned play audiometry in a soundproof room. All tests were administered in random order. The agreement coefficient, sensitivity, specificity, and mean test duration were determined.

Results: A total of 44 children participated in this study. For mild hearing loss screening, the κ coefficients between standard conditioned play audiometry and the PASS-Pro app using TDH39 headphones, Beyerdynamic DT 770 PRO headphones, and generic earmuff headphones were 0.195, 0.290, and 0.261 (P=.02, P=.002, and P=.004), respectively. The sensitivity for all headphones was 50% and the specificity was more than 88%. For moderate hearing loss screening, the κ coefficients were 0.206, 0.272, and 0.235 (all Ps=.001), respectively. The sensitivity for all headphones was 100% and the specificity was more than 92%. There were no statistical differences in sensitivity and specificity between the reference headphone (TDH39), Beyerdynamic DT 770 PRO headphone, and generic earmuff headphones (all Ps >.05). The PASS-Pro app used significantly less time to carry out hearing tests than conditioned play audiometry (P<.001).

Conclusions: The PASS-Pro app, used with generic headphones, is effective for conducting hearing screening tests in preschool children with high sensitivity and specificity.

Trial Registration: Thai Clinical Trials Registry TCTR20201229002; www.thaiclinicaltrials.org/show/TCTR20201229002

(JMIR Mhealth Uhealth 2023;11:e44703) doi:10.2196/44703

KEYWORDS

hearing screening; mobile application; mobile app; headphones; mobile health; mHealth; hearing; headphone; pediatric; child; preschool; audiology; audiometry; audiologist; screening

Introduction

Hearing loss or hearing difficulties is a common cause of disability among people around the world. In 2012, the World Health Organization reported that 360 million people experienced hearing loss, indicating a prevalence of 5.3% in the global population (adults: 9%; children: 1%) [1]. In children aged less than 5 years, the prevalence of hearing loss was 2% (7.5 million) and the incidence of permanent hearing loss in preschool children was 3.4 to 3.56 per 1000 [2-5]. Early detection and intervention can help to improve patients' quality of life [6].

In low- to middle-income countries, hearing difficulties affect people's quality of life more than those living in high-income countries due to limited resources, such as medical staff, audiologists, hearing screening tools, and accessibility to medical treatment [7]. Hearing disabilities also affect preschool children in many ways, such as psychomotor skills, social skills, learning skills, emotional skills, and development skills [8,9]. Therefore, hearing screening tools play an important role in detecting hearing loss in preschool children and serve to prevent the disadvantageous outcomes described above.

Currently, standard hearing screening for children remains a major concern due to the lack of professional pediatric audiologists and soundproof rooms. The standard hearing test that is widely accepted for use among newborns to children aged 3 years is otoacoustic emission or automated auditory brainstem response and the test for children aged 3 to 6 years is the pure tone audiometry sweep test or conditioned play audiometry [10]. These tests require the cooperation of children, as well as professional audiologists, standard equipment, and standard soundproof rooms.

In our previous study [11], we tested the PASS (Preschool Audiometry Screening System) app with standard audiometric TDH39 headphones in 122 children aged 4 to 5 years. We found good overall sensitivity and specificity, indicating the potential of the app for use as a mobile hearing screening tool.

We designed the app for limited-source areas, rural areas, and areas without medical services. However, we found that audiometric TDH39 headphones were inaccessible in the rural area under study. In this study, we further evaluated the diagnostic value of the PASS-Pro app using 3 types of headphones as well as generic headphones. We selected the comparison headphones according to their availability in the Thai market and their low to moderate price range.

The objective of this study was to compare the sensitivity, specificity, κ coefficiency, and time consumption between standard conditioned play audiometry and the PASS-Pro (Preschool Audiometry Screening System–Pro) app using TDH39 headphones, Beyerdynamic DT 770 PRO headphones, and generic earmuff headphones.

Methods

Participants

Children in northeastern Thailand were recruited between November 2020 and February 2021. The inclusion criteria were as follows: children aged 4 to 5 years, who were able to communicate using the Thai language, and who were able to cooperate with the audiologists and medical staff. The exclusion criteria included pathology on the pinna and external auditory canal (which could interfere with headphone insertion), blurred or distorted vision, and blindness. The withdrawal criteria included being uncooperative during the study or incomplete hearing screening tests.

Headphone Specifications

The specifications of the headphones used in this study are described in Table 1.

Table . Headphone specifications.

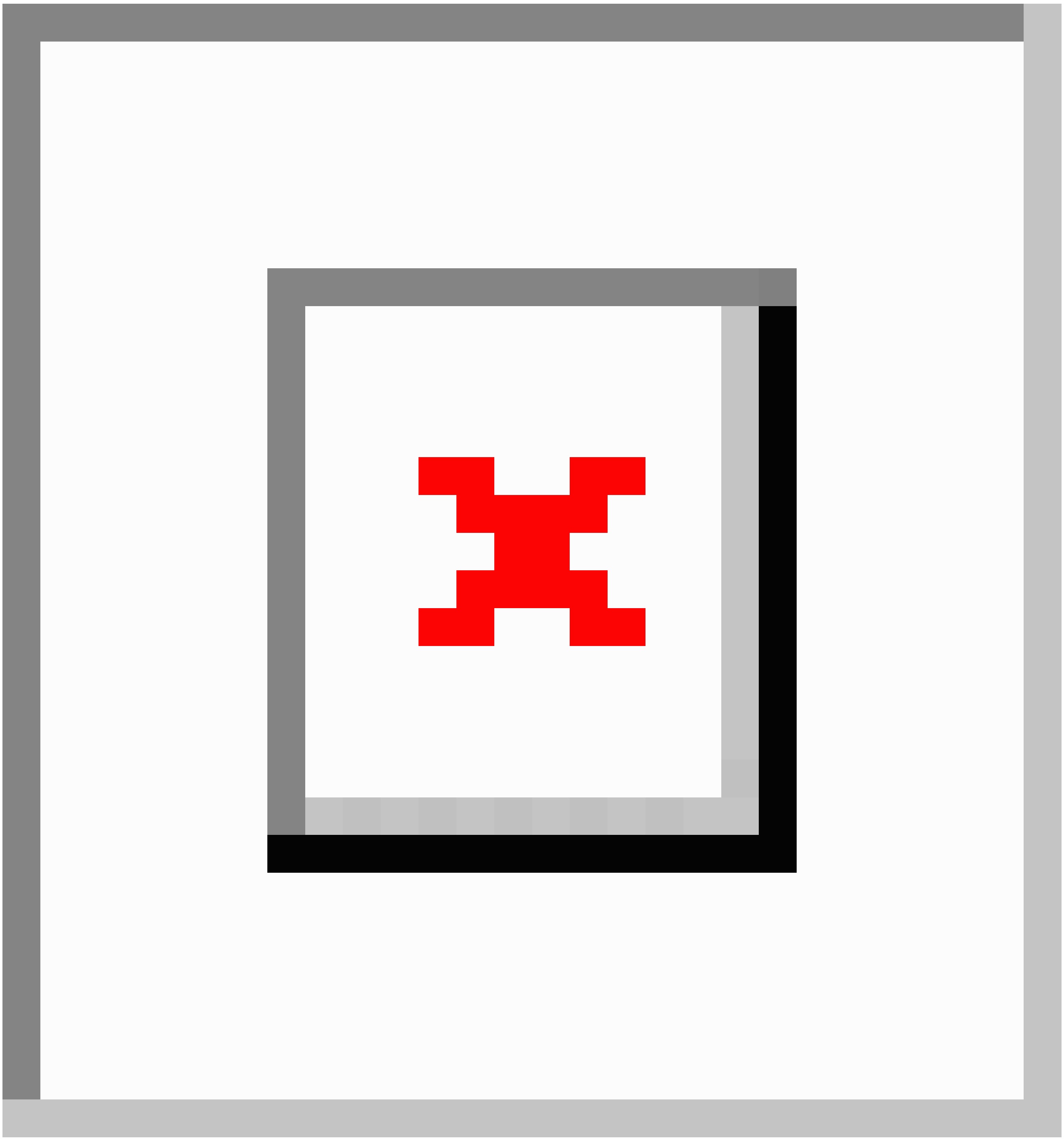
	TDH39 with earmuffs	Beyerdynamic DT 770 PRO	Generic headphones with earmuffs
Impedance	10 Ohms	32 Ohms	32 Ohms
Frequency response	100-8000 Hz	5-350,000 Hz	Up to 4 kHz with an approximate 10-dB attenuation from the maximum output
Sensitivity	108 dB SPL[a]/1 mW at 1 kHz	96 dB SPL	Estimated at 90 dB SPL
Total harmonic distortion	<1%	<0.2%	Approximately 10% at 1 kHz
Ambient noise isolation	30 dB	20 dB	30 dB

[a]SPL: sound pressure level.

Study Flow

All participants received otoscopic examination and hearing tests a total of 4 times: 3 PASS-Pro screening tests using different types of headphones (TDH39, Beyerdynamic DT 770 PRO, generic headphones with earmuffs) in a quiet room and 1 standard conditioned play audiometry in a soundproof room. All tests were randomly allocated to the participants (Figure 1).

Figure 1. The PASS-Pro (Preschool Audiometry Screening System–Pro) app with TDH39 headphones.

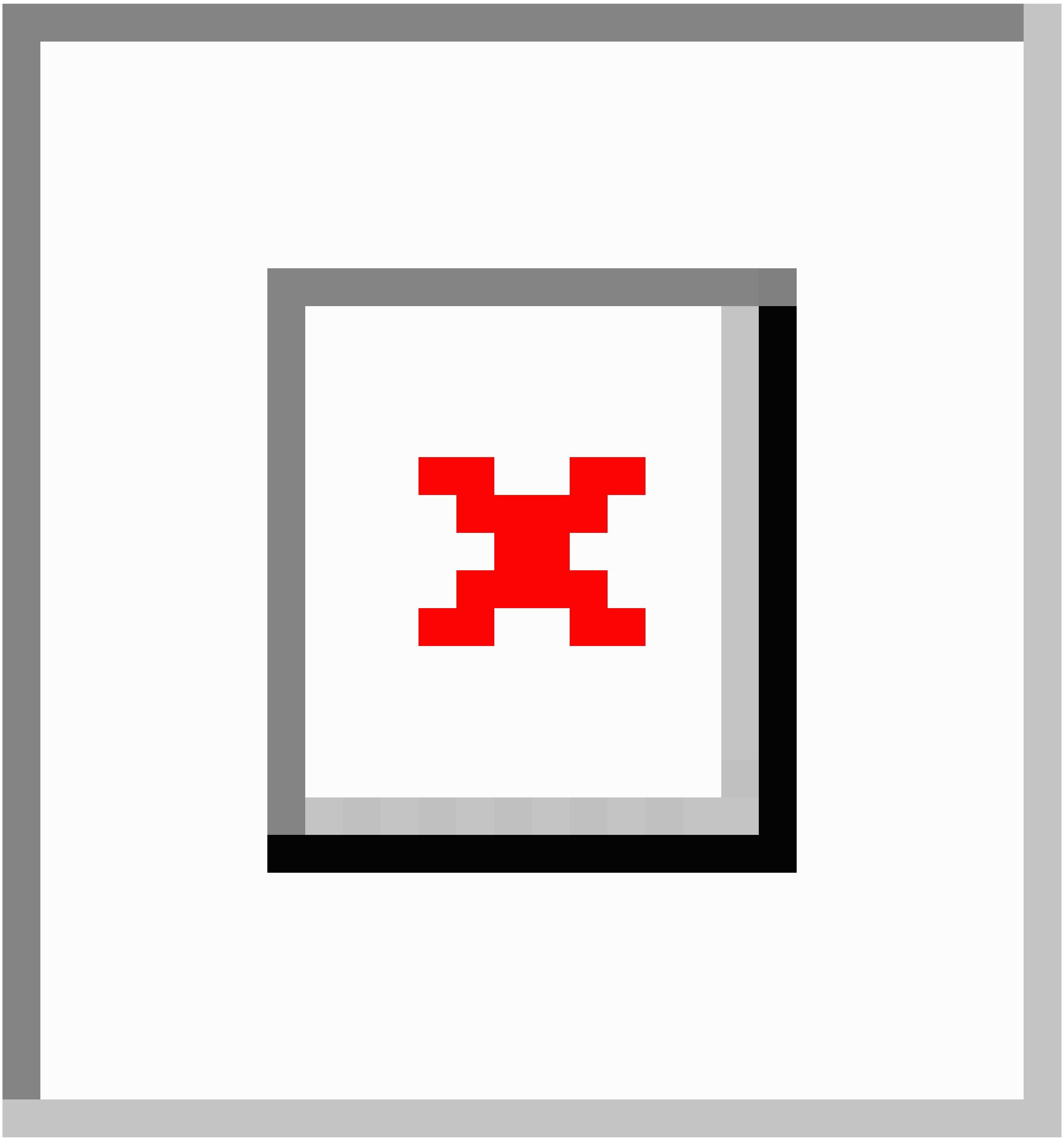


An otolaryngologist examined each child's external ears prior to the intervention. All audiologists in this study were blinded to the results of the hearing tests and the time spent doing each test to avoid bias.

Test Protocol

The PASS-Pro app was accessed through the Samsung Galaxy Tab A 10.5 Android tablet. Starting with the right ear at 40 dB, the app randomly provided 1 set of pictures (6 pictures per set), then sounded a 2-syllable word that corresponded to one of the pictures. The participant had 10 seconds to choose the correct picture after hearing the word. The app recorded whether the answer was correct or not before the next random set of pictures were displayed. The participant had to select the correct picture again for each level of hearing as 3 new sets of pictures were presented. Obtaining 2 out of 3 correct responses meant the participant had passed the hearing level. Subsequently, the app reduced the volume from 40 dB to 30 dB, then to 20 dB. The process was repeated with the left ear (Figure 2).

Figure 2. Test protocol. dB HL: dB hearing level.



The app stopped if the participant chose the correct picture fewer than 2 out of 3 sets of pictures in the same hearing level or did not choose the picture within 10 seconds. This was reported as hearing test failure for that level. The test results showed the hearing threshold and the time taken for each ear.

Reference Standard

The children's hearing threshold was evaluated using air-conduction pure tone audiometry at 0.5, 1, 2, 3, and 4 kHz. Participants with hearing loss, confirmed by the audiometry, will be sent to the otolaryngologist for standard evaluation and treatment.

Test Environment

Hearing screening with the PASS-Pro app was conducted in a standard quiet room at a hospital. The pure tone audiometry was conducted in a soundproof room.

Hearing Level Definition

According to the American Academy of Audiology's childhood hearing screening guidelines [10], a hearing level of ≤20 dB was defined as normal. A hearing level between >20 dB and ≤40 dB was defined as mild hearing loss and a hearing level of >40 dB was defined as moderate hearing loss [10].

Sound Equalization, Speech Signal Measurement, and Calibration

To calibrate the system to prepare for the trial, we first equalized the word files to have an equal root mean square. We then calibrated each word individually at 20, 30, and 40 decibel hearing level (dB HL) using the reference headphones (TDH39).

Through a 6cc coupler, a sound level meter was used to measure the peak power (in A-weighted dB, dBA) outputs from the standard audiometer and the tablet for each word at 20-, 30-, and 40-dB HL settings. The calibration coefficient calculated at each measuring point was essentially an additional gain required that would make the tablet output the word with the same dBA peak when measured using the sound level meter. The Beyerdynamic DT 770 PRO and the generic headphones with earmuffs were calibrated using the same procedure as the TDH39 headphones.

Ethical Considerations

The study was approved by the Khon Kaen University Ethical Committee for Human Research (HE631548) and was registered in the Thai Clinical Trials Registry (TCTR20201229002). Written informed consent was obtained from all participants and participants were given the option to opt out. All data were deidentified. Compensation was provided for transportation and participation in the study.

Statistical Analysis

Sample Size

The sample size was calculated from κ estimation with an expected κ value of 0.5±0.3. A power of 95% and a significance level of .05 were used. The ideal sample size was calculated to be 44, with a 20% dropout rate.

Data Analysis

Statistical analysis was conducted using STATA (StataCorp LLC). The agreement coefficiency, sensitivity, and specificity of the PASS-Pro app were determined. The McNemer test was used to compare the dichotomous data of the headphones with the reference headphones. A paired *t* test was used to compare average test duration of the app with conventional audiometry. For all tests, a *P* value of <.05 was considered to be statistically significant.

Results

A total of 44 children participated in this study, comprising 19 (43%) female and 25 (57%) male participants. No participants dropped out of the study. Using standard conditioned play audiometry, we found that 31 (70%) children had normal hearing in both ears, 9 (21%) had mild or moderate unilateral hearing loss, and 4 (9%) had mild or moderate bilateral hearing loss. An otoscopic examination was performed on all children. Of the 44 children, 4 (45%) had a normal external ear canal and tympanic membrane and 5 (55%) had impacted cerumen.

The agreement between all headphones and standard conditioned play audiometry ranged from 69.52% to 94.32%, indicating good agreement between these tools. For mild hearing loss, the κ coefficients between the PASS-Pro app using TDH39 headphones, Beyerdynamic DT 770 PRO headphones, and generic earmuff headphones versus standard conditioned play audiometry were 0.195, 0.290, and 0.261 (*P*=.02, *P*=.002, and *P*=.004), respectively. For moderate hearing loss, the κ values between TDH39 headphones, Beyerdynamic DT 770 PRO headphones, and generic earmuff headphones versus standard conditioned play audiometry were 0.206, 0.272, and 0.235 (all *P*s=.001), respectively. The κ statistic indicated fair agreement between the headphones and audiometry (Table 2 and Figure 3).

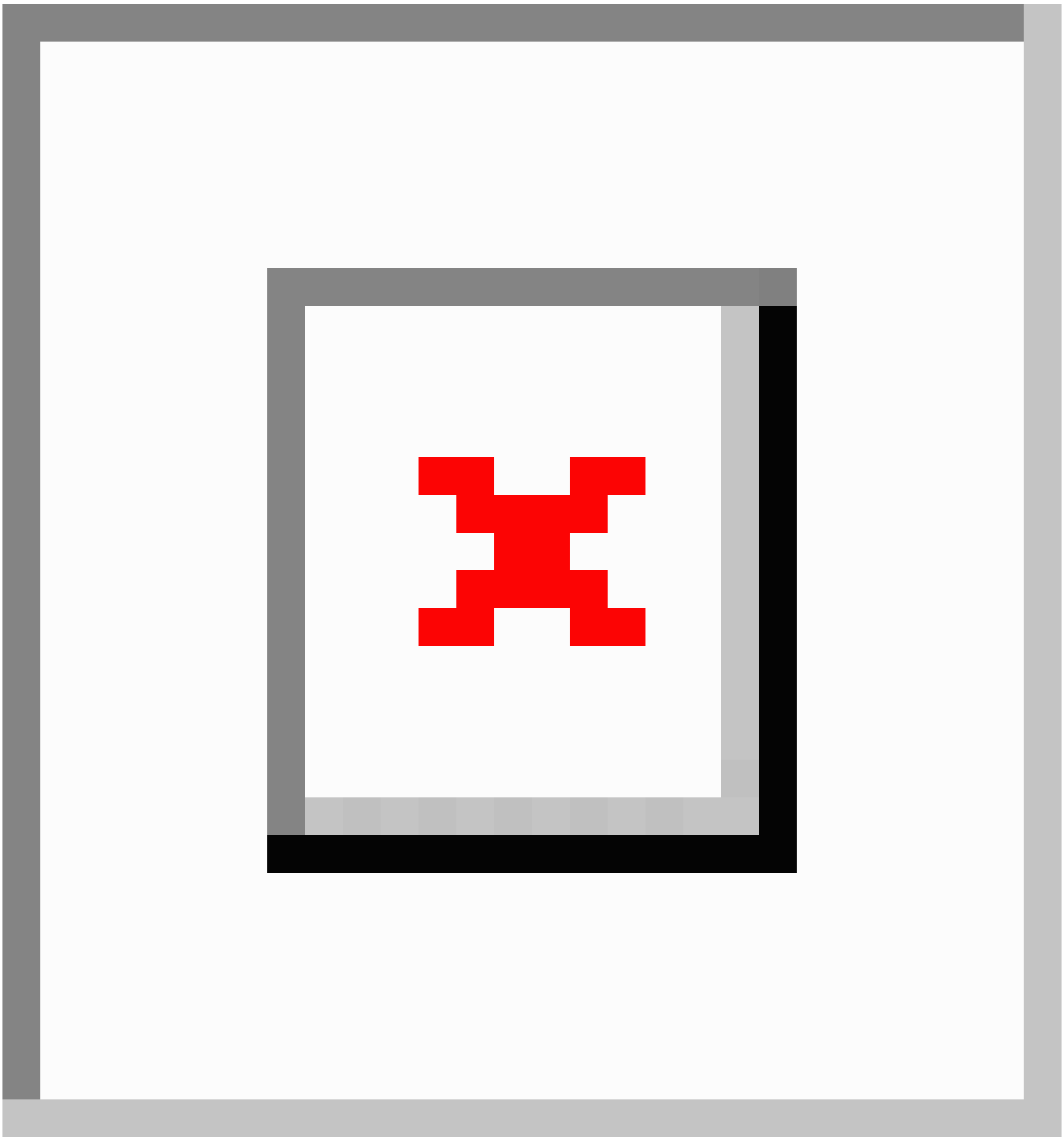
Table . Agreement between the PASS-Pro (Preschool Audiometry Screening System–Pro) app and standard conditioned play audiometry (N=88).

Headphone type and hearing level	Agreement (%)	κ coefficient	*P* value[a]
TDH39 headphones			
Normal hearing (n=69)	73.86	0.181	.04[b]
Mild hearing loss (n=11)	83.06	0.195	.02[b]
Moderate hearing loss (n=8)	92.05	0.206	.001[b]
Beyerdynamic DT 770 PRO headphones			
Normal hearing (n=70)	69.52	0.068	.26
Mild hearing loss (n=12)	87.19	0.290	.002[b]
Moderate hearing loss (n=6)	94.32	0.272	.001[b]
Generic headphones			
Normal hearing (n=62)	70.45	0.201	.02[b]
Mild hearing loss (n=19)	89.77	0.261	.004[b]
Moderate hearing loss (n=7)	91.09	0.235	.001[b]

[a]κ statistic.

[b]Indicates statistical significance.

Figure 3. Pure tone average versus PASS-Pro's (Preschool Audiometry Screening System–Pro) hearing screening levels. dB HL: dB hearing level.



The sensitivity of all headphones in detecting mild hearing loss was 50% and specificity was more than 88% whereas the sensitivity of all headphones in detecting moderate hearing loss was 100% and specificity was more than 92% (Table 3).

Table . Sensitivity and specificity of PASS-Pro (Preschool Audiometry Screening System–Pro) app (N=88).

Headphone type and hearing level		Sensitivity (95% CI)	Specificity (95% CI)	*P* value[a]
TDH39 headphones				
	Normal hearing (n=69)	37.5 (15.2-64.6)	81.9 (71.1-90)	—[b]
	Mild hearing loss (n=11)	50 (6.76-93.2)	88.1 (79.2-94.1)	—
	Moderate hearing loss (n=8)	100 (2.5-100)	92 (84.1-96.7)	—
Beyerdynamic DT 770 PRO headphones				
	Normal hearing (n=70)	25 (7.27-52.4)	81.9 (71.1-90)	.56
	Mild hearing loss (n=12)	50 (6.76-93.2)	92.9 (85.1-97.3)	.21
	Moderate hearing loss (n=6)	100 (2.5-100)	94.3 (87.1-98.1)	.45
Generic headphones				
	Normal hearing (n=62)	50 (24.7-75.3)	75 (63.4-84.5)	.12
	Mild hearing loss (n=19)	50 (6.76-93.2)	91.7 (83.6-96.6)	.37
	Moderate hearing loss (n=7)	100 (2.5-100)	93.1 (85.6-97.4)	.66

[a]McNemar test; comparison with reference headphones (TDH39).

[b]Not applicable.

The average test duration was 579.8 seconds (range 421-1152 seconds) for conditioned play audiometry, 91.27 seconds (range 58-150 seconds) for the PASS-Pro app with TDH39 headphones, 80.39 seconds (range 46-113 seconds) for the PASS-Pro app with Beyerdynamic DT 770 PRO headphones, and 84.39 seconds (range 50-130 seconds) for the PASS-Pro app with generic headphones. The PASS-Pro app used significantly less time than conditioned play audiometry (P<.001).

Discussion

In this study, we found that the agreement between audiometry (the gold standard) and all headphones ranged from 69.52% to 94.32%. The κ statistic found a statistically significant correlation coefficient (all Ps<.05) for all headphones. The sensitivity to detect mild hearing loss was 50% for all headphones while the specificity was more than 80%. Moreover, the sensitivity to detect moderate hearing loss was 100% for all headphones while the specificity was more than 90%. It can be inferred that generic headphones can be used for hearing screening.

Hearing loss among children is an important problem around the world. It significantly affects speech development, language learning skills, the thought process, communication skills, and social skills. Use of hearing screenings for early diagnosis and proper management may enhance quality of life [10,12,13]. However, in low- or middle-income countries, hearing screening tools are difficult to access [14]. There is a need to produce novel screening tools that are user-friendly, inexpensive, accessible, and practical for areas with limited resources to facilitate hearing screening.

A number of studies on hearing screening apps have been conducted, including Audioscope [15], HearCheck [9], SHOEBOX [16], and Tablet Hearing Game Screen [17]. Their sensitivity and specificity ranged from 80% to 91% and 80% to 100%, respectively.

In our previous study [11], the PASS app used with standard TDH39 headphones had a sensitivity and specificity of 76.67 and 95.83, respectively, for detecting mild hearing loss. In this study, we compared the diagnosis value of 3 different headphones. We found that the sensitivity for all headphones was 50% and specificity was more than 80% for mild hearing loss. For for moderate hearing loss, sensitivity for all headphones was 100% and specificity was more than 90%.

The discrepancy between the sensitivity results of 76.67% (95% CI 59.07%-88.21%) from our previous study versus 50% (95% CI 6.76%-93.2%) in this study may have occurred by chance. Our previous study of 122 children aimed to evaluate the sensitivity of the PASS app for hearing screening. However, the current study's main objective was not to evaluate the sensitivity and sensitivity but to evaluate the agreement between the headphones and standard audiometry.

Our results agreed with those of other hearing screening apps that used various types of headphones. For example, Audiometer used earphones or earbuds [14], SHOEBOX Audiometry used earbuds [16], and the uHear and uHearing Test used 3 types of headphones (earbud headphones, supra-aural headphones, and circumaural headphones) [18]. All exhibited high sensitivity and specificity. Table 4 presents the sensitivity and specificity of current hearing screening apps.

Table . Comparison of the sensitivity and specificity of current hearing screening apps.

Measure	PASS-Pro[a] (various head-phones)	HearCheck (Ukoumunne et al [9])	PASS[b] (Yimtae et al [11])	Tablet based (Xiao et al [17])
Sensitivity	50-100	89	76.67	91
Specificity	88.1-93.1	86.5	95.83	73.59

[a]PASS-Pro: Preschool Audiometry Screening System–Pro.

[b]PASS: Preschool Audiometry Screening System.

Although the agreement between audiometry (the gold standard) and all headphones was high, the κ correlation coefficient between these tools indicated fair agreement. This can be explained by Cohen's Kappa Paradox, which is usually found in sensitivity studies. The effects of the paradox arise when participants tend to be classified into one of the possible outcomes. This is either due to the nature of the outcome itself and its high prevalence or because at least one of the evaluators tends to assign more frequently to a specific outcome (ie, the normal hearing group) [19].

Due to the limited number of participants with hearing loss in this study, increasing the sample size in future studies should provide more accurate results. The PASS-Pro app provides only spondee words; the next software update should allow the app to produce pure tone sound for better agreement with standard audiometry.

The standard reference audiometric speaker is typically supplied in the form of headphones. The design of the headphones allows it to be soundproof, suitable for high-frequency audiometry and high passive ambient noise attenuation. However, the headphones are large, heavy, and expensive whereas other types of devices such as earbuds and earphones are more portable and lightweight. Using earbuds or earphones for hearing screening may decrease the sensitivity of the system. However, to our knowledge, no study has been conducted to quantify this sensitivity difference yet.

For future research, researchers should study commercially available portable earphones instead of generic headphones with earmuffs as the former can be directly plugged into the ear canal, are low cost, and are widely available. In this study, we limited the hearing screening test to quiet hospital rooms; therefore, the next study will be tested in a quiet community setting such as a home or school to obtain a larger sample size and emulate a real-world environment.

In conclusion, the combination of the PASS-Pro app and generic headphones is a reliable method for conducting hearing screening tests in preschool children, offering both high sensitivity and specificity.

Acknowledgments

We would like to thank the participants of this study. We also thank the National Electronics and Computer Technology Center (NECTEC) for their provision of the electroacoustics data, the PASS-Pro app, and the calibration tools. This study was funded by the NECTEC, the Faculty of Medicine, Khon Kaen University, Thailand, and the Khon Kaen Ear, Hearing, and Balance Research Group.

Conflicts of Interest

KY, P Thanawirattananit and PK were involved in the development of the PASS-Proapplication. The authors have no further interests to declare.

References

1. WHO global estimates on prevalence of hearing loss. World Health Organization. 2017 May 4. URL: apps.who.int/gb/ebwha/pdf_files/WHA70/A70_34-en.pdf [accessed 2023-10-12]
2. Watkin PM, Baldwin M. Identifying deafness in early childhood: requirements after the newborn hearing screen. Arch Dis Child 2011 Jan;96(1):62-66. [doi: 10.1136/adc.2010.185819] [Medline: 21047829]
3. Piromchai P, Kasemsiri P, Vatanasapt P, Yimtae K, Reechaipichitkul W, Thanaviratananich S. Ear, nose, throat and craniofacial diseases community services initiative of Khon Kaen University. J Med Assoc Thai 2016 Aug;99 Suppl 5:S81-S85. [Medline: 29905458]
4. Piromchai P, Laohasiriwong S, Saesiew P, Teeramaswanitch W, Reechaipichitkul W. Ear, nose, throat, and craniofacial disease screening in primary school: Khon Kaen University 2017 initiative. J Med Assoc Thai 2018 May;101(5):77 [FREE Full text]
5. Piromchai P, Saeseow P, Reechaipichitkul W, Kasemsiri P, Chayaopas N. Prevalence of ear, nose, and throat diseases in the elderly: Khon Kaen University's community service from 2017 to 2018. J Med Assoc Thai 2019;102(6):128 [FREE Full text]

6. Piromchai P, Tanamai N, Kiatthanabumrung S, Kaewsiri S, Thongyai K, Atchariyasathian V, et al. Multicentre cohort study of cochlear implantation outcomes in Thailand. BMJ Open 2021 Nov 29;11(11):e054041. [doi: 10.1136/bmjopen-2021-054041] [Medline: 34845075]
7. Shinn JR, Jayawardena ADL, Patro A, Zuniga MG, Netterville JL. Teacher prescreening for hearing loss in the developing world. Ear Nose Throat J 2021 Jun;100(3_suppl):259S-262S. [doi: 10.1177/0145561319880388] [Medline: 31608682]
8. Shearer AE, Shen J, Amr S, Morton CC, Smith RJ, Newborn Hearing Screening Working Group of the National Coordinating Center for the Regional Genetics Networks. A proposal for comprehensive newborn hearing screening to improve identification of deaf and hard-of-hearing children. Genet Med 2019 Nov;21(11):2614-2630. [doi: 10.1038/s41436-019-0563-5] [Medline: 31171844]
9. Ukoumunne OC, Hyde C, Ozolins M, Zhelev Z, Errington S, Taylor RS, et al. A directly comparative two-gate case–control diagnostic accuracy study of the pure tone screen and HearCheck screener tests for identifying hearing impairment in school children. BMJ Open 2017 Jul 11;7(7):e017258. [doi: 10.1136/bmjopen-2017-017258] [Medline: 28701413]
10. American Academy of Audiology clinical practice guidelines: childhood hearing screening. American Academy of Audiology. 2011 Sep. URL: www.audiology.org/wp-content/uploads/2021/05/ChildhoodScreeningGuidelines.pdf_5399751c9ec216.42663963.pdf [accessed 2023-10-10]
11. Yimtae K, Israsena P, Thanawirattananit P, Seesutas S, Saibua S, Kasemsiri P, et al. A tablet-based mobile hearing screening system for preschoolers: design and validation study. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e186. [doi: 10.2196/mhealth.9560] [Medline: 30355558]
12. Guidelines for the audiologic assessment of children from birth to 5 years of age. American Speech-Language-Hearing Association. 2004. URL: www.chfs.ky.gov/agencies/ocshcn/Documents/AudiologicGuidelinesfortheAssessmentofInfantsandYoungChildrenAAA2012.pdf [accessed 2023-10-10]
13. Childhood hearing loss: strategies for prevention and care. World Health Organization. 2019 Oct 17. URL: www.who.int/publications/i/item/childhood-hearing-loss-strategies-for-prevention-and-care [accessed 2023-10-10]
14. Yousuf Hussein S, Swanepoel DW, Mahomed F, Biagio de Jager L. Community-based hearing screening for young children using an mHealth service-delivery model. Glob Health Action 2018;11(1):1467077. [doi: 10.1080/16549716.2018.1467077] [Medline: 29764328]
15. Bienvenue GR, Michael PL, Chaffinch JC, Zeigler J. The AudioScope: a clinical tool for otoscopic and audiometric examination. Ear Hear 1985;6(5):251-254. [Medline: 4054439]
16. Chu YC, Cheng YF, Lai YH, Tsao Y, Tu TY, Young ST, et al. A mobile phone-based approach for hearing screening of school-age children: cross-sectional validation study. JMIR Mhealth Uhealth 2019 Apr 1;7(4):e12033. [doi: 10.2196/12033] [Medline: 30932870]
17. Xiao L, Zou B, Gao L, Weng M, Lando M, Smith AE, et al. A novel tablet-based approach for hearing screening of the pediatric population, 516-patient study. Laryngoscope 2020 Sep;130(9):2245-2251. [doi: 10.1002/lary.28329] [Medline: 31661567]
18. Szudek J, Ostevik A, Dziegielewski P, Robinson-Anagor J, Gomaa N, Hodgetts B, et al. Can Uhear me now? validation of an iPod-based hearing loss screening test. J Otolaryngol Head Neck Surg 2012 Apr;41 Suppl 1:S78-S84. [Medline: 22569055]
19. Zec S, Soriani N, Comoretto R, Baldi I. High agreement and high prevalence: the paradox of Cohen's Kappa. Open Nurs J 2017 Oct 31;11:211-218. [doi: 10.2174/1874434601711010211] [Medline: 29238424]

Abbreviations

dB HL: dB hearing level
dBA: A-weighted dB
PASS: Preschool Audiometry Screening System
PASS-Pro: Preschool Audiometry Screening System–Pro

Edited by L Buis; submitted 08.12.22; peer-reviewed by M Masalski, S Kaewsiri; revised version received 21.08.23; accepted 18.09.23; published 03.11.23.

Please cite as:
Tananuchittikul P, Yimtae K, Chayaopas N, Thanawirattananit P, Kasemsiri P, Piromchai P
App-Based Hearing Screenings in Preschool Children With Different Types of Headphones: Diagnostic Study
JMIR Mhealth Uhealth 2023;11:e44703
URL: https://mhealth.jmir.org/2023/1/e44703
doi:10.2196/44703

Original Paper

WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial

Hai-Tong Sun[1,2*], BSc; Xiao-Ru Fan[1,2*], BSc; Yu-Zhou Gu[3], MSc; Yong-Heng Lu[4], BSc; Jia-Ling Qiu[1,2], MSc; Qing-Ling Yang[1,2], MSc; Jing-Hua Li[1,2], PhD; Jing Gu[1,2], PhD; Chun Hao[1,2], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-Sen University, Guangzhou, China

[2]Sun Yat-Sen Global Health Institute, Institute of State Governance, Sun Yat-Sen University, Guangzhou, China

[3]Guangzhou Center for Disease Control and Prevention, Guangzhou, China

[4]Lingnan Community Support Center, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Chun Hao, PhD
Department of Medical Statistics, School of Public Health, Sun Yat-Sen University
74 Zhongshan 2nd Rd Yuexiu District
Guangzhou, 510080
China
Phone: 86 87332517
Email: haochun@mail.sysu.edu.cn

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2023/1/e48961/

Abstract

Background: Requesting and disclosing HIV serostatus is associated with a reduction in HIV transmission among men who have sex with men (MSM). However, the reliability of common methods for HIV serostatus request and disclosure is inadequate. Validated approaches for requesting and disclosing HIV serostatus are necessary.

Objective: The objective of this study was to investigate the use of the HIV e-report as authentic evidence of HIV serostatus among the MSM community in Guangzhou, China. Additionally, the study aimed to explore its correlation with HIV serostatus requesting and disclosure receiving behavior.

Methods: This study is a subgroup analysis of a cluster randomized controlled trial (RCT) that enrolled 357 participants during the first year. Participants in this RCT were recruited from the WeChat-based HIV testing service miniprogram developed by Guangzhou Center for Disease Control and Prevention, China. Participants completed web-based questionnaires at baseline and at the month 3 follow-up, which covered sociodemographic characteristics, HIV-related, HIV serostatus requests, receiving HIV serostatus disclosures, and HIV e-report usage.

Results: The WeChat-based HIV e-report was available in Guangzhou when the RCT project started. At the month 3 follow-up, 32.2% (115/357) of participants had their own HIV e-reports, and 37.8% (135/357) of them had received others' HIV e-reports. In all, 13.1% (27/205) and 10.5% (16/153) of participants started to use HIV e-reports to request the HIV serostatus from regular and casual male sex partners, respectively. Moreover, 27.3% (42/154) and 16.5% (18/109) of the regular and casual male sex partners, respectively, chose HIV e-reports to disclose their HIV serostatus. Compared to MSM who did not have HIV e-reports, those who had HIV e-reports and stated, "I had had my own HIV e-report(s) but hadn't sent to others" (multivariate odds ratio 2.71, 95% CI 1.19-6.86; *P*=.02) and "I had had my own HIV e-reports and had sent to others" (multivariate odds ratio 2.67, 95% CI 1.07-7.73; *P*=.048) were more likely to request HIV serostatus from their partners. However, no factor was associated with receiving an HIV serostatus disclosure from partners.

Conclusions: The HIV e-report has been accepted by the MSM community in Guangzhou and could be applied as a new optional approach for HIV serostatus requests and disclosures. This innovative intervention could be effective in promoting infectious disease serostatus disclosure among the related high-risk population.

Trial Registration: ClinicalTrials.gov NCT03984136; https://clinicaltrials.gov/show/NCT03984136

International Registered Report Identifier (IRRID): RR2-10.1186/s12879-021-06484-y

(JMIR Mhealth Uhealth 2023;11:e44513) doi:10.2196/44513

KEYWORDS

behavioral intervention; HIV serostatus disclosure; HIV testing; men who have sex with men; mHealth

Introduction

Men who have sex with men (MSM) are a population bearing a disproportionate burden of HIV infection in China [1]. It accounted for 25.5% of new HIV infections [1], and the prevalence in China reached 6% in 2020 [2]. Having unprotected anal intercourse with partners of unknown HIV status accelerates the rising rates of HIV among this population. To prevent HIV acquisition and transmission in the context of condomless sex, it is crucial that MSM are aware of their own and their partner's HIV status [3]. Such awareness necessitates frequent HIV testing and mutual HIV status disclosure before engaging in condomless sex with partners [4].

Studies indicated that over half of MSM used verbal communication and guessing for HIV serostatus request and disclosure among MSM [5,6]. While taking HIV tests together is a reliable approach to confirm partners' HIV status, some individuals doubt the reliability of HIV self-testing, and self-test kits may not always be readily available. Deception of HIV serostatus in verbal information is prevalent [7] and difficult to confirm. Trusting partners without reliable evidence will lead to an increased risk of HIV infection [6]. It is crucial for MSM to engage in more verified HIV serostatus disclosure before sexual activity, as this can effectively reduce the risk of HIV infection. However, current methods of disclosure are insufficient in meeting this need, indicating a need for more effective strategies.

WeChat, a popular social media app with over 1.2 billion active users [8], similar to Twitter or the mix of WhatsApp and Facebook, isa ubiquitous daily-use app in China [9]. WeChat miniprograms are subapps within the WeChat ecosystem. It has great potential for health intervention research [10]. In Guangzhou city, a unique and well-established WeChat miniprogram of the HIV testing service system in China has been developed by Guangzhou Centers for Disease Control and Prevention (CDC) and the MSM community-based organization Lingnan Partners Community Support Center (hereinafter called "Lingnan Center") [11]. This miniprogram served web-based HIV testing service appointments, web-based-to-offline referral, offline clinic testing, and HIV results e-report delivery after midyear 2019 coinciding with the start of this research project. The HIV result e-report is convenient and well-reserved on the internet. It is an MSM community demand-driven tool which is codeveloped by CDC and Lingnan Center. Due to the use of digital technology, an HIV e-report cannot be forged, ensuring the CDC authenticity of the serostatus results.

Regular or exchangeable HIV e-reports were applied as an intervention in our whole randomized controlled study [12]. In the context of social exchange theory [13], which posits that individuals engage in rational, reciprocal, and fair exchanges in order to achieve desired outcomes, MSM engage in HIV serostatus disclosure as a means of promoting safe sex. By disclosing their own HIV serostatus or requesting their partner's status, these principles promote mutual disclosure. If a partner's HIV serostatus is unknown, the desire for sexual activity may drive MSM to autonomously test for HIV, thus promoting HIV testing behavior.

The objective of this study is to describe the usage of the HIV e-report after it was available in Guangzhou and investigate whether it is associated with promoting HIV serostatus requests and disclosure-related behaviors among this high-risk population.

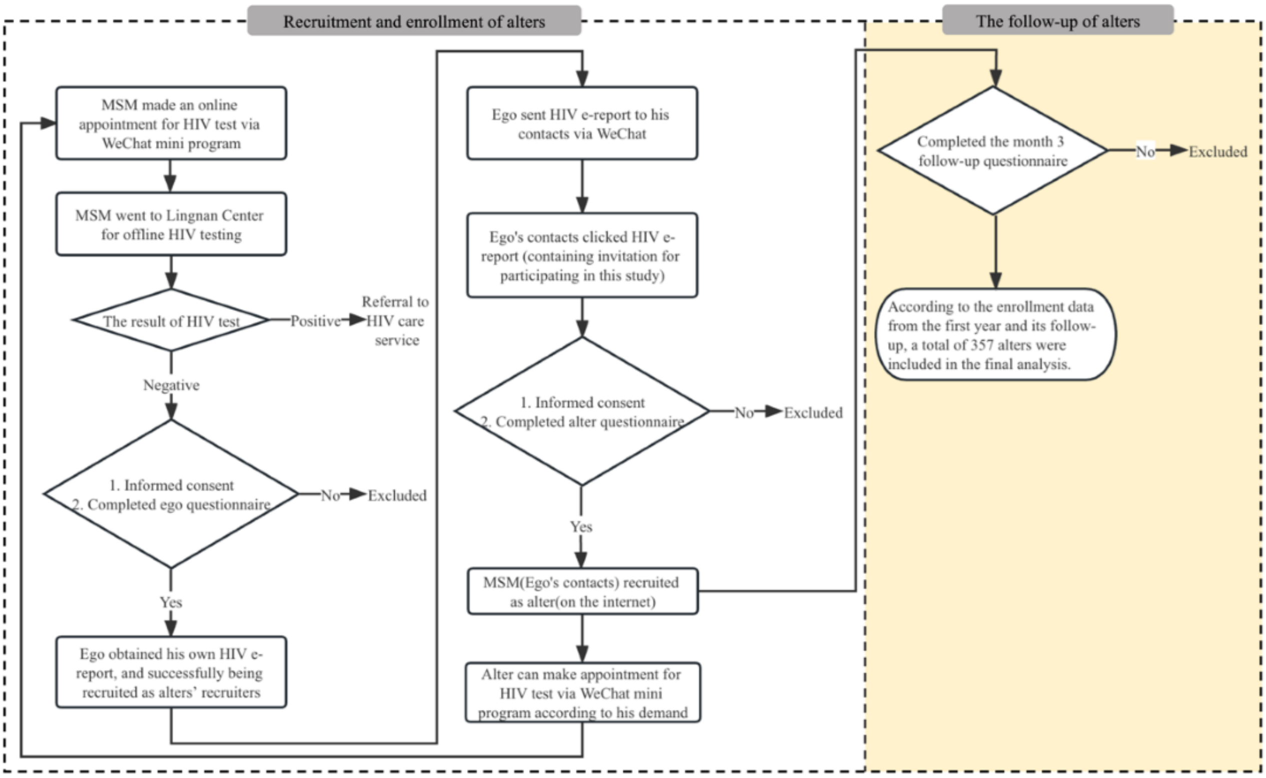
Methods

Study Design

This study is an early-stage subgroup analysis of a cluster randomized controlled trial that was conducted based on the WeChat miniprogram of HIV testing service system in Guangzhou, China (see Figure 1). Guangzhou, the capital city of Guangdong province, where some headquarters of high-tech companies sit, such as Tencent, Huawei, etc, is a megacity with 18 million population in China [14]. On account of the booming economy and tolerant culture toward homosexuality, the earliest internet-based MSM dating platform in China was established in Guangzhou in 1998 [15]. The prevalence of HIV infection in Guangzhou reached 8.27% [16]. This randomized controlled trial aims to increase HIV testing behaviors by using an HIV e-report exchange mechanism among MSM. Further information on the project can be found elsewhere [12].

MSM who were enrolled in the study at the first year and completed the month 3 follow-up were included in the analysis of this study. The HIV e-report was available to MSM in Guangzhou at the beginning of this study.

Figure 1. The recruitment and follow-up of alters. MSM: men who have sex with men.



Recruitment of Participants

The egocentric social network method [17] which involves participants nominating individuals from their social network to participate in a study, was used to recruit participants for this study. Egos were asked to nominate individuals from their social network (alters) to participate. These alters were then recruited to participate in the study.

Figure 1 shows the progress of recruitment. People who went to Lingnan Center to take HIV antibody tests were recruited as "Egos." Lingnan Center is an MSM-friendly clinic which is cooperated with Guangzhou CDC. In 2022, 10,292 HIV tests have been done in Guangzhou among MSM, and over 60% of HIV tests among MSM in Guangzhou are conducted by this clinic [11]. Every ego who tested for HIV at Lingnan Center got his own CDC-certified web-based HIV results report, hereinafter referred to as "HIV e-report." Only MSM who went to Lingnan Center for offline HIV testing can get HIV e-reports. The HIV e-report includes the basic testing information and the test result, which is certified by Guangzhou CDC (see Figure 2). Egos nominated alters by sending HIV e-reports to their contacts via WeChat.

HIV e-report receivers were invited to complete the baseline questionnaire (the link to the questionnaire was attached at the bottom of the HIV e-report) and were recruited as "alters" participants. After 3 months, alters would receive WeChat messages which contain the link to the follow-up questionnaire. Once alters went to Lingnan Center to take HIV antibody tests within 3 months, they would get their own HIV e-reports.

Egos were included in the study if they met the following inclusion criteria: (1) were 18 years of age or older, (2) engaged in anal sex with men, and (3) had confirmed negative HIV status. Alters were included in the study if they met the following inclusion criteria: (1) were 18 years of age or older, (2) were engaged in anal sex with men, (3) planned to reside in Guangzhou for the next year, and (4) had unknown or confirmed negative HIV status. Individuals who were unable to complete the questionnaire for any reason were excluded from the study.

Figure 2. The WeChat-based HIV testing service system (smartphone-based HIV test results report module). The Chinese version image depicted in this figure is the screenshot of the smartphone-based HIV test result report in the WeChat miniprogram. Only Chinese version is available.

Chinese Version English Translation

Measures

Sociodemographic Characteristics

All background characteristics of alters were collected in the baseline questionnaire.

Sociodemographic characteristics collected include age, marital status, local residence, length of stay in Guangzhou, education, and income. Participants were asked about MSM-related characteristics, including sex orientation, sex role, and whether to recruit male sex partners on the internet. We dichotomized age by 25 years and income by 5000 RMB (US $700) according to the median.

HIV-Related Information

HIV-related information was collected at the month 3 follow-up questionnaire. Participants were asked whether they had intervened by any HIV-related programs and whether they had taken up HIV antibody testing and drug use during sex in the past 3 months.

The risk perception of HIV infection was evaluated with 2 questions. One is "How do you think the prevalence of HIV infection in MSM in Guangzhou." Considering the prevalence of HIV infection in Guangzhou at 8.27% [16], responses options are "≥10%" and "<10%" as high perceived risk to HIV infection or not. The other is, "Is someone close to you infected with HIV."

HIV stigma was measured using a 7-item version of the HIV stigma scale [18], designed to measure the extent to which participants anticipated negative interpersonal and intrapersonal consequences were they to contract HIV in the future. All 7 items were rated on a Likert-type scale (1=Strongly Disagree and 4=Strongly Agree). Higher scores are indicative of greater perceived HIV-related stigma. The Cronbach α for HIV stigma was .728.

HIV testing social norms [19] were calculated using 3 survey items rated on a 4-point Likert scale in the web-based survey. All 3 items were rated on a Likert-type scale (1=Strongly Disagree and 4=Strongly Agree). Higher values indicate positive HIV testing social norms. The Cronbach α for HIV testing social norms was .737.

HIV Serostatus Request and Disclosure Receiving From Different Kinds of Male Sex Partners

Participants were asked for information about HIV serostatus in the past 3 months from regular male sex partners and casual male sex partners, respectively. Moreover, we used "or" to generate a variable called "any kinds of male sex partners." First, we asked whether they had regular and casual male sex partners in the past 3 months and whether they had unprotected anal intercourse with their male sex partners. Once they had male sex partners, we asked them how they requested the HIV serostatus of their male sex partners, including "I didn't request," "I requested orally or by message," "I requested by taking HIV test together," "I requested by sending my own HIV e-report," and "I requested by asking for partner's HIV e-report" while HIV e-report was available at the month 3 follow-up. If participants request at the month 3 follow-up, then they were asked how they received the HIV serostatus of their male sex partners by checking the response options (1=I did not receive partner's disclosure, 2=I received partner's disclosure orally or by message [namely, without any evidence], 3=I received partner's disclosure by HIV reports or HIV test kits, and 4=I received partner's disclosure by HIV e-reports).

HIV e-Reports

Participants were asked whether they had HIV e-reports, whether they received others' HIV e-reports, and whether they sent HIV e-reports to others in the past 3 months. HIV e-report has been operating in Guangzhou after the midyear of 2019 when the baseline of this study was initiated. HIV e-reports–related results were only available at the month 3 follow-up.

Statistical Analysis

Descriptive statistics were used to depict participants' sociodemographic characteristics, MSM-related information, HIV-related information, HIV serostatus request, and disclosure receiving from different kinds of male sex partners.

A percentage stacked bar graph was used to describe the manner of HIV serostatus request at baseline and the month 3 follow-up. Another bar graph was used to depict the manner of receiving HIV serostatus disclosures at the month 3 follow-up.

Univariate associations were assessed using binary logistic regression to examine each of the independent variables listed above with the 2 outcomes of HIV serostatus request behavior, which were "Had request the HIV serostatus of any kinds of male sex partners at the month 3 follow-up?" and response behaviors to HIV serostatus request, "Had received HIV serostatus disclosure from any kinds of male sex partners at the month 3 follow-up?" Subsequently, significant variables (P<.05) from the univariate logistic regression analysis were included in the multivariate logistic regression analysis. Multivariate stepwise logistic regression was applied to select the final model.

Measures of association were presented as univariate odds ratio versus multivariate odds ratio (OR_m), with 95% CI. All statistical analyses were performed using R (version 4.2.1) with 2-tailed test. A P<.05 was considered statistically significant. The packages used were dplyr, tableone, ggplot, and glm.

Ethical Considerations

The study protocol was approved by the Ethics Committee of Sun Yat-sen University (Institutional Review Board number 054/19; February 28, 2019). Informed consent was obtained from each ego at the clinic and from each alter on the internet through the WeChat-based system. The study data were collected and stored securely on the servers of the Guangzhou CDC HIV testing service system, WeChat-based database, and web-based questionnaire database, in accordance with relevant data protection laws and regulations. Access to these databases was restricted to the research team in order to ensure the confidentiality and privacy of the data.

Results

Recruitment

From September 2019 to August 2020, a total of 1607 MSM who took HIV testing at the Lingnan Center were invited to participate in this study. Of these, 1295 were recruited as Egos. Egos subsequently sent HIV e-reports to 1782 contacts via WeChat, and 702 of them (response rate, 22.3%, 702/1295) clicked the link to the questionnaire. At baseline, 397 participants were recruited as alters. Of these, 40 were lost at the month 3 follow-up whose HIV e-report information and HIV serostatus disclosure behavior information were missing. As a result, a total of 357 participants were included in the final analysis.

Characteristics of Participants

Out of 357 participants, around half of them were over 25 years of age (195/357, 54.6%), educated above high school (200/357, 56%), and earned more than 5000 RMB (US $700) per month (198/357, 55.5%). Most of them were not married to women (339/357, 95%).

Most participants were homosexual (295/357, 82.6%) and the proportion of self-identified sex role was approximately similar (122/357, 34.2% in insertive; 126/357, 35.3% in receptive; and 109/357, 30.5% in both).

A total of 79% (282/357) of participants took part in HIV-related programs in the past 3 months. HIV stigma scores ranged from8 to 23 and were at a high level (median 19, IQR 17-22) overall. On average, participants' social norm (median 3, IQR 2.67-3) inclined to a positive direction. Further details on participants characteristics participants' characteristics were presented in Table 1.

Table 1. Background characteristics of alters who enrolled at the first project year from Sep 2019 to Aug 2020.

	Alters (N=357), n (%)
Sociodemographic characteristics	
Age (years)	
≤25 years	162 (45.4)
>25 years	195 (54.6)
Currently married to a woman	
No	339 (95)
Yes	18 (5)
Guangzhou permanent resident (Hukou)	
No	106 (29.7)
Yes	251 (70.3)
How long had been in Guangzhou	
≤3 years	151 (42.3)
>3 years	206 (57.7)
Highest education obtained	
High school or below	157 (44)
Above high school	200 (56)
Personal monthly income (5000 RMB=US $700)	
≤5000	159 (44.5)
>5000	198 (55.5)
MSM[a]-related characteristics	
Self-identified sex orientation	
Bisexual, heterosexual, or not sure	62 (17.4)
Homosexual	295 (82.6)
Self-identified sex role	
Insertive only	122 (34.2)
Receptive only	126 (35.3)
Both	109 (30.5)
Male sex partner mostly recruited from the internet	
No	13 (3.6)
Yes	344 (96.4)
HIV-related information[b]	
Had intervened in any HIV-related programs in the past 3 months	
No	75 (21)
Yes	282 (79)
Had taken up HIV antibody testing in the past 3 months	
No	182 (51)
Yes	175 (49)
Had drug use during sex in the past 3 months	
No	279 (78.2)
Yes	78 (21.8)
How do you think the prevalence of HIV infection in MSM in Guangzhou	

	Alters (N=357), n (%)
<10%	79 (22.1)
≥10%	278 (77.9)
Is someone close to you infected with HIV	
No	226 (63.3)
Yes	131 (36.7)

[a]MSM: men who have sex with men.

[b]Median HIV stigma score 19 (IQR 17-22) and median HIV testing social norm score 3 (IQR 2.67-3.00).

HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving

At baseline, HIV serostatus request behaviors contained only 3 types of manners, including "I did not request," "I requested orally or by message," and "I requested by taking HIV test together." The proportions for requesting manners were 15.1% (35/232), 50.0% (116/232), and 34.9% (81/232) in 232 participants with regular male sex partners and were 18.5% (34/184), 59.2% (109/184), and 22.3% (41/184) in 184 participants with casual male sex partners, respectively (Figure 3).

At month 3 follow-up, for all 357 participants, 57.4% (205/357) of them had regular male sex partners, 42.9% (153/357) of them had casual male sex partners, and 73.4% (262/357) of them had either kind of male sex partner in the past 3 months.

After HIV e-reports were available, out of all 357 participants, 32.2% (n=115) of them had their own e-reports, and 37.8% (n=135) of them had received others' HIV e-reports at month 3 follow-up. For those who had their own HIV e-reports, 50.4% (58/115) of them had sent their own HIV e-reports to others. Therefore, 2 new request approaches for HIV serostatus using the HIV e-report emerged; namely "I requested by sending my own HIV e-report" and "I requested by asking for partner's HIV e-report." The proportions of these 2 approaches were 10.7% (22/205) and 2.4% (5/205) toward regular male sex partners, and 7.2% (11/153) and 3.3% (5/153) toward casual male sex partners, respectively. See Figure 3 and Table 2.

Participants' casual sex partners were more likely to disclose HIV serostatus without any evidence (37/109, 33.9%) and with HIV regular reports or HIV test kits (35/109, 32.1%), compared with their regular sex partners (37/158, 24% and 44/158, 28.6%). Instead, participants were more likely to receive HIV serostatus disclosure via e-reports from regular male sex partners (42/158, 27.3%) than casual male sex partners (18/109, 16.5%; Figure 4).

Figure 3. The proportion of MSM's HIV serostatus requesting manners in different types of male sex partners at baseline and the month 3 follow-up. MSM: men who have sex with men.

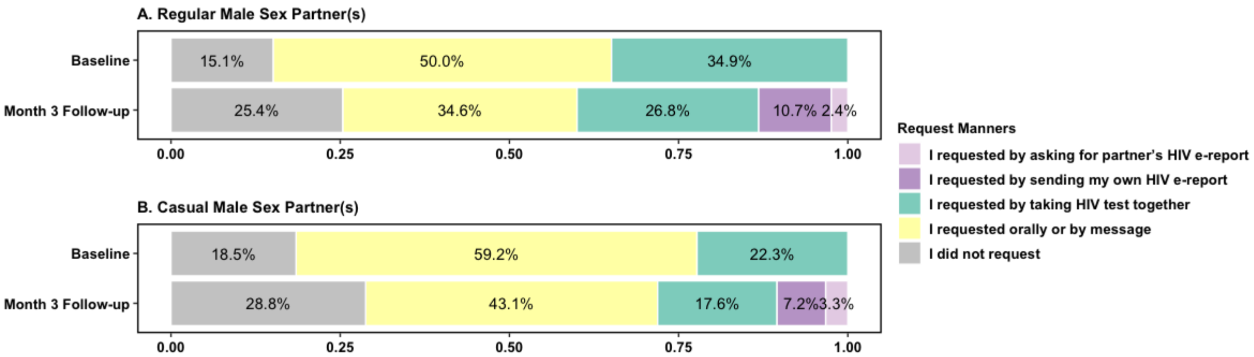


Table 2. HIV serostatus request and disclosure receiving behaviors toward different male sex partners among alters at month 3 follow-up.

	Participants (N=357), n (%)
Regular male sex partner information in the past 3 months	
Whether had regular male sex partners	
No	152 (42.6)
Yes	205 (57.4)
Whether had had UAI[a] with regular male sex partners	
No	60 (29.3)
Yes	145 (70.7)
How did you request the HIV serostatus from regular male sex partners	
I did not request	51 (24.9)
I requested orally or by message	72 (35.1)
I requested by taking HIV test together	55 (26.8)
I requested by sending my own HIV e-report	22 (10.8)
I requested by asking for partner’s HIV e-report	5 (2.4)
How did you receive disclosure of the HIV serostatus from regular male sex partners	
I did not receive partner’s disclosure	31 (20.1)
I received disclosure without any evidence	37 (24)
I received disclosure with HIV reports or HIV test kits	44 (28.6)
I received disclosure with HIV e-reports	42 (27.3)
Casual male sex partner information in the past 3 months	
Whether had casual male sex partners	
No	204 (57.1)
Yes	153 (42.9)
Whether had had UAI with casual male sex partners	
No	29 (19)
Yes	124 (81)
How did you request the HIV serostatus of casual male sex partners	
I did not request	44 (28.8)
I requested orally or by message	66 (43.1)
I requested by taking HIV test together	27 (17.6)
I requested by sending my own HIV e-report	11 (7.2)
I requested by asking for partner’s HIV e-report	5 (3.3)
How did you receive disclosure of the HIV serostatus from casual male sex partners	
I did not receive partner’s disclosure	44 (32.8)
I received disclosure without any evidence	37 (27.6)
I received disclosure with HIV reports or HIV test kits	35 (26.2)
I received disclosure with HIV e-reports	18 (13.4)
Any kinds of male sex partner information in the past 3 months	
Whether had any kinds of male sex partners	
No	95 (26.6)
Yes	262 (73.4)
Whether had UAI with any male sex partners	
No	66 (25.2)

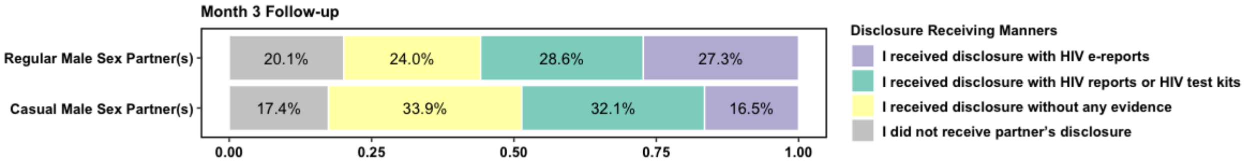
	Participants (N=357), n (%)
Yes	196 (74.8)
How did you request the HIV serostatus of any kinds of male sex partners	
I did not request	67 (25.6)
I requested orally, by message, or by taking HIV test together	160 (61)
I requested by HIV e-reports	35 (13.4)
How did you receive disclosure of the HIV serostatus from any kinds of male sex partners	
I did not receive partner's disclosure	32 (16.4)
I received disclosure with no evidence, HIV reports, or HIV test kits	112 (57.4)
I received disclosure with HIV e-reports	51 (26.2)
HIV e-reports–related information in the past 3 months	
Had sent your own HIV e-reports to others	
I didn't have my own HIV e-reports	242 (67.8)
I had my own HIV e-reports but didn't send to others[b]	57 (16)
I had my own HIV e-reports and sent to others[b]	58 (16.2)
Had received others' HIV e-reports[c]	
No	222 (62.2)
Yes	135 (37.8)

[a]UAI: unprotected anal intercourse.

[b]Had their own e-reports at the month 3 follow-up.

[c]Despite the HIV e-report which ego sent to alter for recruitment.

Figure 4. The proportion of MSM's HIV serostatus disclosure receiving manners in different types of male sex partners at the month 3 follow-up. MSM: men who have sex with men.



Factors Associated With HIV Serostatus Requests and Receiving Disclosures

In univariate logistic analysis, 5 factors were significantly (*P*<.05) associated with HIV serostatus request behavior. Further details were described in Table 3.

The multivariate stepwise regression model showed that HIV e-report–related variables, namely, having had HIV e-reports but didn't send to others (OR_m 2.71, 95% CI 1.19-6.86; *P*=.02; reference: participants who didn't have their own HIV e-reports), having had HIV e-reports and sent to others (OR_m 2.67, 95% CI 1.07-7.73; *P*=.048; reference: participants who didn't have their own HIV e-reports), and having had received others' HIV e-reports (OR_m 2.03, 95% CI 1.04-4.11, *P*=.04) were significantly associated with HIV serostatus request behavior. For variables about HIV-related information in the past 3 months, namely higher HIV testing behavior social norm scores (OR_m 2.13, 95% CI 1.12-4.17; *P*=.02), having had intervened by any HIV-related programs (OR_m 2.28, 95% CI 1.12-4.65; *P*=.02), and having taken up HIV antibody testing (OR_m 1.85, 95% CI 1.00-3.44; *P*=.05) in the past 3 months, remained in the final model as well.

All variables listed in Table 2 were not associated with receiving HIV serostatus disclosures (not tabulated).

Table 3. Univariate and multivariate regression analysis of associated factors with HIV serostatus request behavior. Univariate logistic regressions were conducted for "Had request the HIV serostatus of any kinds of male sex partners at the month 3 follow-up" and "Had any kinds of male sex partners inform you the HIV serostatus at the month 3 follow-up." All variables in were contained in univariate logistic regressions, and those with *P* value of >.10 were excluded. No factor significantly associated with "Had any kinds of male sex partners inform you the HIV serostatus at the month 3 follow-up."

	All, N	Participants, n (%)	Had requested HIV serostatus[a] from any kinds of male sex partners at the month 3 follow-up	
			OR[b] (95% CI)	OR_m[c] (95% CI)
Had sent your own HIV e-reports to other MSM in the past 3 months				
I hadn't had my own HIV e-reports	165	112 (67.9)	Reference	Reference
I had had my own HIV e-reports but hadn't sent to others	47	39 (83)	2.31 (1.05-5.63)	2.71 (1.19-6.86)
I had had my own HIV e-reports and had sent to others	50	44 (88)	3.47 (1.49-9.53)	2.67 (1.07-7.73)
Had received other MSM's HIV e-reports in the past 3 months				
No	164	112 (68.3)	reference	reference
Yes	98	83 (85.7)	2.57 (1.35-4.88)	2.03 (1.04-4.11)
Had taken up HIV antibody testing in the past 3 months				
No	120	79 (65.8)	Reference	reference
Yes	142	116 (81.7)	2.32 (1.31-4.09)	1.85 (1.00-3.44)
Had intervened in any HIV-related programs in the past 3 months				
No	51	31 (60.8)	Reference	reference
Yes	211	164 (77.7)	2.25 (1.18-4.31)	2.28 (1.12-4.65)

[a]Median HIV testing behavior social norm score was 3 (IQR 2.67-3.00) overall, while that for participants who had requested HIV serostatus from any kinds of male sex partners at month 3 follow-up was 1.5 (95% CI 1.03-1.28) (ie, OR group) and 2.13 (95% CI 1.12-4.17) (ie, OR_m group).

[b]OR: univariate odds ratio.

[c]OR_m: multivariate odds ratio. Five variables were put in the multivariate model, and Akaike's information criterion was used to select the variables in this model.

Discussion

Principal Results

e-Reports are a new approach for HIV serostatus request and disclosure for the HIV risk population. MSM chose HIV e-report as theweb-based approach to disclose their own HIV serostatus or to request partner's HIV serostatus with authenticity when it was available in Guangzhou. The most important is that HIV serostatus request behaviors were positively associated with having their own HIV e-report. To the best of our knowledge, this is the first study to discuss the use of HIV e-reports and explore its association. HIV e-report, codeveloped by the MSM community itself, could be considered a novel approach to promote mutual HIV status disclosure before engaging in sexual behaviors among HIV high-risk population, and being capable to be replicated in other countries and regions based on the ability of building information platforms.

The most interesting finding is that MSM who used HIV e-reports were more likely to actively request the HIV serostatus of their male sex partners. As e-report is a new modality in the HIV research area, studies to investigatethe association between HIV e-reports and HIV serostatus request and disclosure behaviors have rarely been reported. The possible reason for this finding is the confidence in the e-report, which is the CDC-certified credible evidence of HIV serostatus that cannot be modified. Holding an HIV e-report may make MSM feel more confident and self-assured when requesting the HIV serostatus of their male sex partners. Therefore, HIV e-report has the potential to be applied as an intervention to control the risk of HIV infection by promoting HIV serostatus disclosure. However, it is important to emphasize that the HIV e-report is not a substitute for condom use. A meta-analysis conducted in 2017 showed that seroadaptive substitution for condoms can increase the risk of HIV infection [20]. This study demonstrates the favorable influence of HIV e-report on HIV-related behaviors in the MSM community in the short term, and the long-term impact of HIV e-report among HIV risk population is expected to be explored in the future.

After the HIV e-report was available, a new proportion of MSM had used the e-report to request their sex partner's HIV serostatus (13.4%, 35/262) and had received their sex partner's e-report as an HIV serostatus disclosure (26.2%, 51/195). In addition to e-report results certificated by CDC, the significance of HIV e-report is that it is driven by MSM community demand, which correspondingly engaged in the development of HIV e-report. MSM designed it because they feel sending out their own HIV e-report is the most natural and credible way to request male sex partners' HIV serostatusas well as disclose their own HIV serostatus. According to the results, HIV e-report is becoming more acceptable within the MSM community.

We found that nearly one-third of MSM did not request the HIV serostatus of their casual male sex partners. A total of 74.9% (197/232) of alters requested their regular sex partner's HIV serostatus, while 71.5% (150/184) of alters requested their casual sex partner's HIV serostatus at baseline. Requesting a partner's HIV serostatus before engaging in sexual activity is an active coping strategy to control the risk of HIV infection. However, most research on HIV serostatus focuses on self-disclosure rather than request behavior [21-25]. Only a few studies have investigated HIV serostatus request behavior, and our data contribute to the literature [5,6]. Some influence factors have been identified in studies, such as perceived high HIV risk, lower HIV-related stigma, and greater engagement with the MSM community [26,27]. HIV serostatus request, as a proactive behavior, which can elevate the importance of consciousness about HIV serostatus disclosure before sex, has ample room for improvement among MSM.

In our study, we found that 62.2% (163/262) of participants reported receiving disclosure from male sex partners. Other studies showed that the HIV serostatus self-disclosure rate among MSM in China was approximately 20% in 2016 and 2018 [22,28]. We hypothesize that this higher disclosure rate may be due to increased awareness of the importance of disclosure over the past 5 years. We did not identify any factors associated with receiving an HIV serostatus disclosure. The possible reason may be that receiving a disclosure is a passive behavior that is primarily influenced by the characteristics of the person who disclosed their status rather than the recipient. Furthermore, our study only collected information from the recipient's perspective, which limits our ability to identify factors that may influence disclosure receiving behavior. Active coping strategies used by MSM, such as promoting the active behavior of requests by expanding the use of HIV e-reports, should be promoted.

Finally, previous studies indicated that HIV testing is positively associated with HIV serostatus request and disclosure [22,27], and our study found corroborating results. Furthermore, MSM who had higher HIV testing social norm scores and had intervened in any HIV-related programs were more likely to request HIV serostatus. This suggests that promoting both health education and HIV serostatus request and disclosure synchronously could be a promising route.

Limitations

This study is subject to several limitations. First, as the baseline and month 3 follow-up questionnaires collected alters' information in the past 3 months, the information and recall bias might exist as other HIV-related publications [29,30]. Second, there might be selection bias since participants were recruited through HIV testers from a local MSM-friendly clinic in Guangzhou and questionnaires were conducted on the mobile app, which led to participants being younger and well educated. Third, this is an observational study. Though we found several factors associated with HIV serostatus request, the causation between them needs further study.

Conclusions

This study indicated that the HIV e-report, the health service tool coproduced by community members, has become acceptable and could be used as a new optional approach for HIV serostatus request and disclosure among populations at high risk of sexually transmitted infections. In particular, the use of HIV e-report has potential influence on promoting HIV serostatus request behaviors. It is anticipated that the e-report approach will have an extended spectrum of coverage to reach more target populations and ultimately accelerate the decline of infectious disease transmission.

Acknowledgments

Y-ZG and CH were both corresponding authors for this manuscript. H-TS and X-RF contributed equally to this manuscript. H-TS wrote the manuscript and conducted all statistical analyses; X-RF conducted baseline and follow-up data collection, quality control, project implementation, and manuscript revision; Y-HL is the coprincipal investigator being responsible for the web-based part of this project, including research design, WeChat-based miniprogram design, management, and supervision; J-LQ and Q-LY participated in the research design, data collection design, project launch, and beginning of baseline data collection, beginning of baseline quality control, and beginning of baseline project implementation; JG and J-HL participated in the research design, data collection design, and part of human and funding resources support during the project implementation; Y-ZG is the coprincipal investigator being responsible for site works, including research design, WeChat-based miniprogram design, clinic and telephone data collection, management, and supervision; and CH is the principal investigator of this project, in the management and leadership of the research project. This study was supported by the National Natural Science Foundation of China (grant #71974212, #71774178, #81803334, #72204059), Guangdong Basic and Applied Basic Research Foundation (grant #2020A1515010737) and The Scientific Research Plan Project of Guangzhou (grant #202201010078).

Conflicts of Interest

None declared.

References

1. Zhang L, Chow EP, Jing J, Zhuang X, Li X, He M, et al. HIV prevalence in China: integration of surveillance data and a systematic review. Lancet Infect Dis 2013;13(11):955-963. [doi: 10.1016/S1473-3099(13)70245-7] [Medline: 24107261]

2. Country factsheets - China. UNAIDS. 2020. URL: https://www.unaids.org/en/regionscountries/countries/china [accessed 2022-06-06]
3. Eaton LA, Kalichman SC, O'Connell DA, Karchner WD. A strategy for selecting sexual partners believed to pose little/no risks for HIV: serosorting and its implications for HIV transmission. AIDS Care 2009;21(10):1279-1288 [FREE Full text] [doi: 10.1080/09540120902803208] [Medline: 20024704]
4. Marcus U, Schink SB, Sherriff N, Jones AM, Gios L, Folch C, Sialon II Network. HIV serostatus knowledge and serostatus disclosure with the most recent anal intercourse partner in a European MSM sample recruited in 13 cities: results from the Sialon-II study. BMC Infect Dis 2017;17(1):730 [FREE Full text] [doi: 10.1186/s12879-017-2814-x] [Medline: 29178847]
5. Horvath KJ, Nygaard K, Simon Rosser BR. Ascertaining partner HIV status and its association with sexual risk behavior among internet-using men who have sex with men. AIDS Behav 2010;14(6):1376-1383 [FREE Full text] [doi: 10.1007/s10461-009-9633-z] [Medline: 19921419]
6. Hotton AL, Gratzer B, Mehta SD. Association between serosorting and bacterial sexually transmitted infection among HIV-negative men who have sex with men at an urban lesbian, gay, bisexual, and transgender health center. Sex Transm Dis 2012;39(12):959-964. [doi: 10.1097/OLQ.0b013e31826e870d] [Medline: 23191950]
7. Lociciro S, Jeannin A, Dubois-Arber F. Men having sex with men serosorting with casual partners: who, how much, and what risk factors in Switzerland, 2007-2009. BMC Public Health 2013;13:839 [FREE Full text] [doi: 10.1186/1471-2458-13-839] [Medline: 24025364]
8. The second quarter performance evaluation in 2022. Tencent. URL: https://static.www.tencent.com/uploads/2022/08/17/edc989fb139819b8e6c024abde716ca0.pdf [accessed 2022-10-09]
9. No authors listed. Science on WeChat. Nat Methods 2020;17(9):863. [doi: 10.1038/s41592-020-0954-1] [Medline: 32873983]
10. Yang X, Kovarik CL. A systematic review of mobile health interventions in China: identifying gaps in care. J Telemed Telecare 2021;27(1):3-22. [doi: 10.1177/1357633X19856746] [Medline: 31319759]
11. Cheng W, Cai Y, Tang W, Zhong F, Meng G, Gu J, et al. Providing HIV-related services in China for men who have sex with men. Bull World Health Organ 2016;94(3):222-227 [FREE Full text] [doi: 10.2471/BLT.15.156406] [Medline: 26966334]
12. Shi Y, Qiu J, Yang Q, Chen T, Lu Y, Chen S, et al. Increasing the HIV testing among MSM through HIV test result exchange mechanism: study protocol for a cluster randomized controlled trial. BMC Infect Dis 2021;21(1):764 [FREE Full text] [doi: 10.1186/s12879-021-06484-y] [Medline: 34362323]
13. Nord WR. Social exchange theory: an integrative approach to social conformity. Psychol Bull 1969;71(3):174-208. [doi: 10.1037/h0027032] [Medline: 4887892]
14. Roberts T, Williams I, Preston J. The Southampton system: a new universal standard approach for port-city classification. Marit Policy Manag 2021;48(4):530-542. [doi: 10.1080/03088839.2020.1802785]
15. Wei C, Herrick A, Raymond HF, Anglemyer A, Gerbase A, Noar SM. Social marketing interventions to increase HIV/STI testing uptake among men who have sex with men and male-to-female transgender women. Cochrane Database Syst Rev 2011(9):CD009337. [doi: 10.1002/14651858.CD009337] [Medline: 21901734]
16. Liang R, Wang C, Zhao P, Luo Q, Tang S, He W, et al. [Analysis of HIV infection status and factors among men who have sex with men in Guangzhou]. Chin J Dis Control Prev 2019;23(12):1482-1486, 1497. [doi: 10.16462/j.cnki.zhjbkz.2019.12.010]
17. O'Malley AJ, Arbesman S, Steiger DM, Fowler JH, Christakis NA. Egocentric social network structure, health, and pro-social behaviors in a national panel study of Americans. PLoS One 2012;7(5):e36250 [FREE Full text] [doi: 10.1371/journal.pone.0036250] [Medline: 22615760]
18. Golub SA, Gamarel KE. The impact of anticipated HIV stigma on delays in HIV testing behaviors: findings from a community-based sample of men who have sex with men and transgender women in New York City. AIDS Patient Care STDS 2013;27(11):621-627 [FREE Full text] [doi: 10.1089/apc.2013.0245] [Medline: 24138486]
19. Pettifor A, MacPhail C, Suchindran S, Delany-Moretlwe S. Factors associated with HIV testing among public sector clinic attendees in Johannesburg, South Africa. AIDS Behav 2010;14(4):913-921. [doi: 10.1007/s10461-008-9462-5] [Medline: 18931903]
20. Purcell DW, Higa D, Mizuno Y, Lyles C. Quantifying the harms and benefits from serosorting among HIV-negative gay and bisexual men: a systematic review and meta-analysis. AIDS Behav 2017;21(10):2835-2843. [doi: 10.1007/s10461-017-1800-z] [Medline: 28573378]
21. Cao B, Saffer AJ, Yang C, Chen H, Peng K, Pan SW, et al. MSM behavior disclosure networks and HIV testing: an egocentric network analysis among MSM in China. AIDS Behav 2019;23(5):1368-1374 [FREE Full text] [doi: 10.1007/s10461-019-02404-z] [Medline: 30680538]
22. Tang W, Liu C, Cao B, Pan SW, Zhang Y, Ong J, SESH Study Group. Receiving HIV serostatus disclosure from partners before sex: results from an online survey of Chinese men who have sex with men. AIDS Behav 2018;22(12):3826-3835 [FREE Full text] [doi: 10.1007/s10461-018-2062-0] [Medline: 29470809]
23. Chen YH, Gilmore HJ, Maleke K, Lane T, Zuma N, Radebe O, et al. Increases in HIV status disclosure and sexual communication between South African men who have sex with men and their partners following use of HIV self-testing kits. AIDS Care 2021;33(10):1262-1269 [FREE Full text] [doi: 10.1080/09540121.2020.1828564] [Medline: 33021097]

24. Grov C, Rendina HJ, Moody RL, Ventuneac A, Parsons JT. HIV serosorting, status disclosure, and strategic positioning among highly sexually active gay and bisexual men. AIDS Patient Care STDS 2015;29(10):559-568 [FREE Full text] [doi: 10.1089/apc.2015.0126] [Medline: 26348322]
25. Noor SW, Rampalli K, Rosser BR. Factors influencing HIV serodisclosure among men who have sex with men in the US: an examination of online versus offline meeting environments and risk behaviors. AIDS Behav 2014;18(9):1638-1650 [FREE Full text] [doi: 10.1007/s10461-014-0774-3] [Medline: 24743960]
26. Murphy DA, de Wit JB, Donohoe S, Adam PC. The need to know: HIV status disclosure expectations and practices among non-HIV-positive gay and bisexual men in Australia. AIDS Care 2015;27(suppl 1):90-98 [FREE Full text] [doi: 10.1080/09540121.2015.1062077] [Medline: 26616130]
27. Cao W, Li J, Sun S, Sturm C, Peng L, Gu J, et al. HIV serostatus disclosure among men who have sex with men in China in the era of U=U and PrEP. AIDS Behav 2022;26(5):1477-1488 [FREE Full text] [doi: 10.1007/s10461-021-03502-7] [Medline: 34697704]
28. Gu Y, Zeng W, Luo Y, Cheng W, Zhong F, Wu H, et al. [Effect of knowledge of sexual partners? HIV status on HIV infection among men who have sex with men in Guangzhou, China]. Chin J AIDS STD 2021;27(01):33-36 [FREE Full text] [doi: 10.13419/j.cnki.aids.2021.01.09]
29. Bavinton BR, Pinto AN, Phanuphak N, Grinsztejn B, Prestage GP, Zablotska-Manos IB, Opposites Attract Study Group. Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. Lancet HIV 2018;5(8):e438-e447. [doi: 10.1016/S2352-3018(18)30132-2] [Medline: 30025681]
30. Sullivan PS, Rosenberg ES, Sanchez TH, Kelley CF, Luisi N, Cooper HL, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. Ann Epidemiol 2015;25(6):445-454 [FREE Full text] [doi: 10.1016/j.annepidem.2015.03.006] [Medline: 25911980]

Abbreviations

CDC: Center for Disease Control and Prevention
MSM: men who have sex with men
ORm: multivariate odds ratio

Edited by L Buis; submitted 22.11.22; peer-reviewed by Y Lu, F Hou; comments to author 27.02.23; revised version received 27.03.23; accepted 30.03.23; published 08.05.23.

Please cite as:
Sun HT, Fan XR, Gu YZ, Lu YH, Qiu JL, Yang QL, Li JH, Gu J, Hao C
WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e44513
URL: https://mhealth.jmir.org/2023/1/e44513
doi:10.2196/44513
PMID:37155223

Original Paper

Assessment of a Digital Symptom Checker Tool's Accuracy in Suggesting Reproductive Health Conditions: Clinical Vignettes Study

Kimberly Peven[1], BSN, MPH, PhD; Aidan P Wickham[1], MEng, MRes, PhD; Octavia Wilks[1], BSc, MPH, MBBCh; Yusuf C Kaplan[1], MD; Andrei Marhol[1], MD, PhD; Saddif Ahmed[1], BSc, MBBChB, MSc, DPhil; Ryan Bamford[1], BSc, MSc; Adam C Cunningham[1], PhD; Carley Prentice[1], BA; András Meczner[2], MD; Matthew Fenech[3], MD, PhD; Stephen Gilbert[4], PhD; Anna Klepchukova[1], MD; Sonia Ponzo[1*], BSc, MSc, PhD; Liudmila Zhaunova[1*], BSc, PhD

[1]Flo Health UK Limited, London, United Kingdom

[2]Your.MD Limited, London, United Kingdom

[3]Una Health GmbH, Hamburg, Germany

[4]Else Kröner Fresenius Center for Digital Health, TUD Dresden University of Technology, Dresden, Germany

[*]these authors contributed equally

Corresponding Author:
Aidan P Wickham, MEng, MRes, PhD
Flo Health UK Limited
27 Old Gloucester Street
London, WC1N 3AX
United Kingdom
Phone: 44 60396823
Email: a_wickham@flo.health

Abstract

Background: Reproductive health conditions such as endometriosis, uterine fibroids, and polycystic ovary syndrome (PCOS) affect a large proportion of women and people who menstruate worldwide. Prevalence estimates for these conditions range from 5% to 40% of women of reproductive age. Long diagnostic delays, up to 12 years, are common and contribute to health complications and increased health care costs. Symptom checker apps provide users with information and tools to better understand their symptoms and thus have the potential to reduce the time to diagnosis for reproductive health conditions.

Objective: This study aimed to evaluate the agreement between clinicians and 3 symptom checkers (developed by Flo Health UK Limited) in assessing symptoms of endometriosis, uterine fibroids, and PCOS using vignettes. We also aimed to present a robust example of vignette case creation, review, and classification in the context of predeployment testing and validation of digital health symptom checker tools.

Methods: Independent general practitioners were recruited to create clinical case vignettes of simulated users for the purpose of testing each condition symptom checker; vignettes created for each condition contained a mixture of condition-positive and condition-negative outcomes. A second panel of general practitioners then reviewed, approved, and modified (if necessary) each vignette. A third group of general practitioners reviewed each vignette case and designated a final classification. Vignettes were then entered into the symptom checkers by a fourth, different group of general practitioners. The outcomes of each symptom checker were then compared with the final classification of each vignette to produce accuracy metrics including percent agreement, sensitivity, specificity, positive predictive value, and negative predictive value.

Results: A total of 24 cases were created per condition. Overall, exact matches between the vignette general practitioner classification and the symptom checker outcome were 83% (n=20) for endometriosis, 83% (n=20) for uterine fibroids, and 88% (n=21) for PCOS. For each symptom checker, sensitivity was reported as 81.8% for endometriosis, 84.6% for uterine fibroids, and 100% for PCOS; specificity was reported as 84.6% for endometriosis, 81.8% for uterine fibroids, and 75% for PCOS; positive predictive value was reported as 81.8% for endometriosis, 84.6% for uterine fibroids, 80% for PCOS; and negative predictive value was reported as 84.6% for endometriosis, 81.8% for uterine fibroids, and 100% for PCOS.

Conclusions: The single-condition symptom checkers have high levels of agreement with general practitioner classification for endometriosis, uterine fibroids, and PCOS. Given long delays in diagnosis for many reproductive health conditions, which lead to increased medical costs and potential health complications for individuals and health care providers, innovative health apps and symptom checkers hold the potential to improve care pathways.

(JMIR Mhealth Uhealth 2023;11:e46718) doi:10.2196/46718

KEYWORDS

women's health; symptom checkers; symptom checker; digital health; chatbot; accuracy; eHealth apps; mobile phone; mobile health; mHealth; mobile health app; polycystic ovary syndrome; gynecology; digital health tool; endometriosis; uterus; uterine; uterine fibroids; vignettes; clinical vignettes

Introduction

Background

Millions of women and people who menstruate worldwide are affected by reproductive health conditions. Endometriosis, uterine fibroids, and polycystic ovary syndrome (PCOS) are among the most common with prevalences estimated at 10%-15%, 20%-40%, and 5%-20%, respectively [1-12]. All 3 conditions have been associated with fertility issues [12-14]. Endometriosis is a condition where endometrial tissue is found outside of the uterus and is typically characterized by painful periods, abnormal bleeding, and chronic pelvic pain, among other symptoms [5,14,15]. Uterine fibroids are benign uterine tumors that can cause a variety of debilitating symptoms, such as heavy menstrual bleeding, pain, and bladder or bowel dysfunction [12,16]. PCOS is a complex endocrine disorder characterized by a variety of symptoms, such as menstrual dysfunction and hirsutism, of differing severity and without a certain etiology [13]. These conditions can have similar presentations; for example, both pain and intermenstrual bleeding are symptoms of endometriosis and uterine fibroids, and additionally, these conditions can coexist in individuals at the same time.

Long diagnostic delays are common for endometriosis, uterine fibroids, and PCOS, with patients reporting receiving a diagnosis between 2 and 12 years from the onset of symptoms [17-22]. Controversy over diagnostic criteria may further complicate or delay final diagnosis [12,23-25]. Another contributing factor to diagnostic delays is a low level of knowledge on reproductive health as affected persons may believe symptoms are normal or hereditary, thus delaying in seeking medical input until symptoms worsen [26]. Delays in diagnosis can lead to worsening of symptoms, further health complications with fertility or psychiatric conditions, and a reduced quality of life [26-31]. Both endometriosis and uterine fibroids severely affect quality of life, everyday functioning, and workplace productivity [32-36]. A common sequela of PCOS is also a lowered quality of life [37] but also includes infertility, type 2 diabetes, and cardiovascular and psychiatric conditions (eg, hypertension, depression, and anxiety) [38].

In addition to risks of developing complications with fertility or psychiatric conditions [27-30], long diagnostic delays are associated with increased health care use and costs [39]. Endometriosis costs an average of US $27,855 per patient annually in the United States alone [40], while overall yearly expenditure for uterine fibroids is estimated to be US $34.4 billion [35]. Further, patients with long diagnostic delays for endometriosis have 60% higher mean all-cause costs compared to those with short delays [39]. Similarly, the economic costs of PCOS on individuals and health care systems are estimated to be US $8 billion per year [8]. As diagnostic costs represent a small proportion of the total economic burden of disease, particularly in light of long diagnostic delays, access to simpler screening processes may be a cost-effective strategy [41].

Innovations in health technologies and mobile apps have the potential to bridge this economic gap, deliver better health outcomes, and improve quality of life. Worldwide, there are more than 6 billion smartphone subscribers [42] and more than 350,000 health-related mobile apps [43]. As such, people increasingly turn to the internet for health information [44-46], with an increasing number of digital health interventions existing to assist with condition diagnosis (eg, check user symptoms against common condition symptoms) [47,48].

Despite the widespread availability and advantages of symptom checker apps, there remains a knowledge gap on the accuracy of many of these tools [49]. Researchers, clinicians, and patient groups are increasingly demanding more rigorous validation and evaluation of digital health solutions, with scientists highlighting the need for evidence generation [50-53]. Case vignette studies represent an established methodology for the evaluation of digital symptom checkers. In such studies, relevant fictitious patient cases are assessed by the symptom checker under investigation, and the output is compared to that of a human expert assessing the same case [54]. However, several scoping reviews have identified significant variability in study designs and reported quantitative measures when assessing symptom checkers, with about half reports describing app characteristics and half examining actual accuracy metrics, which were found to vary greatly [49,55]. A recent review of digital and web-based symptom checkers found diagnostic accuracy of the primary diagnosis varied from 19% to 38% and triage accuracy ranged from 49% to 90% [56]. Even though information on their development and validation is limited and its reliability is in question [47,49], trust in symptom checker apps is high among laypersons [57].

Flo App and Symptom Checker Development

Flo (by Flo Health UK Limited) is a health and well-being mobile app and period tracker for women and people who menstruate, with over 58 million monthly active users [58]. Flo allows users to track their symptoms throughout their menstrual cycle (eg, cramps, menstrual flow, and mood) or pregnancy and postpartum (eg, lochia), as well as general health information like contraceptive use, ovulation or pregnancy test results, water intake, and sleep. Additionally, the app offers personalized, evidence-based, and expert-reviewed content via an in-app library. Further, digital health assistants (chatbots) provide users with information about a range of conditions.

Flo has developed 3 single-condition symptom checker “chatbots” to assess symptoms of reproductive health conditions (endometriosis, uterine fibroids, and PCOS). The decision to focus on these conditions was based on their prevalence, the feasibility of symptom assessment via an app, and the multifactorial impact that these conditions can have (eg, quality of life, productivity, cardiovascular diseases, mental health conditions, and fertility). The symptom checkers use symptom information gained through conversation-like questions and answers as well as symptom or menstrual cycle information previously entered into the app. Users with acute presentations are provided with a list of red flag symptoms (eg, nausea with vomiting, fever, and vaginal bleeding not related to the period) at the beginning of the conversation and are advised to discontinue the conversation with the symptom checker and seek urgent medical advice if their presence is confirmed by the user. After the conversation, the symptom checker gives the user one of two possible outcomes: (1) a strong match for the condition—”You’re experiencing several symptoms typically associated with [condition]” or (2) a weak or no match for the condition—”While you may be experiencing some symptoms of [condition], your combination of symptoms does not strongly indicate it.” An informative summary is available for the user that reiterates which of the user’s symptoms match the presentation of a particular condition as described in the relevant clinical guidelines. This summary can then be used by the user to facilitate any subsequent conversations with their health care provider. The symptom checker is not intended as a diagnostic tool, does not provide medical advice, and users are advised to seek medical input to further investigate any concerns they have.

To ensure medical accuracy and safety during the development of symptom checkers, Flo uses a combination of an in-house medical team and external doctors specializing in the conditions of interest. The medical team builds the chat sequences considering the most relevant signs and symptoms based on the latest medical guidelines and evidence. The chat sequence is medically tested, reviewed, and adjusted in an iterative product development process.

The aim of this study was to determine the accuracy (agreement between clinician and symptom checker) of 3 symptom checkers for endometriosis, uterine fibroids, and PCOS developed using current medical guidelines (Monash, European Society of Human Reproduction and Embryology, and American Academy of Family Physicians) [24,59,60]. To this end, we devised a case vignette study whereby fictional patient cases were assessed for symptoms of the earlier-mentioned conditions by both symptom checkers and medical practitioners. We also aimed to provide a comprehensive illustration of how we created, reviewed, and categorized vignette cases in the predeployment testing and validation of digital health symptom checker tools.

Methods

Vignette Testing

Overview

Clinical case vignettes were created, reviewed, approved, classified, and entered into the symptom checkers by independent general practitioners recruited specifically for this study. Vignette cases needed to encompass presentations of not just endometriosis, fibroids, and PCOS but also other similarly presenting reproductive (eg, amenorrhea) and general health (eg, thyroid disorder) conditions. General practitioners have knowledge of a wide range of condition symptomatology and are typically the first point of contact for a patient in a health care system in the United Kingdom (where the study took place). Therefore, we reasoned that general practitioners were a more suitable choice for vignette creation, review, and classification instead of obstetricians and gynecologists.

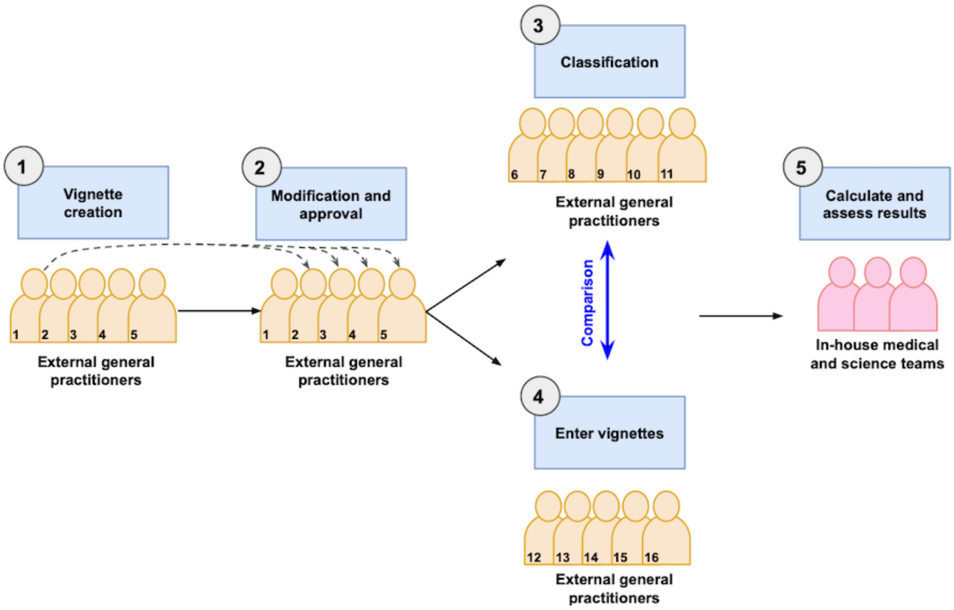
All general practitioners were UK-based with an average of 12 years of clinical experience and were not previously affiliated with Flo. All general practitioners were remunerated for their time. No human subjects, interviews, or patient-doctor transcripts were used in the creation of vignettes; all case vignettes involved in this study are fictitious and were created from each general practitioner’s experience of treating patients with these conditions.

Vignette Creation, Review, and Approval

Five external general practitioners were recruited to independently create clinical case vignettes of simulated users (Figure 1, step 1). These simulated users would be presenting for the first time without any history of diagnosis or treatment for 1 of the 3 conditions of interest, namely, endometriosis, uterine fibroids, or PCOS. Cases were derived from the general practitioners’ clinical experience and the literature. The general practitioners completed a template (Multimedia Appendix 1) for each vignette that contained information on the user’s background, history of presenting condition, medical, surgical, and family history, as well as details on their menstrual cycle and other symptoms. The general practitioners were instructed to create a set number of cases for each of the 3 conditions and for each of the 3 possible outcomes to ensure a spread of severity and condition types: (A) “You’re experiencing specific signs and symptoms commonly associated with [condition]”, (B) “Although you’re experiencing some of the potential signs and symptoms of [condition], they are not specific enough to indicate it strongly,” and (C) “You’re not experiencing any of the signs and symptoms commonly associated with [condition].” The general practitioners were instructed that “A” cases are those for which the user has specific features of the condition, and this differential diagnosis is the most likely cause of their symptoms. “B” and “C” cases are those which are considered to not have the condition. General practitioners were instructed that “B” cases represent users who show either too few or only some specific findings, and a clinician would not think of this condition as the most likely cause for these symptoms. “C” cases represent users who show either too few or nonspecific symptoms, and there would be other differential diagnoses that are more likely to be the cause of the symptoms. Condition-negative cases had other diagnoses such as urinary tract infection, thrush, pregnancy, and functional constipation.

Each vignette was reviewed by a second general practitioner (Figure 1, step 2) who could either approve the vignette as-is or suggest changes to clarify the case. If changes were suggested, the case would then be reviewed, edited, and approved by a third general practitioner who would finalize the case. In total, 24 cases were created for each condition, in line with other single-condition or single-system symptom checker evaluations [61-64].

Figure 1. Vignette study procedure including (1) independent vignette creation by 5 external general practitioners; (2) vignette review, modification, and approval by a second and third general practitioner where required; (3) independent vignette classification by 6 external general practitioners not involved in other stages; (4) entry of vignettes into symptom checkers by 5 external general practitioners not involved in other stages; and (5) analysis of results.



Independent Classification of Vignettes

After vignette approval (Figure 1, step 2), all information related to the intended designation of each vignette was removed: the type of case (A, B, or C above) and any notes about the diagnosis the creator had in mind when creating the vignette were removed from the vignette template. To avoid bias from the case creator when setting the final classification, an additional independent panel of 6 additional external general practitioners not involved in previous steps of the vignette creation was recruited to classify the vignettes (Figure 1, step 3). The classifying general practitioners received a random selection of vignettes, each designated as either an endometriosis vignette, uterine fibroid vignette, or PCOS vignette. For each vignette, the general practitioners reviewed the case and designated the most likely outcome for the specified condition (endometriosis, uterine fibroids, or PCOS) matching the symptom checker wording: (1) a strong match for the condition—"You're experiencing several symptoms typically associated with [condition]" or (2) weak or no match for the condition—"While you may be experiencing some symptoms of [condition], your combination of symptoms does not strongly indicate it." During this step of classification, to ensure there was a shared agreement on the classification of each case, each vignette was reviewed independently by 3 general practitioners; the majority view (at least 2 out of 3) was taken as the "true value" or gold-standard classification for the vignette. While the vignettes were created with 3 levels of categorization for each condition, the classifying general practitioners were not aware of these levels and were asked to make a binary classification for each vignette.

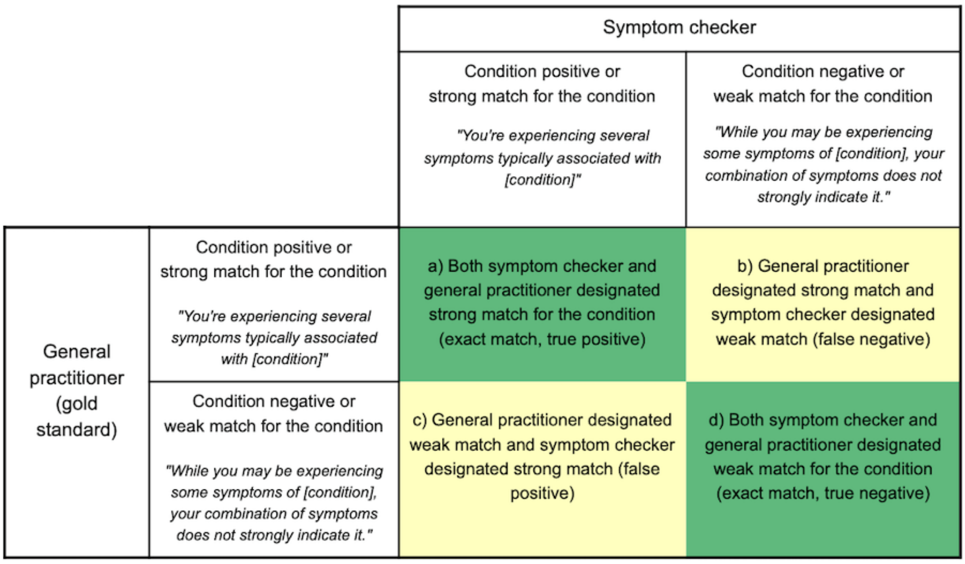
Vignette Entry

An additional set of 5 external general practitioners (not involved in the other steps) were recruited to enter the vignette cases into a prototype of the symptom checkers (Figure 1, step 4). At this stage, the general practitioners were blinded to the condition assigned to the vignette, the classification, and the condition the symptom checker was assessing. If the symptom checker asked a question that was not contained in the vignette, general practitioners were instructed to follow a step-by-step protocol to determine the appropriate answer. First, if the symptom information requested by the symptom checker was specified in the vignette template but not included by the creator, a negative response should be selected (eg, the vignette template specifies pain symptoms should include whether the radiation of pain is present, but the vignette creator does not detail this in their description of pain, then pain radiation should be assumed to be absent). If the information was not part of the template, a neutral response (eg, "I don't know" and "I don't want to answer this question") should be selected. If no neutral response was available, a negative response should be selected. If no negative response was available, the answer mostly within normal limits should be selected (eg, the inputting general practitioner would select a period length of 2-7 days, as opposed to a period length of 1 day or less or a period length of 8 days or more).

Analysis

The final classification set by the independent general practitioner classifiers (Figure 1, step 3) was compared with the outcome of the symptom checker as tested in Figure 1, step 4. Outcomes were arranged in 2-way tables as shown in Figure 2. Accuracy statistics for percent agreement between general practitioner classification and symptom checker, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as detailed: accuracy (percent agreement): (TP+TN)/(TP+TN+FP+FN); sensitivity: TP/(TP+FN); specificity: TN/(FP+TN); PPV: TP/(TP+FP); and NPV: TN/(FN+TN) (Figure 2).

Figure 2. Two-way validation table demonstrating the true positive, true negative, false positive, and false negative cases produced when comparing the symptom checker output to the general practitioner gold standard.



		Symptom checker: Condition positive or strong match for the condition *"You're experiencing several symptoms typically associated with [condition]"*	Symptom checker: Condition negative or weak match for the condition *"While you may be experiencing some symptoms of [condition], your combination of symptoms does not strongly indicate it."*
General practitioner (gold standard)	Condition positive or strong match for the condition *"You're experiencing several symptoms typically associated with [condition]"*	a) Both symptom checker and general practitioner designated strong match for the condition (exact match, true positive)	b) General practitioner designated strong match and symptom checker designated weak match (false negative)
	Condition negative or weak match for the condition *"While you may be experiencing some symptoms of [condition], your combination of symptoms does not strongly indicate it."*	c) General practitioner designated weak match and symptom checker designated strong match (false positive)	d) Both symptom checker and general practitioner designated weak match for the condition (exact match, true negative)

Results

Vignette Cases

Of the total of 24 cases that were created per condition (Table 1), 11-13 cases were classified as a strong match for the condition, and 11-13 cases were classified as a weak match for the condition after final classification by a panel (shown in Figure 1, step 3).

Table 1. Two-way validation table by condition (endometriosis [E], uterine fibroids [UF], and polycystic ovary syndrome [P]).

	Condition positive or strong match for the condition			Condition negative or weak match for the condition			Total		
	E, n	UF, n	P, n	E, n	UF, n	P, n	E, n	UF, n	P, n
General practitioner (gold standard)									
Condition positive or strong match for the condition	9	11	12	2	2	0	11	13	12
Condition negative or weak match for the condition	2	2	3	11	9	9	13	11	12
Total	11	13	15	13	11	9	24	24	24

Accuracy Metrics

Overall, exact matches (percent agreement) between the vignette classification and the symptom checker outcome ranged from 83% (20/24) for endometriosis and uterine fibroids to 88% (21/24) for PCOS (Figure 3 and Table 2). While there were no FN outcomes for PCOS, 8% (6/72) of all cases were falsely identified by the relevant symptom checker as negative for endometriosis and uterine fibroids. FP outcomes ranged from 8% (2/24) for endometriosis and uterine fibroids to 13% (3/24) of all cases for PCOS. An example vignette case showing a TP, TN, FP, and FN case (determined by agreement between general practitioner and symptom checker) is provided for each condition in Multimedia Appendix 2.

Figure 3. Overall symptom checker performance showing the proportion of false-positive outcomes, exact match outcomes, and false-negative outcomes by condition. PCOS: polycystic ovary syndrome.

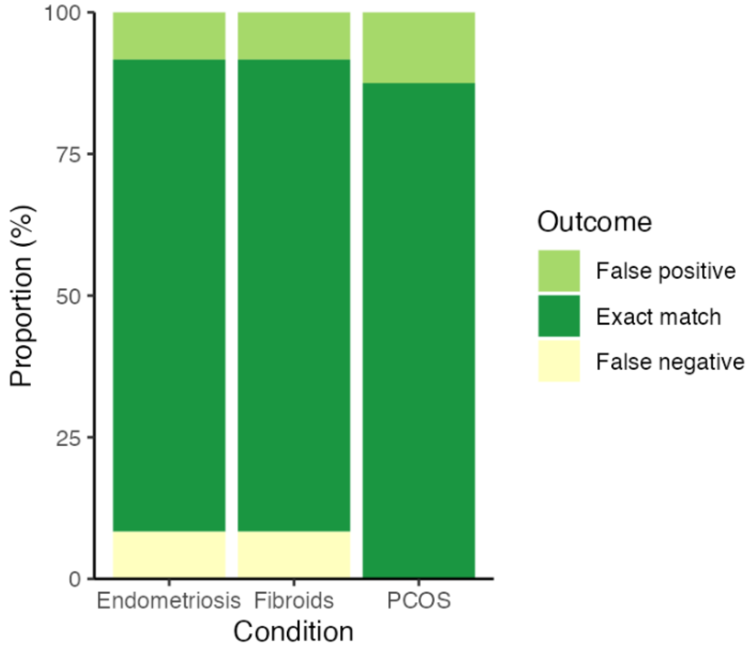


Table 2. Accuracy metrics for endometriosis, fibroids, and polycystic ovary syndrome (PCOS).

Condition	Values, n	Agreement (%)	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)
Endometriosis	24	83.3	81.8	84.6	81.8	84.6
Fibroids	24	83.3	84.6	81.8	84.6	81.8
PCOS	24	87.5	100	75	80	100

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

While sensitivity was very high (100%) for PCOS (Table 2), specificity was high for all 3 conditions (>81%). PPV ranged from 80% for PCOS to 84.6% for uterine fibroids, while NPV ranged from 81.8% for uterine fibroids to 100% for PCOS.

Discussion

Summary

In this study, we provide an example methodology for the creation, review, and classification of vignette cases for the testing and validation of digital health symptom checker tools. The percent agreement between general practitioner-designated vignette cases and 3 single-condition symptom checkers for 3 reproductive health conditions (endometriosis, fibroids, and PCOS) was assessed. We found the designation given to case vignettes by the symptom checkers had high levels of agreement between general practitioner and symptom checker (83.3%-87.5%), sensitivity (81.8%-100%), specificity (75%-84.6%), PPV (80%-84.6%), and NPV (81.8%-100%) when compared to gold standard designation by general practitioners. Overall, these metrics show the high performance of the symptom checkers when tested on robustly designed clinical vignettes.

Comparison With Prior Work

This high accuracy of the identification of reproductive health conditions is particularly important as high rates of diagnostic error are reported by patients. A study of patients with self-reported surgically confirmed endometriosis found that 75.2% of patients reported being misdiagnosed with another physical health or mental health problem by their health care professional [65]. A similar study of patients diagnosed with PCOS found that 33.6% of women reported >2 years time to diagnosis, 47.1% visited ≥3 health professionals before a diagnosis was established, and 64.8% were dissatisfied with the diagnostic process [66]. The use of a tool like a symptom checker could give the user better knowledge and awareness of their symptoms in conversations they may have with their health care provider, leading to a more effective diagnostic pathway. We have shown in previous research that users agree Flo increases their knowledge of the menstrual cycle and facilitates easier conversations with their health care provider [67].

Other vignette studies of multicondition symptom checkers have shown mixed results for accuracy. A study by Gilbert et al [68] comparing urgency advice (ie, triage) from 7 multicondition symptom checker apps and 7 general practitioners to gold-standard vignettes found that the condition suggested first matched the gold standard (ie, M1 accuracy) for 71% of general practitioners and 26% of apps; when broadening to the condition suggested in the top 5 (ie, M5 accuracy), the accuracy of general practitioners rose to 83% and apps to 41%. Another study by Schmieding et al [69] comparing 22 symptom checkers using 45 vignettes found M1 accuracy of 46%, and M10 accuracy was 71%.

The multicondition symptom checkers evaluated by Gilbert et al [68] and Schmieding et al [69] were assessed using vignette cases that covered both common and less-common conditions seen in primary care practice, conditions that affect all body systems, and conditions that have a range of urgency levels. Further, these evaluated symptom checkers are designed to detect a wide range of conditions for a general population. In contrast, this study evaluated single-condition symptom checkers using vignettes specifically designed to represent presentations with specific symptoms of the condition (strong match or condition positive) and presentations with symptoms not specific to the condition (weak match or condition negative). This symptom checker design difference may explain the variation in accuracy found between our symptom checkers (single condition) and other studied symptom checkers (multicondition).

Evaluations of single-condition symptom checkers include a study of 12 web-based symptom checkers for COVID-19 [70] and a study of an app-based symptom checker for PCOS [62]. COVID-19 symptom checkers ranged widely in both sensitivity (14%-94%) and specificity (29%-100%), with only 4 symptom checkers having both sensitivity and specificity above 50% and 2 with both sensitivity and specificity above 75%. Sensitivity and specificity in our symptom checkers were between 75% and 100%. The PCOS symptom checker evaluated by Rodriguez et al [62] reported 12%-25% FP cases and no FN out of 8 cases tested. Our PCOS symptom checker had no FNs and 3 (13%) FP cases out of 24 cases tested. Our PPV and NPV values were 80%-100% for our 3 symptom checkers, suggesting relatively high chances that positively tested cases truly have the condition in question.

With the exception of COVID-19, which has a symptomatology and overall presentation that differs greatly from the reproductive health disorders assessed in this study, digital or app-based symptom checkers for a single condition are uncommon. Symptom-based patient-completed questionnaires and screening tools do exist, including for common reproductive health conditions such as endometriosis or PCOS. A patient self-assessment tool for endometriosis with 21 questions found sensitivity of 76% and specificity of 72%, PPV of 73%, and NPV of 75% [71]. Our endometriosis symptom checker had a similar but slightly higher sensitivity (81.8%), specificity (84.6%), PPV (81.8%), and NPV (84.6%). A 4-item questionnaire for use in the diagnosis of PCOS among women with a primary complaint of infertility had 77% sensitivity and 94% specificity, and a PPV and NPV calculated from their data as 87% and 88%, respectively [72]. Our PCOS symptom checker had higher sensitivity (100%), lower specificity (75%), higher NPV (100%), and lower PPV (80%), prioritizing the identification of cases. It should be noted, however, that our symptom checker is designed to be for a broader population than the 4-item clinical tool, including those who are not trying to get pregnant or experiencing fertility issues. Questionnaires such as these have some limitations. They may not be available to the public and additionally may be subject to more user error (eg, question skipping). App-based symptom checkers, on the other hand, can use historical data from users such as menstrual regularity to improve the accuracy of user answers. Additionally, users cannot accidentally skip questions, and the app will provide a detailed summary of results and recommendations.

It is not uncommon for variation of opinion between groups of general practitioners reviewing vignettes with El-Osta et al [73] reporting classification agreement between 3 general practitioners' primary diagnosis and the intention of the vignette being 32.4%. Each vignette in this study was reviewed independently by 3 different general practitioners, and in 71% (51/72) of cases, all 3 general practitioners agreed with the vignette assignment given at the vignette approval stage (Figure 1, step 2). However, it should be noted that El-Osta et al [73] provided a comparison between primary diagnosis of general practitioners and original vignette intention, whereas this analysis only concerns strong or weak match for known reproductive conditions. All 3 general practitioners agreed with each other (regardless of the vignette intention) for 81% (58/72) of vignette cases. This disagreement between general practitioners and some differences with the symptom checker results are to be expected, particularly when using symptom-based assessment for reproductive health conditions that can be complicated to diagnose, have overlapping symptomatology with other system conditions such as gastrointestinal and urinary conditions, and are often dismissed or considered to be "normal" variations in the menstrual cycle by some. These conditions have a notoriously prolonged time to diagnosis [16-19] and require investigations including imaging. Further, the sensitivity of different testing methods can vary. For example, physical examination for deep infiltrating endometriosis can have poor accuracy and requires imaging [74].

The possible applications of symptom checkers and health apps are far-reaching and could have benefits at the individual user level, health care professional level, and macro or health system level [63,75]. Especially for many reproductive health conditions where the time to diagnosis is currently long and contributes to high health care costs [17,26,66,76], an earlier diagnosis can lead to early treatment and thus decrease complications from untreated conditions and decrease health care costs of treating more advanced disease [27,28,39]. Menstrual cycle details such as cycle length, period length, or flow can be important information for health care providers when diagnosing patients. Health apps can help track cycle details over time and use these details when determining risk for conditions as well as in summary information for users to share with their health care providers (eg, the Flo app provides a "health report" where you can download a summary of symptoms over a period of time, average cycle length, and other details to share with a health care provider). Additionally, as people with symptoms such as heavy bleeding or menstrual pain may believe these are normal or hereditary [26], personalized assessment of symptoms and encouragement to seek further evaluation from a medical professional where appropriate may improve an individual's understanding of their symptoms and health status and decrease time to diagnosis. Our prior research has demonstrated that 58%

of Flo users report improvements in understanding the normality of certain cycle-related symptoms and recognizing the abnormal nature of others, while 1 in 3 Flo users reported that the use of the Flo app improved their communication with their health care provider [67]. Therefore, mobile apps with symptom checkers could identify users with risk factors for certain conditions, educate users about their symptoms, and further encourage conversation with their medical providers.

Strengths and Limitations

Strengths of this study include the use of different groups of independent, external general practitioners unfamiliar with the symptom checkers to create, enter, and classify case vignettes for symptom checker testing. Additionally, vignettes were created with a wide range of symptomatology to ensure the inclusion of borderline presentations as these are notoriously difficult to assess, even for doctors, although they represent a frequent reality as people do not often fit neatly into textbook case presentations. Further, each vignette case was reviewed by an independent, experienced general practitioner and classified by a separate panel viewing the vignettes for the first time. When generating vignette cases that represent typical presentations of a single condition as seen by a general practitioner, there will only be so many permutations of symptomatology that can be generated before repetitions of vignette cases occur; as a result, we created 72 vignette cases in total, 24 for each of our 3 conditions. The number of vignettes needed to evaluate symptom checkers is not well defined [54]. Other vignette symptom checker evaluations have used between 3 and 400 cases for testing, with single-condition or single-system evaluations (eg, mental health, ophthalmology, and PCOS) using fewer cases and multicondition evaluations using larger numbers of cases [61-63,77,78]. Among the 400 vignettes published by Hammoud et al [78], any single condition is only represented by at most 5 cases.

Limitations, however, should be noted. Vignette studies rely on clinical opinion of a small number of general practitioners. An audit study of clinical vignette benchmarking has shown significant variation between groups of general practitioners considering clinical vignettes [73]. To decrease bias from differences in clinical opinion, all cases were blindly reviewed by 3 general practitioners, one-third involved in cases of disagreement. We found agreement between all 3 general practitioners in 81% (58/72) of our cases. Vignettes also rely on the classical presentation of conditions that may present differently in real life or in patients with complex or atypical condition presentations. When creating vignettes, we recognize there is a possibility for bias, expected patterns, or entrenched unknowns in the understanding of each condition's symptomatology. Additionally, although we recognize that patients do not usually present to primary care practitioners with a prespecified suspected diagnosis and that therefore this aspect of the study design does not reflect usual medical practice, these chatbots are not meant to replace the interaction with primary care providers but rather to allow users to review their symptoms in advance of seeing a health care professional. As outlined in the medical guidelines, symptom severity, risk, and prevalence may vary across world regions and ethnicities. Neither the Flo app nor the symptom checker collect data on the user's race or ethnicity, so neither race nor ethnicity were included in the vignette creation process. In addition to this, the vignettes in this study were created by panels of general practitioners based in the United Kingdom and may not provide an accurate representation of symptoms for every cultural context. We recognize the inclusion of such information could help to identify at-risk individuals better.

While we found 100% sensitivity for our PCOS symptom checker, it is likely with a larger sample size and real-life cases, this level of perfect sensitivity will not be maintained. Other changes in accuracy statistics are likely to be seen in real-world use. Further, as real-world users may interpret their symptoms and the questions differently than doctors, future studies including the general population should be carried out to test each symptom checker's performance in the context of real-world deployment. The use of vignette-patient cases is an important part of predeployment algorithm testing for digital symptom checkers [54,79] and is the first stage of our symptom checker evaluation. The next stage in evaluating our symptom checkers will include the use of real-world data such as observational studies of condition diagnosed and undiagnosed people's symptoms and early field-testing of the in-app symptom checkers on users and comparing the output to an official diagnosis from a doctor. Evaluation of symptom checkers and digital health tools should follow multistage processes with increasing exposure to real environments exploring not only effectiveness but also usability and balance between probability of disease and risk of missing a diagnosis [79].

Conclusions

In conclusion, we have described a methodology for creating and classifying vignettes using multiple independent panels of general practitioners for the predeployment testing of digital health symptom checker tools. We found high levels of agreement between general practitioner classification and single-condition symptom checkers for 3 reproductive health conditions (endometriosis, fibroids, and PCOS). Given long delays in diagnosis for many reproductive health conditions, which lead to increased medical costs and potential health complications, innovative health apps and symptom checkers hold the potential to improve care pathways.

Acknowledgments

This study was funded by Flo Health UK Limited.

Conflicts of Interest

KP, APW, OW, YCK, A Marhoi, SA, ACC, AK, SP, and LZ were employees at Flo Health, Inc and have stock ownership in the company. RB, CP, A Meczner, MF, and SG are paid consultants for Flo Health, Inc. SG declares no nonfinancial interests but the following competing financial interests: he has or has had consulting relationships with Una Health GmbH, Lindus Health Ltd, Flo Health UK Limited, Thymia Ltd, and Ada Health GmbH and holds share options in Ada Health GmbH. MF declares no nonfinancial interests but the following competing financial interests: he has a consulting relationship with Flo Health UK Limited and holds share options in Una Health GmbH. A Meczner declares no nonfinancial interests but the following competing financial interests: he is an employee and shareholder at Healthily or Your.MD.

Multimedia Appendix 1
Vignette template.
[DOCX File , 19 KB - mhealth_v11i1e46718_app1.docx]

Multimedia Appendix 2
Example case vignettes with matching and mismatching classification between general practitioners and the symptom checker (SC).
[DOCX File , 62 KB - mhealth_v11i1e46718_app2.docx]

References

1. Azziz R, Carmina E, Chen Z, Dunaif A, Laven JSE, Legro RS, et al. Polycystic ovary syndrome. Nat Rev Dis Primers 2016 Aug 11;2:16057 [FREE Full text] [doi: 10.1038/nrdp.2016.57] [Medline: 27510637]
2. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. Hum Reprod 2016;31(12):2841-2855 [FREE Full text] [doi: 10.1093/humrep/dew218] [Medline: 27664216]
3. Carmina E, Azziz R. Diagnosis, phenotype, and prevalence of polycystic ovary syndrome. Fertil Steril 2006;86(Suppl 1):S7-S8 [FREE Full text] [doi: 10.1016/j.fertnstert.2006.03.012] [Medline: 16798288]
4. Deswal R, Narwal V, Dang A, Pundir CS. The prevalence of polycystic ovary syndrome: a brief systematic review. J Hum Reprod Sci 2020;13(4):261-271 [FREE Full text] [doi: 10.4103/jhrs.JHRS_95_18] [Medline: 33627974]
5. Ellis K, Munro D, Clarke J. Endometriosis is undervalued: a call to action. Front Glob Womens Health 2022;3:902371 [FREE Full text] [doi: 10.3389/fgwh.2022.902371] [Medline: 35620300]
6. Eskenazi B, Warner ML. Epidemiology of endometriosis. Obstet Gynecol Clin North Am 1997;24(2):235-258 [FREE Full text] [doi: 10.1016/s0889-8545(05)70302-8] [Medline: 9163765]
7. Rawson JM. Prevalence of endometriosis in asymptomatic women. J Reprod Med 1991;36(7):513-515. [Medline: 1834839]
8. Riestenberg C, Jagasia A, Markovic D, Buyalos RP, Azziz R. Health care-related economic burden of polycystic ovary syndrome in the United States: pregnancy-related and long-term health consequences. J Clin Endocrinol Metab 2022;107(2):575-585 [FREE Full text] [doi: 10.1210/clinem/dgab613] [Medline: 34546364]
9. Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. BMC Med 2010;8:41 [FREE Full text] [doi: 10.1186/1741-7015-8-41] [Medline: 20591140]
10. Waller KG, Lindsay P, Curtis P, Shaw RW. The prevalence of endometriosis in women with infertile partners. Eur J Obstet Gynecol Reprod Biol 1993;48(2):135-139 [FREE Full text] [doi: 10.1016/0028-2243(93)90254-a] [Medline: 8491333]
11. Stewart EA. Uterine fibroids. Lancet 2001;357(9252):293-298 [FREE Full text] [doi: 10.1016/S0140-6736(00)03622-9] [Medline: 11214143]
12. Khan AT, Shehmar M, Gupta JK. Uterine fibroids: current perspectives. Int J Womens Health 2014;6:95-114 [FREE Full text] [doi: 10.2147/IJWH.S51083] [Medline: 24511243]
13. Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, Yildiz BO. The prevalence and features of the polycystic ovary syndrome in an unselected population. J Clin Endocrinol Metab 2004;89(6):2745-2749 [FREE Full text] [doi: 10.1210/jc.2003-032046] [Medline: 15181052]
14. Bulletti C, Coccia ME, Battistoni S, Borini A. Endometriosis and infertility. J Assist Reprod Genet 2010;27(8):441-447 [FREE Full text] [doi: 10.1007/s10815-010-9436-1] [Medline: 20574791]
15. Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 2005;20(10):2698-2704 [FREE Full text] [doi: 10.1093/humrep/dei135] [Medline: 15980014]
16. Stewart EA, Cookson CL, Gandolfo RA, Schulze-Rath R. Epidemiology of uterine fibroids: a systematic review. BJOG 2017;124(10):1501-1512 [FREE Full text] [doi: 10.1111/1471-0528.14640] [Medline: 28296146]
17. Kiesel L, Sourouni M. Diagnosis of endometriosis in the 21st century. Climacteric 2019;22(3):296-302 [FREE Full text] [doi: 10.1080/13697137.2019.1578743] [Medline: 30905186]

18. Hudelist G, Fritzer N, Thomas A, Niehues C, Oppelt P, Haas D, et al. Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. Hum Reprod 2012;27(12):3412-3416 [FREE Full text] [doi: 10.1093/humrep/des316] [Medline: 22990516]
19. Husby GK, Haugen RS, Moen MH. Diagnostic delay in women with pain and endometriosis. Acta Obstet Gynecol Scand 2003;82(7):649-653 [FREE Full text] [doi: 10.1034/j.1600-0412.2003.00168.x] [Medline: 12790847]
20. Gibson-Helm ME, Lucas IM, Boyle JA, Teede HJ. Women's experiences of polycystic ovary syndrome diagnosis. Fam Pract 2014;31(5):545-549 [FREE Full text] [doi: 10.1093/fampra/cmu028] [Medline: 24925927]
21. Stewart EA, Nicholson WK, Bradley L, Borah BJ. The burden of uterine fibroids for African-American women: results of a national survey. J Womens Health (Larchmt) 2013;22(10):807-816 [FREE Full text] [doi: 10.1089/jwh.2013.4334] [Medline: 24033092]
22. Borah BJ, Nicholson WK, Bradley L, Stewart EA. The impact of uterine leiomyomas: a national survey of affected women. Am J Obstet Gynecol 2013;209(4):319.e1-319.e20 [FREE Full text] [doi: 10.1016/j.ajog.2013.07.017] [Medline: 23891629]
23. Spaczynski RZ, Duleba AJ. Diagnosis of endometriosis. Semin Reprod Med 2003;21(2):193-208. [doi: 10.1055/s-2003-41326] [Medline: 12917789]
24. Becker CM, Bokor A, Heikinheimo O, Horne A, Jansen F, Kiesel L, et al. ESHRE guideline: endometriosis. Hum Reprod Open 2022;2022(2):hoac009 [FREE Full text] [doi: 10.1093/hropen/hoac009] [Medline: 35350465]
25. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Hum Reprod 2018;33(9):1602-1618 [FREE Full text] [doi: 10.1093/humrep/dey256] [Medline: 30052961]
26. Ghant MS, Sengoba KS, Vogelzang R, Lawson AK, Marsh EE. An altered perception of normal: understanding causes for treatment delay in women with symptomatic uterine fibroids. J Womens Health (Larchmt) 2016;25(8):846-852 [FREE Full text] [doi: 10.1089/jwh.2015.5531] [Medline: 27195902]
27. Ballweg ML. Impact of endometriosis on women's health: comparative historical data show that the earlier the onset, the more severe the disease. Best Pract Res Clin Obstet Gynaecol 2004;18(2):201-218 [FREE Full text] [doi: 10.1016/j.bpobgyn.2004.01.003] [Medline: 15157638]
28. Peña AS, Witchel SF, Hoeger KM, Oberfield SE, Vogiatzi MG, Misso M, et al. Adolescent polycystic ovary syndrome according to the international evidence-based guideline. BMC Med 2020;18(1):72 [FREE Full text] [doi: 10.1186/s12916-020-01516-x] [Medline: 32204714]
29. Matsuzaki S, Canis M, Pouly JL, Rabischong B, Botchorishvili R, Mage G. Relationship between delay of surgical diagnosis and severity of disease in patients with symptomatic deep infiltrating endometriosis. Fertil Steril 2006;86(5):1314-1316; discussion 1317 [FREE Full text] [doi: 10.1016/j.fertnstert.2006.03.048] [Medline: 16978622]
30. Witchel SF, Teede HJ, Peña AS. Curtailing PCOS. Pediatr Res 2020;87(2):353-361 [FREE Full text] [doi: 10.1038/s41390-019-0615-1] [Medline: 31627209]
31. Robinson KM, Christensen KB, Ottesen B, Krasnik A. Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: a nationwide Danish study. Qual Life Res 2012;21(9):1519-1525 [FREE Full text] [doi: 10.1007/s11136-011-0077-3] [Medline: 22138966]
32. Álvarez-Salvago F, Lara-Ramos A, Cantarero-Villanueva I, Mazheika M, Mundo-López A, Galiano-Castillo N, et al. Chronic fatigue, physical impairments and quality of life in women with endometriosis: a case-control study. Int J Environ Res Public Health 2020;17(10):3610 [FREE Full text] [doi: 10.3390/ijerph17103610] [Medline: 32455618]
33. Della Corte L, Di Filippo C, Gabrielli O, Reppuccia S, La Rosa VL, Ragusa R, et al. The burden of endometriosis on women's lifespan: a narrative overview on quality of life and psychosocial wellbeing. Int J Environ Res Public Health 2020;17(13):4683 [FREE Full text] [doi: 10.3390/ijerph17134683] [Medline: 32610665]
34. Marsh EE, Al-Hendy A, Kappus D, Galitsky A, Stewart EA, Kerolous M. Burden, prevalence, and treatment of uterine fibroids: a survey of U.S. Women. J Womens Health (Larchmt) 2018;27(11):1359-1367 [FREE Full text] [doi: 10.1089/jwh.2018.7076] [Medline: 30230950]
35. Al-Hendy A, Myers ER, Stewart E. Uterine fibroids: burden and unmet medical need. Semin Reprod Med 2017;35(6):473-480 [FREE Full text] [doi: 10.1055/s-0037-1607264] [Medline: 29100234]
36. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. Hum Reprod 2012;27(5):1292-1299 [FREE Full text] [doi: 10.1093/humrep/des073] [Medline: 22422778]
37. Barnard L, Ferriday D, Guenther N, Strauss B, Balen AH, Dye L. Quality of life and psychological well being in polycystic ovary syndrome. Hum Reprod 2007;22(8):2279-2286 [FREE Full text] [doi: 10.1093/humrep/dem108] [Medline: 17537782]
38. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. Fertil Steril 2016;106(1):6-15 [FREE Full text] [doi: 10.1016/j.fertnstert.2016.05.003] [Medline: 27233760]
39. Surrey E, Soliman AM, Trenz H, Blauer-Peterson C, Sluis A. Impact of endometriosis diagnostic delays on healthcare resource utilization and costs. Adv Ther 2020;37(3):1087-1099 [FREE Full text] [doi: 10.1007/s12325-019-01215-x] [Medline: 31960340]
40. Soliman AM, Yang H, Du EX, Kelley C, Winkel C. The direct and indirect costs associated with endometriosis: a systematic literature review. Hum Reprod 2016;31(4):712-722 [FREE Full text] [doi: 10.1093/humrep/dev335] [Medline: 26851604]

41. Azziz R. Overview of long-term morbidity and economic cost of the polycystic ovary syndrome. In: Azziz R, Nestler JE, Dewailly D, editors. Androgen Excess Disorders in Women Polycystic Ovary Syndrome and Other Disorders, Second Edition. Totowa, NJ: Humana Press; 2007:353-362.
42. Ericsson mobility report. Ericsson. 2022. URL: https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf [accessed 2023-11-15]
43. Digital health trends 2021. IQVIA. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2022-07-11]
44. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-07-12]
45. Yigzaw KY, Wynn R, Marco-Ruiz L, Budrionis A, Oyeyemi SO, Fagerlund AJ, et al. The association between health information seeking on the internet and physician visits (The Seventh Tromsø Study—Part 4): population-based questionnaire study. J Med Internet Res 2020;22(3):e13120 [FREE Full text] [doi: 10.2196/13120] [Medline: 32134387]
46. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
47. Jutel A, Lupton D. Digitizing diagnosis: a review of mobile applications in the diagnostic process. Diagnosis (Berl) 2015;2(2):89-96 [FREE Full text] [doi: 10.1515/dx-2014-0068] [Medline: 29540025]
48. Wetzel AJ, Koch R, Preiser C, Müller R, Klemmt M, Ranisch R, et al. Ethical, legal, and social implications of symptom checker apps in primary health care (CHECK.APP): protocol for an interdisciplinary mixed methods study. JMIR Res Protoc 2022;11(5):e34026 [FREE Full text] [doi: 10.2196/34026] [Medline: 35576570]
49. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018;5(3):95-105 [FREE Full text] [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
50. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. It—Inf Technol 2019;61(5-6):253-263 [FREE Full text] [doi: 10.1515/itit-2019-0019]
51. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
52. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2(1):38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
53. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
54. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A. Online symptom checkers: recommendations for a vignette-based clinical evaluation standard. J Med Internet Res 2022;24(10):e37408 [FREE Full text] [doi: 10.2196/37408] [Medline: 36287594]
55. Pairon A, Philips H, Verhoeven V. A scoping review on the use and usefulness of online symptom checkers and triage systems: how to proceed? Front Med (Lausanne) 2022;9:1040926 [FREE Full text] [doi: 10.3389/fmed.2022.1040926] [Medline: 36687416]
56. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022;5(1):118 [FREE Full text] [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
57. Kopka M, Schmieding ML, Rieger T, Roesler E, Balzer F, Feufel MA. Determinants of laypersons' trust in medical decision aids: randomized controlled trial. JMIR Hum Factors 2022;9(2):e35219 [FREE Full text] [doi: 10.2196/35219] [Medline: 35503248]
58. The world's most popular female health app. Flo Health. URL: https://flo.health/ [accessed 2023-11-23]
59. Teede H, Tay CT, Laven JSE, Dokras A, Moran LJ, Piltonen T, et al. International evidence-based guideline for the assessment and management of polycystic ovary syndrome 2023. Monash University. 2023. URL: https://bridges.monash.edu/articles/online_resource/International_Evidence-based_Guideline_for_the_Assessment_and_Management_of_Polycystic_Ovary_Syndrome_2023/24003834/1 [accessed 2023-11-15]
60. De La Cruz MSD, Buchanan EM. Uterine fibroids: diagnosis and treatment. Am Fam Physician 2017;95(2):100-107 [FREE Full text] [Medline: 28084714]
61. Ćirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. J Med Internet Res 2020;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
62. Rodriguez EM, Thomas D, Druet A, Vlajic-Wheeler M, Lane KJ, Mahalingaiah S. Identifying women at risk for polycystic ovary syndrome using a mobile health app: virtual tool functionality assessment. JMIR Form Res 2020;4(5):e15094 [FREE Full text] [doi: 10.2196/15094] [Medline: 32406861]
63. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]

64. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a popular online symptom checker for ophthalmic diagnoses. JAMA Ophthalmol 2019;137(6):690-692 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.0571] [Medline: 30973602]
65. Bontempo AC, Mikesell L. Patient perceptions of misdiagnosis of endometriosis: results from an online national survey. Diagnosis (Berl) 2020;7(2):97-106 [FREE Full text] [doi: 10.1515/dx-2019-0020] [Medline: 32007945]
66. Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J Clin Endocrinol Metab 2017;102(2):604-612 [FREE Full text] [doi: 10.1210/jc.2016-2963] [Medline: 27906550]
67. Zhaunova L, Bamford R, Radovic T, Wickham A, Peven K, Croft J, et al. Characterization of self-reported improvements in knowledge and health among users of Flo period tracking app: cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e40427 [FREE Full text] [doi: 10.2196/40427] [Medline: 37099370]
68. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
69. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
70. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
71. Cho HH, Yoon YS. Development of an endometriosis self-assessment tool for patient. Obstet Gynecol Sci 2022;65(3):256-265 [FREE Full text] [doi: 10.5468/ogs.21252] [Medline: 35381626]
72. Pedersen SD, Brar S, Faris P, Corenblum B. Polycystic ovary syndrome: validated questionnaire for use in diagnosis. Can Fam Physician 2007;53(6):1042-1047 [FREE Full text] [Medline: 17872783]
73. El-Osta A, Webber I, Alaa A, Bagkeris E, Mian S, Sharabiani MTA, et al. What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. BMJ Open 2022;12(4):e053566 [FREE Full text] [doi: 10.1136/bmjopen-2021-053566] [Medline: 35477872]
74. Bazot M, Lafont C, Rouzier R, Roseau G, Thomassin-Naggara I, Daraï E. Diagnostic accuracy of physical examination, transvaginal sonography, rectal endoscopic sonography, and magnetic resonance imaging to diagnose deep infiltrating endometriosis. Fertil Steril 2009;92(6):1825-1833 [FREE Full text] [doi: 10.1016/j.fertnstert.2008.09.005] [Medline: 19019357]
75. Critchley HOD, Babayev E, Bulun SE, Clark S, Garcia-Grau I, Gregersen PK, et al. Menstruation: science and society. Am J Obstet Gynecol 2020;223(5):624-664 [FREE Full text] [doi: 10.1016/j.ajog.2020.06.004] [Medline: 32707266]
76. Hoeger KM, Dokras A, Piltonen T. Update on PCOS: consequences, challenges, and guiding treatment. J Clin Endocrinol Metab 2021;106(3):e1071-e1083 [FREE Full text] [doi: 10.1210/clinem/dgaa839] [Medline: 33211867]
77. Fleming J, Jeannon JP. Head and neck cancer in the digital age: an evaluation of mobile health applications. BMJ Innov 2020;6(1):13-17. [doi: 10.1136/bmjinnov-2019-000350]
78. Hammoud M, Douglas S, Darmach M, Alawneh S, Sanyal S, Kanbour Y. Avey: an accurate AI algorithm for self-diagnosis. medRxiv. Preprint posted online on March 11, 2022 [FREE Full text] [doi: 10.1101/2022.03.08.22272076]
79. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018;392(10161):2263-2264 [FREE Full text] [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]

Abbreviations

FN: false negative
FP: false positive
NPV: negative predictive value
PCOS: polycystic ovary syndrome
PPV: positive predictive value
TN: true negative
TP: true positive

Edited by L Buis, G Eysenbach; submitted 23.02.23; peer-reviewed by C Vorisek, E Baker, CI Sartorão Filho, S Dixon; comments to author 28.07.23; revised version received 06.09.23; accepted 07.11.23; published 05.12.23.

Please cite as:
Peven K, Wickham AP, Wilks O, Kaplan YC, Marhol A, Ahmed S, Bamford R, Cunningham AC, Prentice C, Meczner A, Fenech M, Gilbert S, Klepchukova A, Ponzo S, Zhaunova L
Assessment of a Digital Symptom Checker Tool's Accuracy in Suggesting Reproductive Health Conditions: Clinical Vignettes Study
JMIR Mhealth Uhealth 2023;11:e46718
URL: https://mhealth.jmir.org/2023/1/e46718
doi:10.2196/46718
PMID:38051574

Original Paper

The Effectiveness of an eHealth Family-Based Intervention Program in Patients With Uncontrolled Type 2 Diabetes Mellitus (T2DM) in the Community Via WeChat: Randomized Controlled Trial

Yuheng Feng[1,2,3], BA; Yuxi Zhao[1,2,3], MD; Linqi Mao[1,2,3], MD; Minmin Gu[4], BA; Hong Yuan[5], BA; Jun Lu[1,2,3], PhD; Qi Zhang[6], PhD; Qian Zhao[1,2,3], BA; Xiaohong Li[1,2,3], PhD

[1]Department of Health Policy and Management, School of Public Health, Fudan University, Shanghai, China

[2]China Research Center on Disability Issues, Fudan University, Shanghai, China

[3]Key Laboratory of Health Technology Assessment, National Health Commission, Fudan University, Shanghai, China

[4]Juyuan New District Community Health Service Center, Jiading District, Shanghai, China

[5]Shanghai Jiading District Center for Disease Control and Prevention, Shanghai, China

[6]School of Community and Environmental Health, Old Dominion University, Norfolk, VA, United States

Corresponding Author:
Xiaohong Li, PhD
Department of Health Policy and Management
School of Public Health
Fudan University
220 Handan Rd
Yangpu District
Shanghai, 200433
China
Phone: 86 13918213586
Email: lixh@fudan.edu.cn

Abstract

Background: Intervention based on family support and risk perception can enhance type 2 diabetes mellitus (T2DM) patients' self-care activities. In addition, eHealth education is considered to improve family members' support for patients with T2DM. However, there is little evidence from rigorously designed studies on the effectiveness of an intervention combining these approaches.

Objective: This randomized controlled trial (RCT) aimed to assess the effectiveness of an eHealth family-based health education intervention for patients with T2DM to improve their glucose control, risk perception, and self-care behaviors.

Methods: This single-center, 2-parallel-group RCT was conducted between 2019 and 2020. Overall, 228 patients were recruited from Jiading District, Shanghai, and randomly divided into intervention and control groups. The intervention group received an eHealth family intervention based on community management via WeChat, whereas the control group received usual care. The primary outcome was the glycated hemoglobin (HbA_{1c}) level of the patients with T2DM, and the secondary outcomes were self-management behavior (general and specific diet, exercise, blood sugar testing, foot care, and smoking), risk perception (risk knowledge, personal control, worry, optimism bias, and personal risk), and family support (supportive and nonsupportive behaviors). A 2-tailed paired-sample *t* test was used to compare the participants at baseline and follow-up within the control and intervention groups. An analysis of covariance was used to measure the intervention effect.

Results: In total, 225 patients with T2DM were followed up for 1 year. After intervention, they had significantly lower HbA_{1c} values (β=–.69, 95% CI –0.99 to –0.39; *P*<.001). They also had improved general diet (β=.60, 95% CI 0.20 to 1.00; *P*=.003), special diet (β=.71, 95% CI 0.34 to 1.09; *P*<.001), blood sugar testing (β=.50, 95% CI 0.02 to 0.98; *P*=.04), foot care (β=1.82, 95% CI 1.23 to 2.42; *P*<.001), risk knowledge (β=.89, 95% CI 0.55 to 1.24; *P*<.001), personal control (β=.22, 95% CI 0.12 to 0.32; *P*<.001), worry (β=.24, 95% CI 0.10 to 0.39; *P*=.001), optimism bias (β=.26, 95% CI 0.09 to 0.43; *P*=.003), and supportive behaviors (β=5.52, 95% CI 4.03 to 7.01; *P*<.001).

Conclusions: The eHealth family-based intervention improved glucose control and self-care activities among patients with T2DM by aiding the implementation of interventions to improve T2DM risk perceptions among family members. The intervention is generalizable for patients with T2DM using health management systems in community health centers.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900020736; https://www.chictr.org.cn/showprojen.aspx?proj=31214

(JMIR Mhealth Uhealth 2023;11:e40420) doi:10.2196/40420

KEYWORDS

public health; type 2 diabetes mellitus; intervention; randomized controlled trial; community health center

Introduction

Background

Worldwide, type 2 diabetes mellitus (T2DM) has become a serious public health problem; it leads to severe health outcomes and creates a heavy economic burden [1,2]. The prevalence of T2DM is rising across all regions [1]. For people aged over 60 years, the prevalence of T2DM exceeds 20%, and it is expected to increase with the aging of the world's population [1].

China has the highest number of people with diabetes and mortality globally [1]. Among all types of diabetes, T2DM accounts for over 90% of diagnoses, and this type of the disease can cause various complications [3]. In 2021, health expenditures in China related to T2DM and complications reached about US $0.17 trillion, accounting for 1% of gross domestic product [1,4]. Therefore, it is urgent to explore practical measures to reduce the occurrence of T2DM and prevent complications from occurring in patients with T2DM.

Poor glucose control is associated with the occurrence of complications [5]. Glycated hemoglobin (HbA_{1c}) can be used to determine glucose control level [6]. HbA_{1c} ≥7% indicates that a patient has poor glucose control [7].

If HbA_{1c} increases by 2%, the risk of T2DM all-cause mortality increases by 40% to 86% in the following 10 to 20 years [8]. However, most Chinese patients with T2DM have poor glucose control [9]. Studies show that many factors influence glucose control, such as psychosocial factors [10], self-care activities [11], risk perception [12], and social support [13].

China released the National Standard for Basic Public Health Services in 2009 [14], stating that the health management of T2DM is one of the main areas of focus and that community health service centers should regularly provide usual care to patients with T2DM, such as face-to-face follow-up, health education, physical examination, and glucose testing. Until 2016, although a large number of patients with T2DM received standardized management at community health service centers [15,16], the effectiveness of usual care was not good, leading to a failure to enhance perceptions among patients with T2DM of the importance of glucose control and improving self-care activities. The effectiveness of usual care is thus unsustainable due to poor understanding and compliance with such aspects of care as regular face-to-face follow-ups and glucose testing [17]. Therefore, it is essential to explore an effective way to implement health education interventions for patients with T2DM at the community level.

Prior Studies

Prior studies have explored various ways to improve self-care activities among patients with T2DM and improve interventions in many aspects.

First, the effectiveness of smartphone-based interventions has been confirmed for intervention platforms, and they can improve glucose control levels and self-care activities [18,19]. The development of technologies and apps that are self-developed or can be downloaded in app stores provides a valuable intervention tool for patients with diabetes [20-22]. Kho et al [23] developed an empirical diabetes app, but developing and popularizing such interventions is expensive, and they are difficult to generalize. Therefore, it is vital to explore cheap and generalizable intervention platforms. In China, WeChat, a free and popular social media platform with about 576 million users as of 2017, is used by people in daily life, [24,25] providing a potential platform to develop a generalized intervention program.

Second, the content of the intervention must focus on self-care activities, because these are important to achieve therapeutic targets and prevent the development of complications [26]; many prior interventions have therefore focused on self-care activities [27,28]. A majority of interventions implemented health education on behaviors such as diet, exercise, and foot care [27]. Self-care activities are directly influenced by T2DM risk knowledge and attitudes, such as risk perception [29]. However, few studies have conducted interventions via comprehensive programs [30]. Exploring a comprehensive intervention program that targets knowledge, health beliefs, and self-care behaviors is essential to improve glucose control among patients with T2DM [29,30], help them participate in community health management, and improve their communication with doctors via family members.

Third, for intervention subjects, prior studies focused on patients with T2DM [31], peers [32], and family members [33]. Family is an essential social support resource for self-care activities among patients with T2DM [34]. However, implementation of family-based intervention is difficult due to low participation rates of family members [35].

Aim

The study aimed to develop an eHealth family-based intervention program targeting knowledge, attitude, and behaviors and validate whether the intervention could improve glucose control among patients with T2DM registered in community health centers through a rigorously designed trial.

Methods

Study Design

According to the study protocol [36], the conceptual framework for this study was that the eHealth family-based intervention via WeChat would improve family members' knowledge and risk perception of T2DM, which could help patients with T2DM to practice self-management, promote participation in community T2DM management, and improve communication with doctors, so that glucose control among patients with T2DM would ultimately be improved.

We conducted a single-center, 2-parallel-group randomized controlled trial that lasted 1 year to evaluate the effectiveness of this family-based intervention. Author LM was responsible for generating a random allocation sequence. The community health center enrolled the participants and assigned them to the intervention group or the control group. The intervention group received the eHealth intervention and usual care. The control group received usual care. The primary outcome was the HbA_{1c} value. The secondary outcomes were self-care activities, risk perception, and family support. Notably, because the intervention was an eHealth intervention, participants knew they would receive the intervention when they provided informed consent; thus, the study participants could not be blinded. The study design strictly followed the Consolidated Standards of Reporting Trials (CONSORT) eHealth checklist (version 1.6.1).

Participants

According to the National Standard for Basic Public Health Services [14], community health service centers should maintain correspondence with patients and determine the prevalence of diabetes. This randomized control trial was conducted in the central area of Jiading District, which includes 2 community health service centers: Jiading Town Community Health Service Center and Juyuan New District Community Health Service Center. A total of 3874 individuals with diabetes were registered at these community health service centers. Of 1650 individuals with recorded HbA_{1c} values, 879 had a value over 7%.

The inclusion criteria for the patients were as follows: (1) they were registered in 1 of the 2 community health service centers in the urban area in Jiading District, (2) they had been diagnosed with type 2 diabetes by a doctor at least 6 months before study enrollment, (3) they were aged 18 to 79 years, (4) their HbA_{1c} level was ≥7%, (5) they had no plans to leave their place of residence in the following 12 months, and (6) they had a family member who could use WeChat and lived with the patient or visited them at least once a week.

The exclusion criteria for the patients were as follows: (1) they had other serious illnesses or illnesses not suitable for this study, (2) they were women who were pregnant or preparing for pregnancy, (3) they were unable to complete the 12-month follow-up for reasons such as moving and not transferring to another health care facility, (4) they were unwilling or unable to provide informed consent, and (5) they were currently participating in another intervention study.

Patients were required to choose 1 family member as a supporter to receive the corresponding intervention. The inclusion and exclusion criteria for the family members were as follows: (1) they were in regular contact with the patient, (2) they were nominated by the patient, (3) they were older than 18 years, and (4) they had never participated in other, similar research.

Patients and Public Involvement

Patients and the public were not involved in our research's design, conduct, reporting, or dissemination plans.

Sample Size Calculation

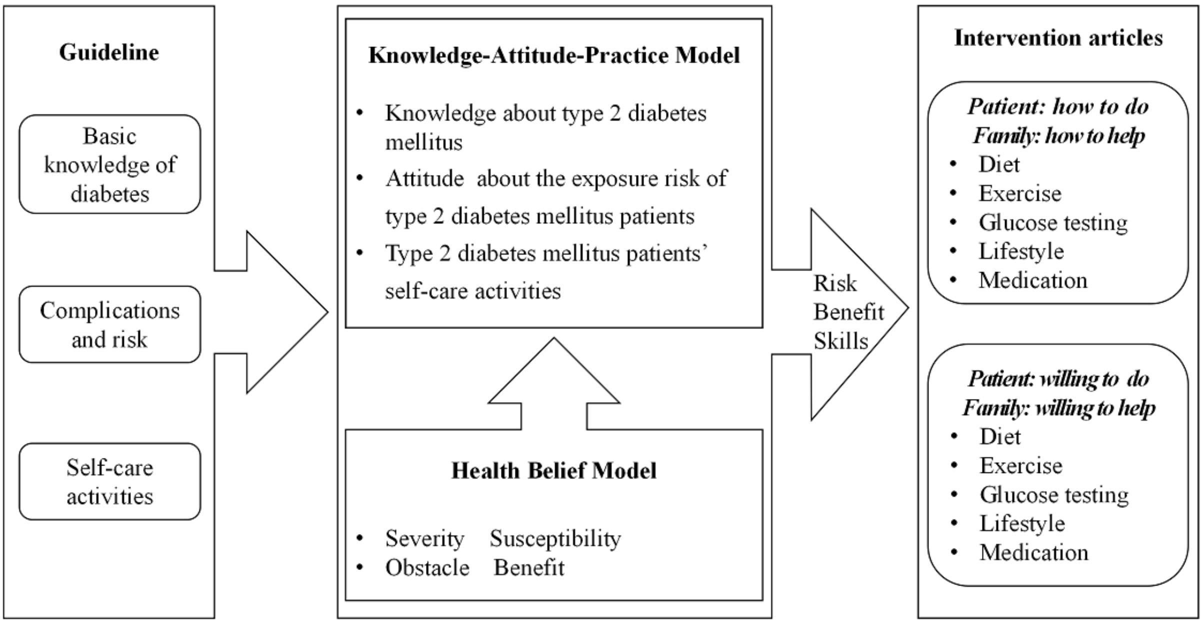
We calculated the sample size based on a formula for comparing 2 means with a ratio of 1:1 between the intervention and control groups. This study used the HbA_{1c} value as the most important indicator. The formula for comparing 2 sample means was as follows:

We set α=.05 and β=.20. δ was the effect size, which ranged from 0.25 to 0.70 [37-39]. In this study, we supposed that the effect size was 0.50, and σ (standard deviation) was 1.2 [36]. Therefore, about 91 participants were required in each group. Considering a dropout rate of 20%, we planned to recruit about 110 patients in each group. The ratio of family members to patients was 1:1.

Intervention Tool

The study established an official WeChat account called Jiading Sugar Steward, which included 3 modules: blood glucose data, complications, and notices. The intervention was implemented by using this account to deliver intervention articles. First, following the Guidelines for the Prevention and Treatment of T2DM, the study searched 4 domains of health intervention, including diabetes knowledge, complications, risk, and self-care activities (eg, diet, exercise, medication, lifestyle, glucose testing, and skills). All intervention articles were then developed based on the Knowledge-Attitude-Practice (KAP) model [40] and the Health Belief Model (HBM) [41]. The KAP model guided the relationship between basic knowledge and risk knowledge about T2DM, behaviors related to improving daily self-care activities among patients with T2DM, and family members' attitudes on reminding patients with T2DM to improve their self-care activities. The HBM guided two aspects of family members' attitudes: (1) knowledge that T2DM can induce significant health complications and (2) knowledge of the benefits of self-care activities and the risk of not performing these activities. We used the KAP model and the HBM to develop 38 comprehensive articles on topics including basic knowledge, skills, and risk knowledge, that is, behaviors and psychology (Figure 1 and Multimedia Appendix 1). All articles were divided into 3 categories: acquisition (A-level articles, scored 4 to 5), familiarity (B-level articles, scored 3), and understanding (C-level articles, scored 1 to 2).

Figure 1. Design of the online intervention articles based on the Knowledge-Attitude-Practice model and the Health Belief Model.



Intervention Process

The intervention was composed of an online intervention and an in-person intervention.

Online Intervention: Family Members

The online intervention implementation included 3 phases. During phase 1 (from January 1 to February 28, 2019), family members of patients with T2DM in the intervention group were asked to follow the aforementioned official WeChat account and add it to their personal WeChat. During phase 2 (from March 1 to May 31, 2019), the study regularly delivered the 38 articles described above (about 3 articles were delivered per week, on Monday, Wednesday, and Friday). A-level articles were sent one-to-one via the official WeChat account and messages. B-level articles were sent by forwarding individual WeChat Moments. C-level articles were released through the official WeChat account (Multimedia Appendix 2). All online intervention articles were delivered at 6 AM or 5 PM on the push day so that subjects could receive the intervention outside of working hours. During phase 3 (June 1, 2019, to February 29, 2020), we measured the effectiveness of the online intervention.

In-Person Intervention: Patients With T2DM

The National Standard for Basic Public Health Services [14] calls for all patients with T2DM to receive on-site health education once every 3 months. For the in-person intervention in this study, which we compared with usual care (diet control, glucose testing, and auxiliary activities during exercise), we developed risk-perception–based health education courses focused on the practice of skills. Taking exercise as an example, we conducted health education on how patients with T2DM should properly exercise to manage their glucose levels. We also educated patients with T2DM on the risks and benefits of self-care activities. An example of a risk is that patients with T2DM cannot spend a long time out of the home because of the possibility of their glucose becoming low. An example of a benefit is that exercise can help control glucose.

The intervention measure was implemented by family practitioners and doctors in the prevention and health section. Family practitioners were mainly responsible for conducting follow-ups every 3 months. Doctors in the prevention and health section were mainly responsible for background operations, maintaining the official WeChat account, and cooperating with family practitioners.

Outcome Measures

Primary Outcome

HbA_{1c} has been clinically used as a gold standard for assessing long-term blood glucose control and reflects blood glucose concentration over approximately 3 months [42,43]. Therefore, we selected HbA_{1c} as the primary outcome. HbA_{1c} was measured at Jiading Central Hospital. A lower HbA_{1c} level means that a patient with T2DM has better glucose control.

Secondary Outcomes

Four secondary outcomes were measured by questionnaires that were confirmed to be suitable for use by Chinese people. The questionnaires covered topics including diabetes self-care activities [44], risk perception of diabetes [45], and family support [46].

Self-care Activities

Self-care activities were measured using the Summary of Diabetes Self-Care Activities scale [47], which measures normal diet, abnormal diet, exercise, blood glucose monitoring, foot care, and smoking. Besides the last item, which is scored from 1 to 2, the other items range from 0 to 7. The higher the final score, the better the patient's self-care behavior. A higher score means that a patient with T2DM has better performance of self-care activities.

Risk Perception

Risk perception was examined using the Risk Perception Survey–Diabetes Mellitus scale [48], which is divided into 2 modules (risk perception and risk knowledge), with 31 entries in 6 dimensions. However, the dimension "comparative environmental risk" was removed because it is unsuitable for people with a Chinese cultural background. This decision was made after expert consultation and interviews with patients with T2DM and their family members. The meanings of the remaining 5 dimensions are shown in Multimedia Appendix 3. The total risk knowledge score is the sum of the 5 items.

Family Support

Family support was evaluated using the Diabetes Family Behavior Checklist [49], which contains 16 items scored from 1 to 5. A score of 1 means that the patient never receives support from their family. A score of 5 means that the family always supports the patient. This indicator includes positive support and negative support. A higher positive support score or lower negative support score mean patients with T2DM have a better family living environment with better family support behaviors.

Data Collection and Randomization

From January 1 to February 28, 2019, the study recruited participants according to the inclusion and exclusion criteria. Eventually, 228 pairs of participants were enrolled in each group. They were randomly allocated to the intervention and control groups at the individual level via a random-number table; 114 pairs of participants were assigned to the intervention and control groups. The intervention measure was then implemented for a period of 12 months.

To measure HbA_{1c}, we collected blood samples at baseline and the 12th month. Laboratory tests were done at the Jiading Central Hospital. Secondary outcomes were collected via on-site questionnaire surveys at baseline and the 12th month. The baseline questionnaire survey was conducted by 4 authors (XL, LM, YZ, and YF), who collected data from patients with T2DM and their family members individually at the community health service centers. At the 12th month, the participants were contacted by telephone and the research team re-collected their data.

Statistical Analysis

Sociodemographic characteristics were summarized for the intervention and control groups. Frequencies and percentages were used for categorical variables and mean (SD) for continuous variables.

To assess the effectiveness of the intervention, a 2-tailed paired-sample *t* test was used to compare the baseline and follow-up data in the intervention and control groups. Changes between the baseline and follow-up period in the intervention and control groups were measured with the 95% CI at baseline and the 12th month. Finally, since HbA_{1c} was the primary outcome, we further analyzed sex subgroups to explore whether the results differed between males and females. Furthermore, we used an analysis of covariance to clarify the intervention effect and included the sex, age, and education of patients with T2DM, the family members' education, and the family members' relationship as covariables.

All data management and analyses were performed using Stata (version 15.0; Stata Corp). We set α=.05 and β=.20; the power was 80%. Statistical significance was set at P<.05.

Ethics Approval

Informed consent was provided by all participants. In addition, the trial was ethically approved by the Medical Research Ethics Committee of the School of Public Health Fudan University (2018-01-0663). All participants provided written informed consent.

Results

Participant Characteristics

In total, 225 patients (113 patients in the intervention group and 112 patients in the control group) completed this 1-year intervention study. Three participants (1 in the intervention group and 2 in the control group) were lost to follow-up. Figure 2 shows the process of inclusion in each analysis, starting with the originally assigned groups.

Among the 225 patients, 48.4% (n=109) were men and 51.6% (n=116) were women, with a mean age of 65.6 (SD 7.1) years. Of the 225 family members nominated by the included patients, 48% (n=108) were men and 52% (n=117) were women, with a mean age of 48.5 years. Spouses were the family members who most commonly lived with the patients. The baseline sociodemographic and clinical characteristics of the patients and their family members are shown in Table 1.

Figure 2. Flow diagram of trial participation.

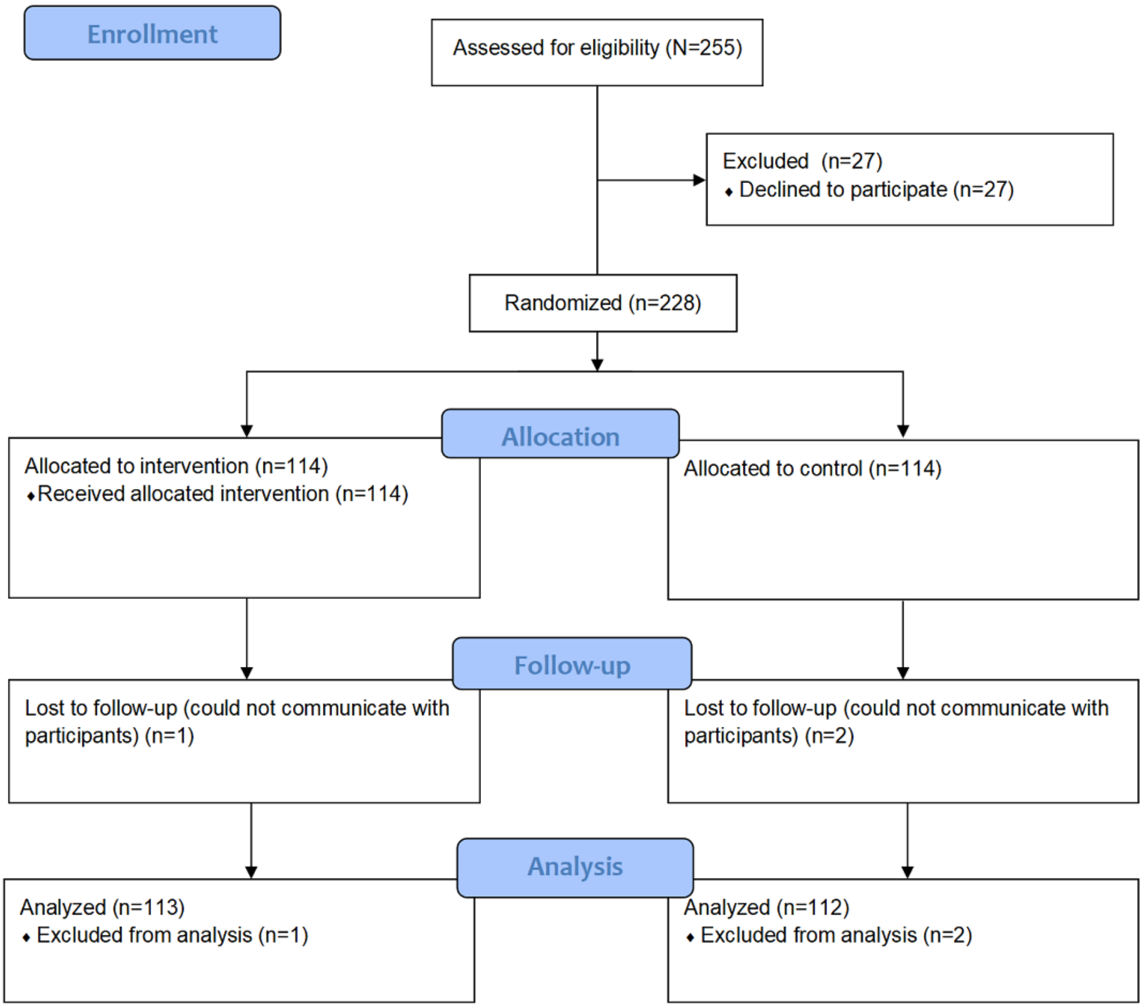


Table 1. Characteristics of the patients and their family members at baseline.

Participants/characteristics	Intervention group (n=113)	Control group (n=112)	*P* value
Patients with type 2 diabetes			
Gender, n (%)			.64
Male	53 (46.9)	56 (50)	
Female	60 (53.1)	56 (50)	
Age (years), mean (SD)	65.7 (6.7)	65.4 (7.5)	.78
Marriage status, n (%)			.20
Married	106 (93.8)	98 (87.5)	
Divorced	2 (1.8)	2 (1.8)	
Widowed	5 (4.4)	12 (10.7)	
Education, n (%)			.59
Illiterate	12 (10.6)	12 (10.7)	
Primary school	26 (23)	21 (21.4)	
Junior school	45 (39.8)	39 (34.8)	
Senior school	21 (18.6)	31 (27.7)	
Undergraduate or college	9 (8)	6 (5.4)	
Postgraduate or above	0 (0)	0 (0)	
Employment status, n (%)			.19
Employed	11 (9.7)	14 (12.5)	
Retired	99 (87.6)	98 (87.5)	
Unemployed	3 (2.7)	0 (0)	
Average household monthly income (US $), n (%)			.38
<416.10	14 (12.4)	15 (13.4)	
416.10-832.19	52 (46)	41 (36.6)	
832.20-1248.30	32 (28.3)	39 (34.8)	
>1248.30	15 (13.3)	17 (15.2)	
Duration of diabetes (years), mean (SD)	12.8 (6.3)	13.6 (7.2)	.37
Family members			
Gender, n (%)			.39
Male	51 (45.1)	57 (50.9)	
Female	62 (54.8)	55 (49.1)	
Age (years), mean (SD)	47.9 (14.5)	49.1 (11.8)	.50
Relationship, n (%)			.61
Spouse	36 (31.9)	38 (33.9)	
Son	26 (23)	28 (25)	
Daughter	7 (6.2)	11 (9.8)	
Daughter-in-low	35 (31)	31 (27.7)	
Son-in-low	1 (0.9)	0 (0)	
Grandson or granddaughter	8 (7.1)	4 (3.6)	
Education, n (%)			.98
Illiterate	1 (0.9)	1 (0.9)	
Primary school	12 (10.6)	5 (4.5)	
Junior school	24 (21.2)	26 (23.2)	

Participants/characteristics	Intervention group (n=113)	Control group (n=112)	*P* value
Senior school	27 (23.9)	39 (34.8)	
Undergraduate or college	47 (41.6)	37 (33)	
Postgraduate or above	2 (1.8)	4 (3.6)	
Chronic disease, n (%)			.07
Yes	70 (61.9)	82 (73.2)	
No	43 (38.1)	30 (26.8)	

Effectiveness Outcomes

In the intervention group, HbA_{1c} level (*P*<.001) and nonsupportive behavior (*P*=.03) decreased and the scores for general diet (*P*<.001), specific diet (*P*<.001), exercise (*P*=.002), blood sugar testing (*P*=.02), foot care (*P*<.001), risk knowledge (*P*<.001), personal control (*P*<.001), worry (*P*=.02), optimism bias (*P*=.03), and supportive behaviors (*P*<.001) improved. In the control group, besides general diet (*P*=.002), there were no statistically significant differences between baseline and follow-up (Table 2).

Table 2. Primary and secondary effectiveness outcomes. The change in score is the follow-up score minus the baseline score. The 95% CIs that do not contain 0 are statistically used to show the intervention effectiveness after an individual received the intervention. *P* values are the result of a paired-sample *t* test to demonstrate whether the difference was statistically significant.

	Intervention group (n=113)				Control group (n=112)			
	Baseline, mean (SD)	Follow-up, mean (SD)	Change (95% CI)	*P* value	Baseline, mean (SD)	Follow-up, mean (SD)	Change (95% CI)	*P* value
Glycated hemoglobin A_{1c} level (%)	7.90 (0.75)	7.30 (1.07)	–0.60 (–0.82 to –0.38)	<.001	7.84 (0.66)	7.99 (1.22)	0.15 (–0.09 to 0.39)	.22
Gender (score)								
Male	7.99 (0.87)	7.46 (1.11)	–0.53 (–0.88 to –0.18)	.004	7.80 (0.57)	7.89 (1.29)	0.08 (–0.28 to 0.44)	.66
Female	7.83 (0.61)	7.16 (1.01)	–0.62 (–0.95 to –0.37)	<.001	7.87 (0.75)	8.09 (1.14)	0.22 (–0.11 to 0.54)	.18
Self–care activities (score)								
General diet	5.36 (2.37)	6.39 (1.09)	1.03 (0.58 to 1.48)	<.001	5.05 (2.36)	5.75 (1.88)	0.70 (0.26 to 1.13)	.002
Specific diet	3.59 (1.73)	4.31 (1.36)	0.72 (0.34 to 1.09)	<.001	3.85 (1.96)	3.62 (1.45)	–0.23 (–0.67 to 0.21)	.31
Exercise	3.55 (1.96)	4.22 (1.99)	0.67 (0.26, 1.07)	.002	3.83 (1.71)	3.95 (1.97)	0.12 (–0.27 to 0.51)	.55
Blood sugar testing	1.53 (1.81)	2.09 (2.07)	0.56 (0.11 to 1.00)	.02	1.77 (2.20)	1.64 (1.66)	–0.13 (–0.51 to 0.26)	.52
Foot care	1.56 (2.46)	3.23 (2.76)	1.67 (1.12 to 2.22)	<.001	1.13 (2.17)	1.35 (1.84)	0.22 (–0.29 to 0.75)	.39
Smoking	0.79 (0.41)	0.85 (0.36)	0.06 (–0.01 to 0.13)	.07	0.82 (0.39)	0.85 (0.36)	0.03 (–0.02 to 0.07)	.26
Risk perception of diabetes (score)								
Risk knowledge	3.32 (1.63)	4.45 (1.13)	1.13 (0.79 to 1.47)	<.001	3.28 (1.43)	3.54 (1.51)	0.26 (–0.08 to 0.60)	.14
Personal control	2.94 (0.40)	3.15 (0.44)	0.21 (0.12 to 0.32)	<.001	2.99 (0.38)	2.96 (0.30)	–0.03 (–0.13 to 0.06)	.44
Worry	2.73 (0.69)	2.91 (0.55)	0.18 (0.03 to 0.33)	.02	2.74 (0.65)	2.68 (0.57)	–0.06 (–0.20 to 0.08)	.42
Optimism bias	2.75 (0.63)	2.92 (0.58)	0.17 (0.01 to 0.32)	.03	2.56 (0.77)	2.66 (0.64)	0.10 (–0.08 to 0.27)	.30
Personal risk	2.22 (0.55)	2.11 (0.62)	–0.11 (–0.25, 0.02)	.10	2.22 (0.60)	2.15 (0.59)	–0.07 (–0.22 to 0.05)	.22
Family support (score)								
Supportive behaviors	20.25 (6.65)	25.69 (6.67)	5.44 (4.07 to 6.81)	<.001	20.71 (7.11)	20.38 (6.02)	–0.33 (–1.42 to 0.76)	.55
Nonsupportive behaviors	29.82 (3.93)	28.71 (4.48)	–1.11 (–2.09 to –0.14)	.03	29.21 (3.30)	28.75 (3.60)	–0.46 (–1.21 to 0.30)	.24

The Effect of the eHealth Family-Based Intervention

Patients with T2DM who participated in the eHealth family-based intervention had significantly lower HbA_{1c} values (β=–.69, 95% CI –.99 to –.39; *P*<.001) and improved scores for general diet (β=.60, 95% CI .20-1.00; *P*=.003), special diet (β=.71, 95% CI .34-1.09; *P*<.001), blood sugar testing (β=.50, 95% CI .02-.98; *P*=.04), foot care (β=1.82, 95% CI 1.23-2.42; *P*<.001), risk knowledge (β=.89, 95% CI .55-1.24; *P*<.001), personal control (β=.22, 95% CI .12-.32; *P*<.001), worry (β=.24, 95% CI .10-.39; *P*=.001), optimism bias (β=.26, 95% CI .09-.43; *P*=.003), and supportive behaviors (β=5.52, 95% CI 4.03-7.01; *P*<.001), as shown in Table 3.

Table 3. Results of analysis of covariance for the effect of the eHealth family-based intervention. The covariables included patient gender, patient age, patient education, family member education, and family member relationship. We only show covariables with statistical implications.

Effective outcomes	β	SE (95% CI)	*P* value
Glycated hemoglobin A_{1c} level			
Intervention	–.69	.15 (–.99 to –.39)	<.001
Baseline hemoglobin A_{1c} level	.26	.11 (.04 to .47)	.02
General diet			
Intervention	.60	.30 (.20 to 1.00)	.003
Baseline general diet	.23	.04 (.14 to .31)	<.001
Special diet			
Intervention	.71	.19 (.34 to 1.09)	<.001
Exercise			
Intervention	.38	.25 (–.13 to .88)	.14
Baseline exercise	.39	.07 (.24 to .53)	<.001
Blood sugar testing			
Intervention	.50	.24 (.02 to .98)	.04
Baseline blood sugar testing	.33	.06 (.20 to .45)	<.001
Foot care			
Intervention	1.82	.30 (1.23 to 2.42)	<.001
Baseline foot care	.22	.07 (.09 to .35)	.001
Patient age	.79	.31 (.17 to 1.41)	.01
Smoking			
Intervention	.02	.04 (–.05 to .09)	.56
Baseline smoking	.57	.05 (.47 to .67)	<.001
Patient gender	.09	.04 (.01 to .17)	.03
Risk knowledge			
Intervention	.89	.18 (.55 to 1.24)	<.001
Baseline risk knowledge	.15	.06 (.03 to .26)	.02
Patient age	.49	.18 (.12 to .89)	.009
Personal control			
Intervention	.22	.05 (.12 to .32)	<.001
Worry			
Intervention	.24	.08 (.10 to .39)	.001
Baseline worry	.15	.06 (.04 to .26)	.009
Optimism bias			
Intervention	.26	.09 (.09 to .43)	.003
Personal risk			
Intervention	–.02	.08 (–.17 to .13)	.81
Baseline personal risk	.28	.07 (.14 to .41)	<.001
Patient age	.23	.08 (.08 to .38)	.004
Family member education	.10	.04 (.02 to .17)	.01
Family member relationship	–.05	.02 (–.10 to –.004)	.03
Supportive behaviors			
Intervention	5.52	.75 (4.03 to 7.01)	<.001

Effective outcomes	β	SE (95% CI)	*P* value
Baseline supportive behaviors	.47	.06 (.36 to .58)	<.001
Nonsupportive behaviors			
Intervention	–.25	.54 (–1.32 to .82)	.65
Baseline nonsupportive behaviors	.28	.07 (.13 to .43)	<.001

Discussion

Principal Findings

The study was a single-center, 2-parallel-group randomized controlled trial to assess the effectiveness of an eHealth family-based intervention. This structured intervention program assessed knowledge, attitude, and behaviors. The difference between the intervention and control groups was whether the family members of the patients with T2DM followed the official WeChat account. This study confirmed that this intervention could enhance glucose control, self-care activities, and risk perception among patients with T2DM. Additionally, the patients were better able communicate with community health service providers and obtain knowledge of diabetes-related risks, benefits, and skills after their family members reminded them of inappropriate self-management behaviors and inadequate risk perception of T2DM. Family support of the patients was also improved.

Comparison With Prior Work

Our principal findings are similar to those of McEwen et al [50], Cai and Hu [51], and Wichit et al [52], who implemented interventions via health education classes, home visits, and telephone calls. These interventions can be classified as in-person [53], a type of delivery that may not only waste a large amount of staff resources and time but may also lack sustainability due to poor compliance. During the intervention process, the intervention tool of the studies mentioned above was a booklet. Although participants received health education, the education model was only on-site, and the education content was the same as that in the booklet. Additionally, telephone-based interventions were completed by telephone. Hemmati Maslakpak et al [54] reported holding on-site health education classes and contacting participants who needed to attend classes at a fixed time. Compared with the characteristics of the aforementioned studies, this study implemented an eHealth family-based intervention via an official WeChat account without time or site limitations. Family members of patients could read the online intervention articles at any time, and the offline intervention only focused on skills practice, in order to enable the family members to complete the tasks by themselves in daily life.

There is another principal finding worth noting: although the eHealth family-based intervention was effective, patients with T2DM who participated in the intervention did not achieve ideal glucose control (HbA_{1c} <7%). This finding is similar to that of a single-arm study of an eHealth intervention program that used an accessible app (Vida Health) as the intervention platform, which is similar to WeChat [55]. The inability of patients with T2DM to achieve ideal control could be explained by the fact that only patients with uncontrolled T2DM (HbA_{1c} >7%) were enrolled in our study. Glycemic control in patients with T2DM is complex and is influenced by various factors, such as medicine adjustment, regular visits to a doctor, and lifestyle; the presence of these factors was confirmed during interviews with doctors and community health service providers. While ideal control was not achieved, patients with T2DM in the intervention group still had significantly lower HbA_{1c} levels than patients in the control group after 12 months, suggesting that the eHealth family-based intervention was an effective way to improve glucose control. In future studies, specific factors influencing glucose control should be considered in combination with an eHealth family-based intervention to educate patients with T2DM and their family members.

Sun et al [19] conducted a study that was similar to ours based on a diet intervention conducted by dietitians among outpatients; their intervention was more effective (HbA_{1c} decreased from 7.84% to 6.84%) than ours (HbA_{1c} decreased from 7.9% to 7.3%). A possible reason is that the health providers were from a tertiary hospital and could provide more professional suggestions, making the patients more willing to comply. However, dietitian-based interventions are difficult to generalize due to the shortage of these professionals. Our study was a community-based intervention, and the community was a national health-management platform for patients with T2DM. Therefore, our community-based intervention is more generalizable in real life.

Yang et al [56] also conducted a study of patients with T2DM and obtained good results; however, effectiveness in their study (a change in HbA_{1c} of 0.3%) was lower than in our study. One possible reason is that their intervention content was based only on the medical guidelines of the Korean Diabetes Association, while our study developed intervention content based on the National Standard for Basic Public Health Services [14] and provided specific family support. Although the Guidelines for the Prevention and Treatment of Type 2 Diabetes Mellitus in China (2020 revision) [3] reported that family members also play an important role in health management for patients with T2DM, this guideline does not provide specific measures. Our study added a specific family member–based health intervention to the basic health intervention in accordance with the guideline. Another possible reason for the differences between this study and the previous one is that we added information on risk perception during the process of health intervention development, which could have improved our health intervention.

New eHealth interventions, with their improved accessibility, have the potential to replace in-person interventions [57]. Despite their potential benefits, we identified certain limitations of eHealth interventions in our study. First, social support can be classified into 4 types: emotional, instrumental, informational, and appraisal [53,58,59]. It is, however, difficult to evaluate how eHealth interventions may contribute to the social support of patients with T2DM [53]. Second, commonly used eHealth interventions, such as text messages, apps, and web-based programs, are known to have a positive impact on the self-management behaviors of patients with T2DM; however, due to limitations related to character count per message, content type, interactivity, cost, accessibility, and internet connection capacity, it is difficult to compare and generalize findings across studies [60].

Considering the aforementioned limitations of eHealth interventions in previous studies, we used WeChat, a common and free app used daily by Chinese people [24,25], to address limitations related to accessibility, cost, and internet connection capacity, although we did not consider instrument support in our study. Additionally, for the delivery of the articles, we classified them into 3 types and delivered them in different ways, which solved the limitation related to character count. For articles in the official WeChat account, different types of content, such as video content, could be linked at the bottom of each article, which solved the limitation related to content type. To evaluate the extent of social support in our study, we sufficiently considered the impact of family members on patients with T2DM when designing the intervention tools by referring to the KAP and HBM conceptual frameworks.

Strengths and Limitations

The major strength of this study is that the eHealth family-based intervention used content usually delivered as part of in-person health education, as set out in the National Standard for Basic Public Health Services. The dropout rate of participants was very low (only 3 pairs of participants).

However, this study also has some limitations. First, the sample was small and only included patients enrolled in 1 city. Future studies should include larger samples from multiple sources. Second, the intervention time was 1 year, and we did not perform continuous follow-up of the patients with T2DM and their family members. Consequently, we could only conclude whether the intervention was effective over a period of time; we could not confirm its long-term effects. Third, participant compliance was not comprehensively documented; for example, we did not determine how long or how often patients with T2DM or their family members read the intervention articles. Fourth, due to a reduced workforce and funding restrictions, we did not collect data at 6 months. Restrictions also prevented us from collecting some indicators that were included in the protocol, such as weight, height, waist circumference, hip circumference, blood pressure, and BMI.

Additionally, during the process of intervention implementation, we revised the inclusion and exclusion criteria that were published in the protocol to remove "no history of diabetes" for the family members and add "never participated in other, similar research." We made this revision because the study aimed to find an intervention program that is suitable for the real world. After the protocol was developed, we came to understand that the family members of patients with T2DM can also develop T2DM. If we had recruited participants based on the inclusion and exclusion criteria in the protocol, a large number of suitable participants would have been lost, because T2DM is a common disease in older adults. The patients with T2DM were often older people whose family members were a spouse, their child, or a son- or daughter-in-law; the mean age of these family members was greater than 45 years. In addition, to avoid influence from other research, we required that the family members had not participated in other research that was similar to our study.

Conclusions

We conducted a structured eHealth family-based intervention program that included T2DM-related knowledge, risks, and skills to enhance the ability and risk perceptions of family members regarding T2DM. This intervention was able to effectively help patients with T2DM to improve their glucose control by promoting participation in community health management and strengthening their self-management behaviors. Overall, this eHealth family-based intervention for patients with uncontrolled T2DM is a promising way to empower family members to support these patients in their endeavors to improve self-management behaviors. Our study also provides information for community health providers to develop health intervention content and mobile health intervention platforms for patients with uncontrolled T2DM.

Acknowledgments

This project was supported by the National Natural Science Foundation of China (grants 72274036 and 71573049), the National Social Science Foundation of China (grant 17ZDA078), and the 111 Project (grant B16031). We would like to thank the Community Health Service Center of Juyuan New District, Jiading District, Shanghai, China, for the trial implementation. We are also very grateful to all the family members and doctors for their time and participation in the study. In addition, we would like to thank Editage for English language editing.

Authors' Contributions

YF contributed to data acquisition, analysis, and interpretation and the conception, design, drafting, and revision of the manuscript. YZ contributed to data acquisition and study design. LM and MG recruited the study participants, designed the study, collected baseline and follow-up data, and suggested the intervention. HY contributed to the provision of important information on chronic noncommunicable disease intervention and guided the implementation of the intervention. JL contributed to the design and revision of the manuscript. Q Zhang provided suggestions on the study design, questionnaires, and intervention. Q Zhao contributed to data interpretation and manuscript revision. XL contributed to data acquisition and interpretation and conception, design, and revision of the manuscript. All authors approved the final draft submitted. YF is the guarantor of this work and, as such, has full access to all the study data and takes responsibility for the integrity of the data and the accuracy of data analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The list of intervention articles' themes.
[DOCX File , 13 KB - mhealth_v11i1e40420_app1.docx]

Multimedia Appendix 2
Online intervention flowchart.
[DOCX File , 122 KB - mhealth_v11i1e40420_app2.docx]

Multimedia Appendix 3
The meanings of each indicator in Risk Perception Survey-Diabetes Mellitus.
[DOCX File , 12 KB - mhealth_v11i1e40420_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3261 KB - mhealth_v11i1e40420_app4.pdf]

References

1. IDF Diabetes Atlas (10th edition). International Diabetes Federation. URL: https://diabetesatlas.org/atlas/tenth-edition/ [accessed 2022-10-17]
2. Bommer C, Sagalova V, Heesemann E, Manne-Goehler J, Atun R, Bärnighausen T, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes Care 2018 May;41(5):963-970. [doi: 10.2337/dc17-1962] [Medline: 29475843]
3. Chinese Diabetes Society. Guideline for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition). Chin J Diabetes Mellitus 2021;13(04):315-409 [FREE Full text] [doi: 10.3760/cma.j.cn115791-20210221-00095]
4. Ding C, Bao Y, Bai B, Liu X, Shi B, Tian L. An update on the economic burden of type 2 diabetes mellitus in China. Expert Rev Pharmacoecon Outcomes Res 2022 Jun;22(4):617-625. [doi: 10.1080/14737167.2022.2020106] [Medline: 34937503]
5. Bain SC, Bekker Hansen B, Hunt B, Chubb B, Valentine WJ. Evaluating the burden of poor glycemic control associated with therapeutic inertia in patients with type 2 diabetes in the UK. J Med Econ 2020 Jan;23(1):98-105 [FREE Full text] [doi: 10.1080/13696998.2019.1645018] [Medline: 31311364]
6. Fellinger P, Rodewald K, Ferch M, Itariu B, Kautzky-Willer A, Winhofer Y. HbA1c and glucose management indicator discordance associated with obesity and type 2 diabetes in intermittent scanning glucose monitoring system. Biosensors (Basel) 2022 Apr 29;12(5):288 [FREE Full text] [doi: 10.3390/bios12050288] [Medline: 35624589]
7. Bonnet F, Chen H, Cooper A, Gomes MB, Ji L, Leigh P, et al. What are the factors associated with long-term glycaemic control in patients with type 2 diabetes and elevated glycated haemoglobin (≥7.0%) at initiation of second-line therapy? Results from the DISCOVER study. Diabetes Obes Metab 2021 Oct;23(10):2336-2343. [doi: 10.1111/dom.14476] [Medline: 34212481]
8. Lind M, Imberg H, Coleman RL, Nerman O, Holman RR. Historical HbA1c values may explain the type 2 diabetes legacy effect: UKPDS 88. Diabetes Care 2021 Jul 07;44(10):2231-2237 [FREE Full text] [doi: 10.2337/dc20-2439] [Medline: 34244332]
9. Li J, Chattopadhyay K, Xu M, Chen Y, Hu F, Chu J, et al. Glycaemic control in type 2 diabetes patients and its predictors: a retrospective database study at a tertiary care diabetes centre in Ningbo, China. BMJ Open 2018 Mar 25;8(3):e019697 [FREE Full text] [doi: 10.1136/bmjopen-2017-019697] [Medline: 29581203]
10. Roy K, Iqbal S, Gadag V, Bavington B. Relationship between psychosocial factors and glucose control in adults with type 2 diabetes. Can J Diabetes 2020 Oct;44(7):636-642. [doi: 10.1016/j.jcjd.2020.01.005] [Medline: 32205074]
11. Roy K, Iqbal S, Gadag V, Bavington B. Relationship between psychosocial factors and glucose control in adults with type 2 diabetes. Can J Diabetes 2020 Oct;44(7):636-642. [doi: 10.1016/j.jcjd.2020.01.005] [Medline: 32205074]
12. Vu A, Turk N, Duru OK, Mangione CM, Panchal H, Amaya S, et al. Association of type 2 diabetes risk perception with interest in diabetes prevention strategies among women with a history of gestational diabetes. Diabetes Spectr 2022;35(3):335-343 [FREE Full text] [doi: 10.2337/ds21-0056] [Medline: 36072816]

13. Mohebi S, Parham M, Sharifirad G, Gharlipour Z, Mohammadbeigi A, Rajati F. Relationship between perceived social support and self-care behavior in type 2 diabetics: A cross-sectional study. J Educ Health Promot 2018;7:48 [FREE Full text] [doi: 10.4103/jehp.jehp_73_17] [Medline: 29693029]
14. National Standard for Basic Public Health Services (Third Edition). National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/jws/s3578/201703/d20c37e23e1f4c7db7b8e25f34473e1b.shtml [accessed 2022-09-25]
15. Fang G, Yang D, Wang L, Wang Z, Liang Y, Yang J. Experiences and challenges of implementing universal health coverage with China's national basic public health service program: Literature review, regression analysis, and insider interviews. JMIR Public Health Surveill 2022 Jul 22;8(7):e31289 [FREE Full text] [doi: 10.2196/31289] [Medline: 35867386]
16. Development of China's public health as an essential element of human rights. The State Council Information Office of the People's Republic of China. URL: http://www.scio.gov.cn/ztk/dtzt/36048/37159/37161/Document/1565159/1565159.htm [accessed 2018-12-15]
17. He M, Gao J, Liu W, Tang X, Tang S, Long Q. Case management of patients with Type 2 diabetes mellitus: a cross-sectional survey in Chongqing, China. BMC Health Serv Res 2017 Feb 11;17(1):129 [FREE Full text] [doi: 10.1186/s12913-017-2039-0] [Medline: 28187724]
18. Aminuddin HB, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: A systematic review and meta-analysis. Int J Nurs Stud 2021 Apr;116:103286. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]
19. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: Randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
20. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
21. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: A systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
22. Zhang Y, Li X, Luo S, Liu C, Xie Y, Guo J, et al. Use, perspectives, and attitudes regarding diabetes management mobile apps among diabetes patients and diabetologists in China: National web-based survey. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12658 [FREE Full text] [doi: 10.2196/12658] [Medline: 30735147]
23. Kho SES, Lim SG, Hoi WH, Ng PL, Tan L, Kowitlawakul Y. The development of a diabetes application for patients with poorly controlled type 2 diabetes mellitus. Comput Inform Nurs 2019 Feb;37(2):99-106. [doi: 10.1097/CIN.0000000000000485] [Medline: 30407213]
24. 40th Statistical Report On Internet Development In China. China Internet Network Information Center. URL: http://cnnic.com.cn/IDR/ReportDownloads/201807/P020180711387563090220.pdf [accessed 2022-09-01]
25. 39th Statistical Report On Internet Development In China. China Internet Network Information Center. URL: http://cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf [accessed 2022-09-01]
26. American Diabetes Association. 1. Improving care and promoting health in populations: standards of medical care in diabetes - 2020. Diabetes Care 2020 Jan;43(Suppl 1):S7-S13 [FREE Full text] [doi: 10.2337/dc20-S001] [Medline: 31862744]
27. Vas A, Devi ES, Vidyasagar S, Acharya R, Rau NR, George A, et al. Effectiveness of self-management programmes in diabetes management: A systematic review. Int J Nurs Pract 2017 Oct;23(5):e12571. [doi: 10.1111/ijn.12571] [Medline: 28758701]
28. Pamungkas RA, Chamroonsawasdi K. Self-management based coaching program to improve diabetes mellitus self-management practice and metabolic markers among uncontrolled type 2 diabetes mellitus in Indonesia: A quasi-experimental study. Diabetes Metab Syndr 2020;14(1):53-61. [doi: 10.1016/j.dsx.2019.12.002] [Medline: 31887715]
29. Karbalaeifar R, Kazempour-Ardebili S, Amiri P, Ghannadi S, Tahmasebinejad Z, Amouzegar A. Evaluating the effect of knowledge, attitude and practice on self-management in patients with type 2 diabetes. Acta Diabetol 2016 Dec;53(6):1015-1023. [doi: 10.1007/s00592-016-0905-6] [Medline: 27624577]
30. Alaofè H, Hounkpatin WA, Djrolo F, Ehiri J, Rosales C. Knowledge, attitude, practice and associated factors among patients with type 2 diabetes in Cotonou, Southern Benin. BMC Public Health 2021 Feb 12;21(1):339 [FREE Full text] [doi: 10.1186/s12889-021-10289-8] [Medline: 33579243]
31. McElfish PA, Long CR, Kohler PO, Yeary KH, Bursac Z, Narcisse M, et al. Comparative effectiveness and maintenance of diabetes self-management education interventions for Marshallese patients with type 2 diabetes: A randomized controlled trial. Diabetes Care 2019 May;42(5):849-858 [FREE Full text] [doi: 10.2337/dc18-1985] [Medline: 30862659]
32. Peimani M, Monjazebi F, Ghodssi-Ghassemabadi R, Nasli-Esfahani E. A peer support intervention in improving glycemic control in patients with type 2 diabetes. Patient Educ Couns 2018 Mar;101(3):460-466. [doi: 10.1016/j.pec.2017.10.007] [Medline: 29055655]

33. Withidpanyawong U, Lerkiatbundit S, Saengcharoen W. Family-based intervention by pharmacists for type 2 diabetes: A randomised controlled trial. Patient Educ Couns 2019 Jan;102(1):85-92. [doi: 10.1016/j.pec.2018.08.015] [Medline: 30150128]
34. McEwen MM, Pasvogel A, Murdaugh C. Effects of a family-based diabetes intervention on family social capital outcomes for Mexican American adults. Diabetes Educ 2019 Jun;45(3):272-286. [doi: 10.1177/0145721719837899] [Medline: 30895881]
35. Baig AA, Benitez A, Quinn MT, Burnet DL. Family interventions to improve diabetes outcomes for adults. Ann N Y Acad Sci 2015 Sep;1353(1):89-112 [FREE Full text] [doi: 10.1111/nyas.12844] [Medline: 26250784]
36. Mao L, Lu J, Zhang Q, Zhao Y, Chen G, Sun M, et al. Family-based intervention for patients with type 2 diabetes via WeChat in China: protocol for a randomized controlled trial. BMC Public Health 2019 Apr 05;19(1):381 [FREE Full text] [doi: 10.1186/s12889-019-6702-8] [Medline: 30953483]
37. Li Z, Chen Q, Yan J, Liang W, Wong WCW. Effectiveness of motivational interviewing on improving care for patients with type 2 diabetes in China: A randomized controlled trial. BMC Health Serv Res 2020 Jan 23;20(1):57 [FREE Full text] [doi: 10.1186/s12913-019-4776-8] [Medline: 31973759]
38. Widayanti AW, Heydon S, Green JA, Norris P. Effectiveness of an intensive community-based intervention for people with type 2 diabetes in Indonesia: A pilot study. Diabetes Res Clin Pract 2021 Nov;181:109087. [doi: 10.1016/j.diabres.2021.109087] [Medline: 34637848]
39. Kong L, Hu P, Zhao Q, Yao H, Chen S. Effect of peer support intervention on diabetes distress in people with type 2 diabetes: A systematic review and meta-analysis. Int J Nurs Pract 2020 Oct;26(5):e12830. [doi: 10.1111/ijn.12830] [Medline: 32167225]
40. Cleland J. A critique of KAP studies and some suggestions for their improvement. Stud Fam Plann 1973 Feb;4(2):42-47. [Medline: 4691025]
41. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
42. Beck RW, Connor CG, Mullen DM, Wesley DM, Bergenstal RM. The fallacy of average: How using HbA1c alone to assess glycemic control can be misleading. Diabetes Care 2017 Aug;40(8):994-999 [FREE Full text] [doi: 10.2337/dc17-0636] [Medline: 28733374]
43. American Diabetes Association. 6. glycemic targets: standards of medical care in diabetes-2018. Diabetes Care 2018 Dec;41(Suppl 1):S55-S64. [doi: 10.2337/dc18-S006] [Medline: 29222377]
44. Tang J, Wu T, Hu X, Gao L. Self-care activities among patients with type 2 diabetes mellitus: A cross-sectional study. Int J Nurs Pract 2021 Dec;27(6):e12987. [doi: 10.1111/ijn.12987] [Medline: 34101931]
45. Nie R, Han Y, Xu J, Huang Q, Mao J. Illness perception, risk perception and health promotion self-care behaviors among Chinese patient with type 2 diabetes: A cross-sectional survey. Appl Nurs Res 2018 Feb;39:89-96. [doi: 10.1016/j.apnr.2017.11.010] [Medline: 29422183]
46. Kang C, Chang S, Chen P, Liu P, Liu W, Chang C, et al. Comparison of family partnership intervention care vs. conventional care in adult patients with poorly controlled type 2 diabetes in a community hospital: a randomized controlled trial. Int J Nurs Stud 2010 Nov;47(11):1363-1373. [doi: 10.1016/j.ijnurstu.2010.03.009] [Medline: 20371056]
47. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950 [FREE Full text] [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
48. Walker EA, Caban A, Schechter CB, Basch CE, Blanco E, DeWitt T, et al. Measuring comparative risk perceptions in an urban minority population: the risk perception survey for diabetes. Diabetes Educ 2007;33(1):103-110. [doi: 10.1177/0145721706298198] [Medline: 17272796]
49. Lewin AB, Geffken GR, Heidgerken AD, Duke DC, Novoa W, Williams LB, et al. The Diabetes Family Behavior Checklist: a psychometric evaluation. J Clin Psychol Med Settings 2005 Dec;12(4):315-322. [doi: 10.1007/s10880-005-7817-x]
50. McEwen MM, Pasvogel A, Murdaugh C, Hepworth J. Effects of a family-based diabetes intervention on behavioral and biological outcomes for Mexican American adults. Diabetes Educ 2017 Jun;43(3):272-285 [FREE Full text] [doi: 10.1177/0145721717706031] [Medline: 28447545]
51. Cai C, Hu J. Effectiveness of a family-based diabetes self-management educational intervention for Chinese adults with type 2 diabetes in Wuhan, China. Diabetes Educ 2016 Dec;42(6):697-711. [doi: 10.1177/0145721716674325] [Medline: 27831523]
52. Wichit N, Mnatzaganian G, Courtney M, Schulz P, Johnson M. Randomized controlled trial of a family-oriented self-management program to improve self-efficacy, glycemic control and quality of life among Thai individuals with Type 2 diabetes. Diabetes Res Clin Pract 2017 Jan;123:37-48 [FREE Full text] [doi: 10.1016/j.diabres.2016.11.013] [Medline: 27918976]
53. Vorderstrasse A, Lewinski A, Melkus GD, Johnson C. Social support for diabetes self-management via eHealth interventions. Curr Diab Rep 2016 Jul;16(7):56. [doi: 10.1007/s11892-016-0756-0] [Medline: 27155606]
54. Hemmati Maslakpak M, Razmara S, Niazkhani Z. Effects of face-to-face and telephone-based family-oriented education on self-care behavior and patient outcomes in type 2 diabetes: a randomized controlled trial. J Diabetes Res 2017;2017:8404328 [FREE Full text] [doi: 10.1155/2017/8404328] [Medline: 29359166]

55. Zimmermann G, Venkatesan A, Rawlings K, Scahill MD. Improved glycemic control with a digital health intervention in adults with type 2 diabetes: Retrospective study. JMIR Diabetes 2021 Jun 02;6(2):e28033 [FREE Full text] [doi: 10.2196/28033] [Medline: 34075880]
56. Yang Y, Lee EY, Kim H, Lee S, Yoon K, Cho J. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: Cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266 [FREE Full text] [doi: 10.2196/16266] [Medline: 32130172]
57. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Diabetes Care 2015 Jul;38(7):1372-1382. [doi: 10.2337/dc15-0730] [Medline: 26048904]
58. Glanz K, Lewis FM, Rimer BK. Health Behavior and Health Education: Theory, Research, and Practice. New York, NY: John Wiley & Sons; 2008.
59. Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997 Jan 28;25(1):95-100. [doi: 10.1046/j.1365-2648.1997.1997025095.x] [Medline: 9004016]
60. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016;9:381-390 [FREE Full text] [doi: 10.2147/DMSO.S95247] [Medline: 27853384]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HbA_{1c}: glycated hemoglobin A_{1c}
HBM: Health Belief Model
KAP: Knowledge-Attitude-Practice
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 20.06.22; peer-reviewed by R Jiwani, V Buss; comments to author 16.09.22; revised version received 09.11.22; accepted 01.03.23; published 20.03.23.

Please cite as:
Feng Y, Zhao Y, Mao L, Gu M, Yuan H, Lu J, Zhang Q, Zhao Q, Li X
The Effectiveness of an eHealth Family-Based Intervention Program in Patients With Uncontrolled Type 2 Diabetes Mellitus (T2DM) in the Community Via WeChat: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e40420
URL: https://mhealth.jmir.org/2023/1/e40420
doi:10.2196/40420
PMID:36939825

Original Paper

Testing Mechanisms of Change for Text Message–Delivered Cognitive Behavioral Therapy: Randomized Clinical Trial for Young Adult Depression

Michael J Mason[1], PhD; J Douglas Coatsworth[1], PhD; Nikola Zaharakis[1], PhD; Michael Russell[2], PhD; Aaron Brown[3], PhD; Sydney McKinstry[1], BA

[1]Center for Behavioral Health Research, College of Social Work, University of Tennessee, Knoxville, TN, United States
[2]Department of Biobehavioral Health, Pennsylvania State University, University Park, PA, United States
[3]College of Social Work, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Michael J Mason, PhD
Center for Behavioral Health Research
College of Social Work
University of Tennessee
Center for Behavioral Health Research
202 Henson Hall
Knoxville, TN, 37996
United States
Phone: 1 18659749148
Email: mmason29@utk.edu

Abstract

Background: Current psychiatric epidemiological evidence estimates that 17% of young adults (aged 18-25 years) experienced a major depressive episode in 2020, relative to 8.4% of all adults aged ≥26 years. Young adults with a major depressive episode in the past year are the least likely to receive treatment for depression compared with other age groups.

Objective: We conducted a randomized clinical trial following our initial 4-week SMS text message–delivered cognitive behavioral therapy (CBT-txt) for depression in young adults. We sought to test mechanisms of change for CBT-txt.

Methods: Based on participant feedback, outcome data, and the empirical literature, we increased the treatment dosage from 4-8 weeks and tested 3 mechanisms of change with 103 young adults in the United States. Participants were from 34 states, recruited from Facebook and Instagram and presenting with at least moderate depressive symptomatology. Web-based assessments occurred at baseline prior to randomization and at 1, 2, and 3 months after enrollment. The primary outcome, the severity of depressive symptoms, was assessed using the Beck Depression Inventory II. Behavioral activation, perseverative thinking, and cognitive distortions were measured as mechanisms of change. Participants were randomized to CBT-txt or a waitlist control condition. Those assigned to the CBT-txt intervention condition received 474 fully automated SMS text messages, delivered every other day over a 64-day period and averaging 14.8 (SD 2.4) SMS text messages per treatment day. Intervention texts are delivered via TextIt, a web-based automated SMS text messaging platform.

Results: Across all 3 months of the study, participants in the CBT-txt group showed significantly larger decreases in depressive symptoms than those in the control group (P<.001 at each follow-up), producing a medium-to-large effect size (Cohen d=0.76). Over half (25/47, 53%) of the treatment group moved into the "high-end functioning" category, representing no or minimal clinically significant depressive symptoms, compared with 15% (8/53) of the control condition. Mediation analysis showed that CBT-txt appeared to lead to greater increases in behavioral activation and greater decreases in cognitive distortions and perseverative thinking across the 3-month follow-up period, which were then associated with larger baseline to 3-month decreases in depression. The size of the indirect effects was substantial: 57%, 41%, and 50% of the CBT-txt effect on changes in depression were mediated by changes in behavioral activation, cognitive distortions, and perseverative thinking, respectively. Models including all 3 mediators simultaneously showed that 63% of the CBT-txt effect was mediated by the combined indirect effects.

Conclusions: Results provide evidence for the efficacy of CBT-txt to reduce young adult depressive symptoms through hypothesized mechanisms. To the best of our knowledge, CBT-txt is unique in its SMS text message–delivered modality, the strong clinical evidence supporting efficacy and mechanisms of change.

Trial Registration: ClinicalTrials.gov NCT05551702; https://clinicaltrials.gov/study/NCT05551702

(JMIR Mhealth Uhealth 2023;11:e45186) doi:10.2196/45186

KEYWORDS

young adults; depression; SMS text message–delivered treatment; cognitive behavioral therapy; randomized clinical trial; mobile health treatment; mHealth treatment; mobile phone

Introduction

Background

In 2020, young adults (ages 18-25 years) in the United States experienced more major depressive episodes than other age groups and were the least likely to receive treatment for depression (eg, talking with a health provider or taking prescription medication) [1].

Even when the rates for the treatment of depression rose nationwide, the rates for the treatment of depression for young adults rose at a significantly lower rate [2]. Paradoxically, more young adults with depression report greater perceived unmet need for mental health services than other age groups [1]. From 2011 to 2019, the most common reason that young adults report for not receiving treatment is cost, followed by "not knowing where to go to receive treatment" [3]. This unmet treatment need has been consistent over the last decade and places this group at an increased risk for substance use disorders, possibly indicating self-medication with substances to reduce depressive symptoms [4].

Mobile Health Treatments

Creative mental health treatment strategies are needed to reach young adults to address this unmet need. Digitally delivered mobile health (mHealth) treatments (or mHealth interventions) show promise for reaching young adults and serving as a clinical tool to address depression in young adults. Because mHealth interventions can increase large-scale access to evidence-based treatments, some countries such as New Zealand are integrating these approaches into their national health service infrastructure [5]. The rapid development and use of these treatments is a promising step for addressing psychiatric disorders. However, there is reason to be cautious regarding the science behind these efforts. In particular, rigorous studies on 2 important features of mHealth interventions are lacking in the research literature: testing treatment dosage and clinical mechanisms of change. First, understanding the appropriate dosage for mHealth interventions is critical to advance an empirical database that has maximum clinical utility. Knowing the correct dosage (how much treatment is needed to be delivered to meet the treatment goals of particular patients) is a primary goal of clinical research. In a recent meta-analysis that assessed the relationship between mHealth intervention length and intervention effect, Lu et al [6] found that interventions of at least 7 weeks' duration had larger effect sizes on anxiety symptom reduction. However, for depression treatment, the optimal dosage remains unclear. Second, understanding the mechanisms of behavior change has not been adequately studied in mHealth interventions. Digital interventions have tremendous potential for identifying mechanisms of behavior change because these can be cost-effectively delivered to large enough samples to sufficiently power studies to detect mechanisms of change, the content is delivered reliably and precisely, and the manipulation of the hypothesized mediators can be strategically sequenced within a study design [7]. Unfortunately, many mHealth studies lack scientific rigor (eg, the lack of a priori hypothesized causal mechanisms of change coupled with an appropriate design), which is necessary to conduct mediational analyses [7]. In addition, although the delivery of mHealth interventions for mental disorders continues to expand, accompanying rigorous randomized clinical trials, including those that test dosage and mechanisms, have not kept pace with this growth [8].

Clinical Foundation

This study was conducted to better understand the mechanisms of SMS text message–delivered cognitive behavioral therapy (CBT-txt) as well as the acceptability of the treatment length. Our recent randomized clinical trial treating 102 young adults in the United States with at least moderate depressive symptomatology found that CBT-txt was acceptable, feasible, and efficacious compared with a waitlist control condition [9]. This 2-month trial had 4 weeks of treatment and tested a single mediator, behavioral activation, with assessments at baseline and at 1- and 2-month follow-ups. The treatment length mirrored the duration of our successful text-delivered cannabis use disorder treatment [10]. The initial CBT-txt trial found that the strongest treatment effect appeared at the 1-month follow-up (immediately after the treatment ended), particularly for participants who began with severe depressive symptoms. Mediation analysis revealed significant indirect treatment effects of increases in behavioral activation on reducing depressive symptoms, suggesting a mechanism of change [9,11]. Based upon these promising findings, we examined the treatment satisfaction data from this study, which revealed participants' interest in receiving more treatment content and an increase in the duration of the intervention. Accordingly, we expanded CBT-txt from 4 to 8 weeks, which is consistent with the average mHealth dosage of 7 to 8 weeks [6,12]. We also expanded the treatment's clinical content to include cognitive distortions and perseverative thinking, 2 additional candidate mediators common to cognitive behavioral therapy (CBT) treatment of depression [13-15].

This Study

Four hypotheses were tested for this follow-up study. The first hypothesis was that participants allocated to the treatment condition would show greater reductions in their depressive symptoms immediately following treatment (2 months after enrollment) and at the 3-month follow-up than participants in the waitlist control condition. Because treatment response has been linked to baseline symptom severity [9,16], we also tested whether treatment effects were moderated by the participants' baseline severity level. The second hypothesis was that treatment effects would be mediated by behavioral activation, such that increases in behavioral activation would reduce depressive symptoms in the treatment condition relative to the controls. The third hypothesis was that treatment effects would be mediated by cognitive distortions, such that decreases in cognitive distortions would reduce depressive symptoms in the treatment condition relative to controls. The fourth hypothesis was that treatment effects would be mediated by perseverative thinking, such that decreases in perseverative thinking would reduce depressive symptoms in the treatment condition relative to the controls. In addition to testing these 4 hypotheses, we were interested in comparing the results (effect sizes and treatment response over time) from this 8-week intervention with those from our prior trial of the 4-week intervention.

Methods

Procedures

In total, 103 young adults (ages 18-25 years) were recruited and enrolled in a 3-month randomized clinical trial. Participants were recruited using age-targeted advertising on Facebook and Instagram for young adults residing in the United States. Recruitment occurred over a 7-week period from July 6 to August 28, 2022. The study was advertised to all those within the study age range (18-25 years) across the United States who used Facebook and Instagram. With the aim of recruiting a geographic, racial and ethnic, and socioeconomically diverse sample, we placed no restrictions on advertisements other than age, used advertisements with a variety of images featuring people of different racial groups, and used a variety of placements (eg, feeds, stories, and reels) to reach the widest swath of potential participants. Interested individuals were directed to a study website where they would read more about the study and answer eligibility screening questions on their phones.

The study was registered at ClinicalTrials.gov (identifier NCT05551702).

Design

The study design was a 2-arm randomized clinical trial with participants allocated to either the experimental condition or the waitlist control condition. Eligibility requirements were (1) age between 18 and 25 years; (2) a score of at least 10 on the Patient Health Questionnaire–9 (PHQ-9) [17], indicating at least moderate depressive symptom severity; (3) access to a smartphone; (4) fluent in English; (5) have not received treatment for depression in the past 3 months; and (6) did not endorse suicidal ideation (SI) on the PHQ-9 measure. All interested participants completed a screening questionnaire assessing eligibility. Those who screened positive for suicide and those who did not meet the study criteria were immediately referred to local and national mental health resources. To ensure that participants who did not qualify for the study still had access to care, participants who were not eligible based on the endorsement of SI had the option of speaking with a project staff member who is a licensed mental health professional for assistance with referral to care. Individuals who wanted to talk with a licensed mental health staff member were contacted within 1 business day and provided contact information to at least 3 mental health providers in their area, including at least 1 provider who offered sliding scale or safety net (ie, no cost) services. In addition, individuals were given the contact information to the National Alliance on Mental Illness location in their area. The National Alliance on Mental Illness provides additional mental health resources, including no-cost support groups for individuals living with mental illness [18].

Remote Data Collection Quality Control

Data were reviewed carefully on a daily basis for quality control. Beginning July 26, 2022, we noticed increased screening activity with similar name patterns. After carefully reviewing the screening data, including the name, mailing address, IP address, phone number with area code and service carrier, and survey responses, we determined that 17 enrolled cases may have been fraudulent. We reached out via text, phone, and email to these 17 cases to confirm their identity but none of the participants responded. All 17 cases were administratively removed from the study on July 28, 2022. We enacted other procedures, such as using Qualtrics (Qualtrics), the web-based survey program that flags responses likely to be from bots, attempts to prevent multiple submissions from a single respondent, and assigns a fraud score indicating the likelihood of a response being fraudulent [19]. Furthermore, at screening, participants must enter a unique phone number, which is then verified using a Twilio mobile phone app.

Eligible individuals were instructed to complete the baseline survey on their phones, where upon completion, they were randomized to either the treatment or waitlist control condition. Randomization was automated by Qualtrics as part of the baseline survey. Randomization was stratified by sex using block randomization with a fixed block size of 10 to reduce bias during randomization and to ensure equal representation of sex across both conditions. The participants completed follow-up assessments at the 1-, 2-, and 3-month follow-ups. Waitlist condition participants were eligible to receive the treatment texts upon the completion of the 3-month follow-up survey. Participants who completed the screening, baseline, and all follow-up assessments received US $150 in Amazon eGift cards.

Text-Delivered Treatment: CBT-txt

Overview

CBT-txt was adapted from an evidenced-based, in-person CBT treatment manual [20] shown to be effective in reducing depressive symptoms [21-23] and adaptable to digital formats [24]. CBT-txt focuses on empowering participants to understand how thoughts, activities, and other people affect their moods. CBT-txt is a fully automated SMS text message–delivered program that initiates a conversation at predetermined times and requires a participant response to activate the subsequent text message. Tailored messages are sent based on the participant's responses or ratings of their depression. For example, a participant may be asked about their engagement with CBT skills and are given 3 response choices: a, b, or c. Each response option activates a different CBT-txt message providing tailored responses. The treatment provides individualized text-based conversations every other day over the course of the intervention.

Theoretical Underpinnings of CBT-txt

The theory underlying CBT-txt is the Generic Cognitive Model, which specifies common cognitive and behavioral processes associated with disorders such as depression [25]. These processes manifest along a continuum from normal adaptive functioning to psychopathology. As individuals experience environmental, psychological, and social stimuli, their attentional response system is activated to determine an adaptive response. The attentional response activates schemas, defined as internally stored representations of stimuli that form underlying structures for organizing perceptions of the world [26]. When the schema is maladaptive or disproportionate relative to the stimuli, the individual experiences psychological problems that can escalate into a psychiatric disorder.

Clinical Structure of CBT-txt

We expanded our initial 4-week CBT-txt treatment [9] into an 8-week intervention. CBT-txt intervention content is guided by CBT core mechanisms [13] and is sequenced using the in-person treatment manual of Muñoz et al [20]. CBT-txt is organized around eight topical areas delivered across 8 weeks as indicated in Table 1: (1) introduction to CBT, (2) automatic thoughts, (3) behavioral activation, (4) automatic thoughts and health, (5) perseverative thinking, (6) cognitive distortions, (7) more on behavioral activation, and (8) social support and summary.

Table 1. SMS text message–delivered cognitive behavioral therapy (CBT) content and structure (474 texts over 8 weeks; mean 14.8 texts per day, every other day).

Week	Days, n	Texts, n	Treatment content
1	4	62	How we think about depression (Intro to CBT, moods, thoughts, feelings)
2	4	72	How thoughts affect moods (Automatic thoughts; ABCD method)
3	4	62	How activities affect moods (Behavioral activation)
4	4	57	How thoughts affect our habits and health (ABCD method & health)
5	4	55	How repetitive negative thinking affects moods (Perseverative thinking)
6	4	53	How cognitive distortions affect our moods (Cognitive distortion)
7	4	59	How activities, goals, and values affect our moods (Behavioral activation)
8	4	54	How other people affect moods (Social support and moods); Summary

Participants assigned to the CBT-txt intervention condition received 474 texts delivered every other day over a 64-day period, averaging 14.8 texts per treatment day. Participants indicated the time of day they would like to receive the intervention texts. Intervention texts were delivered via a web-based automated SMS text messaging platform called TextIt. Project staff programmed TextIt to deliver intervention texts and extract data from the web-based survey platform Qualtrics to automatically personalize intervention texts. Texts were individualized based on data provided by participants in the baseline survey as well as throughout the treatment period.

In addition to the scheduled intervention messages, participants could access automated booster messaging to provide on-demand supportive messages by texting “4MOOD” at any time, as needed. The “4MOOD” messages are organized around topics (eg, cognitive techniques and positive activities) such that participants select a category of message to receive each time they text “4MOOD” for additional support. All text messages end with the response, “If this is a crisis call 911.” If participants want to find out about receiving professional help, they choose option 4 and the program texts them a link to a list of national mental health services, suicide prevention hotlines, and a child and adult abuse hotline. This list includes a link to access Substance Abuse and Mental Health Services Administration’s treatment locator, which offers a database of treatment providers that can be filtered by location. Textbox 1 provides example SMS text messages of CBT-txt as well as an example booster message.

Textbox 1. Example of a texting conversation over 1 day. (The italicized text is autopopulated by programming and is programming logic; it is not shown to the participant. This limited example of SMS text message–delivered cognitive behavioral therapy does not provide the context of previous texting conversations or subsequent conversations that are built on past conversations.)

1. Hi *name*. Before we start, how is your mood right now? (0=very good-10=very depressed) Text 0-10
2. Let's look at how thoughts, what we tell ourselves, affect our mood. We are usually not aware of these negative thinking patterns that influence our mood
3. Text *selftalk* any time to read the 10 common negative thinking patterns. We believe these thoughts because we think them and don't challenge them. Try it now!
4. We actually fool ourselves and create more bad feelings about things that are *simply not true.* Our depressed thinking keeps us in this cycle.
5. Study the list carefully, see which patterns you use. We will refer to these often as these serve as a guide to identify and change our thinking and mood. Text OK
6. Looking at the list, identify at least 1 of the thinking patterns that you've used recently. Text back the number of the pattern. Text 1-10 or UNSURE

If 1-10:

Good job identifying a pattern. *all or nothing thinking* is very common as are most of these. We're unaware of these patterns so we call them automatic thoughts

So, what was the situation that led you to use *all or nothing thinking*? Please describe. Text__

IF UNSURE:

Ok, Look at the list again & think about this some more: http://selftalk.mjmood.com Text READY when you're ready to discuss. Text UNSURE if you're still unsure.

If READY, repeat #6

IF UNSURE:

Ok, we know it can be hard to look at the way we talk to ourselves honestly. Remember that the more you put into this program the more you will get out of it.

1. Carefully observe your automatic thoughts the next few days. Try to challenge them as not accurate or facts, but a result of your being in a depressed mood
2. This is a critical skill for you to learn. It seems simple, but when you practice this you'll be surprised at how much this can change how you feel.
3. Thanks *name*! That ends today's texts. Need mood support? Text: *4MOOD*. This is not being read immediately, if this is a crisis call 911

Fidelity of CBT-txt

We followed previous successful research in developing text-delivered interventions derived from in-person treatments [9,27-29]. The following four steps were used to ensure that the core mechanisms of CBT were incorporated into CBT-txt: (1) selected an evidence-based treatment manual as an adaptation source [20] as well as primary CBT source materials [26], (2) developed texts to match the treatment manual content and structure [20], (3) cross-checked the texts against the core mechanisms of CBT for depression [13], and (4) applied a CBT fidelity scale [30] to rate the texts with an outside expert, Dr John Curry, who provided independent quality assurance scoring. Dr John Curry is a professor of Psychiatry and Behavioral Sciences at Duke University and is a nationally recognized expert in CBT who wrote the protocol for the Treatment of Adolescents with Depression Study [31]. Finally, we applied the results of the fidelity review (average rating of 4 out of 5, with 5 being excellent) and made revisions as needed (provided more overview and summary content).

Participant Safety Protocol

We instituted a participant safety protocol that reviewed all incoming texts for crisis-related words both automatically (ie, autotext review) and with staff reading every text from all participants at least once per day. If participants indicated SI on the Beck Depression Inventory-II (BDI-II) or if their follow-up BDI-II (see the *Measures* section) score increased, they were automatically sent a supportive SMS text message that included national resources. Furthermore, a licensed mental health professional on the staff texted, emailed, and called participants falling into these at-risk categories to ensure safety, ensure the availability of resources, and provide local treatment resources and general support.

Measures

Demographics

Participant age, sex, race and ethnicity, socioeconomic status, family history of depression, and current use of antidepressant medication were collected at the baseline assessment.

Acceptability and Engagement

Participants' acceptability of the intervention was measured via participant-reported satisfaction with the intervention as well as passively collected engagement with the intervention content and features. Satisfaction was assessed using a measure developed in our past work [32]. Two items assessed whether the number of days on which texts were received and the number of texts received each day were (1) *too few*, (2) *just right*, or (3) *too many*. An additional 2 items measured whether the texts were easy to understand and complete and whether the

“4MOOD” booster messages were a helpful option. In addition, 5 items measured perceptions of the helpfulness of the intervention content (Cronbach α=.88), and 6 items assessed participants’ practice of the intervention skills taught (Cronbach α=.78). All of these items were rated on a 5-point Likert scale (1=*strongly disagree* to 5=*stronglyagree*). Engagement with the intervention was measured using data passively collected that were programmed to be automatically gathered during the administration of the intervention, including the number of intervention texts a participant responded to each day and the number of booster SMS text messages requested. Intervention completion was defined as responding to ≥95% of all intervention texts (at least 192 responses out of a total of 198 responses across all intervention days).

Screen of Depression Symptoms

The PHQ-9 was used to determine eligibility [17]. The PHQ-9 is a questionnaire consisting of all 9 criteria for major depressive disorder (MDD). Responses range from “not at all” (score=0) to “nearly every day” (score=3). Item scores are summed for a total score ranging from 0 to 27. The PHQ-9 has good validity, test-retest reliability, and internal consistency. Cronbach α in this study’s sample was .50.

Depression Symptoms and Severity

Baseline and follow-up assessment of depression severity was assessed with BDI-II [33]. There are 21 items, each corresponding with a symptom of depression, and possible scores on each item range from 0 to 3. The scores are summed to obtain a single severity score. A score of 0 to 13 indicates none or minimal depressive symptoms, a score of 14 to 19 indicates mild depressive symptoms, a score of 20 to 28 indicates moderate depressive symptoms, and a score of ≥29 indicate severe depressive symptoms. Cronbach α in this study’s sample at each assessment were .83, .89, .92, and .92, respectively.

Behavioral Activation

Behavioral activation was measured using the Behavioral Activation for Depression Scale, Short Form [34]. The Behavioral Activation for Depression Scale, Short Form is a 9-item scale used to assess activation toward goals during the treatment for depression. Items are scored from 0 to 6, and higher scores indicate greater activation toward goals and less avoidance of tasks. Cronbach α in this study’s sample at each assessment were .70, .70, .82, and .81, respectively.

Perseverative Thinking

Repetitive negative thinking was measured using the Perseverative Thinking Questionnaire (PTQ) [35]. The PTQ is a 15-item questionnaire used to characterize respondents thinking about negative experiences or problems. Items are scored on a 5-point scale from 0=*never* to 4=*almost always* and are summed with higher scores indicating more repetitive negative thinking. Cronbach α in this study’s sample at each assessment was .94, .95, .93, and .97, respectively.

Cognitive Distortion

Cognitive distortion was measured using the Cognitive Distortions Scale (CDS) [36]. The CDS assesses 10 types of thinking biases (eg, catastrophizing or all-or-nothing thinking). Participants rate the frequency of their use of each type of thinking. Items are scored on a 7-point scale from 0=*never* to 7=*all the time* and are summed with higher scores indicating more cognitive distortion. Cronbach α in this study’s sample at each assessment was .92, .94, .93, and .95, respectively.

Statistical Analyses

Latent Change Score Modeling

Latent change score (LCS) analyses were conducted in a structural equation modeling framework using the *lavaan* package in R (R Foundation for Statistical Computing). LCS was used instead of more commonly used modeling methods, such as latent growth modeling and repeated measures ANOVA, because it offers 2 unique capabilities. First, it allows for the estimation of wave-to-wave change in mediators and outcomes adjusted for previous levels, which controls for regression to the mean. Second, it does not force the pattern of change to follow a prespecified shape; wave-to-wave changes are allowed to freely vary across time [37,38]. LCSs are created by (1) specifying an autoregression of each score on its immediately previous time point while fixing the coefficient to 1 and (2) specifying a latent factor with a loading of 1 on the current time point to capture the difference. These LCSs are then used as dependent variables in regressions testing for CBT-txt intervention effects. BDI-II depression scores were centered on their pretreatment mean; CBT-txt was also centered (CBT-txt=0.51; control=−0.49). This centering strategy facilitates the interpretation of model intercepts as a change from the previous month for a person with average pretreatment levels of depression, averaged across conditions. All LCS regressions controlled for both pretreatment levels of depression and prior-month levels of depression to account for the likely association between pretreatment severity and posttreatment change.

Mediation Analysis

Mediation was also tested in LCS models. LCSs were created for (1) the baseline to 3-month change in the mediator (Behavioral Activation for Depression Scale, CDS, and PTQ in separate models) and (2) the baseline to 3-month change in the outcome (BDI-II). Pretreatment mediator and BDI-II scores at baseline were included as covariates in the model. The change in mediators was regressed on CBT-txt (and baseline covariates) to create the *a* path; the change in BDI-II was regressed on the change in the mediator (the *b* path) and CBT-txt (the direct effect or *c’* path). The significance of the indirect effect (*a***b*) was tested using bias-corrected bootstrapped CIs with 10,000 bootstrap draws. The indirect effect tests the hypothesis that CBT-txt influences BDI-II through the mediator.

Ethics Approval

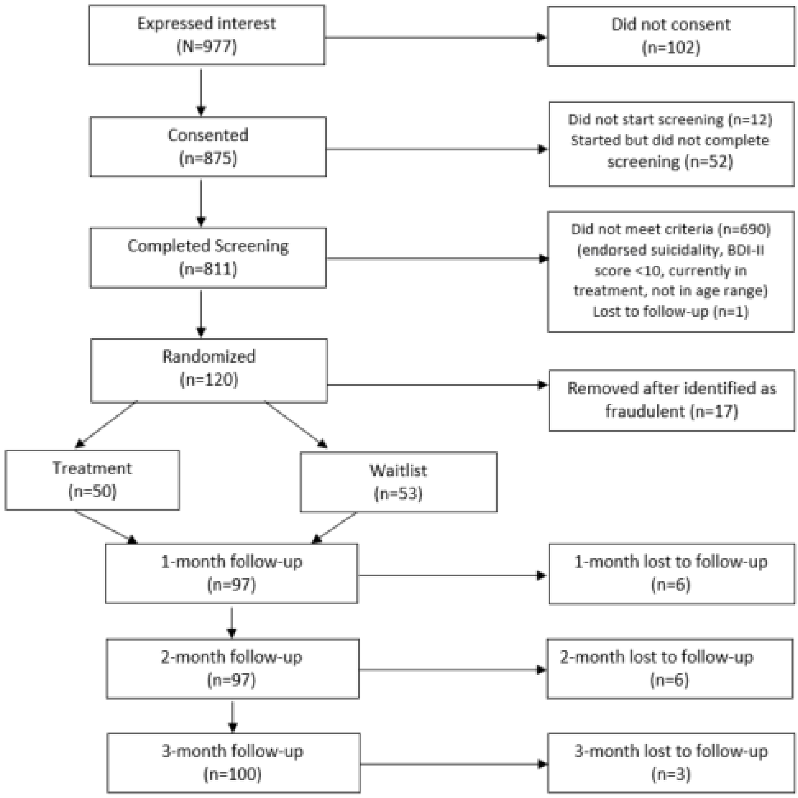
All procedures were approved by the institutional review board of The University of Tennessee (approval UTK IRB-20-06164-FB).

Results

Demographic Results

Our sample size was 103. The participants' mean age was 22 (SD 2.2) years and 84.5% (87/103) were female, and they resided in 34 different states in the United States and the District of Columbia. The sample comprised 15.5% (16/103) Asian, 3.9% (4/103) Black or African American, 7.8% (8/103) Hispanic or Latino, 8.7% (9/103) more than 1 race, and 63.1% (65/103) White participants. While not used analytically, college student status, childhood socioeconomic status ("Has your family ever received food assistance, such as free or reduced lunch, or SNAP benefits?"), and past and current use of antidepression medication were described to characterize the sample. Slightly more than half (55/103, 54.5%) were not enrolled in college part time or full time. Approximately one-third (35/103, 34%) of the participants endorsed a family history of government food assistance, and 42.7% (44/103) of participants had been prescribed antidepressant medication at some time in their life but only 20.4% (21/103) were currently prescribed antidepressant medication. Figure 1 provides details of the enrollment and allocation process via the CONSORT (Consolidated Standards of Reporting Trials) diagram.

Figure 1. CONSORT diagram for study participant flow and condition allocation.



Acceptability and Engagement

Participants endorsed levels of acceptability and engagement similar to our previous pilot of CBT-txt for young adult depression. Two-thirds (69/103, 67%) of the participants completed the intervention (responding to ≥95% of intervention texts), although somewhat lower than in our previous pilot (85%) [9] but still substantially higher than the average response rates for unguided internet-delivered CBT (40%) [39]. On average, participants completed 163.6 (SD 60.2) texts of the total potential 198 responses. Slightly more participants (37/50, 74%) completed the first month of texts (111 total potential responses; mean 95.3, SD 30.6) than the second month of texts (33/50, 66%; mean 68.3, SD 30.9). The participants endorsed high levels of satisfaction with the treatment content and features. More participants agreed or strongly agreed (35/44, 80%) that the intervention texts were more helpful than in our first pilot study (77%), that the number of days of texts received per day was "just right" (32/42, 79%) than previously (76%), and that the number of texts per day was "just right" (36/42, 86%) than in the previous study (72%) [9]. The overall mean of the helpfulness subscale (mean 4.0, SD 0.7) was very similar to that of the prior trial (mean 3.9, SD 0.8) [9]. However, fewer participants (36/44, 82%) reported that the texts were easy to understand and complete than those in the previous trial (93%) [9]. The participants endorsed moderate levels of implementing the skills learned (mean 3.2, SD 0.7), very similarly to our previous study [9]. Slightly more (32/44, 73%) number of participants agreed that the booster messages were a useful option compared with our prior study (68%) [9]. However, fewer participants (20/50, 40%) used booster messages than the previous trial (42%) [9].

LCS Results

Main Effect of CBT-txt on Depression

Figure 2 shows the Beck Depression Inventory means over time for the CBT-txt and control groups. Significant treatment–control group differences were observed at each of the 3 postenrollment follow-ups (P<.001 at each follow-up).

No treatment–control group difference was seen in pretreatment depression, supporting balance across groups in baseline levels. Table 2 shows the results of the LCS model testing whether the monthly changes differed between the CBT-txt and control groups. At each of the 3 follow-ups, young adults in the CBT-txt group showed significantly larger decreases in depression than those in the control group, producing a medium to large effect size (Cohen *d*=0.76).

Figure 2. Mean scores of the Beck Depression Inventory-II scores over time by condition. CBT-txt: SMS text message–delivered cognitive behavioral therapy.

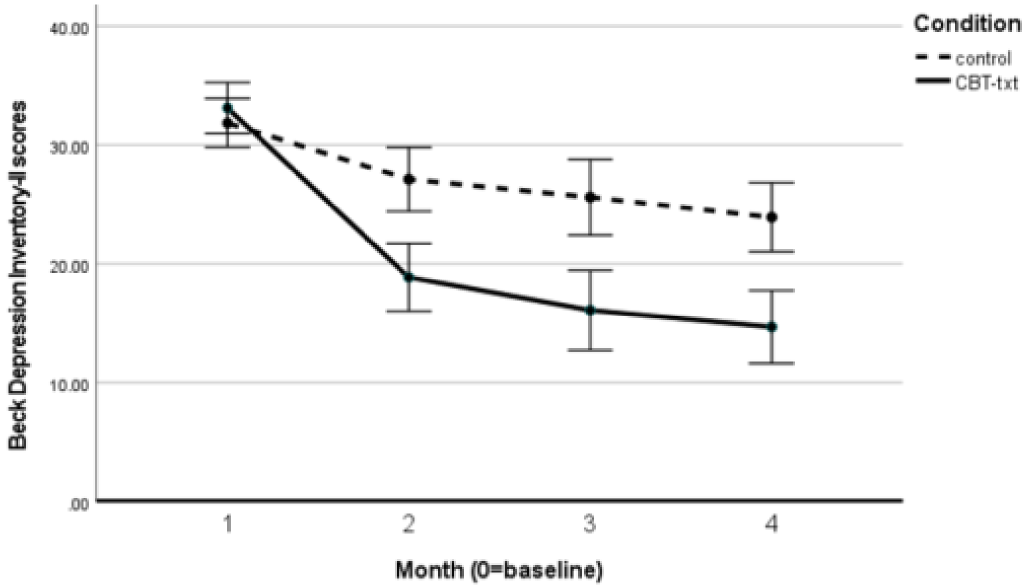


Table 2. Latent change score model results testing SMS text message–delivered cognitive behavioral therapy (CBT-txt) efficacy for depression treatment.

	Estimate	SE	*Z* value	*P* value	95% CI
Baseline depression level					
Intercept	0.00	0.89	0.00	>.99	−1.75 to 1.75
CBT-txt	−0.22	1.78	−0.12	.90	−3.71 to 3.27
Change					
Baseline–1 month					
Intercept	*−9.17*[a]	0.86	−10.63	<.001	−10.86 to −7.48
CBT-txt	*−9.02*	1.73	−5.22	<.001	−12.40 to −5.63
Baseline depression level	*−0.51*	0.10	−5.33	<.001	−0.70 to −0.32
1 month–2 month					
Intercept	*−4.65*	1.20	−3.88	<.001	−7.00 to −2.30
CBT-txt	*−3.79*	1.82	−2.08	.04	−7.36 to −0.21
1-month depression Level	*−0.27*	0.10	−2.87	.004	−0.46 to −0.09
Baseline depression level	*0.20*	0.10	1.97	.049	0.00 to 0.40
2 month–3 month					
Intercept	*−6.68*	1.27	−5.25	<.001	−9.18 to −4.19
CBT-txt	*−4.64*	1.90	−2.44	.02	−8.37 to −0.91
2-month depression level	*−0.50*	0.09	−5.93	<.001	−0.67 to −0.34
Baseline depression level	0.07	0.11	0.69	.49	−0.13 to 0.28

[a]Italicized text indicates *P*<.05.

Clinical Significance

We used 2 measures of clinical significance. First, the number of participants at the end of the trial with high-end state functioning as defined by Jacobson and Truax [40] was measured. High-end state functioning is defined as a participant having a BDI-II score between 0 and 13. By examining the BDI-II severity levels based on total scores (*none to minimal* *depressive symptoms*=0-13, *mild depressive symptoms*=14-19, *moderate depressive symptoms*=20-28, and *severe depressive* symptoms=29-63), treatment responses can be understood clinically. Thus, scores between 0 and 13 represent high-end state functioning owing to the elimination of symptoms used to classify participants with MDD. Over half (25/47, 53%) of the treatment group moved to the "none to minimal" category, compared with 15% (8/53) of the control group. Second, we used a reliable change index (RCI), which is a measure of how much change has occurred during the course of treatment [40]. Our study produced an RCI of 4.46 with 95% confidence, which is twice the minimum required RCI of 1.96. This RCI represents 2 SDs of clinical change for the treatment group based on the baseline to 3-month follow-up BDI-II scores. Table 3 displays the depression severity level percentages at the 3-month assessment for comparison by experimental conditions.

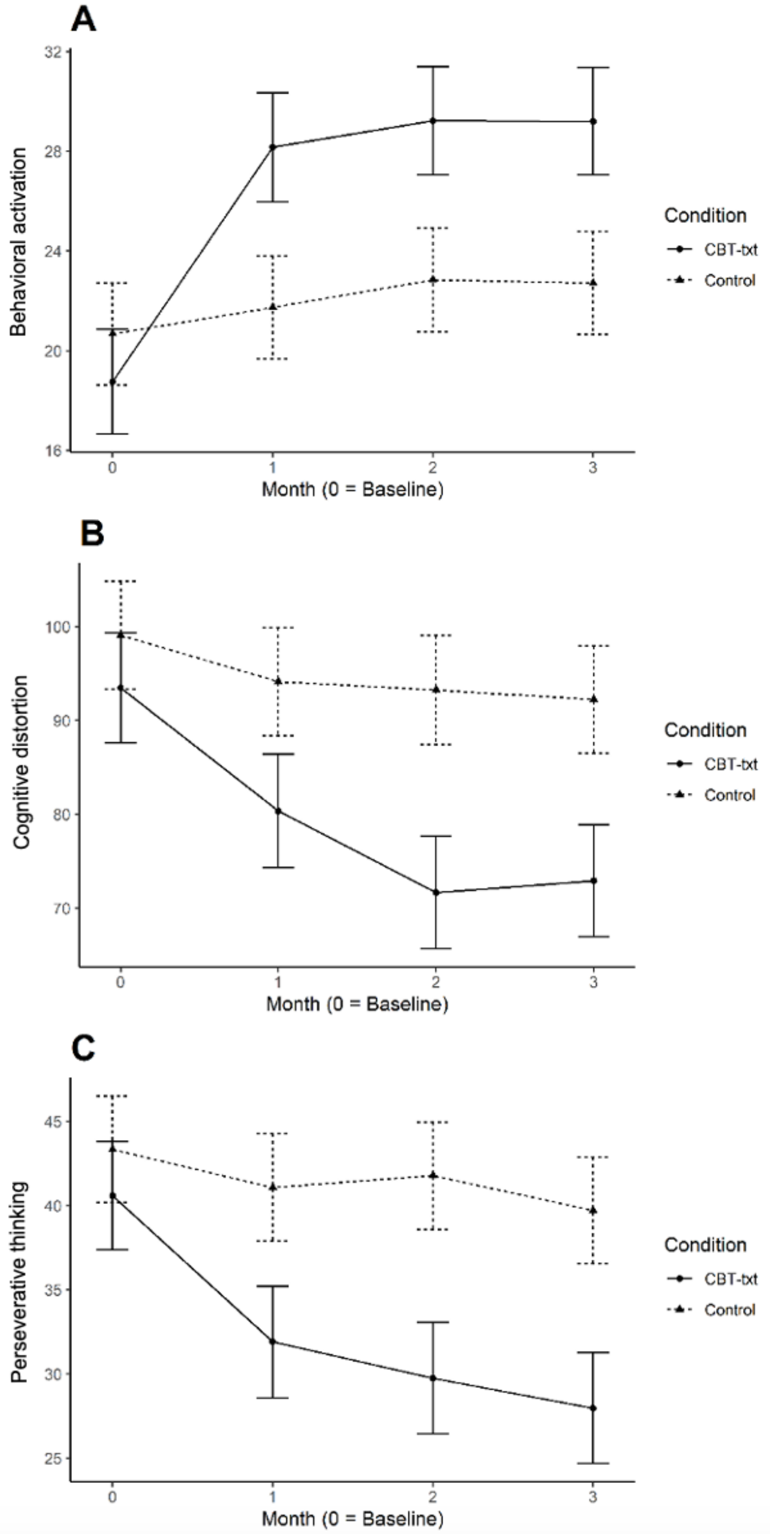
Table 3. Depression severity level (BDI-II categories) percentages at 3 months by condition.

Depression severity levels	CBT-txt[a]	Control
None to minimal	53%	15%
Mild	17%	5%
Moderate	13%	45%
Severe	17%	34%

Main Effect of CBT-txt on Mechanisms

Figure 3 shows the means of the hypothesized treatment mechanisms (behavioral activation, cognitive distortion, and perseverative thinking in panels A, B, and C, respectively) over time for the CBT-txt and control groups. None of the pretreatment means were significantly different between the CBT-txt and control groups. Significant treatment–control group differences were observed at each of the follow-up waves for each of the 3 mechanisms, with higher behavioral activation and lower cognitive distortions and perseverative thinking observed among CBT-txt versus control participants. Tables S1-S3 in Multimedia Appendix 1 show the LCS results for the 3 mechanisms. The largest and most consistently significant treatment effects were observed at the 1-month follow-up. Significantly larger 1- to 2-month decreases in cognitive distortions and perseverative thinking were seen for CBT-txt versus control participants; no significant treatment–control group differences in 1- to 2-month change in behavioral activation were seen. No significant treatment–control differences in the 2- to 3-month change were observed for any of the mechanisms.

Figure 3. Mean scores of (A) behavioral activation (Behavioral Activation for Depression Scale), (B) cognitive distortion (Cognitive Distortions Scale), and (C) perseverative thinking (Perseverative Thinking Questionnaire), measures overtime by condition. BDI: Beck Depression Inventory; CBT-txt: SMS text message–delivered cognitive behavioral therapy.



LCS Mediation Results

Table 4 shows *a*, *b*, *c'*, indirect, and total effects for mediation models testing indirect effects of CBT-txt on change in depression through change in the 3 hypothesized mechanisms (Tables S4-S6 in Multimedia Appendix 1 show full model results for LCS mediations). Significant direct and indirect effects of CBT-txt on depression were observed in all 3 models. CBT-txt appeared to lead to greater increases in behavioral activation and greater decreases in cognitive distortions and perseverative thinking across the 3-month follow-up period, which were then associated with larger baseline to 3-month decreases in depression. The size of the indirect effects was substantial: 57%, 41%, and 50% of the CBT-txt effect on changes in depression were mediated by changes in behavioral activation, cognitive distortions, and perseverative thinking, respectively.

Models testing all mediators simultaneously were also conducted and both combined and independent indirect effects were estimated (Table 5). The combined indirect effect, representing the combined action of all 3 mediators, was significant (estimate=−6.07, 95% CI −8.95 to −3.26) and explained 63% of the treatment effect on depression. Independent indirect effects were nonsignificant for behavioral activation (estimate=−2.09, 95% CI −4.92 to 0.24) and cognitive distortions (estimate=−1.14, 95% CI −3.40 to 1.20) but were significant for perseverative thinking (estimate=−2.83, 95% CI −5.80 to −0.57). Behavioral activation, cognitive distortions, and perseverative thinking independently explained 22%, 12%, and 30% of the CBT-txt effect on depression, respectively.

Table 4. Latent change score mediation model results testing SMS text message–delivered cognitive behavioral therapy (CBT-txt) efficacy for depression through mediators.

	Estimate	SE	*Z* value	*P* value	95% CI
CBT-txt > behavioral activation > depression					
Paths					
CBT-txt > BADS[a] change (*patha*)	*7.03* [b]	1.68	4.20	<.001	3.75 to 10.30
BADS change > depression change (*path b*)	*−0.75*	0.11	−6.78	<.001	−0.96 to −0.53
CBT-txt > depression change (*path c′*)	*−4.01*	1.83	−2.19	.03	−7.69 to −0.49
Indirect effect (*a*b*)	*−5.27*	—[c]	—	—	−8.69 to −2.54
Total effect (*a*b + c'*)	*−9.28*	—	—	—	−13.19 to −5.09
Percent mediated ([indirect/total] × 100%)	56.8	—	—	—	
CBT-txt > cognitive distortion > depression					
Paths					
CBT-txt > CDS[d] change (*patha*)	*−15.98*	4.06	−3.94	<.001	−23.99 to −7.92
CDS change > depression change (*path b*)	*0.25*	0.04	6.30	<.001	0.17 to 0.32
CBT-txt > depression change (*path c′*)	*−5.67*	2.11	−2.69	.007	−9.85 to −1.58
Indirect effect (*a*b*)	*−4.01*	—	—	—	−6.51 to −2.04
Total effect (*a*b + c'*)	*−9.68*	—	—	—	−13.67 to −5.35
Percent mediated ([indirect/total] × 100%)	41.4	—	—	—	
CBT-txt > perseverative thinking > depression					
Paths					
CBT-txt > PTQ[e] change (*patha*)	*−10.33*	2.40	−4.30	<.001	−14.95 to −5.63
PTQ change > depression change (*path b*)	*0.50*	0.06	8.43	<.001	0.38 to 0.61
CBT-txt > depression change (*path c′*)	*−5.18*	1.79	−2.89	.004	−8.64 to −1.57
Indirect effect (*a*b*)	*−5.12*	—	—	—	−8.05 to −2.66
Total effect (*a*b + c'*)	*−10.30*	—	—	—	−14.02 to −6.19
Percent mediated ([indirect/total] × 100%)	49.7	—	—	—	—

[a]BADS: Behavioral Activation for Depression Scale.

[b]Italicized text indicate *P*<.05.

[c]Not available.

[d]CDS: Cognitive Distortions Scale.

[e]PTQ: Perseverative Thinking Questionnaire.

Table 5. Latent change score mediation model results testing SMS text message–delivered cognitive behavioral therapy (CBT-txt) efficacy for depression through multiple mediators.

Paths	Estimate	SE	*Z* value	*P value*	95% CI
CBT-txt > BADS[a] 2-month change (*patha1*)	*7.29* [b]	1.68	4.34	<.001	3.99 to 10.63
CBT-txt > CDS[c] 2-month change (*patha2*)	*−18.72*	3.75	−4.99	<.001	−26.40 to −11.63
CBT-txt > PTQ[d] 2-month change (*patha3*)	*−11.28*	2.33	−4.85	<.001	−16.05 to −6.83
BADS 2-month change > depression 3-month change (*path b1*)	−0.29	0.16	−1.80	.07	−0.57 to 0.06
CDS 2-month change > depression 3-month change (*path b2*)	0.06	0.06	1.02	.31	−0.07 to 0.17
PTQ 2-month change > depression 3-month change (*path b3*)	*0.25*	0.11	2.34	.02	0.04 to 0.46
CBT-txt > depression 3-month change (*path c′*)	−3.52	2.12	−1.66	.10	−7.66 to 0.66
Indirect effect, BADS (*a1*b1*)	−2.09	—[e]	—	—	−4.92 to 0.24
Indirect effect, CDS (*a2*b2*)	−1.14	—	—	—	−3.40 to 1.20
Indirect effect, PTQ (*a3*b3*)	−2.83	—	—	—	−5.80 to −0.57
Indirect effect, combined (*a1*b1 + a2*b2 + a3*b3*)	−6.07	—	—	—	−8.95 to −3.26
Total effect (*a1*b1 + a2*b2 + a3*b3 + c'*)	−9.59	—	—	—	−13.43 to −5.31

[a]BADS: Behavioral Activation for Depression Scale.

[b]Italicized text indicate *P*<.05.

[c]CDS: Cognitive Distortions Scale.

[d]PTQ: Perseverative Thinking Questionnaire.

[e]Not available.

Discussion

Principal Findings

The findings from this investigation contribute to the mHealth literature by providing convincing evidence that text-delivered CBT treatment can significantly and consistently reduce depressive symptoms in young adults. These findings also specify 3 treatment mechanisms that each explain a significant portion of the treatment effect when separately introduced into the mediation analysis models and even larger effects when combined. These results support the specification of 3 candidate therapeutic mechanisms of change within CBT-txt.

All 4 of our hypotheses were supported in this trial's findings. Our primary hypothesis tested the efficacy of CBT-txt to reduce depressive symptoms, relative to the control condition. The results supported this hypothesis across all 3 months of the study, revealing a strong treatment effect. The mediation analyses also supported our hypotheses regarding the mechanisms of change within the CBT-txt treatment structure. Separate models showed that each mediator accounted for a substantial proportion of the CBT-txt treatment effect on depression. Models with all mediators included showed that, combined, these mechanisms explained 63% of the CBT-txt treatment effect, with changes in perseverative thinking showing the strongest independent indirect effect.

Although we did not experimentally test dosage effects in this trial, this study was a follow-up to our initial CBT-txt trial [9], where participants indicated wanting more texts and more detail on CBT. Thus, an important general research question was related to the treatment effect and acceptability of an 8-week versus a 4-week treatment. The direct treatment effect improved over the first trial (Cohen *d*=−0.63 to Cohen *d*=−0.76), and significant treatment effects were found at 3 months after enrollment compared with 1 month after enrollment in the initial 4-week trial. The largest changes in mechanism scores occurred during the first month of treatment. This may be an indication that these mechanisms are new or novel to participants and, as such, they may be more likely to engage with the treatment to quickly reduce their MDD symptoms. As noted, acceptability and satisfaction improved compared with the first trial. Specifically, the content was deemed helpful and the number of days of treatment and the number of texts per day appeared acceptable, supporting the increase in treatment dosage and thus aligning CBT-txt with the average length of mHealth treatment for depression.

Finally, these findings provide further support for recruiting young adults with depression via Facebook and Instagram, providing an efficient pathway toward engaging this hard-to-reach and underserved population. We recruited a sample of 103 participants in 7 weeks with excellent rates of retention, satisfaction, and engagement. These findings also support the use of CBT-txt as part of a continuum of care. For example, CBT-txt could serve as a form of pretreatment while individuals are placed on waitlists to be seen by clinicians. This approach could quickly reduce the severity of symptoms, particularly for individuals who are experiencing moderate to severe depressive symptoms. Targeting those with the most severe depression with CBT-txt may provide rapid symptom relief, which could then be followed up by a clinician. For some, CBT-txt may be enough; for others, it may serve as a jump start to their treatment and others may find it useful to combine CBT-txt while seeing a clinician for therapy. Supporting this latter idea, it is noteworthy that 278 participants were excluded from the study for reporting current or recent treatment for depression, suggesting that some young adults may recognize a need for supplemental options to traditional depression treatment. Social media recruitment may also reach young adults seeking preventive or early intervention services. Of note, 110 participants were not eligible for the study, as they reported only minimal (25/110, 22.7%) or mild (85/110, 77.3%) depression symptoms on the PHQ-9.

Limitations

The study results should be considered in light of the following limitations. First, although this study was structured as a follow-up pilot study, having a larger sample size and longer follow-up periods may strengthen confidence in the findings. Second, the sample was 84% female, limiting the generalization across biological sexes. Although the rates of major depressive episodes are double for female young adults compared with male young adults (22.9% vs 11.1%) [1], the sample imbalance raises a methodological question as to how best to engage male young adults in treatment. More research is needed that tests varying recruitment methods, branding, and advertising content, including images and language, to engage more male young adults. Third, the control condition was a nonactive comparator and a waitlist control. The study findings could be strengthened by comparing CBT-txt against an active comparator. Fourth, experimental dosage testing could be conducted in future studies to determine the most clinically useful, acceptable, and efficient dosage. A larger multiarm study to test varying dosage levels and associated outcomes could be clinically and scientifically useful. Fifth, participants were not confirmed to have an MDD diagnoses, as we used the PHQ-9 as a screening and inclusion criterion. Finally, the sample was limited to individuals without SI. Given that this is very common among patients with MDD, this exclusion criterion altered the composition of the sample. Future studies with a more robust clinical infrastructure are needed to safely enroll patients with MDD and SI. Mobile interventions are not a replacement for standard depression treatment but may be used to quickly reduce symptoms while patients are on the waiting list, as an adjunct treatment component, or as part of a stepped-care model.

The results of this trial provide further support for CBT-txt as an efficacious, efficient, and reliable form of treatment to address depressive symptomatology among young adults. To the best of our knowledge, CBT-txt is unique in its SMS text message–delivered modality and in its strong clinical evidence supporting efficacy. While acknowledging the study limitations and the potential treatment effect fluctuations due to the sample size, these results build upon our first trial’s promising findings and provide more empirical evidence for the treatment format, delivery modality, and clinical possibilities of using CBT-txt at scale.

Acknowledgments

This research was funded by the Betsey R Bush endowment for Children and Families at Risk and Behavioral Health Research, Center for Behavioral Health Research, College of Social Work, University of Tennessee.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full model results for latent change score mediations.
[DOCX File , 43 KB - mhealth_v11i1e45186_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 99 KB - mhealth_v11i1e45186_app2.pdf]

References

1. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2020 national survey on drug use and health (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2021. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2023-03-02]
2. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. Pediatrics 2016 Dec;138(6):e20161878 [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]
3. Lu W, Bessaha M, Muñoz-Laboy M. Examination of young US adults' reasons for not seeking mental health care for depression, 2011-2019. JAMA Netw Open 2022 May 02;5(5):e2211393 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.11393] [Medline: 35536582]
4. Mason MJ, Keyser-Marcus L, Snipes D, Benotsch E, Sood B. Perceived mental health treatment need and substance use correlates among young adults. Psychiatr Serv 2013 Sep 01;64(9):871-877. [doi: 10.1176/appi.ps.201200159] [Medline: 23677415]

5. Aemissegger V, Lopez-Alcalde J, Witt CM, Barth J. Comparability of patients in trials of eHealth and face-to-face psychotherapeutic interventions for depression: meta-synthesis. J Med Internet Res 2022 Sep 14;24(9):e36978 [FREE Full text] [doi: 10.2196/36978] [Medline: 36103217]
6. Lu SC, Xu M, Wang M, Hardi A, Cheng AL, Chang SH, et al. Effectiveness and minimum effective dose of app-based mobile health interventions for anxiety and depression symptom reduction: systematic review and meta-analysis. JMIR Ment Health 2022 Sep 07;9(9):e39454 [FREE Full text] [doi: 10.2196/39454] [Medline: 36069841]
7. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
8. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
9. Mason MJ, Coatsworth JD, Zaharakis N, Russell M, Wallis D, Brown A, et al. Treating young adult depression with text-delivered cognitive behavioral therapy: a pilot randomized clinical trial. Behav Ther 2023 Mar;54(2):315-329. [doi: 10.1016/j.beth.2022.09.005] [Medline: 36858762]
10. Mason MJ, Zaharakis NM, Moore M, Brown A, Garcia C, Seibers A, et al. Who responds best to text-delivered cannabis use disorder treatment? A randomized clinical trial with young adults. Psychol Addict Behav 2018 Nov;32(7):699-709. [doi: 10.1037/adb0000403] [Medline: 30265057]
11. Furukawa TA, Suganuma A, Ostinelli EG, Andersson G, Beevers CG, Shumake J, et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. Lancet Psychiatry 2021 Jun;8(6):500-511 [FREE Full text] [doi: 10.1016/S2215-0366(21)00077-8] [Medline: 33957075]
12. Health Quality Ontario. Internet-delivered cognitive behavioural therapy for major depression and anxiety disorders: a health technology assessment. Ont Health Technol Assess Ser 2019 Feb 19;19(6):1-199 [FREE Full text] [Medline: 30873251]
13. Powers MB, de Kleine RA, Smits JA. Core mechanisms of cognitive behavioral therapy for anxiety and depression: a review. Psychiatr Clin North Am 2017 Dec;40(4):611-623. [doi: 10.1016/j.psc.2017.08.010] [Medline: 29080589]
14. Quigley L, Dozois DJ, Bagby RM, Lobo DS, Ravindran L, Quilty LC. Cognitive change in cognitive-behavioural therapy pharmacotherapy for adult depression: a longitudinal mediation analysis. Psychol Med 2019 Nov;49(15):2626-2634. [doi: 10.1017/S0033291718003653] [Medline: 30560738]
15. Spinhoven P, van Hemert AM, Penninx BW. Repetitive negative thinking as a mediator in prospective cross-disorder associations between anxiety and depression disorders and their symptoms. J Behav Ther Exp Psychiatry 2019 Jun;63:6-11. [doi: 10.1016/j.jbtep.2018.11.007] [Medline: 30551055]
16. Persons JB, Thomas C. Symptom severity at week 4 of cognitive-behavior therapy predicts depression remission. Behav Ther 2019 Jul;50(4):791-802. [doi: 10.1016/j.beth.2018.12.002] [Medline: 31208688]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. National alliance on mental illness homepage. National Alliance on Mental Illness. URL: https://nami.org/Home [accessed 2023-01-12]
19. Fraud detection. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/ [accessed 2023-04-20]
20. Muñoz RF, Miranda J. Individual Therapy Manual for Cognitive-Behavioral Treatment of Depression. Santa Monica, CA: RAND Corporation; 2000.
21. Meredith LS, Jackson-Triche M, Duan N, Rubenstein LV, Camp P, Wells KB. Quality improvement for depression enhances long-term treatment knowledge for primary care clinicians. J Gen Intern Med 2000 Dec;15(12):868-877 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.91149.x] [Medline: 11119183]
22. Ngo VK, Sherbourne C, Chung B, Tang L, Wright AL, Whittington Y, et al. Community engagement compared with technical assistance to disseminate depression care among low-income, minority women: a randomized controlled effectiveness study. Am J Public Health 2016 Oct;106(10):1833-1841. [doi: 10.2105/AJPH.2016.303304] [Medline: 27552274]
23. Sherbourne CD, Wells KB, Duan N, Miranda J, Unützer J, Jaycox L, et al. Long-term effectiveness of disseminating quality improvement for depression in primary care. Arch Gen Psychiatry 2001 Jul;58(7):696-703. [doi: 10.1001/archpsyc.58.7.696] [Medline: 11448378]
24. Dino F, Zandie R, Abdollahi H, Schoeder S, Mahoor M. Delivering cognitive behavioral therapy using a conversational social robot. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2019 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); Nov 03-08, 2019; Macau, China. [doi: 10.1109/iros40897.2019.8968576]
25. Beck AT, Haigh EA. Advances in cognitive theory and therapy: the generic cognitive model. Annu Rev Clin Psychol 2014;10:1-24. [doi: 10.1146/annurev-clinpsy-032813-153734] [Medline: 24387236]
26. Beck AT. Depression: Clinical, Experimental and Theoretical Aspects. New York City, NY: Harper and Row; 1967.

27. Mason MJ, Sabo R, Zaharakis NM. Peer network counseling as brief treatment for urban adolescent heavy cannabis users. J Stud Alcohol Drugs 2017 Jan;78(1):152-157 [FREE Full text] [doi: 10.15288/jsad.2017.78.152] [Medline: 27936376]
28. Mason MJ, Zaharakis NM, Russell M, Childress V. A pilot trial of text-delivered peer network counseling to treat young adults with cannabis use disorder. J Subst Abuse Treat 2018 Jun;89:1-10 [FREE Full text] [doi: 10.1016/j.jsat.2018.03.002] [Medline: 29706169]
29. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
30. Curry JF, Wells KC. Striving for effectiveness in the treatment of adolescent depression: cognitive behavior therapy for multisite community intervention. Cogn Behav Pract 2005 Mar;12(2):177-185. [doi: 10.1016/s1077-7229(05)80023-9]
31. Curry J, Rohde P, Simons A, Silva S, Vitiello B, Kratochvil C, TADS Team. Predictors and moderators of acute outcome in the Treatment for Adolescents with Depression Study (TADS). J Am Acad Child Adolesc Psychiatry 2006 Dec;45(12):1427-1439. [doi: 10.1097/01.chi.0000240838.78984.e2] [Medline: 17135988]
32. Mason MJ, Moore M, Brown A. Young adults' perceptions of acceptability and effectiveness of a text message-delivered treatment for cannabis use disorder. J Subst Abuse Treat 2018 Oct;93:15-18. [doi: 10.1016/j.jsat.2018.07.007] [Medline: 30126536]
33. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
34. Manos RC, Kanter JW, Luo W. The behavioral activation for depression scale-short form: development and validation. Behav Ther 2011 Dec;42(4):726-739. [doi: 10.1016/j.beth.2011.04.004] [Medline: 22036000]
35. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
36. Covin R, Dozois DJ, Ogniewicz A, Seeds PM. Measuring cognitive errors: initial development of the Cognitive Distortions Scale (CDS). Int J Cogn Ther 2011 Sep;4(3):297-322. [doi: 10.1521/ijct.2011.4.3.297]
37. Coman EN, Picho K, McArdle JJ, Villagra V, Dierker L, Iordache E. The paired t-test as a simple latent change score model. Front Psychol 2013 Oct 10;4:738 [FREE Full text] [doi: 10.3389/fpsyg.2013.00738] [Medline: 24124419]
38. Ghisletta P, McArdle JJ. Teacher's corner: latent curve models and latent change score models estimated in R. Struct Equ Model 2012;19(4):651-682 [FREE Full text] [doi: 10.1080/10705511.2012.713275] [Medline: 25505366]
39. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
40. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]

Abbreviations

BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CBT-txt: SMS text message–delivered cognitive behavioral therapy
CDS: Cognitive Distortions Scale
CONSORT: Consolidated Standards of Reporting Trials
LCS: latent change score
MDD: major depressive disorder
mHealth: mobile health
PHQ-9: Patient Health Questionnaire–9
PTQ: Perseverative Thinking Questionnaire
RCI: reliable change index
SI: suicidal ideation

Edited by L Buis; submitted 17.01.23; peer-reviewed by A Aguilera, S Darrow, K Vernmark; comments to author 21.03.23; revised version received 24.04.23; accepted 06.06.23; published 11.07.23.

Please cite as:
Mason MJ, Coatsworth JD, Zaharakis N, Russell M, Brown A, McKinstry S
Testing Mechanisms of Change for Text Message–Delivered Cognitive Behavioral Therapy: Randomized Clinical Trial for Young Adult Depression
JMIR Mhealth Uhealth 2023;11:e45186
URL: https://mhealth.jmir.org/2023/1/e45186
doi:10.2196/45186
PMID:37432723

Original Paper

Virtual Digital Psychotherapist App–Based Treatment in Patients With Methamphetamine Use Disorder (Echo-APP): Single-Arm Pilot Feasibility and Efficacy Study

Tianzhen Chen[1*], MD, PhD; Liyu Chen[1*], MPsy; Shuo Li[1], MPsy; Jiang Du[1], MD, PhD; Hang Su[1], MD, PhD; Haifeng Jiang[1], MD, PhD; Qianying Wu[1], MPsy; Lei Zhang[1], MPsy; Jiayi Bao[1], MPsy; Min Zhao[1,2,3], MD, PhD

[1]Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Shanghai Key Laboratory of Psychotic Disorders, Shanghai, China

[3]Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Sciences, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Min Zhao, MD, PhD
Shanghai Mental Health Center
Shanghai Jiao Tong University School of Medicine
600 South Wan Ping Road, Shanghai
Shanghai, 200030
China
Phone: 86 2134773146
Email: drminzhao@smhc.org.cn

Abstract

Background: Substance use disorder is one of the severe public health problems worldwide. Inequitable resources, discrimination, and physical distances limit patients' access to medical help. Automated conversational agents have the potential to provide in-home and remote therapy. However, automatic dialogue agents mostly use text and other methods to interact, which affects the interaction experience, treatment immersion, and clinical efficacy.

Objective: The aim of this paper is to describe the design and development of Echo-APP, a tablet-based app with the function of a virtual digital psychotherapist, and to conduct a pilot study to explore the feasibility and preliminary efficacy results of Echo-APP for patients with methamphetamine use disorder.

Methods: Echo-APP is an assessment and rehabilitation program developed for substance use disorder (SUD) by a team of clinicians, psychotherapists, and computer experts. The program is available for Android tablets. In terms of assessment, the focus is on the core characteristics of SUD, such as mood, impulsivity, treatment motivation, and craving level. In terms of treatment, Echo-APP provides 10 treatment units, involving awareness of addiction, motivation enhancement, emotion regulation, meditation, etc. A total of 47 patients with methamphetamine dependence were eventually enrolled in the pilot study to receive a single session of the Echo-APP–based motivational enhancement treatment. The outcomes were assessed before and after the patients' treatment, including treatment motivation, craving levels, self-perception on the importance of drug abstinence, and their confidence in stopping the drug use.

Results: In the pilot study, scores on the Stages of Change Readiness and Treatment Eagerness Scale and the questionnaire on motivation for abstaining from drugs significantly increased after the Echo-APP–based treatment (P<.001, Cohen d=–0.60), while craving was reduced (P=.01, Cohen d=0.38). Patients' baseline Generalized Anxiety Disorder-7 assessment score (β=3.57; P<.001; 95% CI 0.80, 2.89) and Barratt Impulsiveness Scale (BIS)–motor impulsiveness score (β=–2.10; P=.04; 95% CI –0.94, –0.02) were predictive of changes in the patients' treatment motivation during treatment. Moreover, patients' baseline Generalized Anxiety Disorder-7 assessment score (β=–1.607; P=.03; 95% CI –3.08, –0.14), BIS—attentional impulsivity score (β=–2.43; P=.004; 95% CI –4.03, –0.83), and BIS—nonplanning impulsivity score (β=2.54; P=.002; 95% CI 0.98, 4.10) were predictive of changes in craving scores during treatment.

Conclusions: Echo-APP is a practical, accepted, and promising virtual digital psychotherapist program for patients with methamphetamine dependence. The preliminary findings lay a good foundation for further optimization of the program and the promotion of large-scale randomized controlled clinical studies for SUD.

(JMIR Mhealth Uhealth 2023;11:e40373) doi:10.2196/40373

KEYWORDS

tablet; Android program; substance use disorder; methamphetamine use disorder; digital agent; virtual digital human

Introduction

Substance abuse and dependence seriously endanger public health worldwide. Globally, 3.5 million people die from alcohol and illicit drug abuse each year [1]. China is also facing huge challenges brought on by illegal drugs, tobacco, and alcohol. According to the latest report, there are more than 300 million tobacco users, 123 million people drink alcohol excessively, and 1.8 million people use illegal drugs in China [2-4]. Substance dependence brings a series of physical and psychological damage, resulting in a large economic burden of disease. Psychotherapy is one of the most important treatment methods for drug dependence currently, including cognitive behavioral therapy (CBT), meditation, motivational enhancement therapy (MET) [5,6]. It can help improve treatment motivation, ameliorate emotional disorders, and decrease cravings that patients face. However, many people with substance dependence do not receive proper treatment, and less than 20% of patients have received standard treatment [7]. The main reasons include shortage of professionals, fear of discrimination, and economic and transportation constraints.

Artificial intelligence (AI) and robotics can help solve the substance dependence treatment dilemma. These technologies have been gradually applied to various mental health scenarios such as emotion regulation, evaluation and treatment of mental diseases, treatment efficacy prediction, and rehabilitation management. By simulating the "intelligent brain," an intelligent machine or program that responds in a manner similar to human intelligence can simulate psychotherapists in the field of psychological assessment, diagnosis, and treatment, which can save labor costs, realize remote intervention, and improve professional use. In the past period, automatic conversational agents have received attention [8]. Automated conversational agents can provide services similar to a therapist or physician, but without the need for human assistance. Studies have shown that text-based dialogue agents have good user engagement and are effective in the treatment of mental symptoms [9-11]. By playing the role of "psychotherapists," AI devices reduce the patient's sense of shame and fear of being discriminated against and increase the possibility of patients revealing their true feelings [12]. On the other hand, the widespread use of the internet and intelligent terminals makes electronic medical care based on apps and AI technology have good application prospects. In 2021, the number of smartphone-owning users in China reached 950 million [13]. The development of a mental health program suitable for mobile smart terminals will further enhance the interest of users.

AI has been applied in related fields such as insomnia, anxiety, depression, schizophrenia, substance dependence, and other diseases [14-18]. Among them, some programs are self-administered, that is, after self-assessment, the program provides mental health–related education [19]. Some programs are conversational agents. Research evidence for conversational agent interventions for addressing psychological problems is growing rapidly and has the potential in terms of acceptability and effectiveness [12]. In substance dependence studies, digital interventions have also been found to reduce substance use behaviors [20] and have the potential to reduce the economic burden of disease from substance use disorder (SUD). These apps mainly use text, animation, or dialogue to provide services to patients [8]. Although these programs are reported to be effective, there are still challenges such as poor interaction experience and high dropout rate due to the lack of a more realistic image. Recent studies suggested a potential improvement in treatment effects when incorporated with the virtual image. For instance, Yokotani and colleagues [21] found that virtual agent (with digital image) has advantages in participants' disclosure of sex-related symptoms. The study by Philip [22] suggested that the smartphone-based virtual agent is feasible in screening patients with sleep complaints and provides acceptable behavior advice [22]. Virtual agents that show better look and feel achieve better user experience [23]. However, most of the virtual agent psychotherapists (with digital image) currently constructed are not vivid enough, with rigid expressions and movements. Besides, few studies have used computer techniques to construct a virtual digital psychotherapist image to provide psychological support services in the field of substance use disorder.

Based on these considerations, we propose, design, and develop a mobile- or tablet-based app (Echo-APP) for the assessment and treatment of substance dependence. Echo-APP, which features a virtual digital human image of a psychotherapist, is a key-based interactive conversational agent program that provides general mental health education, addiction-related symptom tracking and recording, and customized comprehensive psychotherapy. In this study, we describe the development and design of Echo-APP and aim to evaluate the feasibility and preliminary efficacy of Echo-APP through a single-arm pre-post–treatment design.

Methods

Development of a Virtual Digital Psychotherapist App (Echo-APP)

The design and development of Echo-APP started in 2019. The app development was proposed by the Addiction Research Group of the Shanghai Mental Health Center and was optimized through discussions with clinicians, patients with drug dependence, technician, and computer scientists from Mofa Technology Corporation. The multidisciplinary team of this project team meets regularly to form collaborations based on the necessary development techniques and patient-centered design-centric tenet. The core development period is from June 2019 to August 2020, and the first version of the app is finalized. The Echo-APP currently developed is in Chinese.

Echo-APP has been developed for the HUAWEI Android tablet above the application programming interface level 14 (version 4.0). Java (Oracle Corporation) and Python (Python Software Foundation) were used as the programming language, and Visual Studio (Microsoft Corporation) was used as the main development tool. In addition, the image of the virtual digital psychotherapist is developed based on digital human technology. Operationally, the main menu allows psychotherapists or social workers to add new subjects and refer to patients' assessment and treatment records. A unique code is added to each patient. Three modules of "Psychological Assessment," "Treatment Options," and "Homework" are accessed via the buttons on the menu, "Start Assessment," "Start Therapy," and "Homework." The process of assessment and treatment is completed by the virtual digital psychotherapist interacting with the patient. The image of the virtual digital psychotherapist (named "Xiaoying") is shown in Figure 1.

Figure 1. Image of the virtual digital psychotherapist in Echo-APP.

Assessment Module

The current version of Echo-APP could investigate patients' baseline demographic information, drug use history, emotional state, impulsivity trait, and treatment motivation. The interface of the assessment module is shown in Multimedia Appendix 1.

Demographic information is collected by self-designed scales, including the patient's age, gender, education level, employment status, and drug use characteristics.

The patient's emotional status is assessed by Generalized Anxiety Disorder-7 (GAD-7) and Patient Health Questionnaire-9 (PHQ-9).

Impulsivity characteristics are assessed by the Barratt Impulsiveness Scale (BIS)–11. BIS-11 has 30 items and can be divided into 3 dimensions (attention impulsivity, motor impulsivity, and nonplanning impulsivity).

Treatment motivation is assessed by the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES), which is self-assessed. This scale evaluates the treatment motivation from the following 3 aspects: patients' awareness of drug use, attitudes toward behavior change, and ambivalent attitudes toward drug dependence. There are 19 items, and the items on the scale are rated on a 5-point Likert scale (each item ranging from "strongly disagree" to "strongly agree"). The questionnaire on motivation for abstaining from drugs is also used to evaluate treatment motivation. The scale has a total of 36 items, which are divided into the following 5 subdimensions: "tending to rehabilitation-internal motivation," "tending to rehabilitation-external motivation," "avoiding abuse-internal motivation," "avoiding abuse-external motivation," and "confidence in abstaining from drugs." The total score of treatment motivation is summed up by the scores of each dimension. The higher the scores, the stronger the treatment motivation.

Visual analogue scale (VAS) is used to assess the self-perception of the importance of drug abstinence (simply called "IMPORTANCE"), their confidence in stopping drug use (simply called "CONFIDENCE"), and their psychological craving for drug use (simply called "CRAVING"). The VAS scale ranges from 0 mm to 100 mm.

For the items of "IMPORTANCE" and "CONFIDENCE," "0" means "doesn't matter at all," and "100" means very important.

For “CRAVING,” “0” corresponds to “no craving,” and “100” represents “highest craving intensity ever experienced for drug.”

The validity and reliability of the Chinese version of the following scale have been confirmed previously: GAD-7, PHQ-9, SOCRATES, BIS-11, and questionnaire of motivation for abstaining from drugs [24-28].

During the assessment, the virtual digital psychotherapist reads the questions, and the patients select options on the tablet screen.

Treatment Module

The treatment module of Echo-APP can provide patients with a variety of treatment options, aiming to help patients enhance their motivation to stop drug use, reduce drug cravings, strengthen self-control to avoid relapse, strengthen emotional management skills, and enhance individual and social functions (Multimedia Appendix 2).

The whole module includes the following 10 treatment units: (1) strengthening the motivation of drug withdrawal, (2) recognition of drug cravings and incentives, (3) high-risk situations identification and coping skill, (4) dealing with negative cognition, (5) understanding of emotions, (6) stress management, (7) understanding of family conflicts, (8) preventing relapse, (9) mindfulness, and (10) awareness of positive attitude and well-being (Multimedia Appendix 3). Based on the patient's assessment results, the treatment unit that meets the needs of the patient’s condition will be selected to provide to the patient.

Each treatment unit follows a structured setting for CBT. At the beginning of each treatment unit, there will be a brief introduction to the treatment goals and themes of the unit. After that, it will enter the formal treatment session, in which various common tools of CBT [29,30] (eg, thinking record sheet, analysis sheet of drug use pros and cons, risk level evaluation sheet for external factor, alternative behavior selection sheet), mindfulness-based relapse prevention technique [31] (integrated into treatment unit 6, 8, and 9), motivational enhancement therapy [32] (integrated into treatment unit 1), and related knowledge popularization (eg, the damage of different drugs and the antecedent, behavior, and consequence theory of emotion) will be applied.

Homework Module

The homework module corresponds to the treatment module. There are 10 themed homework modules. After each treatment unit is completed, patients will be assigned their homework and asked to complete it offline (Multimedia Appendix 4).

Feedback

After completing each assessment, Echo-APP will provide patients with assessment reports and treatment recommendations. For each treatment unit, Echo-APP will first ask the patient about their treatment experience and their change (eg, treatment motivation and confidence to deal with negative emotions), and then provide the summary report of this treatment unit. The report form is shown in Multimedia Appendix 5.

Instruction Booklet

Although the clinical assessment and treatment implemented in Echo-APP can be useful for patients with substance dependence, good guidance is still required for initial use. To this end, in order to make the use of Echo-APP more practical, safer, and more extensive, we provide the Echo-APP instruction manual (Chinese version) and will provide an English version in the future. With a user-friendly and tailored interface, Echo-APP is easy to learn, and this manual will be an effective tool for patient self-learning. This manual shows the interfaces you may encounter in Echo-APP and what you need to do to complete the interface for each assessment or treatment.

Study Design

This study consists of 3 parts, the first of which describes the development and technical details of Echo-APP. Then, to assess the efficacy and feasibility of the Echo-APP–based assessment and treatment, we conducted a preliminary study in patients with methamphetamine dependence. Specifically, we conducted a single-arm self-control pilot study with the treatment unit “strengthening the motivation of drug withdrawal.”

Ethical Considerations

This study was carried out in accordance with the principles of the Declaration of Helsinki and approved by the institutional review board and the ethics committee of the Shanghai Mental Health Center (approval number: 2020-92). The participants provided their informed consent before the study. The study flow diagram is shown in Multimedia Appendix 6.

Participants

A total of 49 patients were recruited from the Shanghai Drug Rehabilitation Center. Eligible patients were diagnosed with methamphetamine use disorder by psychiatrists based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria. The inclusion criteria were as follows: (1) met the DSM-5 criteria for methamphetamine use disorder, (2) aged 18-55 years, and (3) normal or corrected-to-normal vision and audition. The exclusion criteria were as follows: (1) with severe cognitive deficits or impairments; (2) with serious physical or neurological illness or a diagnosis of any other psychiatric disorder under DSM-5 criteria (except for nicotine use disorder); and (3) inability to understand and operate the app instructions. Two patients stopped from participating in the study after the screening. Therefore, 47 patients were finally included in the analysis.

Treatment Settings

All participants received one session of treatment unit, “strengthening the motivation of drug withdrawal” (Echo-APP–based MET), which was provided by a virtual digital therapist. The contents of the Echo-APP–based MET include an introduction to the damage of drugs, analysis and comparison of the pros and cons of drug abuse, sharing of stories and experiences of people who have successfully abstained from drugs, and improvement of motivation for change. The session duration is about 30 to 45 minutes. During the treatment, there will be a real psychotherapist familiar with the operation of Echo-APP to provide the necessary operation guidance for the patients. Other than that, the real psychotherapist does not provide any therapy during treatment.

Outcome Measures

The primary treatment outcome was the change in SOCRATES score. The secondary outcomes include the following: (1) score change on the questionnaire of motivation for abstaining from drugs; (2) score change of the VAS items "IMPORTANCE," "CONFIDENCE," and "CRAVING."

Adverse Effect

During assessment and treatment, no participant reported significant discomfort or adverse effects.

Statistical Analysis

To identify the treatment efficacy of Echo-APP, paired 2-tailed *t* test was used to analyze the changes in scale scores before and after treatment. For those clinical outcomes that have changed significantly before and after treatment (ie, SOCRATES, questionnaire of motivation for abstaining from drugs, IMPORTANCE, CONFIDENCE, and CRAVING), the general linear regression model was used to analyze potential factors affecting treatment efficacy. For each of the general linear regression models, the change in the scale scores (ie, SOCRATES, questionnaire of motivation for abstaining from drugs, IMPORTANCE, CONFIDENCE, and CRAVING) was used as the dependent variable, and baseline demographic information (ie, age, education, current marital status), drug use history, emotional state (ie, GAD-7 and PHQ-9), and impulsivity characteristics (ie, BIS-11) were included as independent variables in the model. The backward method was used to screen the independent variables. All the above statistical analyses were finished using SPSS 20.0 (IBM Corp).

Results

Baseline Information

Baseline demographic characteristics and clinical data are presented in Table 1. The average age of the patients is 38.85 (SD 8.08) years. The patients' accumulated months of methamphetamine use is 99.89 (SD 56.71) months. Of the 47 participants, 12 (26%) use methamphetamine due to psychological craving.

Table 1. Demographic information and drug use history for all participants (N=47) in a study of Echo-APP–based Motivational Enhancement Therapy for methamphetamine use disorder.

Characteristics	Values
Demographics	
Age (years), mean (SD)	38.85 (8.08)
Education, n (%)	
Less than 7 years	1 (2)
7-9 years	18 (38)
10-12 years	17 (36)
More than 12 years	11 (23)
Currently married, n (%)	16 (34)
Drug use history	
Accumulated months of methamphetamine use, mean (SD)	99.89 (56.71)
Methamphetamine use dosage (grams) per day, mean (SD)	0.71 (0.42)
Methamphetamine use frequency, n (%)	
Use every day	20 (43)
3-5 days per week	7 (15)
1 day per week	9 (19)
1-3 days per month	11 (23)
Methamphetamine use reason, n (%)	
Craving	12 (26)
Others	35 (74)
Baseline clinical measures, mean (SD; range)	
Craving	18.09 (26.41; 0-100)
Awareness of the importance of drug abstinence	77.13 (32.47; 0-100)
Confidence in drug abstinence	83.09 (23.26; 0-100)
SOCRATES[a]	65.13 (15.23; 26-91)
Questionnaire of motivation for abstaining from drugs	160.79 (19.48; 109-180)
PHQ-9[b]	5.53 (4.84; 0-19)
GAD-7[c]	3.79 (4.26; 0-16)
BIS-MI[d]	73.13 (17.73; 34-100)

[a]SOCRATES: Stages of Change Readiness and Treatment Eagerness Scale.

[b]PHQ-9: Patient Health Questionnare-9.

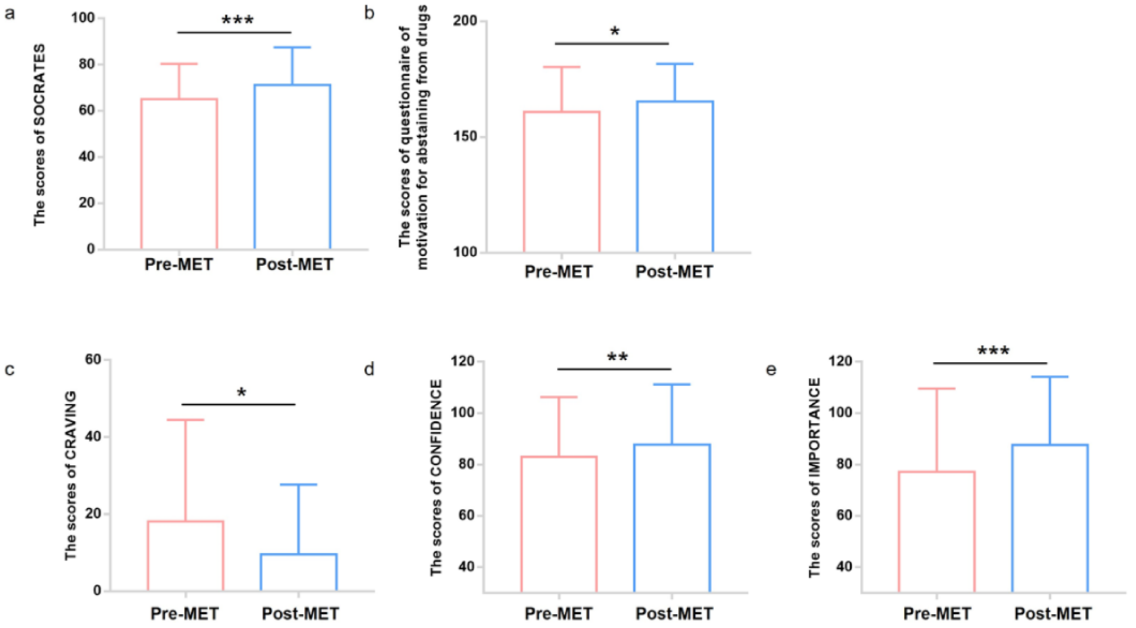
[c]GAD-7: Generalized Anxiety Disorder-7.

[d]BIS-MI: Barratt Impulsiveness Scale—motor impulsivity.

The Efficacy of Echo-App–Based Motivational Enhancement Therapy

The Echo-app–based MET treatment has brought significant improvement in patients' treatment motivation as assessed on the SOCRATES scales (*P*<.001, Cohen *d*=–0.60) and the questionnaire of motivation for abstaining from drugs (*P*=.045, Cohen *d*=–0.30). Besides, it had a significant effect on the reduction of self-reported craving (*P*=.01, Cohen *d*=0.38; Figure 2 and Multimedia Appendix 7).

Figure 2. The changes of the clinical measures and effect sizes (Cohen d) during the Echo-APP–based motivation enhancement treatment (MET). (a) The changes of the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) scores. (b) The changes of the questionnaire of motivation for abstaining from drugs. (c) The changes of visual analogue scale score of “CRAVING.” (d) The changes of visual analogue scale score of “CONFIDENCE.” (e) The changes of visual analogue scale score of “IMPORTANCE.” *<.05, **<.01, ***<.001.



Pre – post Cohen *d*

SOCRATES	MOTIVATION	CRAVING	CONFIDENCE	IMPORTANCE
0.60	–0.30	0.38	–0.45	–0.50

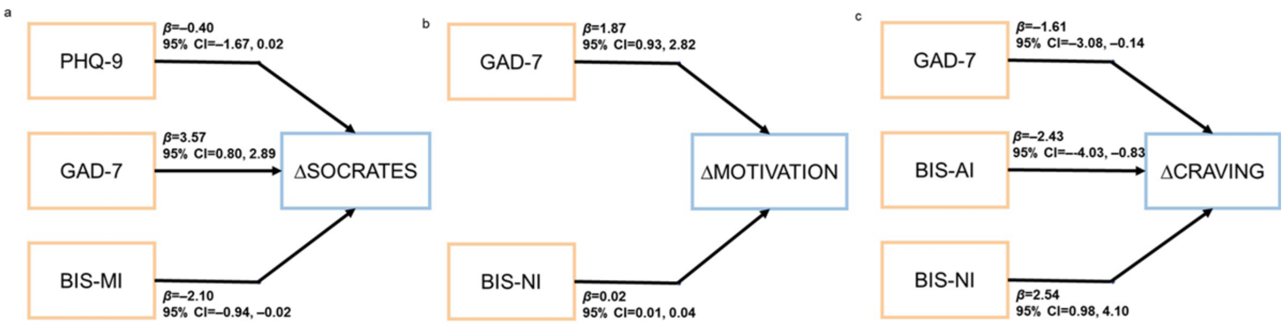
Factors Affecting Echo-APP–Based Treatment Efficacy

To explore if the patients’ baseline characteristics may influence the treatment efficacy of Echo-APP–based MET, we investigated their associations. For the treatment motivation, the baseline GAD-7 score (β=3.57; *P*<.001; 95% CI 0.80, 2.89) and BIS—motor impulsivity scores (β=–2.10; *P*=.04; 95% CI –0.94, –0.02) were predictive of SOCRATES change after MET, and the baseline GAD-7 score (β=1.87; *P*<.001; 95% CI 0.93, 2.82) and abstinence duration (β=0.02; *P*=.01; 95% CI 0.01, 0.04) were predictive of questionnaire of motivation for abstaining from drugs change after MET (Figure 3).

For psychological craving, the baseline GAD-7 score (β=–1.607; *P*=.03; 95% CI –3.08, –0.14), BIS—attentional impulsivity scores (β=–2.43; *P*=.004; 95% CI –4.03, –0.83), and BIS—nonplanning impulsivity (β=2.54; *P*=.002; 95% CI 0.98, 4.10) were predictive of self-reported craving reductions after MET.

None of the baseline characteristics were relative to the changes in the visual analogue scale score of “CONFIDENCE” and “IMPORTANCE.”

Figure 3. The regression analyses predicting change of clinical measures in patients with methamphetamine dependence. (a) Regression analyses predicting an increase in Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) score. (b) Regression analyses predicting an increase in score of questionnaire of motivation for abstaining from drugs. (c) Regression analyses predicting a decrease in craving score. BIS-AI: Barratt Impulsiveness Scale—attention impulsivity; BIS-MI: Barratt Impulsiveness Scale—motor impulsivity; BIS-NI: Barratt Impulsiveness Scale—nonplanning impulsivity; GAD-7: Generalized Anxiety Disorder; PHQ-9: Patient Health Questionnaire-9.



Discussion

Overview

The purpose of this study was to describe the design of Echo-APP, which is the virtual digital psychotherapist app, and conduct a preliminary evaluation of the efficacy of the Echo-APP–based intervention. The study found that 1 session of Echo-APP treatment can enhance patients' treatment motivation and reduce psychological cravings.

The Design of Echo-APP

The core work of this study was to build standardized self-assessment and high-quality psychotherapy tools that are in line with the needs of patients with SUD. Therefore, medical staff and patients can use this professional digital tool for SUD treatment more conveniently and with more interest. Based on this consideration, we refer to the opinions of patients with SUD, discuss existing digital health programs with doctors, psychotherapists, and computer experts, and design new digital health program for patients with SUD. Our group mainly focuses on several points, including the following: (1) accessibility and convenience of the digital program, (2) standardization and effectiveness of assessment and treatment, and (3) optimizing the user-program interaction experience and enhancing immersion. Previous studies have found that both web-based and app-based text conversational agents are effective for patients [12], and app-based programs have advantages in improving accessibility [33]. However, dialogue agent tools that use text or animation to interact have their disadvantages, such as poor interaction and immersion. Besides, most digital programs also offer fewer treatment options, and patients are passively accepted [34]. In this context, we developed the first virtual digital psychotherapist program, Echo-APP, for patients with SUD, which is a more economical and interactive app that runs on the tablet platform.

This study describes the design and development process of Echo-APP and evaluates the utility of Echo-APP in a single-arm design study. From a technical point of view, Echo-APP provides a more friendly operation interface. Patients are able to complete the assessment and treatment process through interaction with a virtual digital psychotherapist. According to the needs of patients with SUD, Echo-APP has 10 units, which can provide patients with comprehensive treatment. Throughout the assessment and treatment process, Echo-APP provides readable feedback reports for patients, doctors, and psychotherapists to clearly know the patient's condition changes and treatment efficacy.

Treatment Efficacy of Echo-APP

This study suggested that a single session of Echo-APP treatment can enhance patients' treatment motivation and reduce psychological cravings. The design of this Echo-APP–based MET program is referred to the MET protocol of our group, which was found to significantly improve the treatment outcome of patients who are dependent on heroin [35]. Echo-APP also had an effect on reducing psychological cravings, which may be related to the improvement of treatment motivation [36]. To explore the factors that may affect the treatment effect, we analyzed the patient's baseline characteristics, and the results suggested that the patient's emotional state and impulsive personality characteristics may affect the effectiveness of the Echo-APP-based MET treatment. Therefore, the multi-unit comprehensive intervention system provided by Echo-APP may have better applicability. It is possible to improve the overall treatment effect through targeted intervention for different symptoms, which has been verified in other studies [37,38]. Although the results are promising, this work is a preliminary study, and further standardized evaluation of Echo-APP is still needed, and a multicenter randomized controlled clinical study is required to explore the efficacy and adverse reactions of Echo-APP.

Limitations

This study inevitably has some limitations. Firstly, the current assessment of Echo-APP is mainly based on scales, which may have similar shortcomings as paper scales. We plan to develop an effective addiction symptom assessment paradigm to reduce the subjective assessment content. Secondly, this study is a single-arm design, which can only provide a preliminary result of the therapeutic value of Echo-APP, whereas the most rigorous way to ensure the effectiveness of the developed APP would be to conduct a randomized controlled clinical study in comparison with a traditional psychotherapy. This randomized study is in progress and in its early stages. Thirdly, this study is only conducted on patients with methamphetamine use disorder, and the popularity of the study results needs to be improved, including further research on people who use legal drugs such as tobacco and alcohol.

Conclusion

In this study, we introduced the design and development of Echo-APP and preliminarily validated the efficacy of the Echo-APP–based treatment for patients with methamphetamine use disorder. This work fills the current lack of addiction treatment programs with the virtual digital psychotherapist on the market. In the future, we will continue to optimize the function of Echo-APP and carry out large-sample multicenter clinical controlled studies to verify the effect of Echo-APP and benefit more patients with SUD.

Acknowledgments

The authors would like to thank the technical support of computer scientists from Mofa Technology. This work was supported by National Nature Science Foundation (81771436, 82130041, 82201650, 82171483); Shanghai Municipal Science and Technology Major Project (2018SHZDZX05); Shanghai Shenkang Hospital Development Center (SHDC2020CR3045B); Shanghai Rising-star Cultivation Program (22YF1439200); Capability Promotion Project for Research-oriented Doctor at SMHC (2021-YJXYS-01);

Brain Science and Brain-Like Intelligence Technology (2021ZD0202105); Shanghai "the Hospital Garden Star" Funding for training of young medical professionals (20224Z0017); Shanghai Clinical Research Center for Mental Health (19MC1911100); Shanghai Engineering Research Center of Intelligent Addiction Treatment and Rehabilitation (19DZ2255200); Shanghai Key Laboratory of Psychotic Disorders (13DZ2260500); and Lingang Lab (Grant LG202106-03-01).

Authors' Contributions

LYC, JD, HS, QYW, LZ, and JYB conducted the research; TZC and LYC analyzed data; HFJ and JD provided the clinical support; MZ, TZC, and LYC designed the research. TZC drafted the manuscript. All authors have contributed to the interpretation of data, critically revised the manuscript, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interface of the assessment module.
[PNG File , 140 KB - mhealth_v11i1e40373_app1.png]

Multimedia Appendix 2
The interface of the treatment module.
[PNG File , 800 KB - mhealth_v11i1e40373_app2.png]

Multimedia Appendix 3
Ten treatment units in the treatment module.
[DOCX File , 16 KB - mhealth_v11i1e40373_app3.docx]

Multimedia Appendix 4
The interface of the homework module.
[PNG File , 431 KB - mhealth_v11i1e40373_app4.png]

Multimedia Appendix 5
The feedback report form provided by Echo-APP.
[PNG File , 673 KB - mhealth_v11i1e40373_app5.png]

Multimedia Appendix 6
Flow diagram of the study.
[PNG File , 19 KB - mhealth_v11i1e40373_app6.png]

Multimedia Appendix 7
The changes of the clinical measures during the Echo-app–based motivation enhancement treatment.
[DOCX File , 15 KB - mhealth_v11i1e40373_app7.docx]

References

1. Global status report on alcohol and health. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639-eng.pdf?sequence=1&isAllowed=y [accessed 2022-09-01]
2. 2020 Blue Book on Moderate Drinking and Happy Life in China. China Alcoholic Drinks Association. 2020. URL: https://baijiahao.baidu.com/s?id=1683938035522034446 [accessed 2022-08-18]
3. 2020 China Drug Use Situation Report. National Narcotics Control Commission. 2021. URL: https://www.mps.gov.cn/n2255079/n6865805/n7355741/n7355780/c8044518/content.html [accessed 2021-12-05]
4. China Report on the Health Hazards of Smoking. National Narcotics Control Commission. 2020. URL: http://www.gov.cn/xinwen/2021-05/30/content_5613994.htm [accessed 2021-12-05]
5. McGovern R, Newham J, Addison M, Hickman M, Kaner E. The effectiveness of psychosocial interventions at reducing the frequency of alcohol and drug use in parents: findings of a Cochrane Review and meta-analyses. Addiction 2022 Oct;117(10):2571-2582 [FREE Full text] [doi: 10.1111/add.15846] [Medline: 35188313]
6. DiClemente CC, Corno CM, Graydon MM, Wiprovnick AE, Knoblach DJ. Motivational interviewing, enhancement, and brief interventions over the last decade: A review of reviews of efficacy and effectiveness. Psychol Addict Behav 2017 Dec;31(8):862-887. [doi: 10.1037/adb0000318] [Medline: 29199843]

7. Key substance use and mental health indicators in the United States: results from the 2018 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2019. URL: https://www.icpsr.umich.edu/web/NAHDAP/studies/32722/versions/V6 [accessed 2022-06-01]
8. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
9. Fitzsimmons-Craft EE, Chan WW, Smith AC, Firebaugh M, Fowler LA, Topooco N, et al. Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial. Int J Eat Disord 2022 Mar;55(3):343-353. [doi: 10.1002/eat.23662] [Medline: 35274362]
10. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
11. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
12. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec;18(sup1):37-49. [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
13. O'Dea S. Number of smartphone users by leading countries in 2021 (in millions). Statista. 2022. URL: https://www.statista.com/statistics/748053/worldwide-top-countries-smartphone-users [accessed 2022-06-01]
14. Su S, Wang Y, Jiang W, Zhao W, Gao R, Wu Y, et al. Efficacy of Artificial Intelligence-Assisted Psychotherapy in Patients With Anxiety Disorders: A Prospective, National Multicenter Randomized Controlled Trial Protocol. Front Psychiatry 2021;12:799917 [FREE Full text] [doi: 10.3389/fpsyt.2021.799917] [Medline: 35126211]
15. Cortes-Briones JA, Tapia-Rivas NI, D'Souza DC, Estevez PA. Going deep into schizophrenia with artificial intelligence. Schizophr Res 2022 Jul;245:122-140 [FREE Full text] [doi: 10.1016/j.schres.2021.05.018] [Medline: 34103242]
16. Andrillon T, Solelhac G, Bouchequet P, Romano F, Le Brun M, Brigham M, et al. Revisiting the value of polysomnographic data in insomnia: more than meets the eye. Sleep Med 2020 Feb;66:184-200. [doi: 10.1016/j.sleep.2019.12.002] [Medline: 31978862]
17. Zhang F, Wu C, Jia C, Gao K, Wang J, Zhao H, et al. Artificial intelligence based discovery of the association between depression and chronic fatigue syndrome. J Affect Disord 2019 May 01;250:380-390. [doi: 10.1016/j.jad.2019.03.011] [Medline: 30877861]
18. Ahn W, Vassileva J. Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. Drug Alcohol Depend 2016 Apr 01;161:247-257. [doi: 10.1016/j.drugalcdep.2016.02.008] [Medline: 26905209]
19. Xu X, Chen S, Chen J, Chen Z, Fu L, Song D, et al. Feasibility and Preliminary Efficacy of a Community-Based Addiction Rehabilitation Electronic System in Substance Use Disorder: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 Apr 16;9(4):e21087 [FREE Full text] [doi: 10.2196/21087] [Medline: 33861211]
20. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
21. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. Computers in Human Behavior 2018 Aug;85:135-145. [doi: 10.1016/j.chb.2018.03.045]
22. Philip P, Dupuy L, Morin CM, de Sevin E, Bioulac S, Taillard J, et al. Smartphone-Based Virtual Agents to Help Individuals With Sleep Concerns During COVID-19 Confinement: Feasibility Study. J Med Internet Res 2020 Dec 18;22(12):e24268 [FREE Full text] [doi: 10.2196/24268] [Medline: 33264099]
23. Curtis RG, Bartel B, Ferguson T, Blake HT, Northcott C, Virgara R, et al. Improving User Experience of Virtual Health Assistants: Scoping Review. J Med Internet Res 2021 Dec 21;23(12):e31737 [FREE Full text] [doi: 10.2196/31737] [Medline: 34931997]
24. He X, Li C, Qian J, Cui H, Wu W. Reliability and validity of a generalized anxiety disorder scale in general hospital outpatients. Shanghai Archives of Psychiatry 2010;22(4):200-203. [doi: 10.3969/j.issn.1002-0829.2010.04.002]
25. Li XY, Phillips MR, Dong X, Zhang YL, Yang SJ, Tong YS, et al. 2.5-D DC resistivity modeling considering flexibility and accuracy. Chinese Mental Health Journal 2011;25(8):610-615. [doi: 10.3969/j.issn.1000-6729.2011.08.013]
26. Wu XY, Zeng H, Li L, Chen Y. Designing questionnaire on motivation for abstaining from drugs. Chinese Journal of Drug Dependence 2008;17(6):465-468 [FREE Full text]
27. Wang W, Bian Q, Zhao Y, Li X, Wang W, Du J, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2014;36(5):539-544 [FREE Full text] [doi: 10.1016/j.genhosppsych.2014.05.021] [Medline: 25023953]
28. Yeh M. A psychometric evaluation of the Chinese version of the stage of change-readiness and treatment-eagerness scale. J Clin Nurs 2009 Sep;18(17):2494-2502. [doi: 10.1111/j.1365-2702.2008.02592.x] [Medline: 19207792]

29. Zhong N, Yuan Y, Chen H, Jiang H, Du J, Sun H, et al. Effects of a Randomized Comprehensive Psychosocial Intervention Based on Cognitive Behavioral Therapy Theory and Motivational Interviewing Techniques for Community Rehabilitation of Patients With Opioid Use Disorders in Shanghai, China. J Addict Med 2015;9(4):322-330. [doi: 10.1097/ADM.0000000000000139] [Medline: 26083955]
30. Thoma N, Pilecki B, McKay D. Contemporary Cognitive Behavior Therapy: A Review of Theory, History, and Evidence. Psychodyn Psychiatry 2015 Sep;43(3):423-461. [doi: 10.1521/pdps.2015.43.3.423] [Medline: 26301761]
31. Witkiewitz K, Marlatt GA, Walker D. Mindfulness-Based Relapse Prevention for Alcohol and Substance Use Disorders. J Cogn Psychother 2005 Jul 01;19(3):211-228. [doi: 10.1891/jcop.2005.19.3.211]
32. Min Z. The Basic Skills of Psychological Intervention for Drug Abuse: Brief Introduction of Motivational Intervention. Chinese Journal of Drug Abuse Prevention and Treatment 2006;12(6):311-314. [doi: 10.3969/j.issn.1006-902X.2006.06.001]
33. Ali EE, Chan SSL, Poh HY, Susanto YA, Suganya T, Leow JL, et al. Design Considerations in the Development of App-Based Oral Anticancer Medication Management Systems: a Qualitative Evaluation of Pharmacists' and Patients' Perspectives. J Med Syst 2019 Feb 06;43(3):63. [doi: 10.1007/s10916-019-1168-x] [Medline: 30726518]
34. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
35. Wang J, Jiang H, Jiang D, Dong A, Wang J, Jin-Mao F. Randomized control trial on improving effect of methadone maintenance treatment by enhancing motivation. Journal of Shanghai Jiaotong University 2015;35(10):1497-1501. [doi: 10.3969/j.issn.1674-8115.2015.10.013]
36. Creswell KG, Skrzynski CJ. The Influence of Smoking Motivation on the Associations Among Cigarette Craving, Attentional Bias to Smoking Cues, and Smoking Behavior. Nicotine Tob Res 2021 Aug 29;23(10):1727-1734 [FREE Full text] [doi: 10.1093/ntr/ntab028] [Medline: 33599777]
37. Sarter L, Heider J, Kirchner L, Schenkel S, Witthöft M, Rief W, et al. Cognitive and emotional variables predicting treatment outcome of cognitive behavior therapies for patients with medically unexplained symptoms: A meta-analysis. J Psychosom Res 2021 Jul;146:110486. [doi: 10.1016/j.jpsychores.2021.110486] [Medline: 33879330]
38. Chen T, Su H, Jiang H, Li X, Zhong N, Du J, et al. Cognitive and emotional predictors of real versus sham repetitive transcranial magnetic stimulation treatment response in methamphetamine use disorder. J Psychiatr Res 2020 Jul;126:73-80. [doi: 10.1016/j.jpsychires.2020.05.007] [Medline: 32422456]

Abbreviations

AI: artificial intelligence
BIS: Barratt Impulsiveness Scale
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD-7: Generalized Anxiety Disorder-7
MET: motivational enhancement therapy
PHQ-9: Patient Health Questionnaire-9
SOCRATES: Stages of Change Readiness and Treatment Eagerness Scale
SUD: substance use disorder
VAS: visual analogue scale

Edited by L Buis; submitted 17.06.22; peer-reviewed by E Kierstead, K Yokotani; comments to author 08.09.22; revised version received 15.11.22; accepted 20.12.22; published 31.01.23.

Please cite as:
Chen T, Chen L, Li S, Du J, Su H, Jiang H, Wu Q, Zhang L, Bao J, Zhao M
Virtual Digital Psychotherapist App–Based Treatment in Patients With Methamphetamine Use Disorder (Echo-APP): Single-Arm Pilot Feasibility and Efficacy Study
JMIR Mhealth Uhealth 2023;11:e40373
URL: https://mhealth.jmir.org/2023/1/e40373
doi:10.2196/40373
PMID:36719727

Original Paper

Improving Children's Sleep Habits Using an Interactive Smartphone App: Community-Based Intervention Study

Arika Yoshizaki[1], MA; Emi Murata[1], PhD; Tomoka Yamamoto[1], MA; Takashi X Fujisawa[2,3], PhD; Ryuzo Hanaie[1], OTR, PhD; Ikuko Hirata[4], MD; Sayuri Matsumoto[5], MD, PhD; Ikuko Mohri[1,3], MD, PhD; Masako Taniike[1,3], MD, PhD

[1]Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Suita, Osaka, Japan
[2]Research Center for Child Mental Development, University of Fukui, Yoshida-gun, Fukui, Japan
[3]United Graduate School of Child Development, Osaka University, Suita, Osaka, Japan
[4]Department of Pediatrics, Osaka University Hospital, Suita, Osaka, Japan
[5]Higashiosaka City Health Center, Higashiosaka, Osaka, Japan

Corresponding Author:
Arika Yoshizaki, MA
Molecular Research Center for Children's Mental Development
United Graduate School of Child Development
Osaka University
2-2-D5 Yamadaoka
Suita, Osaka, 5650871
Japan
Phone: 81 6 6879 3863
Email: arika@kokoro.med.osaka-u.ac.jp

Abstract

Background: Sleep problems are quite common among young children and are often a challenge for parents and a hinderance to children's development. Although behavioral therapy has proven effective in reducing sleep problems in children, a lack of access to professionals who can provide effective support is a major barrier for many caregivers. Therefore, pediatric sleep experts have begun developing apps and web-based services for caregivers. Despite the substantial influence of cultural and familial factors on children's sleep, little effort has gone into developing cultural or family-tailored interventions.

Objective: This study aimed to examine the effectiveness of the interactive smartphone app "Nenne Navi," which provides culturally and family-tailored suggestions for improving sleep habits in young Japanese children through community-based long-term trials. The study also aimed to investigate the association between app-driven improvements in sleep and mental development in children.

Methods: This study adopted a community-based approach to recruit individuals from the Higashi-Osaka city (Japan) who met ≥1 of the following eligibility criteria for sleep problems: sleeping after 10 PM, getting <9 hours of nighttime sleep, and experiencing frequent nighttime awakenings. A total of 87 Japanese caregivers with young children (mean 19.50, SD 0.70 months) were recruited and assigned to the app use group (intervention group) or the video-only group (control group). Both groups received educational video content regarding sleep health literacy. The caregivers in the intervention group used the app, which provides family-tailored suggestions, once per month for 1 year.

Results: A total of 92% (33/36) of the caregivers in the app use group completed 1 year of the intervention. The participants' overall evaluation of the app was positive. The wake-up time was advanced (base mean 8:06 AM; post mean 7:48 AM; $F_{1,65}$=6.769; P=.01 and sleep onset latency was decreased (base mean 34.45 minutes; post mean 20.05 minutes; $F_{1,65}$=23.219; P<.001) significantly in the app use group at the 13th month compared with the video-only group. Moreover, multiple regression analysis showed that decreased social jetlag (β=−0.302; P=.03) and increased sleep onset latency SD (β=.426; P=.02) in children predicted a significant enhancement in the development of social relationships with adults. At 6 months after the completion of the app use, all the caregivers reported continuation of the new lifestyle.

Conclusions: The present findings suggest that the app "Nenne Navi" has high continuity in community use and can improve sleep habits in young Japanese children and that interventions for sleep habits of young children may lead to the enhancement of children's social development. Future studies must focus on the effectiveness of the app in other regions with different regional characteristics and neuroscientific investigations on how changes in sleep impact brain development.

(JMIR Mhealth Uhealth 2023;11:e40836) doi:10.2196/40836

KEYWORDS

infant sleep; app; mHealth; mobile health; behavioral intervention; sleep health; social implementation; mobile phone

Introduction

Background

The Centers for Disease Control and Prevention, the National Public Health agency of the United States, described sleep deprivation as a "public health epidemic" linked to a wide range of medical issues, including hypertension, diabetes, depression, obesity, and cancer [1]. Sleep deprivation can impede not only physical health but also mental health and development.

Sleep problems are quite prevalent among young children, regardless of their cultural origins [2]. An international pediatric task force stated that insufficient sleep among children is a major public health concern [3]. According to a meta-analysis of sleep, cognitive, and behavioral problems in school-aged children (aged 5 to 12 years), inadequate sleep quality or quantity during childhood can affect daytime functioning, cognitive development, and health [4].

Current neuroscience research shows that the process of synaptic pruning occurs during rapid eye movement sleep [5], which demonstrates the importance of sleep from a developmental perspective. Recent cohort studies have shown that children who sleep less during infancy and early childhood are at a higher risk for hyperactivity and lower cognitive functioning later in their lives [6]. In addition, extant literature suggests that the early years (up to 3 years of age) are a sensitive period in which sleep can impact development [6,7]. A Norwegian cohort study investigating the link between sleep in early years and later development showed that short sleep duration and frequent nocturnal awakenings among toddlers aged 1.5 years were associated with the development of both internalizing and externalizing problems at 5 years of age [8]. Sivertsen et al [9] also found that short sleep duration (≤10 hours) and frequent (≥3) nightly awakenings at 1.5 years of age predicted the development of depressive symptoms at 8 years of age. A recent Australian cohort study also reported that sleep problems at 4 to 5 years of age are associated with internalizing difficulties through 12 to 13 years of age [10].

In addition, various studies have demonstrated that children with neurodevelopmental disorders are likely to have more sleep problems and that sleep parameters are associated with the severity of the symptoms of developmental disorders [11-13]. A recent article proposed a novel view on attention-deficit/hyperactivity disorder, whereby a part of the symptoms of attention-deficit/hyperactivity disorder were linked to chronic sleep disorders, with delayed circadian rhythm suggested as the underlying mechanism [14]. Furthermore, several studies on children with neurodevelopmental disorders have indicated that improvements in their sleep problems could lead to improvements in their behavioral problems [15-17]. However, in the context of the evidence regarding the association between sleep problems and developmental trajectories, there is a lack of clarity on whether sleep problems and the predictors of developmental disorders originally coexist or whether sleep problems in early childhood impact developmental trajectories [18].

Therefore, intervention studies examining early childhood are needed to determine whether the improvement of sleep in early childhood can prevent adverse outcomes and enhance healthy developmental trajectories [19]. We hypothesized that improving sleep problems in early childhood would result in better developmental trajectories, which is the ultimate goal of our research.

Furthermore, children's sleep problems are associated with parental stress, family conflict, and maternal depressive symptoms [20-22] and are a risk factor for maltreatment [23]. The "Common risk assessment tool for child consultation centers and municipalities related to child abuse" issued by the Japanese Ministry of Health, Labour and Welfare in 2017 recommends that early support is needed for children with unstable sleep-wake rhythms and difficulty in sleeping [24]. The research also suggested that interventions for improving children's sleep and developing good sleep habits in early childhood are likely to improve the quality of life for the whole family.

Previous studies support the substantial impact of cultural factors on children's sleep habits [25-29]. In addition to cultural factors, caregivers' lifestyles also impact children's sleep habits [30]. Japanese infants and young children are reported to have the shortest sleep duration among the infants and young children among the 17 countries where the survey was conducted [31]. Our previous study revealed that 30.8% of preschool children are sleep deprived and that 56.7% of the caregivers of children who sleep <8 hours at night rated their children's sleep as "good" [30]. It has been suggested that the current situation may be the result of Japan's unique sleep culture, which values working hard over getting enough sleep; the working environment of caregivers; the living environment; and other complex factors along with a lack of sleep literacy. Furthermore, Japan has witnessed a rapid increase in the use of electronic devices and a shift to a more nighttime lifestyle in recent years, which threaten the sleep health and development of Japanese children. As infancy and early childhood are the sensitive periods for sleep, it is necessary for children to develop adequate sleep habits during their early childhood. However, given the various factors involved, improving the sleep health of Japanese toddlers can be quite challenging.

Recent findings suggest that parental factors both predict the outcomes of and are predicted by behavioral interventions for infant sleep problems [32]. Sviggum et al [33] suggested that early and customized guidance for caregivers, with a focus on revealing and acknowledging their experiences with sleep problems in their children, is essential in helping caregivers deal with the challenges [33]. Shetty et al [34] focused on daytime parenting and found that permissive or inconsistent daytime parenting practices were associated with more severe sleep problems [34].

Therefore, we can conclude that suggestions for caregivers should include guidance on daytime parenting practices and be tailored to the unique experiences of families by considering sociopsychological factors such as culture, values, family and housing environment, and the working situations of caregivers. This is particularly the case for Asia, where people have a habit of cosleeping; hence, it is necessary to provide culturally sensitive suggestions. Previous findings have suggested that there are various familial factors that can affect children's sleep, such as sleep environment, wake-up time, delayed or irregular mealtime, screen time, physical activity, and irregular or late bedtime of parents [28,35-39].

In Japan, guidance on sleep and childcare has traditionally been provided through face-to-face consultations at public health care centers. However, as the number of dual-working families has increased, such consultations have become more difficult for caregivers, thereby limiting the availability of guidance. Recent studies have shown the efficacy of web-based and mobile health (mHealth) interventions for sleep problems in infants and young children [31,40,41]. However, these devices developed in Western countries cannot be used in Japan without substantial modifications to account for the cultural differences in sleep habits, such as cosleeping and sleeping on futon mattresses. Considering these background and previous reports, we developed a smartphone app called "Nenne Navi," which facilitates interaction between caregivers and pediatric sleep experts to improve sleep habits in young Japanese children [42]. This app provides culturally and family-tailored advice to each family to make small behavioral changes in their lifestyles through the Plan-Do-Check-Act cycle (Multimedia Appendix 1). We previously conducted a community-based trial for 1 year. The aim of this study was to examine the app's long-term continuity and effectiveness in improving children's sleep habits and development and parental cognition and behavior.

A Priori Hypotheses

The app demonstrates intervention adherence (continuity) for long-term use in the community-based trials and long-term effectiveness in improving the sleep habits of young children and the parenting efficacy of their caregivers.

Methods

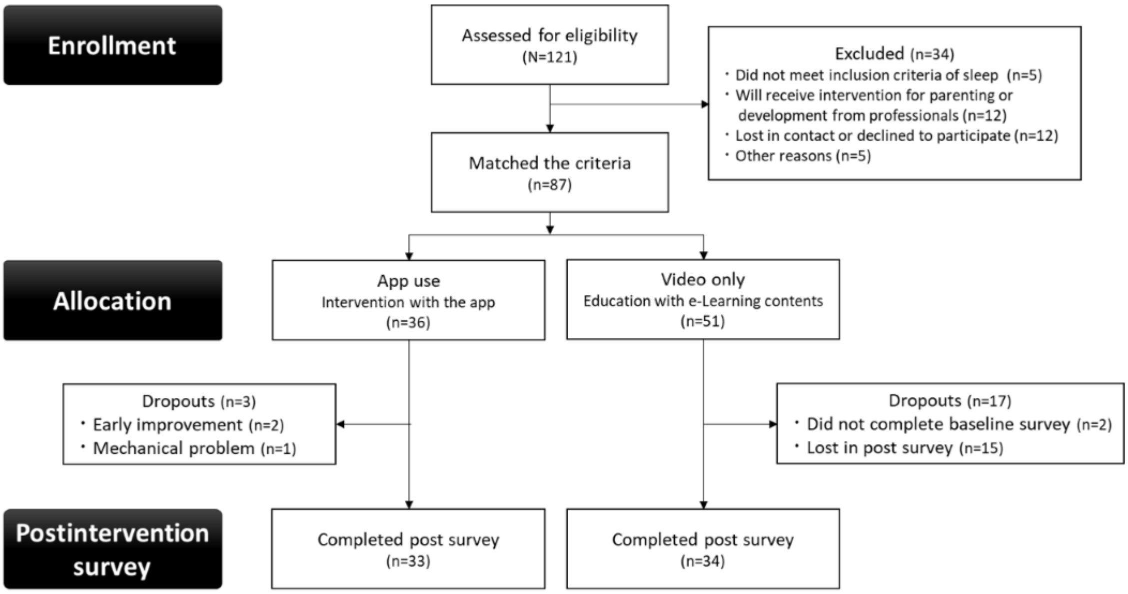
Ethics Approval and Consent to Participate

This study was approved by the Osaka University Clinical Research Review Committee (CRB5180007) on January 23, 2017, before the start of the study. All the study procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki. At the beginning of the baseline assessment, the participants received detailed information about the study's goals and procedures and were informed about the underlying data protection. Written consent was obtained from all the participants individually. All participants received a coupon for books worth JAP ¥5000 (US $42) upon the completion of the trial. The amount was set to be increased to up to US $100 for the app group depending on their contribution. The participants in the app group were notified that they would receive additional rewards according to their app use but were not informed of the exact amount to avoid impacting the results.

Participants

A total of 87 Japanese caregivers (all mothers) with young children (mean 19.50, SD 0.70 months) from the Higashi-Osaka city were recruited over a 6-month period (September 2017 to March 2018) and assigned to either the app use group (intervention group) or the video-only group (control group) based on their preference; those without a preference were randomly assigned to either intervention. The Higashi-Osaka city is an urban area in the Western part of Japan with approximately 3000 births per year. It was confirmed that the sleep-wake patterns of the young children in this city were comparable with the national average in Japan (mean wake-up time: 7:12 AM, SD 0:58; mean bedtime: mean 9:20 PM, SD 0:54; Taniike et al, unpublished data, August 2016). Our study targeted caregivers with children who had completed the developmental health checkup at 1.5 years of age, as the app was designed for this age group. The inclusion criteria were that the children faced at least 1 of the following sleep problems: (1) bedtime later than 10 PM, (2) <9 hours of nighttime sleep, and (3) frequent night awakenings. In addition, the caregivers of the children needed to be fluent in Japanese, not currently be a participant of any interventions for parenting or children's development, and be willing to participate in the study. The following inclusion criteria had to be met for the app use group: possession of a mobile device (iOS [Apple Inc] or Android [Google LLC]) with internet access and the willingness to install the "Nenne Navi" app on the mobile device. The following inclusion criteria had to be met for the video-only group: ability to access the internet to watch the video content and for record data. A total of 34 participants were excluded, as they did not meet the inclusion criteria. Supported devices included the iPhone, iPad, and iPod touch (Apple Inc) with iOS 8.0 or a later version, and Android devices with Android OS 4.3 or a later version. Both groups received educational video content regarding sleep health literacy. The caregivers in the app use group cooperated in the follow-up evaluation 6 months after the completion of the app use. A flow diagram of this study is shown in Figure 1.

Figure 1. Flow diagram of the study.



Measures

"Nenne Navi" App: Acceptability and Safety in Use

The Nenne Navi app was developed by pediatric sleep experts at the pediatric sleep clinic at Osaka University Hospital to positively influence caregivers' behavior to ensure healthy sleep habits among their young children. The pilot trial of the app showed that there were no major problems with the system and that the usability and acceptability of the app was sufficient. Details of the system design were previously reported elsewhere by Yoshizaki et al [42]. The trademark was registered on July 20, 2018, and the registration number is 606435 "Nenne Navi" (Osaka University; MT, IM, Yoko Aoi, and AY). The patent number is 6920731.

The caregivers were asked to respond to approximately 36 questions regarding sleep-related lifestyles such as wake-up time, bedtime, nap time, screen time, daytime activity, dinner time, and bath time for 8 consecutive days via smartphone. Refer to Yoshizaki et al [42] for all the items in the app (Multimedia Appendix 2). The data were then sent to the Osaka University Virtual Server Hosting Service developed at Cybermedia Center, Osaka University, which is equipped with network security measures, such as access restrictions, encrypted communication, monitoring unauthorized access, and backup, to withstand cyberattacks. The system configuration was as follows: Ubuntu18+ apache 2.7, PHP 7, Postgre SQL9.

Individualized (Culturally and Family-Tailored) Intervention: Small Steps, Autonomous Choice of Behavioral Experiment, and Encouragement to Support Caregivers' Motivation for Better Compliance

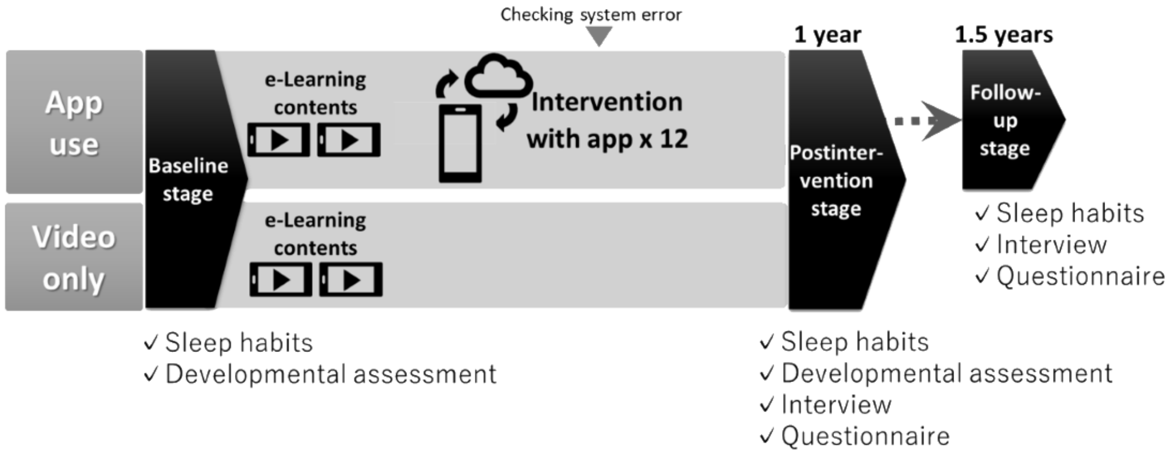
The pediatric sleep expert team consisted of 3 pediatricians and 2 psychologists, who analyzed the information entered by the caregivers and sent various types of practical advice to each caregiver. This app provides culturally sensitive advice on the parent-child lifestyle, such as wake-up time, bedtime, nap time, daytime activities, media use, dinner time, bath time, bedtime routine, and cosleeping habits, as problems with children's sleep habits are caused by a variety of interrelated factors.

The app was designed with the ability to set individualized goals in accordance with individual users' home lives; for example, the app sent personalized advice such as "Try to finish dinner before 8:00 p.m." instead of "Try to have dinner earlier" to deliver specific and optimal goals to caregivers in small progressive steps based on the concept of behavioral therapy. The app does not expect caregivers to obey the advice; conversely, it was designed to send various pieces of advice and suggestions, from which caregivers could choose one to implement temporarily without much effort. Examples of the advice are presented in the study by Yoshizaki et al [42].

One of the features of the app allows caregivers to report the advice they have chosen and whether they have tried it on a monthly basis through the app. This enables the monitoring of caregivers' spontaneous commitment to behavioral change, and the pediatric sleep experts can check the caregivers' degree of compliance. A feedback message (of approximately 150 letters in English) was sent to each caregiver every month through the app to provide them with positive feedback on the improvements they have made in their lifestyle compared with the previous months. In addition, an icon in the app that displays an egg will gradually hatch and grow into a beautiful bird as users use the app. The app also has the function of plotting data into a graph. Caregivers can select their parameters of concern (eg, night awakenings and morning mood) in addition to the time their children fall asleep and see the changes visually. In addition to these personalized interventions and supports, the app provides all users with educational videos and tips on basic sleep and parenting literacy, specifically regarding positive daytime activities to create good sleep.

The design of the intervention study is shown in Figure 2.

Figure 2. Intervention design of the study.



Sleep Parameters

Data on sleep parameters was collected once a month for 8 consecutive days from baseline to the postintervention stage (1 year after) and at follow-up (1.5 years after) for the app use group, whereas for the video-only group, these data were collected for 8 consecutive days at baseline and 1 year after.

In addition to data on typical sleep parameters such as wake-up time, bedtime, nighttime sleep duration, and sleep onset latency, including their SDs, this study used data on social sleep restriction and social jetlag, as both these factors have been noted to be associated with physical health, mental health, and daytime performance. Social sleep restriction and social jetlag are defined as the difference in sleep duration between weekdays and weekends and shifts in sleep timing between weekdays and weekends, respectively.

Measure of Children's Development

The parent-rated Kinder Infant Development Scale (KIDS) [43] was administered to all caregivers at the baseline and postintervention stages. The KIDS is a parent-rated questionnaire that was created in Japan to assess various aspects of the development of infants and toddlers, including motor, language, and social relationships.

Questionnaires and Interviews

A questionnaire was administered and semistructured interviews regarding the changes in parenting efficacy after app use were conducted following the trial to identify the effect of the app and the feasibility of its use in the community. The participants from both groups were asked to describe their impression of the educational video content and the changes in parenting efficacy. The participants in the app use group were additionally asked about their impression of the app, such as about the factors that motivated continued use. The participants responded to items asking them to evaluate the app and the video content using a 5-point scale (5=satisfied, 4=moderately satisfied, 3=neutral, 2=moderately dissatisfied, and 1=dissatisfied) and free descriptions. The participants responded to items asking them to evaluate the changes in parenting efficacy using another 5-point scale (5=very improved, 4=moderately improved, 3=unchanged, 2=moderately worsened, and 1=very worsened).

Data Analysis

We conducted a 2-tailed *t* test to identify the between-group differences in sleep habits and sleep-related lifestyles of the participants in each group at baseline. We then conducted a 2-way ANOVA (group × time) to identify the efficacy of the app in improving children's sleep habits (ie, bedtime, wake-up time, nighttime sleep, and sleep onset latency) via the community-based trial. A multiple regression analysis was conducted to evaluate the effects of improvement in sleep on young children's development. Repeated measures analysis of variance was conducted to examine the maintenance of the effects of app use at follow-up. Data analysis was performed using SPSS statistics (version 26.0; IBM Corp).

Results

Demographic Information of the Study Participants

The demographic information of the study participants is presented in Multimedia Appendix 2. There were no between-group differences, except in fathers' educational status.

Improvement in Sleep-Wake Patterns in Children

The sleep habits and sleep-related lifestyles of the participants at baseline are provided in Table 1.

The app use group displayed significantly later wake-up times, longer sleep onset latency, larger SD for wake-up time, bedtime, and nighttime sleep duration, larger SD for sleep onset latency, and larger social jetlag compared with the video-only group at baseline. Furthermore, the app use group displayed a tendency to have more bedroom feeding habits and delayed wake-sleep rhythm. It is possible that a bias existed in which the caregivers of children with poorer sleep habits preferred the app use group.

Table 1. Sleep habits and sleep-related lifestyles of the participants in each group at baseline[a].

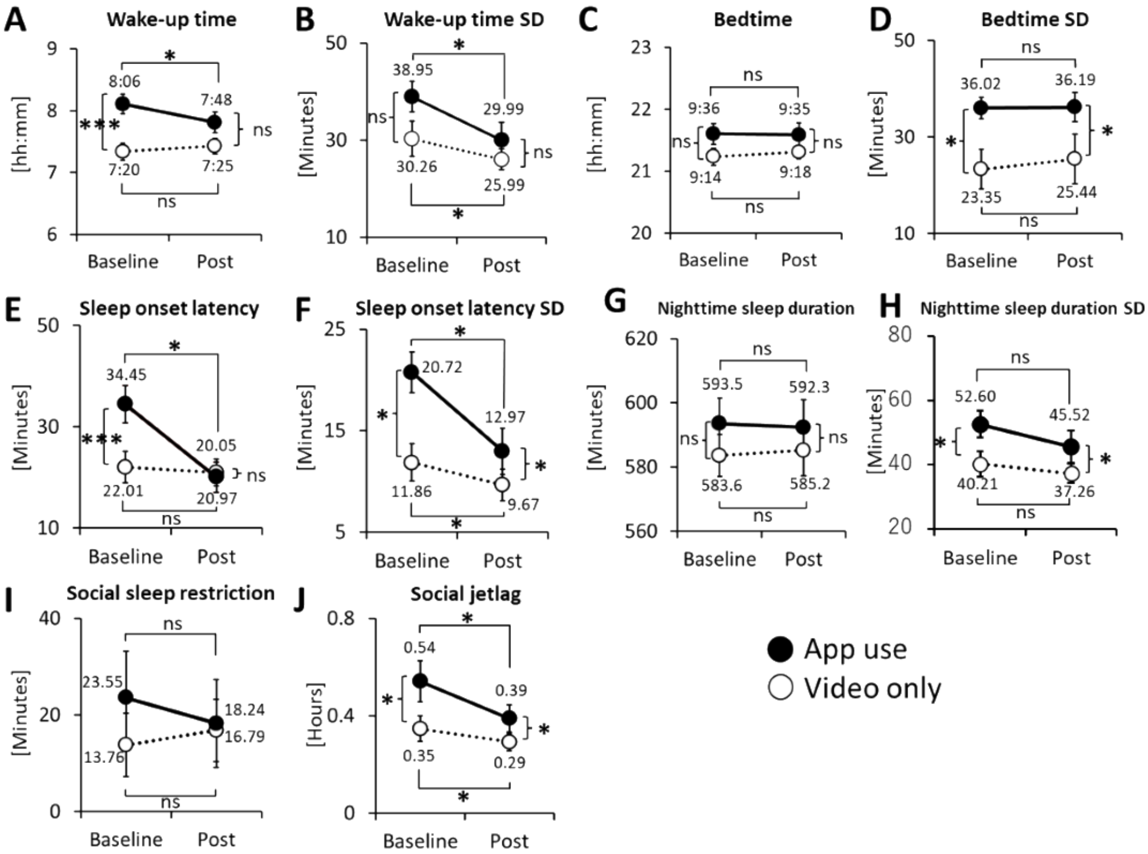
	App use group, mean (SD)	Video-only group, mean (SD)	*P* value
Wake-up time (time, minute)	8:06 AM (0:55)	7:20 AM (0.48)	*<.001*[b]
Wake-up time SD (minutes)	38.95 (18.51)	30.26 (21.33)	.08
Bedtime (time, minute)	9:36 PM (0:59)	9:14 PM (0:49)	.097
Bedtime SD (minutes)	36.02 (23.71)	23.35 (13.16)	*.01*
Sleep onset latency (minutes)	34.45 (21.45)	22.01 (18.45)	*.01*
Sleep onset latency SD (minutes)	20.72 (11.70)	11.86 (11.00)	*.002*
Nighttime sleep duration (minutes)	593.52 (46.55)	583.59 (38.94)	.35
Nighttime sleep duration SD (minutes)	53.93 (25.16)	39.23 (21.44)	*.01*
Number of awakenings after sleep onset	0.99 (1.03)	0.76 (1.13)	.40
Nap starting time (time, minute)	1:41 PM (1:17)	1:22 PM (0:50)	.23
Nap starting time SD (minutes)	97.20 (54.71)	80.47 (37.92)	.15
Nap ending time (time, minute)	3:53 PM (1:20)	3:26 PM (0:57)	.12
Nap ending time SD (minutes)	86.35 (37.70)	78.09 (38.81)	.38
Nap duration (minutes)	106.52 (35.98)	104.79 (25.07)	.82
Nap duration SD (minutes)	39.08 (21.37)	38.77 (17.19)	.95
Total sleep duration (minutes)	700.00 (50.50)	688.35 (42.64)	.31
Total sleep duration SD (minutes)	59.15 (24.78)	48.81 (24.84)	.09
Television-viewing time (minutes)	118.70 (100.35)	87.47 (69.98)	.15
End of television-viewing time after 4 PM (time, minute)	8:37 PM (1:45)	8:31 PM (1:17)	.79
Smartphone-use time (minutes)	11.82 (18.00)	7.62 (19.06)	.36
End of smartphone-use time (time, minute)	5:35 PM (3:13)	5:44 PM (2:33)	.89
Outdoor play in the morning (%)	57.42 (37.65)	68.00 (37.93)	.26
End of dinner (time, minute)	7:22 PM (0:48)	7:15 PM (0:43)	.52
End of bathing (time, minute)	7:54 PM (1:27)	7:34 PM (1:41)	.38
Breastfeeding in the bed (%)	32.48 (45.95)	11.35 (28.76)	*.03*
Media use in the bed (%)	2.15 (6.32)	0.85 (4.97)	.35
Caregiver wake-up time (time, minute)	7:39 AM (0:53)	6:39 AM (0:48)	*<.001*
Caregiver bedtime (time, minute)	11:30 PM (1:20)	10:52 PM (1:20)	.06
Caregiver sleep onset latency (minutes)	39.59 (26.18)	31.31 (30.91)	.39
Caregiver nighttime sleep duration (minutes)	449.88 (54.11)	435.62 (77.23)	.24

[a]Differences between groups were tested using the 2-tailed *t* test.

[b]Italicized values indicate significance.

A summary of the results is presented in Figure 3. Two-way ANOVAs were conducted to determine the effects of the group (the app use group and the video-only group) and intervention (baseline and post) on the sleep habit variables. First, significant interactions between the effects of the group and intervention were confirmed for wake-up time (Figure 3A; $F_{1,65}$=5.748; *P*=.02) and sleep onset latency (Figure 3E; $F_{1,65}$=12.389; *P*<.001), with only the app use group showing reductions for both outcomes (*P*=.05). Next, significant main effects of both the group and intervention were confirmed for sleep onset latency SD (Figure 3F; group: $F_{1,65}$=8.037, *P*=.006; intervention: $F_{1,65}$=7.969, *P*=.006) and social jetlag (Figure 3J; group: $F_{1,65}$=4.264, *P*=.04; intervention: $F_{1,65}$=4.709, *P*=.03). Furthermore, only the main effect of the intervention was observed for wake-up time SD (Figure 3B; $F_{1,65}$=6.413; *P*=.01), whereas only the main effect of the group was observed for bedtime SD (Figure 3D; group: $F_{1,65}$=6.991; *P*=.01) and nighttime sleep duration SD (Figure 3H; group: $F_{1,65}$=5.510; *P*=.02). Finally, there were no significant main effects or interactions for bedtime (Figure 3C; $F_{1,65}$=2.429; *P*=.12), nighttime sleep duration (Figure 3G; group: $F_{1,65}$=.743; *P*=.39), or social sleep restriction (Figure 3I; $F_{1,65}$<0.449; *P*=.51).

Figure 3. Children's sleep-wake patterns by group.



Feasibility of Using Nenne Navi in the Community

There were no dropouts at the 6-month point, and only 8% (3/36) of the caregivers in the app use group dropped out after 1 year of the intervention. A total of 6% (2/36) of caregivers decided to finish the intervention prematurely, as they determined that they had "improved enough" at the 6-month point, based on the disappearance of night awakening (mean: more than 5 times per night before the intervention to none after), reduction of sleep onset latency as the changes for each of the 2 children who finished intervention prematurely (mean 27.9 minutes to "immediately"), reduction of sleep onset latency (mean: from 20 minutes to 5 minutes), and bedtime refusal. Overall, 3% (1/36) of caregivers dropped out of the intervention owing to internet connectivity issues on her smartphone. The mean score for the evaluation of the app was 4.303 (SD 0.969) and that for the evaluation of the video content was 4.545 (SD 0.782). In the feedback section on the app, the participants mentioned the following as the factors that encouraged continued app use (multiple answers allowed): (1) 91% (30/33) of the caregivers mentioned "caring and encouraging text messages to caregivers," (2) 64% (21/33) mentioned "individually tailored advice," (3) 18% (6/33) mentioned "growth of the bird icon," and (4) 9% (3/33) mentioned "nothing in particular," and (5) 3% (1/33) mentioned "other reasons." Regarding their perception of the appropriate duration of app use, 42% (14/33) of the caregivers answered that a period of around 6 months would be appropriate, 52% (17/33) answered "10–12 months," and 6% (2/33) answered "13 months or longer." At the postintervention stage, of the 33 caregivers, 1 (3%) caregiver stated, "At first, I thought it would be impossible to achieve the goal of turning off the TV at 7:00 PM. But as I worked on other goals and adjusted my lifestyle, I gradually started to turn off the TV earlier and earlier, and finally I made it!!"

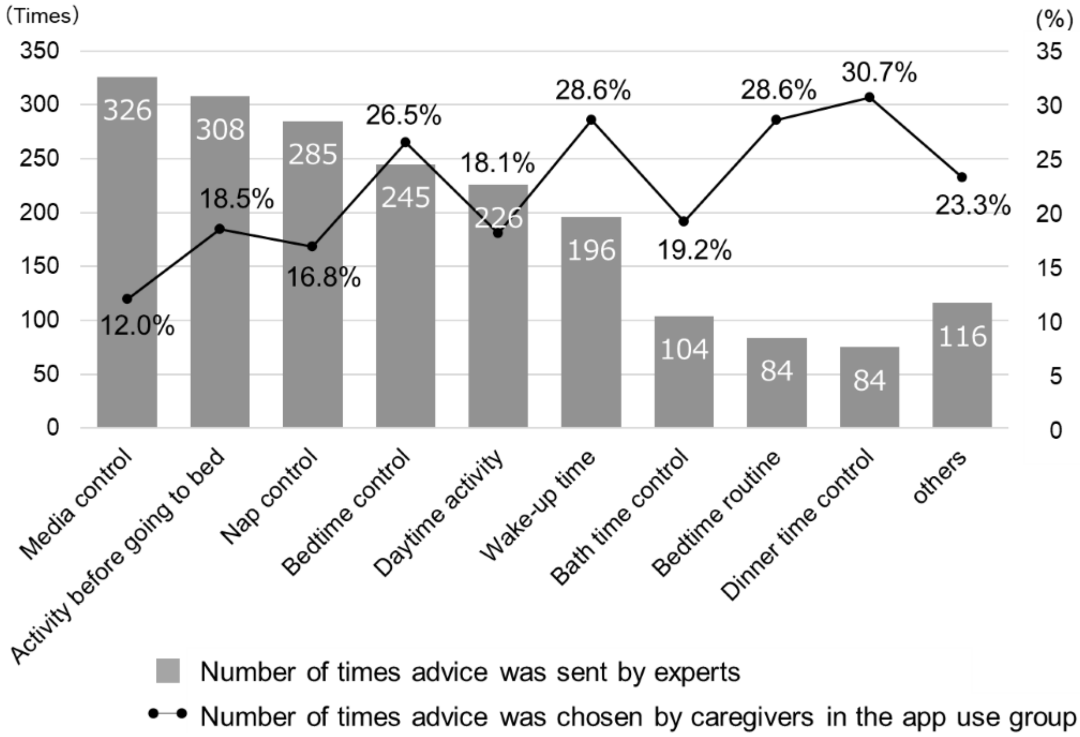
Improvement in Parenting Efficacy

In terms of the improvement in the parenting efficacy of the participants, the mean score was 3.8 (SD 1.1) in the app use group and 3.3 (SD 1.0) in the video-only group. The results of the independent *t* test comparing the 2 groups were as follows: t_{62}=1.996, *P*=.05. In the app use group, 32% (10/31) of the caregivers rated their parenting efficacy after intervention as "very improved" and 32% (10/31) as "moderately improved," whereas in the video-only group, 6% (2/33) of the caregivers rated their parenting efficacy after intervention as "very improved" and 46% (15/33) as "moderately improved." A chi-square test was performed to compare the proportion of caregivers in both groups who responded for the score of change in parenting efficacy. The results showed a trend for differences in response between the groups (χ^2_4=8.3; *P*=.08; ϕ=0.361). A residual analysis revealed that the caregivers of the app use group selected the "very improved" option significantly more than the video-only group (*P*=.01).

Association Between the Content of and Preference for Advice

Figure 4 shows data on the advice sent by the pediatric sleep expert team to the caregivers and the advice chosen by the caregivers. The app offers a variety of advice concerning the daily lives of children and parenting behavior. There was a discrepancy between the advice that the experts tended to focus on (send frequently) and the advice that the caregivers tended to choose to try. The most commonly sent advice was about media control at home; however, it was the advice that the caregivers chose the least. The caregivers were most likely to select the advice about controlling dinner time. The “Others” category included advice on household chores (eg, chores around bedtime should be done after the children have slept or next morning), iron intake for children who are suspected of having restless leg syndrome, receiving help from family members, and adopting measures to maintain a healthy lifestyle. When the advice was sent, the following cultural issues were taken into account: bed sharing with caregivers and siblings, breastfeeding while lying in bed as a bedtime routine, awaiting father’s return home for dinner or bathing, and breastfeeding at midawakening while lying in bed.

Figure 4. The advice was sent by the experts and chosen by the caregivers to try.



Effects of Changes in Sleep Habits on Mental Development

We performed a multiple regression analysis to examine the relationship between the changes in sleep habits and mental development to clarify the effects of the changes in sleep habits from baseline to postintervention stage on children’s development. Developmental age scores in each group, as assessed using KIDS, at baseline and postintervention stage at baseline and postintervention stage are shown in Multimedia Appendix 3. Table 2 shows the results of the multiple regression analysis. The changes (Δ) in sleep parameters from baseline to the postintervention stage were entered as predictors. Multiple regression analysis showed that the model was significant for social relationships with adults ($F_{1,64}$=2.399; P=.02; R^2=0.303) but not for social relationships with children ($F_{1,64}$=1.118; P=.37; R^2=0.169). Increased sleep onset latency SD and decreased social jetlag were significant predictors of the development of social relationships with adults (β=.426, t_{54}=2.521, P=.02; β=−0.302, t=−2.883, P=.03).

Table 2. Multiple regression analysis of social development with sleep habit variables as predictors.

Independent variables	B (SE; 95% CI)	β	*t* test (*df*)	*P* value
Wake-up time (Δ)	−2.092 (4.011; −10.130 to 5.946)	−0.281	−0.522 (54)	.60
Wake-up time SD (Δ)	0.025 (0.037; −0.048 to 0.098)	.107	0.681 (54)	.50
Bedtime (Δ)	3.606 (4.085; −4.581 to 11.794)	.434	0.883 (54)	.38
Bedtime SD (Δ)	−0.033 (0.035; −0.104 to 0.038)	−0.165	−0.941 (54)	.35
Sleep onset latency (Δ)	0.043 (0.072; −0.101 to 0.188)	.160	0.601 (54)	.55
Sleep onset latency SD (Δ)	*0.145 (0.039; 0.068 to 0.223)* [a]	*.426*	*2.521 (54)*	*.02*
Nighttime sleep duration (Δ)	0.042 (0.061; −0.081 to 0.165)	.354	0.686 (54)	.50
Nighttime sleep duration SD (Δ)	0.022 (0.035; −0.048 to 0.091)	.119	0.633 (54)	.53
Social sleep restriction (Δ)	−0.019 (0.010; −0.040 to 0.001)	−0.250	−1.912 (54)	.06
Social jetlag (Δ)	*−4.166 (1.445; −7.053 to 1.278)*	*−0.302*	*−2.883 (54)*	*.03*

[a]Italicized values indicate significance.

Long-term Effects of the App Intervention

A total of 28 (78% of caregivers in the baseline population) caregivers from the app use group participated in the follow-up, including the individual interview months after the end of the intervention. We conducted a repeated measures ANOVA of each sleep parameter at baseline, postintervention, and follow-up time points to examine the follow-up effects of app use. No correction was performed for the missing values. Multimedia Appendix 4 shows the sleep-wake patterns of the children in the app use group at these 3 time points. Wake-up time progressively and significantly advanced from baseline to follow-up. Wake-up time SD, sleep onset latency, sleep onset latency SD, and social jet lag were also significantly shortened between baseline and follow-up. There were no sleep parameters that showed a significant deterioration at follow-up.

As for the caregivers' subjective responses in the follow-up, about 80% (22/28) of the users who participated in the follow-up answered that "the effect of improving sleep habits is still continuing," implying a maintenance of the effect of the app. Furthermore, about 30% (8/28) of app users who participated in the follow-up answered that they began to offer advice to other caregivers around them who were facing problems with their children's sleep.

Discussion

Principal Findings

This study investigated the effectiveness of the interactive smartphone app "Nenne Navi," which provides culturally tailored and family-tailored suggestions for improving the sleep habits of young Japanese children. The results of a long-term trial with community populations suggested that "Nenne Navi" has a very high continuity and that the use of the app can contribute to the promotion of changes in parenting behavior, healthy sleep habits, and development in children. It is also suggested that the improvement in sleep habits resulting from the use of this app might be maintained after the end of the intervention, although further examination of the long-term effects is needed. One of the goals of remote interventions such as mHealth or computerized cognitive behavioral therapy is to minimize dropouts and maintain intervention adherence—high dropout rates, which can reach up to 80%, have been one of the main issues plaguing remote cognitive behavioral therapy [44-47]. Longer programs were associated with higher dropout rates in previous research [48]. The continuity of Nenne Navi was remarkably high in the community-based trial, as there was no dropout at the 6-month mark, and only 8% (3/36) of the participants in the app use group dropped out after 1 year. This high compliance may be explained by the design of the app, which endeavored to promote the confidence and empowerment of caregivers. Some of the design aspects to which the high intervention compliance can be attributed are (1) a reciprocal intervention design instead of a 1-way instruction based on valid suggestions and information; (2) thorough caregiver support function throughout the intervention, such as encouraging and motivating text messages for caregivers; (3) adequate sleep literacy via educational video content; (4) an intervention system that can increase the caregivers' proactive commitment by sending advice several times corresponding to each individual situation and letting the user choose 1 item "to try"; (5) providing small-step multiple advice that can help bridge the gap between "proper literacy" and "caregivers' limitations" while supporting parenting efficacy; and (6) the use of unique design features, such as the growth of the egg icon in response to use and a graphical representation of the parameters selected by individual caregivers. These elements conform to the suggestions of Whittall et al [49]. Surprisingly, as we confirmed the factors that motivated the continued use of the app, many caregivers reported that "caring and encouraging text messages" were the driving force behind their continued use.

Significant between-group differences in sleep-wake patterns were observed at baseline. Overall, the app use group displayed worse sleep habits at baseline. These differences may be explained by a tendency among caregivers more troubled by their children's sleep problems to prefer being recruited to the app use group. Surveys in Japan have shown that wake-up time does not advance in the natural developmental process without intervention among children in the age range covered in this study [50,51]. Regarding the significant within-group change from baseline to the postintervention stage, the improvement in wake-up time and sleep onset latency in the app use group from baseline to the postintervention stage was owing to individual intervention and not the natural course.

The improvement in wake-up time SD, sleep onset latency SD, and social jetlag in both groups may have been the result of the education on "constant sleep rhythm" provided by the educational video delivered to both groups, which led to an improvement in rhythm irregularity.

Overall, this app was effective in improving wake-up rhythm and reducing the difficulty in falling asleep. The improved wake-up rhythm could be explained by the fact that it is relatively easy for mothers to wake up their children at a certain time, as they do not need the cooperation of other family members and already have practice in doing this. Intervention on wake-up time would be important because delayed wake-up time leads to delayed timing of nighttime melatonin secretion as well as napping rhythms, which will inevitably lead to reduced sleep pressure and delayed bedtime at night. In addition, improvement in waking time is also considered beneficial for adaptation to social life. Furthermore, the shortened sleep onset latency was attributed to the increased sleep pressure caused by the advancement of the wake-up rhythm, as well as the guidance provided by the app on establishing bedtime routines and sedative ways of spending time before going to bed, which are relatively new concepts in Japan. Nonetheless, effects on bedtime and extension of sleep time were not observed in this study. It might have been difficult to advance bedtime, as many factors are required to accomplish this. For example, several caregivers reported that it was difficult to obtain the cooperation of other family members in controlling the media use to advance bedtime or bath time. In addition, because many children in Japan go to bed together with their caregivers, the bedtime tended to be late, as the caregivers (the mothers in most cases) had to finish all the household chores before sleeping. However, it must be noted that the trial was conducted in a single community in a relatively urban area and, therefore, may reflect living conditions specific to the region, such as a greater exposure to light at night.

Consequently, although the children in the app use group made significant advances in wake-up time without advances in bedtime, their nighttime sleep duration was not shortened because of decreased sleep onset latency. As sleep duration did not change, there may have been no change in social sleep restriction. The development of further strategies to improve bedtime and sleep duration is an important focus area for future intervention studies.

Allen et al [52] conceptualized the elements of adequate or good sleep for children in a review of studies examining sleep regularity, bedtime routines, quiet or noise comfort or lights, media use, activities, and family conflict. They identified that there are many factors that impact children's sleep. Recent findings suggest that there are multiple barriers for caregivers to adjust parenting behavior to create better sleep habits or reduce sleep problems [49]. The regional characteristics suggest that Japanese caregivers face relatively more barriers to accessing professional help when their children face trouble sleeping, as there are only a few pediatric sleep specialists in Japan. Thus, Japanese caregivers have little access to sleep literacy or solutions to their children's problems. The dissemination of "children's sleep literacy" is an urgent and critical social issue. In addition, the need for remote intervention tools has been further heightened by the COVID-19 pandemic. The COVID-19 pandemic has created new barriers for families struggling to raise their young children and might leave many families isolated. Thus, a means for pediatric sleep specialists to safely communicate up-to-date knowledge without face-to-face contact to caregivers can aid in bridging these barriers and bring many benefits to families.

This app helps caregivers to overcome some of the possible barriers through its unique design incorporating integrated support and small-step care intervention strategies. It is necessary for busy caregivers and those who lack adequate help or knowledge regarding childcare to achieve appropriate levels of sleep literacy and receive positive feedback to improve their children's sleep habits. Many caregivers will benefit from this design aimed at empowering them to change their parenting behavior. Bradway et al [53] pointed out that it is essential to focus on the impact on users' self-efficacy and engagement when judging the success, usefulness, and potential benefits of mHealth in health intervention research [53]. Our findings also suggest that designing the app to increase user engagement and self-efficacy contributed to high continuity of use and effectiveness.

Interestingly, there was a discrepancy between the advice that the experts selected on the basis of scientific priority and the advice that the caregivers chose to try. The greater discrepancy in the media control goal might accurately reflect the reality of parenting. Sleep and media use in young children have been a concern for many sleep professionals overseas [54,55]. However, in recent years, children have been reported to be exposed to media devices at increasingly younger ages, and the guidelines on screen time published by the American Academy of Pediatrics [56] appear to be ignored by approximately 90% of caregivers [57]. Some studies suggest that excessive media use may negatively impact brain development [57-59]. Media control at home is a major concern in Japan, similar to other high-income countries, and is often not practiced despite caregivers' knowledge of the potential harm caused by excessive media use. Social transitions, such as an increasing number of nuclear families and dual-earner families, necessitate caregivers to rely on visual media to keep their children busy so that they can focus on household chores. Nonetheless, interviews with caregivers revealed that one of the most effective pieces of advice for improving sleep habits was media control (data not shown). Although the issue may be difficult to solve, this study clarifies that it is an area that professionals should focus on to help "build the bridge" to improve children's sleep. As this app emphasizes the empowerment of caregivers and their own spontaneous behavioral changes, we designed it to build up "what I can do now" scenarios one by one. Regarding advice that is considered important by experts for improvement but not chosen by caregivers, we may be able to help them by providing practical tips or other support for carrying it out; additionally, we should be considerate of the possibility that there may be some limitations (eg, housing conditions) depending on familial situations.

In accordance with the previous reports that sleep in childhood was associated with later socioemotional problems [8,60], we focused on the social relationships with children and adults as a socioemotional developmental index. We found that decreased social jetlag and increased sleep onset latency SD in children predicted significant enhancement of social development. Although there are many indications of the adverse health effects of social jetlag in adolescents and adults [61,62], the results of this study suggest that close attention should be paid to social jetlag in young children from a developmental perspective. A recent study reported that social jetlag is negatively associated with serum brain-derived neurotrophic factor levels, which play an important role in neuronal maintenance, plasticity, and neurogenesis [63]. The age of the participants in this study, from 1.5 years to 2.5 years, is regarded as an age of remarkable socioemotional development (critical or sensitive period) in the trajectories of brain development [64,65]. The age of 2 years has been defined as a time when the restructuring of the parent-child relationship progresses from the perspective of a developmental theory of parent-child attachment [66]. Our results suggested the possibility that a reduction in social jetlag during this period might play a role in the enhancement of social development. Further research is needed to explain why the social jetlag could be related to social development in young children. Our results also suggested that the increased sleep onset latency SD might be also related to social development with adults. One possible explanation for this association is that as caregivers begin to modify their living situations and bedtime routines to accommodate better sleep, sleep onset latency might range from being very short (successful days) to long (unsuccessful days for any reason), which could have contributed to the current results. This association has not yet been clarified in many aspects and needs to be further investigated in the future.

In Japan, children's sleep habits are strongly influenced by the sleep habits of their caregivers owing to the cosleeping lifestyle. Fukumizu et al [67] suggested that the cosleeping habit and bedtime irregularity were associated with sleep-related nighttime crying in Japanese children. It is necessary to increase awareness among families and help parents make changes in parenting to ensure that their children are not negatively impacted by irregular sleep habits. The relationship between increased sleep onset latency SD and a promotion of social development with adults remains to be clarified; however, there are some possible explanations. First, it may have been related to caregivers' efforts to change parenting behaviors to help children sleep. For example, some caregivers discontinued breastfeeding as a bedtime routine and started reading picture books instead. Although the range of sleep onset latency increased and then varied temporarily, it may have had a positive impact on the parent-child relationship. Alternatively, children who were sleep deprived and fell asleep immediately at a later bedtime at baseline went to bed earlier with varied sleep onset latency.

The use of an interactive design in the app demonstrated its similarity to precision medicine, which can identify the issues in each family and provide optimized suggestions. Recent findings suggested that there are multiple barriers or reasons for caregivers not seeking help for children's sleep problems, indicating that deferential help-seeking interventions are needed depending on the barriers or problem severity [68]. Increasing the choice of caregiver education and interventions, such as video-based caregiver education, app-based individualized remote interventions adopted in this study, traditional face-to-face interventions, telemedicine, and specialized outpatient services, will benefit more families.

Study Strengths and Limitations

This study adopted a community-based approach, which is closer to the real world and could be applied to diverse social service settings and target audiences. This study shows that the app is successfully designed for reciprocal interaction between caregivers and pediatric sleep experts to promote caregivers' behavioral changes to ensure healthy sleep habits among young children. The design of interactive interventions that allow for stepped care, with a specific focus on culturally and family-tailored interventions, and consistent empowerment and support for caregivers resulted in a very high continuity of use while enhancing parenting efficacy. The techniques used in this study may be applicable to other medical or health care domains.

Nonetheless, this study had some limitations. The primary limitation of this study is the moderate sample size. Another limitation is that randomization was not adopted owing to the municipality's preference for the equality of its citizens. As the design of randomized controlled trials could not be adopted, we cannot exclude the possibility that the improvement in sleep variables in the app use group included the effect of regression to the mean. Therefore, it would be difficult to determine the general effectiveness of the app in this study. We assume that a certain intervention effect was observed after the intervention because significant differences in many sleep variables relative to the video-only group disappeared; however, it is essential to confirm the effect of improvement by randomized controlled trials in the future. Therefore, we should be careful not to overgeneralize the results of this study. Data at the follow-up point were obtained only from the app use group; therefore, the sleep habits of the video-only group at that point are unknown. In addition, because the data in this study were reported by the caregivers, there is a possibility of reporting bias, although the accuracy of the sleep rhythm data entered into the app was confirmed in a previous study by Yoshizaki et al [42]. Furthermore, this community-based trial was conducted in a single community in an urban area. We also note that it is still unclear why changes in sleep parameters predicted accelerated development. Although this study did not include an analysis of the association between the objective parameters of app use such as access history and adherence, the focus on objective parameters should be important for understanding adherence and future development of the app. Further studies should be conducted considering these limitations.

Conclusions

This study confirmed the long-term continuity of the use of the app and its efficacy in improving sleep habits. In addition, its effects on follow-up maintenance with long-term intervention in the community-based trial was also confirmed. The use of the Nenne Navi app was associated with improved sleep habits and parenting behavior, suggesting an enhanced parenting efficacy in caregivers. The participants' feedback demonstrated that this effect was supported by the advice that empowered caregivers while encouraging family-tailored, small-step changes in parenting behavior.

The app is expected to be used in sleep medicine and parental education in Japan and is expected to contribute to the expansion of sleep health literacy among families with young children. This app will continue to be implemented in the community as a culturally and individually sensitive, caregiver-supportive sleep education tool. Furthermore, the app could ultimately contribute to improvements in sleep habits and healthy development among Japanese children. Further research must focus on neuroscience to confirm whether this early sleep intervention would lead to more desirable brain development.

Acknowledgments

This study was supported by the Osaka University Center of Innovation program (JPMJCE1310), Japan Society for the Promotion of Science (JSPS) grant 16H03273 (to MT) and 17K17855 (to AY), and the Commissioned Research of the National Institute of Information and Communications Technology, Japan. The authors would like to thank all the families who participated in the study; staff of Healthcare Center at Higashi-Osaka city, Japan; Dr Takafumi Kato (Department of Oral Physiology, Graduate School of Dentistry at Osaka University); Professor Yuko Ohno (Department of Health Science at Osaka University); Dr Masaya Tachibana (United Graduate School of Child Development at Osaka University); Kyoko Makita (Molecular Research Center for Children's Mental Development at Osaka University); Yuka Miki (Molecular Research Center for Children's Mental Development at Osaka University); Dr Yoko Aoi (Department of Oral Physiology, Graduate School of Dentistry at Osaka University); Tetsuya Tajima (Department of Health Science at Osaka University); Shiho Okada (Molecular Research Center for Children's Mental Development at Osaka University); and Dr Shigeyuki Matsuzawa (Molecular Research Center for Children's Mental Development at Osaka University) for their substantial contribution to the study.

Conflicts of Interest

AY, EM, IH, IM, and MT received licensing fees of this app from Panasonic Advanced Technology Development Co, Ltd. AY, IM, and MT received patent royalties' fees of this app from Panasonic Advanced Technology Development Co, Ltd.

Multimedia Appendix 1
Screenshots and the interactive Plan-Do-Check-Act cycle of the app. (A) Snapshot of the app. (B) Screenshots of the top page and data input page. (C) e-Learning content. (D) Interactive Plan-Do-Check-Act cycle of the app.
[PNG File , 392 KB - mhealth_v11i1e40836_app1.png]

Multimedia Appendix 2
Demographic information of the study participants.
[DOCX File , 18 KB - mhealth_v11i1e40836_app2.docx]

Multimedia Appendix 3
Developmental age scores for the Kinder Infant Development Scale in each group at the baseline and postintervention stages.
[DOCX File , 18 KB - mhealth_v11i1e40836_app3.docx]

Multimedia Appendix 4
Children's sleep-wake patterns by group at the baseline, postintervention, and follow-up stages.
[DOCX File , 89 KB - mhealth_v11i1e40836_app4.docx]

References

1. Sleep deprivation described as a serious public health problem. American Association for the Advancement of Science. 2014 Mar 14. URL: https://www.aaas.org/news/sleep-deprivation-described-serious-public-health-problem [accessed 2022-05-01]
2. Sadeh A, Mindell J, Rivera L. "My child has a sleep problem": a cross-cultural comparison of parental definitions. Sleep Med 2011 May;12(5):478-482. [doi: 10.1016/j.sleep.2010.10.008] [Medline: 21478050]
3. Mindell JA, Emslie G, Blumer J, Genel M, Glaze D, Ivanenko A, et al. Pharmacologic management of insomnia in children and adolescents: consensus statement. Pediatrics 2006 Jun;117(6):e1223-e1232. [doi: 10.1542/peds.2005-1693] [Medline: 16740821]
4. Astill RG, Van der Heijden KB, Van Ijzendoorn MH, Van Someren EJ. Sleep, cognition, and behavioral problems in school-age children: a century of research meta-analyzed. Psychol Bull 2012 Nov;138(6):1109-1138. [doi: 10.1037/a0028204] [Medline: 22545685]

5. Li W, Ma L, Yang G, Gan WB. REM sleep selectively prunes and maintains new synapses in development and learning. Nat Neurosci 2017 Mar;20(3):427-437 [FREE Full text] [doi: 10.1038/nn.4479] [Medline: 28092659]
6. Touchette E, Côté SM, Petit D, Liu X, Boivin M, Falissard B, et al. Short nighttime sleep-duration and hyperactivity trajectories in early childhood. Pediatrics 2009 Nov;124(5):e985-e993. [doi: 10.1542/peds.2008-2005] [Medline: 19841107]
7. Touchette E, Petit D, Séguin JR, Boivin M, Tremblay RE, Montplaisir JY. Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. Sleep 2007 Sep;30(9):1213-1219 [FREE Full text] [doi: 10.1093/sleep/30.9.1213] [Medline: 17910393]
8. Sivertsen B, Harvey AG, Reichborn-Kjennerud T, Torgersen L, Ystrom E, Hysing M. Later emotional and behavioral problems associated with sleep problems in toddlers: a longitudinal study. JAMA Pediatr 2015 Jun;169(6):575-582. [doi: 10.1001/jamapediatrics.2015.0187] [Medline: 25867179]
9. Sivertsen B, Harvey AG, Reichborn-Kjennerud T, Ystrom E, Hysing M. Sleep problems and depressive symptoms in toddlers and 8-year-old children: a longitudinal study. J Sleep Res 2021 Feb;30(1):e13150 [FREE Full text] [doi: 10.1111/jsr.13150] [Medline: 32743857]
10. Quach JL, Nguyen CD, Williams KE, Sciberras E. Bidirectional associations between child sleep problems and internalizing and externalizing difficulties from preschool to early adolescence. JAMA Pediatr 2018 Feb 05;172(2):e174363 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.4363] [Medline: 29204659]
11. Hirata I, Mohri I, Kato-Nishimura K, Tachibana M, Kuwada A, Kagitani-Shimono K, et al. Sleep problems are more frequent and associated with problematic behaviors in preschoolers with autism spectrum disorder. Res Dev Disabil 2016;49-50:86-99. [doi: 10.1016/j.ridd.2015.11.002] [Medline: 26672680]
12. Wajszilber D, Santiseban JA, Gruber R. Sleep disorders in patients with ADHD: impact and management challenges. Nat Sci Sleep 2018 Dec 14;10:453-480 [FREE Full text] [doi: 10.2147/NSS.S163074] [Medline: 30588139]
13. Lee SH, Kim HB, Lee KW. Association between sleep duration and attention-deficit hyperactivity disorder: a systematic review and meta-analysis of observational studies. J Affect Disord 2019 Sep 01;256:62-69. [doi: 10.1016/j.jad.2019.05.071] [Medline: 31158717]
14. Bijlenga D, Vollebregt MA, Kooij JJ, Arns M. The role of the circadian system in the etiology and pathophysiology of ADHD: time to redefine ADHD? Atten Defic Hyperact Disord 2019 Mar;11(1):5-19. [doi: 10.1007/s12402-018-0271-z] [Medline: 30927228]
15. Wei JL, Mayo MS, Smith HJ, Reese M, Weatherly RA. Improved behavior and sleep after adenotonsillectomy in children with sleep-disordered breathing. Arch Otolaryngol Head Neck Surg 2007 Oct;133(10):974-979. [doi: 10.1001/archotol.133.10.974] [Medline: 17938319]
16. Wei JL, Bond J, Mayo MS, Smith HJ, Reese M, Weatherly RA. Improved behavior and sleep after adenotonsillectomy in children with sleep-disordered breathing: long-term follow-up. Arch Otolaryngol Head Neck Surg 2009 Jul;135(7):642-646. [doi: 10.1001/archoto.2009.74] [Medline: 19620583]
17. Murata E, Mohri I, Kato-Nishimura K, Iimura J, Ogawa M, Tachibana M, et al. Evaluation of behavioral change after adenotonsillectomy for obstructive sleep apnea in children with autism spectrum disorder. Res Dev Disabil 2017 Jun;65:127-139 [FREE Full text] [doi: 10.1016/j.ridd.2017.04.012] [Medline: 28514706]
18. Huhdanpää H, Morales-Muñoz I, Aronen ET, Pölkki P, Saarenpää-Heikkilä O, Paunio T, et al. Sleep difficulties in infancy are associated with symptoms of inattention and hyperactivity at the age of 5 years: a longitudinal study. J Dev Behav Pediatr 2019;40(6):432-440 [FREE Full text] [doi: 10.1097/DBP.0000000000000684] [Medline: 31166249]
19. Morales-Muñoz I, Broome MR, Marwaha S. Association of parent-reported sleep problems in early childhood with psychotic and borderline personality disorder symptoms in adolescence. JAMA Psychiatry 2020 Dec 01;77(12):1256-1265 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1875] [Medline: 32609357]
20. Byars KC, Yeomans-Maldonado G, Noll JG. Parental functioning and pediatric sleep disturbance: an examination of factors associated with parenting stress in children clinically referred for evaluation of insomnia. Sleep Med 2011 Oct;12(9):898-905. [doi: 10.1016/j.sleep.2011.05.002] [Medline: 21940206]
21. Kelly RJ, El-Sheikh M. Marital conflict and children's sleep: reciprocal relations and socioeconomic effects. J Fam Psychol 2011 Jun;25(3):412-422. [doi: 10.1037/a0023789] [Medline: 21553963]
22. Lam P, Hiscock H, Wake M. Outcomes of infant sleep problems: a longitudinal study of sleep, behavior, and maternal well-being. Pediatrics 2003 Mar;111(3):e203-e207. [doi: 10.1542/peds.111.3.e203] [Medline: 12612272]
23. Kerr SM, Jowett SA, Smith LN. Preventing sleep problems in infants: a randomized controlled trial. J Adv Nurs 1996 Nov;24(5):938-942. [doi: 10.1111/j.1365-2648.1996.tb02929.x] [Medline: 8933253]
24. Common risk assessment tool for child consultation centers and municipalities related to child Abuse. The Japanese Ministry of Health, Labour and Welfare. 2017. URL: http://www.mhlw.go.jp/file/06-Seisakujouhou-11900000-Koyoukintoujidoukateikyoku/0000161641.pdf?2017/9/26 [accessed 2022-06-06]
25. Zhang J, Li A, Fok T, Wing Y. Roles of parental sleep/wake patterns, socioeconomic status, and daytime activities in the sleep/wake patterns of children. J Pediatr 2010 Apr;156(4):606-12.e5. [doi: 10.1016/j.jpeds.2009.10.036] [Medline: 20022339]
26. Mindell JA, Sadeh A, Kwon R, Goh DY. Cross-cultural differences in the sleep of preschool children. Sleep Med 2013 Dec;14(12):1283-1289. [doi: 10.1016/j.sleep.2013.09.002] [Medline: 24269649]

27. Jenni OG, O'Connor BB. Children's sleep: an interplay between culture and biology. Pediatrics 2005 Jan;115(1 Suppl):204-216. [doi: 10.1542/peds.2004-0815B] [Medline: 15866854]
28. Owens JA. Introduction: culture and sleep in children. Pediatrics 2005 Jan;115(1 Suppl):201-203. [doi: 10.1542/peds.2004-0815a] [Medline: 15875290]
29. McLaughlin Crabtree V, Beal Korhonen J, Montgomery-Downs HE, Faye Jones VF, O'Brien LM, Gozal D. Cultural influences on the bedtime behaviors of young children. Sleep Med 2005 Jul;6(4):319-324. [doi: 10.1016/j.sleep.2005.02.001] [Medline: 15978515]
30. Mitsuboshi T, Kato-Nishimura K, Simizu S, Matumoto S, Ganno Y, Inoue E, et al. The factor which have an influence on a sleep habits and the sleep of Japanese infants. J Child Health 2012;71:808-816.
31. Mindell JA, Sadeh A, Wiegand B, How TH, Goh DY. Cross-cultural differences in infant and toddler sleep. Sleep Med 2010 Mar;11(3):274-280. [doi: 10.1016/j.sleep.2009.04.012] [Medline: 20138578]
32. Kahn M, Livne-Karp E, Juda-Hanael M, Omer H, Tikotzky L, Anders TF, et al. Behavioral interventions for infant sleep problems: the role of parental cry tolerance and sleep-related cognitions. J Clin Sleep Med 2020 Aug 15;16(8):1275-1283 [FREE Full text] [doi: 10.5664/jcsm.8488] [Medline: 32279703]
33. Sviggum G, Sollesnes R, Langeland E. Parents' experiences with sleep problems in children aged 1-3 years: a qualitative study from a health promotion perspective. Int J Qual Stud Health Well-being 2018 Dec;13(1):1527605 [FREE Full text] [doi: 10.1080/17482631.2018.1527605] [Medline: 30289029]
34. Shetty J, Newton AT, Reid GJ. Parenting practices, bedtime routines, and consistency: associations with pediatric sleep problems. J Pediatr Psychol 2022 Feb 03;47(1):49-58. [doi: 10.1093/jpepsy/jsab072] [Medline: 34343320]
35. Fukuda K, Hasegawa T, Kawahashi I, Imada S. Preschool children's eating and sleeping habits: late rising and brunch on weekends is related to several physical and mental symptoms. Sleep Med 2019 Sep;61:73-81. [doi: 10.1016/j.sleep.2019.03.023] [Medline: 31288980]
36. Kohyama J, Shiiki T, Ohinata-Sugimoto J, Hasegawa T. Potentially harmful sleep habits of 3-year-old children in Japan. J Dev Behav Pediatr 2002 Apr;23(2):67-70. [doi: 10.1097/00004703-200204000-00001] [Medline: 11943967]
37. Kohyama J, Ono M, Anzai Y, Kishino A, Tamanuki K, Moriyama K, et al. Factors associated with sleep duration among pupils. Pediatr Int 2020 Jun;62(6):716-724. [doi: 10.1111/ped.14178] [Medline: 32012386]
38. Komada Y, Adachi N, Matsuura N, Mizuno K, Hirose K, Aritomi R, et al. Irregular sleep habits of parents are associated with increased sleep problems and daytime sleepiness of children. Tohoku J Exp Med 2009 Oct;219(2):85-89 [FREE Full text] [doi: 10.1620/tjem.219.85] [Medline: 19776524]
39. Hense S, Barba G, Pohlabeln H, De Henauw S, Marild S, Molnar D, et al. Factors that influence weekday sleep duration in European children. Sleep 2011 May 01;34(5):633-639 [FREE Full text] [doi: 10.1093/sleep/34.5.633] [Medline: 21532957]
40. Mindell JA, Du Mond CE, Sadeh A, Telofski LS, Kulkarni N, Gunn E. Efficacy of an internet-based intervention for infant and toddler sleep disturbances. Sleep 2011 Apr 01;34(4):451-458 [FREE Full text] [doi: 10.1093/sleep/34.4.451] [Medline: 21461323]
41. Leichman ES, Gould RA, Williamson AA, Walters RM, Mindell JA. Effectiveness of an mHealth intervention for infant sleep disturbances. Behav Ther 2020 Jul;51(4):548-558 [FREE Full text] [doi: 10.1016/j.beth.2019.12.011] [Medline: 32586429]
42. Yoshizaki A, Mohri I, Yamamoto T, Shirota A, Okada S, Murata E, et al. An interactive smartphone app, Nenne Navi, for improving children's sleep: pilot usability study. JMIR Pediatr Parent 2020 Dec 01;3(2):e22102 [FREE Full text] [doi: 10.2196/22102] [Medline: 33122163]
43. Hashimoto K, Sakamoto N, Takekoh M, Ikeda N, Kato K, Honda M, et al. Validity of the family-rated Kinder Infant Development Scale (KIDS) for children. Pediat Therapeut 2013;03(02):1000153 [FREE Full text] [doi: 10.4172/2161-0665.1000153]
44. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013 Apr 15;13:113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
45. Andersson G, Cuijpers P. Pros and cons of online cognitive-behavioural therapy. Br J Psychiatry 2008 Oct;193(4):270-271. [doi: 10.1192/bjp.bp.108.054080] [Medline: 18827286]
46. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
47. Takeshima N, Watanabe N. Computerized/Internet psychotherapy for mental disorder: qualitative review. Jpn J Gen Hosp Psychiatry 2014;26(3):245-254 [FREE Full text] [doi: 10.1007/s11920-010-0146-2]
48. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]

49. Whittall H, Kahn M, Pillion M, Gradisar M. Parents matter: barriers and solutions when implementing behavioural sleep interventions for infant sleep problems. Sleep Med 2021 Aug;84:244-252. [doi: 10.1016/j.sleep.2021.05.042] [Medline: 34182352]
50. Dai4kai Youjinoseikatsu Anketo Kokunaichousa Houkokusho 2010. Benesse Wducational Research and Development institute. URL: https://berd.benesse.jp/jisedaiken/research/pdf/research13_8.pdf%2012/29/2021 [accessed 2022-03-28]
51. Heisei 22 nendo Youji Kenkoudo Chousa Houkokusyo (Report of survey on health status of toddlers). Nihon Shouni Hoken Kyokai (The Japanese Society of Child Health). Tokyo, Japan: Nihon Shouni Hoken Kyokai; 2011. URL: shorturl.at/jnFWY [accessed 2022-03-28]
52. Allen SL, Howlett MD, Coulombe JA, Corkum PV. ABCs of SLEEPING: a review of the evidence behind pediatric sleep practice recommendations. Sleep Med Rev 2016 Oct;29:1-14. [doi: 10.1016/j.smrv.2015.08.006] [Medline: 26551999]
53. Bradway M, Gabarron E, Johansen M, Zanaboni P, Jardim P, Joakimsen R, et al. Methods and measures used to evaluate patient-operated mobile health interventions: scoping literature review. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e16814 [FREE Full text] [doi: 10.2196/16814] [Medline: 32352394]
54. Bellagamba F, Presaghi F, Di Marco M, D'Abundo E, Blanchfield O, Barr R. How infant and toddlers' media use is related to sleeping habits in everyday life in Italy. Front Psychol 2021 Mar 22;12:589664 [FREE Full text] [doi: 10.3389/fpsyg.2021.589664] [Medline: 33828502]
55. Hackl-Wimmer S, Eglmaier M, Eichen L, Rettenbacher K, Macher D, Walter-Laager C, et al. Effects of touchscreen media use on toddlers' sleep: insights from longtime ECG monitoring. Sensors (Basel) 2021 Nov 12;21(22):7515 [FREE Full text] [doi: 10.3390/s21227515] [Medline: 34833593]
56. Council on Communications and Media. Children, adolescents, and the media. Pediatrics 2013 Nov;132(5):958-961. [doi: 10.1542/peds.2013-2656] [Medline: 28448255]
57. Domingues-Montanari S. Clinical and psychological effects of excessive screen time on children. J Paediatr Child Health 2017 Apr;53(4):333-338. [doi: 10.1111/jpc.13462] [Medline: 28168778]
58. Chonchaiya W, Pruksananonda C. Television viewing associates with delayed language development. Acta Paediatr 2008 Jul;97(7):977-982. [doi: 10.1111/j.1651-2227.2008.00831.x] [Medline: 18460044]
59. Takeuchi H, Taki Y, Hashizume H, Asano K, Asano M, Sassa Y, et al. The impact of television viewing on brain structures: cross-sectional and longitudinal analyses. Cereb Cortex 2015 May;25(5):1188-1197. [doi: 10.1093/cercor/bht315] [Medline: 24256892]
60. Ranum BM, Wichstrøm L, Pallesen S, Falch-Madsen J, Halse M, Steinsbekk S. Association between objectively measured sleep duration and symptoms of psychiatric disorders in middle childhood. JAMA Netw Open 2019 Dec 02;2(12):e1918281 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18281] [Medline: 31880797]
61. Wong PM, Hasler BP, Kamarck TW, Muldoon MF, Manuck SB. Social jetlag, chronotype, and cardiometabolic risk. J Clin Endocrinol Metab 2015 Dec;100(12):4612-4620 [FREE Full text] [doi: 10.1210/jc.2015-2923] [Medline: 26580236]
62. Cespedes Feliciano EM, Rifas-Shiman SL, Quante M, Redline S, Oken E, Taveras EM. Chronotype, social jet lag, and cardiometabolic risk factors in early adolescence. JAMA Pediatr 2019 Nov 01;173(11):1049-1057 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.3089] [Medline: 31524936]
63. Saitoh K, Furihata R, Kaneko Y, Suzuki M, Takahashi S, Uchiyama M. Association of serum BDNF levels and the BDNF Val66Met polymorphism with the sleep pattern in healthy young adults. PLoS One 2018 Jun 26;13(6):e0199765 [FREE Full text] [doi: 10.1371/journal.pone.0199765] [Medline: 29944703]
64. Feldman R. Sensitive periods in human social development: new insights from research on oxytocin, synchrony, and high-risk parenting. Dev Psychopathol 2015 May;27(2):369-395. [doi: 10.1017/S0954579415000048] [Medline: 25997760]
65. Early Years Study 1, Reversing the Real Brain Drain. Early Years Study. 1999. URL: https://earlyyearsstudy.ca/early-years-study-1/ [accessed 2022-06-15]
66. Bowlby J. Attachment and Loss: Volume I: Attachment. London, UK: The Hogarth Press; 1969.
67. Fukumizu M, Kaga M, Kohyama J, Hayes MJ. Sleep-related nighttime crying (yonaki) in Japan: a community-based study. Pediatrics 2005 Jan;115(1 Suppl):217-224. [doi: 10.1542/peds.2004-0815C] [Medline: 15866855]
68. Newton AT, Corkum PV, Blunden S, Reid GJ. Influences on help-seeking decisions for behavioral child sleep problems: why parents do and do not seek help. Clin Child Psychol Psychiatry 2021 Jan;26(1):207-221. [doi: 10.1177/1359104520963375] [Medline: 33054355]

Abbreviations

KIDS: Kinder Infant Development Scale
mHealth: mobile health

Edited by G Eysenbach, T Leung; submitted 07.07.22; peer-reviewed by S Michael, K Shamlian; comments to author 29.07.22; revised version received 16.12.22; accepted 09.01.23; published 10.02.23.

Please cite as:
Yoshizaki A, Murata E, Yamamoto T, Fujisawa TX, Hanaie R, Hirata I, Matsumoto S, Mohri I, Taniike M
Improving Children's Sleep Habits Using an Interactive Smartphone App: Community-Based Intervention Study
JMIR Mhealth Uhealth 2023;11:e40836
URL: https://mhealth.jmir.org/2023/1/e40836
doi:10.2196/40836
PMID:36641237

Original Paper

Clinic-Integrated Smartphone App (JomPrEP) to Improve Uptake of HIV Testing and Pre-exposure Prophylaxis Among Men Who Have Sex With Men in Malaysia: Mixed Methods Evaluation of Usability and Acceptability

Roman Shrestha[1,2], PhD, MPH; Frederick L Altice[2], MD; Antoine Khati[1], MD; Iskandar Azwa[3], MD; Kamal Gautam[1], MPH; Sana Gupta[1]; Patrick Sean Sullivan[4], PhD; Zhao Ni[5], PhD; Adeeba Kamarulzaman[3], MD; Panyaphon Phiphatkunarnon[6], MSc; Jeffrey A Wickersham[2], PhD

[1]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States
[2]AIDS Program, Yale School of Medicine, New Haven, CT, United States
[3]Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[4]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States
[5]Yale School of Nursing, West Haven, CT, United States
[6]Love Foundation, Bangkok, Thailand

Corresponding Author:
Roman Shrestha, PhD, MPH
Department of Allied Health Sciences
University of Connecticut
358 Mansfield Road Unit 1101
Storrs, CT, 06269
United States
Phone: 1 8604862834
Email: roman.shrestha@uconn.edu

Abstract

Background: HIV disproportionately affects men who have sex with men (MSM). In Malaysia, where stigma and discrimination toward MSM are high, including in health care settings, mobile health (mHealth) platforms have the potential to open new frontiers in HIV prevention.

Objective: We developed an innovative, clinic-integrated smartphone app called JomPrEP, which provides a virtual platform for Malaysian MSM to engage in HIV prevention services. In collaboration with the local clinics in Malaysia, JomPrEP offers a range of HIV prevention (ie, HIV testing and pre-exposure prophylaxis [PrEP]) and other support services (eg, referral to mental health support) without having to interface face to face with clinicians. This study evaluated the usability and acceptability of JomPrEP to deliver HIV prevention services for MSM in Malaysia.

Methods: In total, 50 PrEP-naive MSM without HIV in Greater Kuala Lumpur, Malaysia, were recruited between March and April 2022. Participants used JomPrEP for a month and completed a postuse survey. The usability of the app and its features were assessed using self-report and objective measures (eg, app analytics, clinic dashboard). Acceptability was evaluated using the System Usability Scale (SUS).

Results: The participants' mean age was 27.9 (SD 5.3) years. Participants used JomPrEP for an average of 8 (SD 5.0) times during 30 days of testing, with each session lasting an average of 28 (SD 38.9) minutes. Of the 50 participants, 42 (84%) ordered an HIV self-testing (HIVST) kit using the app, of whom 18 (42%) ordered an HIVST more than once. Almost all participants (46/50, 92%) initiated PrEP using the app (same-day PrEP initiation: 30/46, 65%); of these, 16/46 (35%) participants chose PrEP e-consultation via the app (vs in-person consultation). Regarding PrEP dispensing, 18/46 (39%) participants chose to receive their PrEP via mail delivery (vs pharmacy pickup). The app was rated as having high acceptability with a mean score of 73.8 (SD 10.1) on the SUS.

Conclusions: JomPrEP was found to be a highly feasible and acceptable tool for MSM in Malaysia to access HIV prevention services quickly and conveniently. A broader, randomized controlled trial is warranted to evaluate its efficacy on HIV prevention outcomes among MSM in Malaysia.

Trial Registration: ClinicalTrials.gov NCT05052411; https://clinicaltrials.gov/ct2/show/NCT05052411

International Registered Report Identifier (IRRID): RR2-10.2196/43318

(JMIR Mhealth Uhealth 2023;11:e44468) doi:10.2196/44468

KEYWORDS

men who have sex with men; mHealth; HIV prevention; pre-exposure prophylaxis; mobile phone; Malaysia; MSM; mobile health; HIV; prevention; usability; acceptability; sexual minority; gay; homosexual

Introduction

Men who have sex with men (MSM) are disproportionately affected by HIV in Malaysia and accounted for 63% of new HIV diagnoses in 2021, a proportion that has been increasing over the past decade [1,2]. This pattern requires implementation of more effective HIV prevention in MSM, yet in Malaysia, as in many low- and middle-income countries (LMICs), MSM often do not adequately access evidence-based HIV prevention (and treatment). Gaps in prevention and treatment are due, in part, to high levels of social stigma and discrimination against MSM. In Malaysia, these factors are heightened further because same-sex sexual behaviors are criminalized [2-4]. Other factors also contribute to low uptake of services, including sexual networks that have evolved through social networking apps. Transmission potential is heightened by behavioral or biological factors, including condomless sex, multiple concurrent sexual partners, substance use, and mental health problems (eg, depression, anxiety) that act synergistically to increase the HIV risk in this group [5-12].

Routine HIV testing and expanded use of pre-exposure prophylaxis (PrEP) would drastically reduce the population-level burden of HIV [13-16]. Uptake of these evidence-based tools, however, is suboptimal among Malaysian MSM. For example, recent data suggest that 55% of MSM reported not having tested for HIV in the past 6 months, and approximately 30% reported they had never been tested [17-19]. Additionally, only 18.3% of MSM with indications for PrEP reported ever using it, despite high awareness and willingness to use PrEP [20,21]. This low uptake is partly explained by the need to maintain meaningful engagement with the health care system to access these services. Yet, it is often difficult for MSM in Malaysia to find culturally appropriate health care services due to known barriers, such as discomfort and distrust associated with disclosing sexual behavior to providers for fear of ramifications [22]. As such, there is a need for innovative strategies to improve access to HIV prevention services for Malaysian MSM.

Mobile health (mHealth), particularly smartphone apps, holds great promise for HIV prevention [18,23-26], especially when linked to accessible HIV testing and PrEP. App-based interventions can help overcome multilevel barriers, given their ability to anonymously reach and engage populations that are disenfranchised from existing prevention efforts and offer "real-time" delivery and rapid scalability of programs at relatively low implementation costs. In Malaysia, smartphone use among MSM is nearly universal, and MSM report a strong preference for app-based HIV prevention programs [18,19]. Although app-based interventions are evolving and promote the HIV prevention continuum, most, if not all, are limited to high-income countries and none provide comprehensive HIV prevention services [27]. Furthermore, many of these emerging apps deploy an online-to-offline (O2O) strategy [28,29], where clients eventually must be seen in person. In these cases, the anonymity afforded through the online experience ends during the clinical encounter where individuals must be linked to their testing results and medication prescriptions. An important innovation in the traditional O2O and clinic-based models would be to keep the entire care process in the virtual space (online). However, such strategies have yet to be developed and assessed.

To address this unmet need, we developed JomPrEP (where "Jom" means "let us" in Bahasa Malaysia), a clinic-integrated smartphone app, designed to provide HIV prevention services for MSM in Malaysia. The development of JomPrEP has been described previously [30]. In brief, it is adapted from HealthMindr, an app previously demonstrated to increase HIV testing and PrEP uptake among MSM in the United States [25]. In collaboration with local Malaysian clinics, JomPrEP offers a virtual platform for Malaysian MSM to access a range of HIV prevention (ie, HIV testing and PrEP) and other support services (eg, referral to mental health support) without having to interface face to face with clinicians. It includes several on-demand features, including scheduling and managing appointments in person or through e-consultation, communicating with the clinical team (ie, chat), home-based testing, accessing test results, ordering health products, discrete door-to-door delivery, timely notifications, a points-based reward system for completing activities within the app, and a multimedia resource center. Here, we report findings from beta testing of the recently developed JomPrEP app to evaluate its usability and acceptability.

Methods

Study Design and Settings

We conducted beta testing of JomPrEP to assess its usability and acceptability among MSM living in the Greater Kuala Lumpur region, Malaysia. Beta testing of an app helps to identify any final areas for improvement [31]. We hypothesized that beta testing for 30 days of observation (N=50) would allow us to evaluate the app's design, functionality, and usability. The sample size of 50 was determined based on the pragmatics of recruitment and the need to examine feasibility [32-34].

We partnered with the Centre of Excellence for Research in AIDS (CERiA) at the University of Malaya, Kuala Lumpur, Malaysia, to conduct this study. Working closely with several other local and international institutions, CERiA conducts innovative and interdisciplinary research that combines epidemiological, biomedical, and sociobehavioral approaches, focusing on the implementation of HIV prevention and treatment. As part of JomPrEP integration with existing clinics, we partnered with 2 local clinics—the Red Clinic (private clinic) and the Community Health Care Clinic (nongovernmental organization [NGO]-based clinic)—to provide clinical services (eg, HIV testing, sexually transmitted infection [STI] testing, PrEP services) virtually via the app.

Study Participants and Recruitment

Eligibility criteria included (1) being 18 years or older; (2) identifying as a cis-gender man; (3) self-reporting an HIV-negative or HIV status unknown at screening; (4) not having used PrEP previously (ie, PrEP naive); (5) self-reporting evidence of being at risk for HIV acquisition, as defined by the World Health Organization PrEP clinical guidelines [35]; (6) owning a smartphone; and (7) currently residing in the Greater Kuala Lumpur region.

In total, 50 participants were recruited between March and April 2022 using in-person and online recruitment strategies. For in-person recruitment, flyers were distributed to potential participants as well as posted at local partner organizations (eg, clinics, lesbian, gay, bisexual, transgender [LGBT]–friendly community-based organizations). Additionally, we used various general and MSM-specific social media platforms as venues for participant recruitment. These included placing advertisements in geosocial networking (GSN) apps popular among MSM in Malaysia (ie, Hornet) as well as posting study flyers on Malaysian MSM–focused Facebook pages. Interested individuals who clicked on an advertisement were directed to the study website [36], where they were presented with a brief description of the study and web-based screening.

Procedures

After meeting enrollment criteria, eligible participants were asked to provide electronic informed consent for study participation, followed by undergoing a baseline assessment. Study staff then assisted enrolled participants with downloading JomPrEP and provided them with brief instructions on the purpose of the app and an overview of how to use it. To restrict access to JomPrEP to the study participants, participants were provided with a single-use registration code needed to gain access to the app. Upon downloading the app, participants were asked to complete an onboarding process, which included creating log-in credentials. They were then redirected to the JomPrEP landing screen (home screen), which contains several icons representing key app functions (Table 1). Screenshots of the app are available in Multimedia Appendix 1. Participants were requested to keep and use the app for 30 days and encouraged to use all app features. On day 30, participants completed a posttest survey and were asked to provide a synthesis of issues regarding the app (ie, exit interviews). For exit interviews, 20 (40%) participants were randomly sampled and interviews were conducted until data saturation was reached. The 1-on-1 sessions were conducted online via licensed videoconferencing software. Participants were given the choice of turning on or off their cameras and were asked to use a pseudonym/nickname.

Participants received point-based rewards (known as JomPrEP points [JPP]) for completing specific activities or meeting milestones via the app (eg, 100 JPP for baseline and follow-up assessment each, 50 for completing app onboarding, 20 for an HIVST in-app order, 50 for completing a lab test, 50 for completing an e-consultation for PrEP, 30 for an in-app PrEP order, 50 for optimal PrEP adherence; maximum points that could be earned: 740 or US $18.50). Participants were allowed to redeem points for cash at any point during the study period (10 JPP=RM 1, or US $0.24).

Table 1. Main features of JomPrEP.

Features	Description
Customizable Home page	• Visual presentation using avatars and pseudonyms
HIVST[a]	• Allows users to order an HIVST kit (Orasure) • Allows users to upload HIVST results for verification purposes to facilitate posttest linkage to HIV prevention or treatment services • Includes multimedia (ie, text-, picture-, and video-based) content on how to use the HIVST kit and interpret the result
PrEP[b] Express	• Provides users with a fast and convenient way to start PrEP (and state on it) • Includes a sequential pathway for the users to get on PrEP: HIV risk assessment (provides tailored recommendations based on the user's response); choose preferred clinic (allows users to choose between the participating clinics for PrEP); choose PrEP type (allows users to choose between different PrEP prescription modalities, ie, same day[c] vs traditional[d]); schedule appointments (allows users to choose their preferred date and time for phlebotomy and e-consultation[e]); PrEP medication delivery (allows users to choose their preferred method of getting PrEP: pickup at the pharmacy vs discrete door-to-door delivery)
Orders	• Allows users to monitor past and current orders (HIVST kit, PrEP) • Allows users to track their current orders in real time (via application programming interface [API] integration of courier tracking) • Provides notifications on the status of their orders
Labs	• Allows users to view their lab test results • Allows users to receive timely notifications about new lab results
Appointments	• Allows users to view details about past and future appointments and make any necessary changes (eg, reschedule, cancel) • Allows users to meet with a doctor virtually (ie, e-consultation)
Mental Health	• Allows users to self-screen for depression and receive results • Provides users with community resources and support services (ie, a list of mental health service providers) • Provides users with a personalized referral letter to mental health services to facilitate rapid linkage[f] • Allows users to keep track of previous mental health–screening results and access referral letters
MedManager	• Allows users to set and receive personalized medication reminders • Provides users with reminders to refill their prescriptions • Allows users to view visual reports of their medication adherence
MoodTracker	• Allows users to keep track of their mood daily • Provides users with a visual display of their mood over time
Messages (ie, chat)	• Allows users to send and receive nonurgent medical questions to the clinic and research staff
Resources	• Provides users with multimedia (text-, picture-, video-based) content of an array of relevant information (eg, HIV, PrEP, substance use, mental health)
News	• Provides users with the latest health news updated regularly
Reward points	• Allows users to accumulate points for completing activities in the app • Allow users to track and redeem points for cash or in-app purchases
JomPrEP Clinic Dashboard	• Includes a web-based dashboard for the clinics affiliated with the JomPrEP app • Allows staff members of the affiliated clinics (ie, doctors, nurses, pharmacists, front-desk staff) to access the dashboard to facilitate patient care for JomPrEP app users • In the absence of an electronic health record (EHR) in the local setting, functions more like the EHR system

[a]HIVST: HIV self-testing.

[b]PrEP: pre-exposure prophylaxis.

[c]Receive PrEP at the first doctor visit (no need to wait for lab results).

[d]Get PrEP after lab results are complete (need a follow-up doctor visit to review lab results).

[e]Applies only to those who choose traditional PrEP.

[f]Includes a referral letter for mental health.

Ethical Considerations

The Institutional Review Board at the University of Connecticut approved this study (H22-0049), with an institutional reliance agreement with the University of Malaya. Eligible participants provided electronic informed consent for study participation.

Assessments

Participant Characteristics

All assessments (ie, baseline and follow-up) were conducted virtually and self-administered using Qualtrics. We collected participant demographic and baseline characteristics, including age, ethnicity, educational status, relationship status, income, housing status, depressive symptoms [37], substance use and sexual history, HIV/STI-testing practices, and past use of PrEP and postexposure prophylaxis (PEP).

JomPrEP App Evaluation

After 30 days of use, participants were asked to assess the app's features, usability, design, content, and functionality using 2 Likert scales. JomPrEP acceptability was assessed using the System Usability Scale (SUS) [38], a validated measure that assesses the subjective usability of an app. Scores range from 0 to 100, with scores of ≥50 indicating that the app is acceptable [38]. We also collected app analytics, such as the number of log-ins, session duration, pages visited, and frequency and duration of use of app components, to determine usability. Additionally, data on the uptake of HIV testing and PrEP, mental health screening and referral to mental health support services, and use of the HIVST kit for those who placed in-app orders were extracted from the web-based JomPrEP Clinic Dashboard.

Finally, we conducted exit interviews with 20 (40%) participants to obtain feedback on app functionality, technical performance, errors and software bugs encountered, overall experience using the app, feedback for further refinement, and subjective impact of the app on HIV testing and PrEP uptake. One-on-one interviews were conducted by research staff virtually using videoconferencing technology. Interviews were recorded and transcribed for analysis.

Analytical Plan

All quantitative data were managed and analyzed using IBM SPSS Statistics version 28. Means for continuous variables and frequencies for categorical variables were calculated to describe the participants at baseline. App usability and acceptability were based on descriptive statistics from the app analytics and acceptability measure. For example, evaluation responses are reported as the percentage of users who completed the posttest survey. SUS results are reported as an aggregate score, with a score of ≥50 indicating that the app is acceptable [39,40]; the percentage of participants with scores ≥50 is also reported. Descriptive statistics of app analytics were used to examine app engagement and are reported as the mean with the range for time and action measurements. For qualitative data, all the exit interviews were audio-recorded, transcribed, and analyzed. The comments and issues were grouped and categorized according to common themes relative to specific app functions by 3 coders (including 2 senior coders) and agreed upon by all authors. Dedoose version 9.0.54 was used throughout to assist in data management and analysis.

Results

Participant Characteristics

The mean age of the 50 participants was 27.9 years (range 21-45 years), with most being single (36/50, 72%), Malay (26/50, 52%), university graduates (34/50, 68%) and living in a house/apartment with other people (36/50, 72%). Almost all participants reported having been tested for HIV (49/50, 98%), and 39/50 (78%) participants had done so in the past 6 months. Of all 50 participants, 26 (52%) reported using HIVST and only 5 (10%) had used PrEP previously. Regarding sexual behaviors in the past 6 months, 47/50 (94%) participants reported anal sex with another man, while only 16/50 (32%) participants reported consistent condom use, 4/50 (8%) reported having engaged in sexualized drug use, and 9/50 (18%) reported having engaged in group sex (Table 2).

Table 2. Characteristics of participants (N=50).

Variables	Frequency
Age (years), mean (SD)	27.9 (5.3)
Ethnicity (Malaya), n (%)	
No	24 (48)
Yes	26 (52)
University graduate[a], n (%)	
No	16 (32)
Yes	34 (68)
Relationship status, n (%)	
Single	36 (72)
Partner	14 (28)
Monthly income (RM/US $), mean (SD)	3553.40 (2985.90)/837.97 (704.14)
Living status, n (%)	
Alone	14 (28)
Living with others	36 (72)
Tested for HIV (past 6 months), n (%)	
No	11 (22)
Yes	39 (78)
Ever used an HIVST[b] kit, n (%)	
No	24 (48)
Yes	26 (52)
Previously diagnosed with STI[c], n (%)	
No	27 (54)
Yes	23 (46)
Ever used PrEP[d], n (%)	
No	45 (90)
Yes	5 (10)
Ever used PEP[e], n (%)	
No	46 (92)
Yes	4 (8)
Perceived HIV risk, n (%)	
None	6 (12)
Low	25 (50)
Moderate	15 (30)
High	4 (8)
Ever injected drugs, n (%)	
No	49 (98)
Yes	1 (2)
Engaged in anal sex (past 6 months), n (%)	
No	3 (6)
Yes	47 (94)
HIV serodiscordant relationship (past 6 months), n (%)	

Variables	Frequency
No	47 (94)
Yes	3 (6)
Consistent condom use (past 6 months), n (%)	
No	34 (68)
Yes	16 (32)
Engaged in group sex (past 6 months), n (%)	
No	41 (82)
Yes	9 (18)
Engaged in sexualized drug use[f] (past 6 months), n (%)	
No	46 (92)
Yes	4 (8)

[a]Includes college, university, and professional degrees.

[a]HIVST: HIV self-testing.

[c]STI: sexually transmitted infections (eg, gonorrhea, chlamydia, syphilis).

[d]PrEP: preexposure prophylaxis.

[e]PEP: postexposure prophylaxis.

[f]Use of psychoactive substances (eg, amphetamines, 3,4-methylene dioxymethamphetamine [MDMA]) before or during sexual activity.

Uptake of HIV Prevention Services

During the 30-day beta-testing phase of JomPrEP, 42/50 (84%) participants ordered an HIVST kit using the app. Almost all (46/50, 92%) participants used the app to get on PrEP. Specifically, 30/46 (65%) participants chose same-day PrEP versus traditional PrEP, and the majority of them picked up the PrEP medication at the pharmacy (28/46, 61%). Additionally, 44/50 (88%) participants used the online assessment tool to screen for depression, and 39/44 (89%) of them met the criteria for moderate-to-severe depressive symptoms [37] and were provided with a referral letter (Table 3).

Table 3. Participants' uptake of HIV testing and PrEP[a] services using JomPrEP (N=50).

Service usage	Frequency
HIV testing	
Ordered HIVST[b] kit	42 (84)
Verified HIVST results[c]	40 (95)
Linked to PrEP services (n=46, 92%)	
Traditional[d] PrEP delivery	16 (35)
Same-day[e] PrEP delivery	30 (65)
Completed phlebotomy	46 (100)
Completed e-consultation[f]	16 (35)
Completed in-person consultation	30 (65)
Picked-up PrEP medication at pharmacy	28 (61)
PrEP medication delivered at home	18 (39)
Mental health screening	44 (88)

[a]PrEP: preexposure prophylaxis.

[b]HIVST: HIV self-testing.

[c]HIVST result verified by providing an image of the result via the app.

[d]Receive PrEP after lab results are complete (need a follow-up doctor visit to review lab results).

[e]Receive PrEP at the first doctor visit (no need to wait for lab results).

[f]Applies only to those who choose same-day PrEP.

JomPrEP App Evaluation

During the beta-testing phase, 29/50 (58%) participants were Android users, while the remainder (21/50, 42%) were iOS users. Usability measures by participants included app use, with an average of 8 (SD 5.0, range 2-18) unique visits over 30 days, with an average duration of 28 (SD 38.9) minutes per session. The app had a mean of 34.9 (SD 14.7) daily users, with 939.3 (SD 597.9) daily page views, 63.4 (SD 28.5) daily sessions on average, and consistent returning visits (eg, >10: 29/50, 58%; 6-10: 22/50, 44%).

The mean acceptability score was 73.8 (SD 10.1) on the SUS, well above the minimum criteria (≥50) set for the acceptability of the app [39,40], with all participants reporting acceptability scores of >50. Almost all participants reported that they were satisfied with JomPrEP (46/50, 92%) and that the app was useful in addressing their HIV prevention needs (49/50, 98%); see Table 4.

When participants were asked about future app use, most said they were likely to continue using the app as part of their HIV prevention plan (42/50, 84%), would download the app if publicly available (43/50, 86%), and would recommend the app to their friends or colleagues (50/50, 100%). Most participants felt confident in in-app security (43/50, 86%), including autologout after 5 minutes of inactivity (44/50, 88%), an email and password log-in (42/50, 84%), a 4-digit personal identification number (36/50, 72%), and the app name and icon not associated with HIV (32/50, 64%); see Table 5.

Participants found JomPrEP to be easy to use and felt confident that they would be able to learn how to use it quickly and without technical assistance (Table 6).

Table 4. Participants' rating of satisfaction with using JomPrEP features (N=50).

Activity ("How satisfied are you with the following features of the JomPrEP app?")	Participants, n (%)[a]
Ordering an HIVST[b] kit	47 (94)
Ordering PrEP[c] medication	46 (92)
Reward system (earning and redeeming points)	43 (86)
Completing the mental health screener	42 (84)
Chat with clinical or research staff	42 (84)
Booking appointments (blood draw, consultation with doctor)	42 (84)
Keeping track of upcoming and past appointments	42 (84)
Tracking order status (ie, HIVST, PrEP)	40 (80)
MoodTracker (track mood daily)	40 (80)
Online consultation with the doctor (e-consultation)	40 (80)
Reviewing laboratory test results	38 (76)
Resources/News Center	36 (72)
Notifications from the app	36 (72)
MedManager (receive medication reminders, track medication use)	30 (60)

[a]Very satisfied and extremely satisfied (not included: not at all satisfied, slightly satisfied, moderately satisfied).

[b]HIVST: HIV self-testing.

[c]PrEP: preexposure prophylaxis.

Table 5. Participants' rating of the usefulness of JomPrEP features (report their use; N=50).

Activity ("How much do you agree that the use of the JomPrEP app…")	Participants, n (%)[a]
Assisted in getting tested for HIV	50 (100)
Assisted in getting started on PrEP[b]	50 (100)
Made access to medical records easier (ie, test results, appointments)	49 (98)
Made access to HIV testing much easier	49 (98)
Helped to understand the risk of getting HIV	46 (92)
Motivated to get on PrEP	46 (92)
Motivated to get tested for STIs[c]	46 (92)
Helped to understand whether PrEP would be a good fit	46 (92)
Helped to get in touch with the clinic staff (via chat messages)	44 (88)
Helped to get the latest information about HIV	43 (86)
Helped to understand mental health needs	42 (84)
Made access to PrEP much easier	40 (80)

[a]Agree and strongly agree (not included: strongly disagree, disagree, neither agree nor disagree).
[b]PrEP: preexposure prophylaxis.
[c]STI: sexually transmitted infection.

Table 6. Participants' rating of the level of difficulty of JomPrEP app features (N=50).

Activity ("How easy or hard was it to do the following tasks on the JomPrEP app?")	Participants, n (%)[a]
Ordering an HIVST[b] kit	49 (98)
Ordering PrEP[c] medication	47 (94)
Booking appointments (blood draw, consultation with doctor)	46 (92)
Reward system (earning and redeeming points)	44 (88)
Completing a mental health screener	43 (86)
Customizing profile page (eg, avatar, password, address, name)	43 (86)
Tracking the order status (ie, HIVST, PrEP)	42 (84)
Creating an account (onboarding process)	40 (80)
Reviewing laboratory test results	40 (80)
MoodTracker (track mood daily)	40 (80)
Chat with clinical or research staff	38 (76)
Find relevant information about HIV prevention	35 (70)
Online consultation with a doctor (e-consultation)	33 (66)
MedManager (receive medication reminders, track medication use)	25 (50)

[a]Very easy (not included: very difficult, difficult, neutral).
[b]HIVST: HIV self-testing.
[c]PrEP: preexposure prophylaxis.

Exit Interviews

In follow-up exit interviews (n=20, 40%), participants indicated a high level of acceptability for the content, interface, and features of JomPrEP. Participants found the app to be user friendly, easy to navigate, and with a good layout. Participants also appreciated the ability to earn reward points for using specific app features, facilitating user engagement and retention.

…it's straightforward, it's user friendly, it's easy to use…I think the critical part for me is actually the ease of use of the app.

Because of convenience. It's like having a mini-doctor. It's much easier for you to get tested, instead of going to a clinic and stuff.

The point and rewards system are very interesting and attractive.

Participants noted that ordering HIVST kits via the app was straightforward and that the multimedia instructions helped them use HIVST kits and interpret test results.

> *I tried self-test kits from other sources, comparing the experience using this and also ordering elsewhere, I think JomPrEP was very prompt, the delivery was, I think, the next day and self-test kit was easy to use.*

> *The feature I used most in the app is the one that allowed me to order a self-testing kit. It's super convenient because first, order placement is very easy to do and the second one because the delivery is very fast…And they have very clear instructions on how to do it and get the result. Instructions to do the screening at home are very clear. And to upload the result is easy as well.*

Furthermore, participants endorsed that the app helped them initiate PrEP use and maintain optimal adherence by facilitating a safe and stigma-free virtual platform to access PrEP services. Participants commented on the relevant information presented in the app, and many noted that they "didn't know anything about [PrEP] until [they] used the app."

> *There's a lot of information that makes me want to take PrEP more, because of the useful information and why I need to take the PrEP. This app is very like a one-stop center to take the PrEP…inside, you can book a consultation, view your result and you can directly order the PrEP. So, this helped me more easily to get the PrEP.*

> *I have been thinking of getting PrEP but didn't know much about it (process, the cost, etc). The app makes everything more transparent. And when you take it, there's a reminder every day so you can set up the clock, so you don't forget.*

> *When I first came here, I am not originally from here, I was trying to find PrEP, and I had a hard time finding it. When I used this app, it made things much easier. I don't need to worry if the clinic is judgmental.*

Participants also provided suggestions to improve the app and specific feedback on additional resources and features that they found interesting and helpful. For example, participants suggested that the app include an option to make an appointment with a mental health counselors and support groups and the ability to connect with other JomPrEP users through private messaging or discussion forums. Participants shared occasional issues with lagging app response time, difficulty setting up reminder notifications, and missing notifications. A few participants indicated that the test result feature of the app was a little challenging to use and required multiple clicks to view the results. A few participants also noted that some of the information in the app is repetitive or is not updated frequently. One participant recommended that the app allow the users to make the app more discreet (eg, the ability to change app icons).

Participants indicated that they would continue to use the JomPrEP app after the final version is released to the public.

> *Yes, definitely. I would use it because it's easier to put my appointment and view my lab results. I don't have to have it in a hardcopy form, easily accessible to my smartphone, and I could easily order my HIV self-testing kit as well.*

> *I will continue to use it. And I think the JomPrEP app is very useful for me in terms of ordering the PrEP and booking e-consultations.*

Discussion

Principal Findings

Using innovative tools, such as mHealth, in public health programming and the health care system can help bridge gaps in the adoption of needed health and prevention services, particularly among underserved populations [41-44]. In this study, we sought to investigate the usability and acceptability of JomPrEP, a clinic-integrated smartphone app, as an additional platform to promote routine HIV testing and PrEP uptake among MSM in Malaysia. Our findings demonstrated that Malaysian MSM will use a smartphone app to virtually access HIV prevention services and that such an app is acceptable to this at-risk group, as indicated by the participants' empiric use of the app.

Comparison With Prior Work

Prior studies have demonstrated several apps for HIV prevention and treatment efforts to be promising and cost-effective strategies to reach and engage stigmatized and hidden populations, such as MSM [25,45-47]. In Malaysia, the use of mobile technology over the past decades has grown markedly, particularly among MSM, with a mobile phone penetration rate of 97.5% and an internet penetration rate of 71.1% [18,19,48]. Importantly, our beta testing of JomPrEP revealed that an overwhelming majority of men used the app to receive HIV prevention services: ordering an HIVST kit (84%) and getting on PrEP (94%). Participants reported that they were satisfied and comfortable using JomPrEP and would recommend it to friends or colleagues. These findings indicate the potential utility of JomPrEP for Malaysian MSM to promote HIV prevention services.

One of the key innovations on JomPrEP includes incorporating on-demand features, such as home-based HIVST, e-consultations, and discrete door-to-door delivery, to provide a scalable model for remote HIV prevention services delivery in the LMIC setting. Although the users would still be required to visit a laboratory for clinical testing, this would not require them to be face-to-face with their clinicians. Moreover, the platform allows users to self-assess their HIV risk, consult online with clinicians from the participant clinics, and have their medication delivered to their preferred location, thus minimizing the need for in-person interactions with the clinician. This represents a significant and much-needed innovation over traditional clinic-based and O2O models of HIV service delivery to keep at-risk individuals wedded into the virtual clinical ecosystem and boost the uptake of clinical services [28,29]. This is particularly important in LMIC settings, such as Malaysia, as the virtual platform allows users to bypass barriers to care for marginalized populations and feel safer and less vulnerable to potential legal or social harm (eg, by reducing face-to-face interactions with providers).

Prior research has documented low user retention and a lack of sustained use after adoption as key challenges to the effectiveness of existing app-based interventions [49]. Obtaining high engagement and retention is necessary to maintain the integrity and long-term sustainability of effective mHealth interventions [50]. Strategies to integrate other features that do not include individual input, such as passive data collection using inputs from their smartphones or unobtrusive wearable devices, may strengthen the features of the app.

Results from our beta testing, however, revealed that MSM were actively engaged in the app and that retention was excellent through the beta testing. Although it is possible that the perfect retention rate could be because of the shorter follow-up time, it is likely that the incorporation of additional components, such as the ability to customize profiles, personalized messages, and gaming elements (ie, the ability to “level up,” earn and redeem points), may have allowed for enhanced user engagement. In recent years, the utility of gamification features (eg, challenges, tasks, rewards, badges, leaderboards) in nontraditional gaming contexts has increased significantly, thus providing opportunities for greater user engagement in mHealth interventions [51,52]. As confirmed in the exit interviews, the overall high engagement and usage of key features of JomPrEP suggest that an app-based intervention, such as JomPrEP, has a high degree of feasibility to ensure equitable access to HIV testing and PrEP services for MSM in Malaysia.

Although there was consensus on the usability and acceptability of JomPrEP, with no significant differences between different subgroups of Malaysian MSM, the app would benefit from continued refinement to address some of the shortcomings identified by men. For example, most participants who completed screening for depressive symptoms (88%) received referral letters to seek care offline (ie, outside the app). Given the focus of JomPrEP to offer holistic HIV prevention services within the online ecosystem, it would be important for the app to incorporate online consultation with mental health counselors and linkage to support groups via the app. Additionally, as part of the continued effort to ensure the safety and security of users, it would be important that JomPrEP incorporate added security measures, including 2-factor authentication and a discreet app icon (DAI). The availability of a DAI allows users to replace the default JomPrEP app log on their phone with another symbol (of their choice). This will help protect users when there is a possibility that someone may accidentally look at users’ phones and recognize that they have an app that might link them to the HIV or lesbian, gay, bisexual, transgender, queer, and others (LGBTQ+) community. A study conducted with MSM of Malaysia also highlighted the importance of privacy and confidentiality features in the mobile apps targeted for HIV prevention and treatment to minimize harm and safeguard users' privacy and confidentiality [53]. Furthermore, it is important that the app be available for users outside of Kuala Lumpur, the capital city, and be linked to both private and government clinics/hospitals. This will ensure widespread implementation of JomPrEP to scale-up HIV prevention services for MSM across Malaysia.

Limitations

The results of this study should be viewed in the context of the limitations. First, this pilot study included a small sample size and short-term follow-up (ie, 30-days) and used a single-arm design that is commonly used in beta testing; therefore, it was not powered or designed to evaluate efficacy. Second, our participants were subject to selection bias across several dimensions. For example, we recruited men using Facebook or a dating app (ie, Hornet) who may have been more comfortable using mobile apps than other men (ie, hidden MSM). Moreover, participants were already engaged, at least in part, due to their prior high levels of HIV testing. Participants were enrolled in the Greater Kuala Lumpur area only, potentially limiting the generalizability of the findings. Third, social desirability bias may have led participants to speak more positively about their app experience during the survey and exit interviews. This was in part observed by participants who responded favorably to app features, but our usability testing had not confirmed they used the specific feature. Finally, HIVST kits and PrEP services were free to the participants, which may have led to an overestimation of the actual uptake of HIV testing and PrEP services.

Future Directions

Further research is warranted to examine the implementation of JomPrEP in a more real-world setting. Regardless of these limitations, we believe that our findings carry important implications for efforts to improve the uptake of HIV testing and PrEP services among Malaysian MSM using an app-based intervention.

Conclusion

Overall, the JomPrEP app represents a feasible and acceptable tool for Malaysian MSM to access HIV prevention services. Importantly, it incorporates several on-demand features to support the remote delivery of HIV prevention services, thus representing a significant innovation on traditional clinic-based and O2O service delivery models [28,29]. The reported outcomes are promising and indicate the benefits of systematically implementing this platform to foster HIV prevention efforts in LMICs, such as Malaysia, where MSM are disenfranchised from existing prevention efforts [2-4]. A large-scale randomized controlled trial is warranted to establish the efficacy of JomPrEP among this at-risk group.

Acknowledgments

This work was supported by a career development and research award from the National Institute on Drug Abuse (K01 DA051346) and a research award from the Fogarty International Center (R21TW011665) for Roman Shrestha.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
JomPrEP app screenshots.
[PDF File (Adobe PDF File), 776 KB - mhealth_v11i1e44468_app1.pdf]

References

1. Ministry of Health Malaysia. Global AIDS Monitoring 2020: Malaysia HIV/AIDS Progress Report. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/Laporan_Global_AIDS_Monitoring_2020_new.pdf [accessed 2023-01-29]
2. Ministry of Health Malaysia. Global AIDS Monitoring: Country Progress Report - Malaysia. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/20211130_MYS_country_report_2021.pdf [accessed 2023-01-29]
3. Earnshaw VA, Jin H, Wickersham JA, Kamarulzaman A, John J, Lim SH, et al. Stigma toward men who have sex with men among future healthcare providers in Malaysia: would more interpersonal contact reduce prejudice? AIDS Behav 2016 Jan 1;20(1):98-106 [FREE Full text] [doi: 10.1007/s10461-015-1168-x] [Medline: 26324078]
4. Ni Z, Shrestha R, Earnshaw VA, Tee YC, Altice FL, Azwa I, et al. Exploring Malaysian physicians' intention to discriminate against gay, bisexual, and other men who have sex with men patients. LGBT Health 2022 Oct 14:Online ahead of print. [doi: 10.1089/lgbt.2021.0452] [Medline: 36251945]
5. United Nations Office on Drugs and Crime. World Drug Report 2016. URL: https://www.unodc.org/wdr2016/ [accessed 2023-01-30]
6. Chooi W, Mohd Zaharim N, Desrosiers A, Ahmad I, Yasin MAM, Syed Jaapar SZ, et al. Early initiation of amphetamine-type stimulants (ATS) use associated with lowered cognitive performance among individuals with co-occurring opioid and ATS use disorders in Malaysia. J Psychoactive Drugs 2017 Jun 29;49(4):326-332 [FREE Full text] [doi: 10.1080/02791072.2017.1342152] [Medline: 28661714]
7. Desrosiers A, Chooi W, Zaharim NM, Ahmad I, Mohd Yasin MA, Syed Jaapar SZ, et al. Emerging drug use trends in Kelantan, Malaysia. J Psychoactive Drugs 2016 May 25;48(3):218-226. [doi: 10.1080/02791072.2016.1185553] [Medline: 27224011]
8. Singh D, Chawarski MC, Schottenfeld R, Vicknasingam B. Substance abuse and the HIV situation in Malaysia. J Food Drug Anal 2013 Dec;21(4):S46-S51 [FREE Full text] [doi: 10.1016/j.jfda.2013.09.033] [Medline: 25278737]
9. Kanter J, Koh C, Razali K, Tai R, Izenberg J, Rajan L, et al. Risk behaviour and HIV prevalence among men who have sex with men in a multiethnic society: a venue-based study in Kuala Lumpur, Malaysia. Int J STD AIDS 2011 Jan 01;22(1):30-37. [doi: 10.1258/ijsa.2010.010277] [Medline: 21364064]
10. Lim SH, Akbar M, Wickersham JA, Kamarulzaman A, Altice FL. The management of methamphetamine use in sexual settings among men who have sex with men in Malaysia. Int J Drug Policy 2018 May;55:256-262 [FREE Full text] [doi: 10.1016/j.drugpo.2018.02.019] [Medline: 29605540]
11. Nemoto T, Xie H, Iwamoto M, Piroth K, Hazratzai M, Teh YK. HIV risk behaviors and psychological well-being among men who have sex with men (MSM) in Kuala Lumpur, Malaysia. AIDS Educ Prev 2022 Oct;34(5):427-439. [doi: 10.1521/aeap.2022.34.5.427] [Medline: 36181498]
12. Ng RX, Guadamuz TE, Akbar M, Kamarulzaman A, Lim SH. Association of co-occurring psychosocial health conditions and HIV infection among MSM in Malaysia: implication of a syndemic effect. Int J STD AIDS 2020 May 16;31(6):568-578. [doi: 10.1177/0956462420913444] [Medline: 32299293]
13. Sullivan PS, Lyons MS, Czarnogorski M, Branson BM. Routine screening for HIV infection in medical care settings: a decade of progress and next opportunities. Public Health Rep 2016 Jan 01;131(1_suppl):1-4. [doi: 10.1177/00333549161310s101]
14. Patel D, Williams WO, Heitgerd J, Taylor-Aidoo N, DiNenno EA. Estimating gains in HIV testing by expanding HIV screening at routine checkups. Am J Public Health 2021 Aug;111(8):1530-1533. [doi: 10.2105/ajph.2021.306321]
15. Chen Y, Fang C, Shih M, Lin K, Chang S, Wu Z, et al. Routine HIV testing and outcomes: a population-based cohort study in Taiwan. Am J Prev Med 2022 Mar;62(2):234-242 [FREE Full text] [doi: 10.1016/j.amepre.2021.07.010] [Medline: 34649736]
16. Estcourt C, Yeung A, Nandwani R, Goldberg D, Cullen B, Steedman N, et al. Population-level effectiveness of a national HIV preexposure prophylaxis programme in MSM. AIDS 2021 Mar 15;35(4):665-673 [FREE Full text] [doi: 10.1097/QAD.0000000000002790] [Medline: 33290298]

17. Shrestha R, Alias H, Wong LP, Altice FL, Lim SH. Using individual stated-preferences to optimize HIV self-testing service delivery among men who have sex with men (MSM) in Malaysia: results from a conjoint-based analysis. BMC Public Health 2020 Nov 25;20(1):1777 [FREE Full text] [doi: 10.1186/s12889-020-09832-w] [Medline: 33238941]
18. Shrestha R, Maviglia F, Altice FL, DiDomizio E, Khati A, Mistler C, et al. Mobile health technology use and the acceptability of an mHealth platform for HIV prevention among men who have sex with men in Malaysia: cross-sectional respondent-driven sampling survey. J Med Internet Res 2022 Jul 25;24(7):e36917 [FREE Full text] [doi: 10.2196/36917] [Medline: 35877172]
19. Shrestha R, Lim SH, Altice FL, Copenhaver M, Wickersham JA, Saifi R, et al. Use of smartphone to seek sexual health information online among Malaysian men who have sex with men (MSM): implications for mHealth intervention to increase HIV testing and reduce HIV risks. J Community Health 2020 Mar 2;45(1):10-19 [FREE Full text] [doi: 10.1007/s10900-019-00713-x] [Medline: 31375976]
20. Lim SH, Mburu G, Bourne A, Pang J, Wickersham JA, Wei CKT, et al. Willingness to use pre-exposure prophylaxis for HIV prevention among men who have sex with men in Malaysia: findings from an online survey. PLoS One 2017 Sep 13;12(9):e0182838 [FREE Full text] [doi: 10.1371/journal.pone.0182838] [Medline: 28902857]
21. Eger WH, Adaralegbe A, Khati A, Azwa I, Wickersham JA, Osborne S, et al. Exploring drivers of pre-exposure prophylaxis uptake among gay, bisexual, and other men who have sex with men in Malaysia. Int J STD AIDS 2022 Aug 30;33(9):821-828. [doi: 10.1177/09564624221106535] [Medline: 35772943]
22. Lim SH, Alias H, Kien JKW, Akbar M, Kamarulzaman A, Wong LP. A qualitative study of HIV "test-and-treat" experience among men who have sex with men in Malaysia. AIDS Educ Prev 2019 Jun;31(3):193-205. [doi: 10.1521/aeap.2019.31.3.193] [Medline: 31145000]
23. Sullivan PS, Hightow-Weidman L. Mobile apps for HIV prevention: how do they contribute to our epidemic response for adolescents and young adults? Mhealth 2021 Apr;7:36-36 [FREE Full text] [doi: 10.21037/mhealth-20-71] [Medline: 33898605]
24. Kawichai S, Songtaweesin WN, Wongharn P, Phanuphak N, Cressey TR, Moonwong J, et al. A mobile phone app to support adherence to daily HIV pre-exposure prophylaxis engagement among young men who have sex with men and transgender women aged 15 to 19 years in Thailand: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Apr 21;10(4):e25561 [FREE Full text] [doi: 10.2196/25561] [Medline: 35451976]
25. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
26. Goedel WC, Mitchell JW, Krebs P, Duncan DT. Willingness to use mobile phone apps for HIV prevention among men who have sex with men in London: web-based survey. JMIR Mhealth Uhealth 2017 Oct 11;5(10):e153 [FREE Full text] [doi: 10.2196/mhealth.8143] [Medline: 29021132]
27. Nelson K, Perry NS, Horvath KJ, Smith LR. A systematic review of mHealth interventions for HIV prevention and treatment among gay, bisexual, and other men who have sex with men. Transl Behav Med 2020 Oct 12;10(5):1211-1220 [FREE Full text] [doi: 10.1093/tbm/ibaa007] [Medline: 33044531]
28. Anand T, Nitpolprasert C, Phanuphak N. Online-to-offline models in HIV service delivery. Curr Opin HIV AIDS 2017;12(5):447-457. [doi: 10.1097/coh.0000000000000403]
29. Anand T, Nitpolprasert C, Trachunthong D, Kerr SJ, Janyam S, Linjongrat D, et al. A novel online-to-offline (O2O) model for pre-exposure prophylaxis and HIV testing scale up. J Int AIDS Soc 2017 Jan 01;20(1):21326. [doi: 10.7448/ias.20.1.21326]
30. Shrestha R, Wickersham JA, Khati A, Azwa I, Ni Z, Kamarulzaman A, et al. Clinic-integrated mobile health intervention ("JomPrEP" app) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: protocol for an intervention development and multiphase trial. JMIR Res Protoc 2022 Dec 21;11(12):e43318 [FREE Full text] [doi: 10.2196/43318] [Medline: 36542425]
31. Naeem F, Syed Y, Xiang S, Shokraneh F, Munshi T, Yang M, et al. Development, testing and reporting of mobile apps for psycho-social interventions: lessons from the pharmaceuticals. J Med Diagn Meth 2015;4(4):191. [doi: 10.4172/2168-9784.1000191]
32. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
33. Wuest J, Merritt-Gray M, Dubé N, Hodgins MJ, Malcolm J, Majerovich JA, et al. The process, outcomes, and challenges of feasibility studies conducted in partnership with stakeholders: a health intervention for women survivors of intimate partner violence. Res Nurs Health 2015 Mar 12;38(1):82-96 [FREE Full text] [doi: 10.1002/nur.21636] [Medline: 25594917]
34. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
35. World Health Organization. WHO Implementation Tool for Pre-exposure Prophylaxis (PrEP) of HIV Infection. URL: https://apps.who.int/iris/handle/10665/255890 [accessed 2023-01-30]
36. JomPrEP. URL: https://jomprep.com/# [accessed 2023-01-30]
37. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

38. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
39. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
40. Brooke J. SUS-a quick and dirty usability scale. Usability Eval Ind 1996;189(194):4-7. [doi: 10.1201/9781498710411-35]
41. Sharma S, Kumari B, Ali A, Yadav R, Sharma A, Sharma K, et al. Mobile technology: a tool for healthcare and a boon in pandemic. J Family Med Prim Care 2022;11(1):37. [doi: 10.4103/jfmpc.jfmpc_1114_21]
42. Agarwal S, Alhuwail D, John O. mHealth solutions in global health applications and implementation in low- and middle-income countries. In: Kickbusch I, Ganten D, Moeti M, editors. Handbook of Global Health. Berlin, Germany: Springer; 2021:1747-1774.
43. Fortuin J, Salie F, Abdullahi LH, Douglas TS. The impact of mHealth interventions on health systems: a systematic review protocol. Syst Rev 2016 Nov 25;5(1):200 [FREE Full text] [doi: 10.1186/s13643-016-0387-1] [Medline: 27884180]
44. Chén OY, Roberts B. Personalized health care and public health in the digital age. Front Digit Health 2021 Mar 30;3:595704 [FREE Full text] [doi: 10.3389/fdgth.2021.595704] [Medline: 34713084]
45. Horvath KJ, Lammert S, MacLehose RF, Danh T, Baker JV, Carrico AW. A pilot study of a mobile app to support HIV antiretroviral therapy adherence among men who have sex with men who use stimulants. AIDS Behav 2019 Nov 15;23(11):3184-3198 [FREE Full text] [doi: 10.1007/s10461-019-02597-3] [Medline: 31309348]
46. Daniels J, Lane T, Struthers H, Maleke K, Moges W, McIntyre J, et al. Assessing the feasibility of smartphone apps for HIV-care research with MSM and transgender individuals in Mpumalanga, South Africa. J Int Assoc Provid AIDS Care 2017 Aug 09;16(5):433-439 [FREE Full text] [doi: 10.1177/2325957417724207] [Medline: 28791915]
47. Zhang A, Reynolds NR, Farley JE, Wang X, Tan S, Yan J. Preferences for an HIV prevention mobile phone app: a qualitative study among men who have sex with men in China. BMC Public Health 2019 Mar 12;19(1):297 [FREE Full text] [doi: 10.1186/s12889-019-6617-4] [Medline: 30866880]
48. Malaysian Communications and Multimedia Commission. Internet Users Survey: 2017. URL: https://www.mcmc.gov.my/skmmgovmy/media/general/pdf/mcmc-internet-users-survey-2017.pdf [accessed 2023-01-30]
49. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
50. Bremer W, Sarker A. Recruitment and retention in mobile application-based intervention studies: a critical synopsis of challenges and opportunities. Inform Health Soc Care 2022 Jun 03:1-14. [doi: 10.1080/17538157.2022.2082297] [Medline: 35656732]
51. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
52. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Mar 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]
53. Khati A, Wickersham JA, Rosen AO, Luces JRB, Copenhaver N, Jeri-Wahrhaftig A, et al. Ethical issues in the use of smartphone apps for HIV prevention in Malaysia: focus group study with men who have sex with men. JMIR Form Res 2022 Dec 23;6(12):e42939 [FREE Full text] [doi: 10.2196/42939] [Medline: 36563046]

Abbreviations

CERiA: Centre of Excellence for Research in AIDS
DAI: discreet app icon
EHR: electronic health record
HIVST: HIV self-testing
JPP: JomPrEP points
LMIC: low- and middle-income country
mHealth: mobile health
MSM: men who have sex with men
O2O: online-to-offline
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection
SUS: System Usability Scale

Edited by L Buis; submitted 20.11.22; peer-reviewed by JH Huang, M Torres; comments to author 09.12.22; revised version received 15.12.22; accepted 25.01.23; published 16.02.23.

Please cite as:
Shrestha R, Altice FL, Khati A, Azwa I, Gautam K, Gupta S, Sullivan PS, Ni Z, Kamarulzaman A, Phiphatkunarnon P, Wickersham JA
Clinic-Integrated Smartphone App (JomPrEP) to Improve Uptake of HIV Testing and Pre-exposure Prophylaxis Among Men Who Have Sex With Men in Malaysia: Mixed Methods Evaluation of Usability and Acceptability
JMIR Mhealth Uhealth 2023;11:e44468
URL: https://mhealth.jmir.org/2023/1/e44468
doi:10.2196/44468
PMID:36795465

Original Paper

Acceptability and Utility of a Smartphone App to Support Adolescent Mental Health (BeMe): Program Evaluation Study

Judith J Prochaska[1], MPH, PhD; Yixin Wang[1], PhD; Molly A Bowdring[1], PhD; Amy Chieng[1], BA; Neha P Chaudhary[2], MD; Danielle E Ramo[2], PhD

[1]Stanford Prevention Research Center, Department of Medicine, Stanford University, Palo Alto, CA, United States

[2]BeMe Health, Miami, FL, United States

Corresponding Author:
Judith J Prochaska, MPH, PhD
Stanford Prevention Research Center
Department of Medicine
Stanford University
3180 Porter Drive Room A105
Palo Alto, CA, 94304-1212
United States
Phone: 1 650 724 3608
Fax: 1 650 725 6247
Email: jpro@stanford.edu

Abstract

Background: Adolescents face unprecedented mental health challenges, and technology has the opportunity to facilitate access and support digitally connected generations. The combination of digital tools and live human connection may hold particular promise for resonating with and flexibly supporting young people's mental health.

Objective: This study aimed to describe the BeMe app-based platform to support adolescents' mental health and well-being and to examine app engagement, usability, and satisfaction.

Methods: Adolescents in the United States, aged 13 to 20 years, were recruited via the web and enrolled between September 1 and October 31, 2022. App engagement, feature use, clinical functioning, and satisfaction with BeMe were examined for 30 days. BeMe provides content based on cognitive behavioral therapy, dialectical behavior therapy, motivational interviewing, and positive psychology; interactive activities; live text-based coaching; links to clinical services; and crisis support tools (digital and live).

Results: The average age of the sample (N=13,421) was 15.04 (SD 1.7) years, and 56.72% (7612/13,421) identified with she/her pronouns. For the subsample that completed the in-app assessments, the mean scores indicated concern for depression (8-item Patient Health Questionnaire mean 15.68/20, SD 5.9; n=239), anxiety (7-item Generalized Anxiety Disorder Questionnaire mean 13.37/17, SD 5.0; n=791), and poor well-being (World Health Organization–Five Well-being Index mean 30.15/100, SD 16.1; n=1923). Overall, the adolescents engaged with BeMe for an average of 2.38 (SD 2.7) days in 7.94 (SD 24.1) sessions and completed 11.26 (SD 19.8) activities. Most adolescents engaged with BeMe's content (12,270/13,421, 91.42%), mood ratings (13,094/13,421, 97.56%), and interactive skills (10,098/13,421, 75.24%), and almost one-fifth of the adolescents engaged with coaching (2539/13,421, 18.92%), clinical resources (2411/13,421, 17.96%), and crisis support resources (2499/13,421, 18.62%). Overall app engagement (total activities) was highest among female and gender-neutral adolescents compared with male adolescents (all *P*<.001) and was highest among younger adolescents (aged 13-14 years) compared with all other ages (all *P*<.001). Satisfaction ratings were generally high for content (eg, 158/176, 89.8% rated as helpful and 1044/1139, 91.66% improved coping self-efficacy), activities (5362/8468, 63.32% helpful and 4408/6072, 72.6% useful in coping with big feelings), and coaching (747/894, 83.6% helpful and 747/894, 83.6% improved coping self-efficacy). Engagement (total activities completed) predicted the likelihood of app satisfaction (*P*<.001).

Conclusions: Many adolescents downloaded the BeMe app and completed multiple sessions and activities. Engagement with BeMe was higher among female and younger adolescents. Ratings of BeMe's content, activities, and coaching were very positive for cognitive precursors aimed at reducing depression and anxiety and improving well-being. The findings will inform future app development to promote more sustained engagement, and future evaluations will assess the effects of BeMe on changes in mental health outcomes.

(JMIR Mhealth Uhealth 2023;11:e47183) doi:10.2196/47183

KEYWORDS

adolescents; mobile app; depression; anxiety; resilience; digital intervention; digital mental health; mobile phone

Introduction

Background

The state of adolescent mental health has been steadily declining, and the COVID-19 pandemic ushered in a new set of challenges. Across the 10 years preceding the pandemic, feelings of persistent sadness and hopelessness as well as suicidal thoughts and behaviors increased by approximately 40% among young people, according to the Centers for Disease Control and Prevention's Youth Risk Behavior Surveillance System [1,2]. In 2021, about 4 in 10 American high school students reported feeling persistently sad or hopeless over the last year, and 1 in 11 attempted suicide [3]. Since 2019, emergency department visits for mental health conditions among adolescents aged 13 to 17 years have increased annually [4]. A 2021 report from the Surgeon General called adolescent mental health a national crisis and made a series of recommendations to support youth, including empowering youth and their families to recognize, manage, and learn from difficult emotions; ensuring that every child has access to high-quality, affordable, and culturally competent mental health care; and promoting equitable access to technology that supports the well-being of children and youth [5].

Adolescence is the developmental period that begins with the onset of puberty and is defined by the American Academy of Pediatrics as spanning ages 11 to 21 years; it is a phase during which a young individual transitions from childhood to adulthood [6]. Throughout this developmental period, adolescents experience unique stressors, including social transitions, physical and emotional changes, and other potentially overwhelming life challenges, which may benefit from adolescent-targeted psychological guidance and support. Although there are a multitude of psychotherapeutic approaches that have a strong evidence base for enhancing adolescent mental health [7-9], many young people experience barriers in accessing traditional in-person therapy (eg, stigma, finances, transportation barriers, and inconvenient appointment times) [10-12], and half of the children diagnosed with a mental health condition in the United States do not receive treatment [13]. Therefore, alternative methods for care service delivery are required.

Digital health tools have a significant potential to increase adolescents' access to mental health support. Although early studies on web-based mental health services failed to demonstrate effectiveness in increasing help-seeking behavior among adolescents [14], the surge in mobile app development over the past decade has changed the landscape of digital mental health. Recent estimates suggest that there are >40,000 health and medical apps on leading app sites [15]. With 95% saturation of smartphone use among adolescents [16], mobile apps in particular may have the ability to provide desirable, accessible, and affordable support to adolescents when and where they are most likely to consume this support.

A recent systematic review indicated that most existing digital interventions (mobile and otherwise) are not evidence based, and there is inconclusive support for their effectiveness on mental health concerns other than anxiety and depression [17]. However, a separate, recent review and meta-analysis of 80 studies describing 83 mobile health interventions based on evidence-based principles found symptom improvement for a variety of psychological disorders, improved general well-being, and reduced distress [18]. Indeed, evidence-based digital services may offer the best opportunity to enhance outcomes among adolescents. For example, brief coping skills delivered digitally have been shown to improve mood and support young people in managing difficult moods without finding them unbearable, thereby preventing mental health challenges [19]. In addition, positive psychology skills focused on enhancing positive emotions and reducing negative emotions, when delivered via a smartphone app, can buffer older adolescents against loneliness and depression, improve sleep quality, and aid adjustment to college [20].

Mixed findings in the literature for adolescent mental health digital interventions may also be partly because of the combined assessment of bot coaching with human support [18]. The combination of digital tools and live human connection may hold particular promise for resonating with and flexibly supporting adolescent mental health. The addition of synchronous human support through skills-based coaching appears to improve adherence to and outcomes from digital mental health interventions as well as lower dropout among adolescents [17]. For example, live counselor contact was associated with improved clinical outcomes from a smoking cessation intervention delivered via Facebook to young adults [21]. Peer-to-peer counseling from paraprofessionals in a digital environment aided the delivery of evidence-based skills to undergraduate college students [22]. Tools that combine digital coping–skills delivery with live human connection for extra support and crisis management have the potential to help adolescents across a range of mental health challenges.

Despite the profound potential of digital health tools to address gaps in adolescent mental health care, initiation and continued app engagement remains suboptimal. Although some digital mental health tools have reported positive user engagement and adherence under certain conditions [17,23], many apps fail to retain users beyond initial registration [24], with evidence of poor user experience and negative perceptions of app usefulness among adolescents [25,26]. In recent years, there has been increased advocacy for the critical importance of involving people with lived experience in treatment development and, specifically, involving young people in the development of digital health products for adolescents [27,28]. Focus groups with adolescents have identified interest in strengths-based mobile health coaching and structured, supported web-based peer-to-peer interactions [29]. For digital mental health tools, adolescents have also identified interest in self-directed learning, multimedia (eg, audio and video components), and content diversity as opposed to focusing on a singular health issue [30]. Designing the look and functionality of mobile mental health tools according to adolescent preferences is likely to enhance reach and engagement.

With a purposeful focus on adolescent experience in its app design, BeMe Health, a digital mental health company, engages a Teen Advisory Board in partnership with adolescent clinical scientists and technology product and safety experts. The BeMe app was designed to support adolescent mental health by combining digital support through content and interactive care activities, live human connection through skills-based coaching with paraprofessionals, early identification and facilitation of clinical services as needed, and in-app digital crisis support tools linked to live crisis support as needed. Skills and coaching support in the BeMe app are based on interventions that have a strong and sound evidence base with adolescents, including cognitive behavioral therapy [31], dialectical behavior therapy [32], acceptance and commitment therapy [33], mindfulness-based self-compassion [34], positive psychology [35], and motivational interviewing [36].

Objectives

The goal of this first large-scale evaluation of the BeMe app was to assess acceptability and utility among adolescents throughout the United States. We examined adolescent use patterns over 30 days on the BeMe platform, satisfaction and perceived helpfulness across BeMe's features, and predictors of app engagement.

Methods

Participant Recruitment

Deidentified data were obtained within the BeMe app from adolescents aged 13 to 20 years between September 1 and October 31, 2022, when the beta version of the app included all basic features to characterize and support adolescent mental health. Participant use of the BeMe app was tracked for 30 days from the day of enrollment. Participants learned about BeMe from various channels, including unpaid posts on social media platforms, social media or other web-based platform advertisements (maximum daily budget was US $1000), or from organic channels such as word of mouth from other adolescents. Social media posts included adolescent-centric images and phrases that shared BeMe's focus on adolescent mental health and highlighted specific features (eg, coaching). Posts were linked to BeMe's website (BeMe [37]) that shared links to download the app in the iOS or Android app stores.

Upon downloading the app, adolescents registered on the app and were asked to read and accept BeMe's Terms of Service and Privacy Policy, written in adolescent-facing language and at a 5th-grade reading level. The terms of service (Multimedia Appendix 1) indicated that the data would be used to improve BeMe's services. The adolescents then completed a profile and were instructed to use the app freely. Assessments were embedded throughout the app experience. To avoid coercion that would affect app engagement and because all metrics used in this study were embedded in the intervention rather than requiring separate time to be spent on completion, adolescents were not given any compensation for using BeMe. Prior research with adolescents and parents has questioned whether the use of incentives for adolescents might lead to invalid results through misrepresentation [38-40].

Ethical Considerations

The project was deemed to be a program evaluation exempt from review by Stanford's Institutional Review Board. Program evaluations follow a systematic method for collecting and analyzing information with the intent of answering questions about the effectiveness and efficiency of a specific program, in this case a digital health program.

Intervention

Overview

The BeMe platform was designed to improve well-being and address the mental health needs of all teens across the specific need and acuity spectra. BeMe was designed to provide both preventive skills that promote resilience and thriving among all teens as well as provide interventions for common clinical symptoms such as depression and anxiety; support healthy habit formation and behavior change (eg, improve sleep and reduce or quit substance use); and support linkage to clinical services and crisis support for those teens who need it. BeMe accepts any teen on its platform (there are no exclusion criteria for enrollment other than age), and it functions as a primary, secondary, and tertiary prevention program depending on adolescents' needs at enrollment.

BeMe was designed by a combination of adolescent advisors and experts in the fields of behavioral science, adolescent clinical intervention, medicine, crisis support, mobile apps, and child- and adolescent-focused technology products. During the design phase, BeMe's Teen Advisory Board had 68 members across the United States, all of whom had lived experience with the topics addressed in the app and some with clinical symptoms of anxiety, depression, or other common mental health concerns. The adolescents informed the look, tone, and design of BeMe's overall app as well as its specific features. BeMe's app was open to users aged ≥13 years in accordance with data privacy and protection regulations. Caregiver consent was not required for enrollment in the BeMe app; however, it was required for an adolescent to engage in clinical services through the BeMe app. The beta version of the BeMe app contained the following 5 main features (refer to Multimedia Appendix 2 for samples): content, activities, coaching, clinical services, and crisis support.

Content

Content emphasized coping and resilience-building skills based on evidence-based strategies, including cognitive behavioral therapy [31], dialectical behavior therapy [32], acceptance and commitment therapy [33], mindfulness-based self-compassion [34], positive psychology [35], and motivational interviewing [36]. Content included resilience and coping skills for all adolescents, skills for coping with specific mental health conditions (eg, depression and anxiety), and skills specific to adolescent-centric identities and challenges (eg, dealing with a breakup, challenges with friends or caregivers, and school stress). Content was multimodal, including text screens, videos with and without music or voice-overs, carousels with text or images, and images with dynamic features. Content was tagged based on the primary focus of each piece (ie, validation, general psychoeducation, skill building, and inspiring joy), the evidence-based strategy it drew upon, and the theme (eg, friend fights) and was assigned a learning objective (eg, to improve communication). The tagging system enabled the labeling of specific content pieces from the overall content library so that adolescents could navigate to different types of content based on preferences and needs.

Activities

Activities provided interactive tools for practicing resilience and coping skills. Designed collaboratively by BeMe and its Teen Advisory Board, activities included: (1) mood ratings, (2) interactive skills, and a (3) a community-based skill. Mood ratings included both *simple* and more complex *interactive* versions. Simple mood ratings included a single item with 4 responses ranging from positive ("I'm great") to negative ("I'm really struggling"). Interactive ratings used the phone's camera and digital stickers to encourage teens to take a selfie and label their mood and shared these data back with adolescents in a separate journal section of the app. Interactive skills based on cognitive and behavioral coping skills, mindfulness practice, and positive psychology skills, including those described by Bruehlman-Senecal et al [20] and others. A community-based skill was designed to practice sending and receiving "good vibes" with other adolescents (to promote a sense of community without actual social features).

Coaching

Coaching was delivered by BeMe coaches live in the app via text messaging. The coaches were graduate- and undergraduate-level paraprofessionals trained in a program developed by BeMe's clinical leadership team. Training included a combination of didactic training, role playing, and supervisor observation before working on the platform as well as ongoing weekly individual and group supervision and live supervision during all shifts. Adolescents could message at any time of the day or night with a preset message or an open-ended message of their choice. During the study period, BeMe coaches responded to adolescents for 14 hours per day. All adolescents could message an unlimited number of times on any topic of their choice during the study period. After a coaching session, adolescents received a feedback form with session ratings. Coaches tagged conversations with any number of topics designed by BeMe's clinical leadership, or as "other."

Clinical Services

Adolescents could connect to a licensed therapist through the BeMe app. Successful linkage to treatment required registration, verified parental consent, and scheduling via telephone or text messages.

Crisis Support

Adolescents could self-navigate or be directed by a coach to crisis support services, including 3 live options: a crisis hotline staffed 24/7, the Crisis Text Line, or the Trevor Project Crisis Support services. Adolescents could also complete a digital safety plan based on evidence-based suicide prevention support tools for adolescents [41].

The beta version was not designed to be used in any specific pattern or length. Adolescents were given free access to the app and could navigate through any or all features as they chose.

Measures

Sample Characteristics

Upon enrollment in the BeMe app, adolescents reported their age, preferred pronouns, interests from a list codeveloped with BeMe's Teen Advisory Board, and "onboarding topics" adolescents indicated that they would like to explore on the BeMe app. Adolescents could select any number of interests and topics.

Clinical Functioning

At any point in their BeMe journey, adolescents could opt to complete assessments of their clinical functioning regarding anxiety, depression, stress, and overall well-being. The tools were selected based on their evidence of use with adolescents. There was no particular time in the engagement process in which the assessments were administered, and participants self-selected to complete any, all, or none of the assessments and could repeat the assessments. The first instance of a completed assessment was analyzed in this study to characterize adolescent functioning. With reference to the past 2 weeks, adolescents completed the 7-item Generalized Anxiety Disorder Questionnaire [42] with scores ranging from 0 to 21; scores from 5 to 9 indicated mild anxiety, scores from 10 to 14 indicated moderate anxiety, scores from 15 to 19 indicated moderately severe anxiety, scores >15 indicated moderately severe anxiety, and scores >20 indicated severe anxiety. Adolescents could also complete the 8-item Patient Health Questionnaire (PHQ-8) adolescent version [43]. Again, with reference to the past 2 weeks, PHQ-8 total scores ranged from 0 to 24; scores from 5 to 9 indicated mild depression, 10 to 14 indicated moderate depression, 15 to 19 indicated moderately severe depression, and >19 indicated severe depression. The 4-item Perceived Stress Scale [44] assessed participants' self-reported stress level over the past month on a scale ranging from 0 ("never") to 4 ("very often"; total score range 0-16). The World Health Organization–Five Well-being Index (WHO-5), widely used with adolescents [45], was used to assess overall well-being during the past 2 weeks. The WHO-5 scores range from 0 to 25 and are multiplied by 4 to yield a well-being score between 0 and 100. Scores ≤50 indicate poor well-being, and scores <28 indicate depression, including in adolescent samples [46].

Engagement

Engagement was measured at the individual user level and the app feature level. Heterogeneity in engagement metrics has been reported in previous studies on mental health apps [47]. In this study, we used standard engagement metrics to report on at least minimal use (ie, days with any engagement on the app), number of times on the app (ie, number of unique sessions), and number of unique features used (ie, activities) [48]. Engagement was also calculated at the app feature level as the number of times each feature was used in a month among those who engaged in that feature. Content engagement was defined as the completion of the entirety of a piece of content (eg, scrolling through all screens of a carousel or watching an entire video). Engaged content was summarized by an intervention skillset coded as acceptance and commitment therapy, behavioral activation, cognitive and behavioral therapy, dialectical behavioral therapy, emotion regulation and distress tolerance skills, interpersonal effectiveness skills, mindfulness practice, motivational interviewing techniques, positive psychology skills, and trauma-informed care skills. Activity engagement was summarized for drawing practice, mindfulness skills, distress tolerance or emotion regulation skills with the phone's camera, or movement-based skills with the phone's camera. Interactive mood rating feature moods were tallied for each of the 21 moods across all activities completed. The mood list was developed by the study's last authors (NC and DR) in collaboration with BeMe's Teen Advisory Board. The coaching topics were coded by the BeMe coaches at the end of each session. Topic data were only available for approximately half of the study period (October 7, 2022, to November 30, 2022), so the sample of coaching sessions available for analysis was not the full sample of completed sessions during this time. The number of times clinical supports and safety plan resources (digital safety plan and crisis support services) were accessed was tallied. For the adolescents who completed a digital safety plan, we computed the average number of endorsed reasons for living, crisis warning signs, ways to make their environment safe, and coping skills.

Satisfaction and Helpfulness

Satisfaction and helpfulness were measured at the feature level. After select pieces of content created by BeMe, adolescents were asked to rate the content on perceived helpfulness and confidence that they would use the skill outside of the BeMe app (self-efficacy). After select pieces of content, adolescents rated whether they felt hopeful (hope), learned something about themselves (self-identity), the content was helpful to their self-esteem (self-esteem), felt less alone (social connection), or felt relaxed and grounded (relaxation; all yes or no). Activities were rated similarly based on perceived helpfulness and utility in coping with a big feeling. The percentage of polls with yes responses was tallied for each response type. Coaching sessions were rated on a 6-point scale (0-5 stars), perceived helpfulness, and self-efficacy (yes or no).

Analyses

Sample Characteristics and Use Patterns

Descriptive statistics were used to characterize adolescent demographics for the full sample, clinical functioning among the subset that self-selected to complete the assessments, and overall app engagement patterns.

Relationship Between Individual Characteristics and Engagement

Logistic regression was used to evaluate predictors of overall app engagement (ie, total activities accessed). The dependent variable was the number of activities completed in 30 days, coded as high or low based on the sample median (7 activities). The independent variables (IVs) were pronouns, age, and overall well-being (WHO-5 total score). *P* values and CIs determine whether the association between overall app engagement and each term in the model is statistically significant. All IVs were included in the model at once because we lacked an underlying theory to guide model selection. This approach was adopted to prevent bias toward small *P* values and large parameter estimates [49].

Satisfaction With BeMe's Content and Features

Engagement with content and activities was evaluated at the feature level, and the proportion of "yes" responses was computed for each content or activity type. Coaching satisfaction was evaluated among the proportion of adolescents who engaged with coaching at least once. The patterns of engagement, topics, and ratings were computed for this subsample.

Relationship Between Individual Characteristics and Satisfaction

A logistic regression model was used to evaluate the predictors of satisfaction and helpfulness. The dependent variable was the presence of at least 1 "yes" response on a postactivity survey (eg, after a piece of content or activity). IVs were pronouns, age, the WHO-5 total score, and total activities completed dichotomized, consistent with the strategy described in Relationship Between Individual Characteristics and Engagement. *P* values and CIs determine whether the association between satisfaction and helpfulness and each term in the model was statistically significant.

Power Estimation

Planned analyses included at least 2 regression models (1 predicting engagement and 1 predicting satisfaction), with up to 6 IVs, including age, preferred pronouns, depression, anxiety, overall well-being, and perceived stress. A multivariate regression analysis with 6 IVs testing a partial R^2 of 0.1 in each IV with an error probability of 0.05 and 95% power requires a sample size of at least 195. A sample size of up to 2000 was deemed sufficient to detect a small effect size for each IV in the 2 models. Relatively low sample sizes in depression, anxiety, and well-being measures compared with other IVs yielded their exclusion from the final models; thus, we expect the final sample size of >13,000 to be adequately powered to report both the final regression models and a series of frequency and proportion results shared in this study.

Results

Sample Characteristics and Use Patterns

In the 2-month enrollment period, 13,421 adolescents were enrolled in the BeMe app. The characteristics of the sample are presented in Table 1.

Table 1. Sample characteristics (N=13,421).

Characteristics	Values
Age (years)	
Mean (SD)	15.04 (1.7)
Median (IQR)	15 (14-16)
13-14, n (%)	5977 (44.53)
15-16, n (%)	4537 (33.81)
17-18, n (%)	2425 (18.07)
19-20, n (%)	482 (3.59)
Pronouns, n (%)	
She/her	7612 (56.72)
He/him	1468 (10.94)
They/them	1391 (10.36)
Other/no response	2950 (21.98)
Interests, n (%)	
Music	10,966 (81.71)
Art	7821 (58.27)
Food	7231 (53.88)
Animals	7102 (52.92)
Beauty	6721 (50.08)
Fashion	6193 (46.14)
Reading	5941 (44.27)
Nature	5750 (42.84)
Photography	5395 (40.20)
Writing	5383 (40.11)
Gaming	5362 (39.95)
LGBTQIA[a]	5331 (39.72)
Travel	4874 (36.32)
Dance	4408 (32.84)
Anime	4157 (30.97)
Sports	3858 (28.75)
Science	2547 (18.98)
Entrepreneurship	1327 (9.89)
Climate	1131 (8.43)
Auto	768 (5.72)
Goals for using BeMe, n (%)	
Boosting happiness	10,049 (74.88)
Building relationships	9000 (67.06)
Dealing with stressors	9337 (69.57)
Discovering identity	6757 (50.35)
Finding ways to cope	8025 (59.79)
Living mindfully	5706 (42.52)
Managing mood	9314 (69.40)
Navigating life transition	4923 (36.68)

Characteristics	Values
Depression symptoms (PHQ[b] score; n=238)	
Mean (SD)	15.69 (5.91)
Median (IQR)	17 (13-20)
No depression (scores: 0-4), n (%)	18 (7.56)
Mild (score: 5-9), n (%)	17 (7.14)
Moderate (score: 10-14), n (%)	51 (21.43)
Moderately severe (score: 15-19), n (%)	85 (35.71)
Severe (score: 20+), n (%)	67 (28.15)
Anxiety symptoms (GAD[c] score; n=791)	
Mean (SD)	13.37 (5.01)
Median (IQR)	14 (10-17)
None to normal (score: 0-4), n (%)	52 (6.57)
Mild (score: 5-9), n (%)	125 (15.80)
Moderate (score: 10-14), n (%)	259 (32.74)
Severe (score: 15-21), n (%)	355 (44.88)
Overall well-being (WHO-5[d] score; n=1322)	
Mean (SD)	30.15 (16.06)
Median (IQR)	28 (20-40)
≤50, n (%)	1172 (88.65)
≤28, n (%)	747 (56.51)
Perceived stress (PSS[e] score; n=638)	
Mean (SD)	10.62 (2.56)
Median (IQR)	11 (9-12)
≥6, n (%)	623 (97.65)

[a]LGBTQIA: lesbian, gay, bisexual, transgender, queer, intersex, asexual, and similar minority.

[b]PHQ: Patient Health Questionnaire.

[c]GAD: Generalized Anxiety Disorder Questionnaire.

[d]WHO-5: WHO-5 Well-being Index.

[e]PSS: Perceived Stress Scale.

The average age of the adolescents was 15.04 (SD 1.7) years, and they more often identified with female pronouns (7612/13,421, 56.72%) than male pronouns (1468/13,421, 10.94%), gender-neutral pronouns (1391/13,421, 10.36%), or other or decline to answer (2950/13,421, 21.98%). Onboarding topics selected by a majority of adolescents were boosting happiness (10,049/13,421, 74.88%), dealing with stressors (9337/13,421, 69.57%), managing mood (9314/13,421, 69.40%), building relationships (9000/13,421, 67.06%), and generally finding ways to cope (8025/13,421, 59.79%); 50.35% (6757/13,421) of the adolescents selected discovering your identity, and although less common, over a third selected living mindfully (5706/13,421, 42.52%) and navigating a life transition (4923/13,421, 36.68%). The most common interests endorsed by the adolescents were music (10,966/13,421, 81.71%), art (7821/13,421, 58.27%), food (7231/13,421, 53.88%), and animals (7102/13,421, 52.92%).

Engagement

Table 2 displays overall and feature-level engagement in the BeMe app over 30 days. On average, the adolescents engaged with BeMe for >2 days (mean 2.38, SD 2.72; median 1, IQR 1-3; range 1-30), in 8 sessions (mean 7.94, SD 24.14; median 3, IQR 2-7; range 1-1750), and completed >11 activities (mean 11.26, SD 19.81; median 7, IQR 4-12; range 1-776). Adolescents were most likely to engage in simple mood ratings (13,094/13,421, 97.56%), content (12,270/13,421, 91.42%), and interactive skills (10,098/13,421, 75.24%). Almost one-fifth of the adolescents completed a coaching chat session (2539/13,421, 18.92%), explored clinical resources (2499/13,421, 18.62%), and engaged in safety planning (1129/13,421, 8.41%).

Table 2. Overall and specific feature engagement (N=13,421).

Variable	Values, mean (SD)	Values, median (IQR)	Values, n (%)
Overall app engagement[a]			
Days engaged (range 1-30)	2.38 (2.72)	1 (1-3)	13,421 (100)
1	1 (0)	1 (1-1)	6962 (51.87)
2-3	2.35 (0.48)	2 (2-3)	4233 (31.54)
≥4	6.77 (4.35)	5 (4-8)	2226 (16.59)
App sessions (range 1-1750)	7.94 (24.14)	3 (2-7)	13,421 (100)
Activities (range 1-776)	11.26 (19.81)	7 (4-12)	13,421 (100)
Specific feature engagement[a]			
Content views (range 1-505)	4.59 (11.36)	2 (1-5)	12,270 (91.42)
Simple mood rating (range 1-128)	2.61 (3.49)	2 (1-3)	13,094 (97.56)
Interactive skills (range 1-75)	2.58 (3.09)	2 (1-3)	10,098 (75.24)
Interactive mood rating (range 1-80)	1.82 (2.57)	1 (1-2)	6677 (49.75)
Community skills (range 1-73)	1.75 (2.72)	1 (1-2)	1616 (12.04)
Coach session (range 1-20)	1.47 (1.27)	1 (1-1)	2539 (18.92)
Clinical service resource clicks (range 1-19)	1.40 (0.97)	1 (1-1)	2411 (17.96)
Crisis resource views (range 1-25)	1.52 (1.25)	1 (1-2)	2499 (18.62)
Safety plan completed (yes or no)	N/A[b]	N/A	1129 (8.41)

[a]Teens were not given guidance as to how to navigate through BeMe's features, and there were no limits on the number of activities and interactions with each feature (eg, content and mood rating) possible during the 30-day study period.

[b]N/A: not applicable.

Overall, 91.53% (12,285/13,421) of adolescents accessed 66,345 pieces of content (content views: mean 4.59, SD 11.35, median 2, IQR 1-5; range 1-505). Adolescents engaged most with content grounded in dialectical behavior therapy emotion regulation and distress tolerance skills (16,920/66,345, 25.5% of all content viewed), positive psychology skills (14,346/66,345, 21.62% of all content viewed), cognitive and behavioral therapy coping skills (16,920/66,345, 25.5%), and mindfulness-based self-compassion skills (8650/66,345, 13.04%). Lower engagement was found with content grounded in acceptance and commitment therapy skills (5669/66,345, 8.54%), dialectical behavioral therapy interpersonal effectiveness and social skills (5810/66,345, 8.76%), motivational interviewing skills (1693/66,345, 2.55%), behavioral activation (1091/66,345, 1.64%), and skills grounded in trauma-informed care (774/66,345, 1.17%).

Among those adolescents who engaged in each type of interactive activity, engagements were, on average, as follows: 2.61 (SD 3.49; median 2, IQR 1-3) simple mood ratings, 2.58 (SD 3.09; median 2, IQR 1-3) interactive skills, 1.82 (SD 2.57; median 1, IQR 1-2) interactive mood ratings, and 1.75 (SD 2.72; median 1, IQR 1-2) community-based skills. Considering specifically interactive skills, adolescents primarily engaged with those that used the phone's camera to complete a distress tolerance skill (24,944/44,698, 55.81%) or a pleasurable activity (11,963/44,698, 26.76%). Considering interactive mood ratings (n=19,800 rating completed), intraclass correlation in a 2-way mixed effects model using a consistency definition showed a significant correlation among mood ratings across individuals (intraclass correlation 0.427, CI 0.42-0.44; P<.001), so frequency was tallied for all mood ratings. The most frequently identified moods were low arousal moods, including lonely (2315/19,800, 11.69%), low (2280/19,800, 11.52%), relaxed (2160/19,800, 10.91%), hurt (2074/19,800, 10.47%), bored (1947/19,800, 9.83%), and depressed (1755/19,800, 8.86%). The least endorsed moods were those with higher arousal, including pumped (368/19,800, 1.86%), furious (390/19,800, 1.97%), excited (413/19,800, 2.09%), and energetic (447/19,800, 2.26%). Other endorsed moods included chill (1681/19,800, 8.49%), insecure (1460/19,800, 7.37%), anxious (1223/19,800, 6.18%), grateful (1129/19,800, 5.7%), happy (1147/19,800, 5.79%), content (1086/19,800, 5.48%), motivated (1031/19,800, 5.21%), cheerful (957/19,800, 4.83%), rejected (943/19,800, 4.76%), frustrated (805/19,800, 4.07%), and pissed (521/19,800, 2.63%).

Among the adolescents who engaged with coaching, they averaged 1.47 (SD 1.27; median 1, IQR 1-1) sessions. The topics were coded for 1817 coaching sessions during the study period. Anxiety was the most common topic raised in coaching (649/1817, 35.72%) and was more than twice as frequent as the next most popular topic. The relationship concerns of different types were also common: romantic (304/1817, 16.73%), friend (264/1817, 14.53%), and family (187/1817, 10.29%). School-related topics (eg, workload and managing school stress) were raised in 14.42% (262/1817) of the coaching conversations. Adolescents also sought coaching on self-esteem (160/1817, 8.81%); depression and sadness (153/1817, 8.42%); body image (80/1817, 4.4%); issues related to lesbian, gay, bisexual, transgender, queer, intersex, asexual, and similar minority (LGBTQIA) identity (73/1817, 4.02%); anger (61/1817, 3.36%); suicidality (47/1817, 2.59%); bullying (46/1817, 2.53%); and self-harm (37/1817, 2.04%). The less common topics were related to learning new skills (35/1817, 1.93%), grief (29/1817, 1.60%), loneliness (19/1817, 1.05%), neurodiversity (15/1817, 0.83%), abuse or neglect (11/1817, 0.61%), harassment or assault (9/1817, 0.5%), racism or discrimination owing to ethnicity (2/1817, 0.11%), and substance use (2/1817, 0.11%).

The adolescents who sought out safety and clinical resources selected on average 1.52 (SD 1.25; median 1, IQR 12) safety resources (crisis supports and digital safety plan) and 1.35 (SD 0.80; median 1, IQR 1-1) clinical resources. Almost half (1129/2499, 45.18%) of those who engaged with the safety resources completed a digital safety plan, averaging 5.07 (SD 2.62) out of 12 reasons for living, 7.05 (SD 3.64) out of 17 crisis warning signs, 2.25 (SD 1.96) out of 8 ways to make their environments safe, and 3.69 (SD 2.28) out of 9 coping skills to use when in crisis.

Predictors of BeMe App Engagement

To examine the predictors of overall app engagement, with high versus low engagement (<7 vs >7 activities completed over 30 days) as the outcome, we first ran a logistic model including pronouns, age, and overall well-being (WHO-5 total score). The WHO-5 measure was not predictive in the model (*P*=.61) and had a lot of missing data (completed by 1322 adolescents); therefore, we reran the model with only pronouns and age as predictors (Figure 1; Table 3). Predictors were pronouns in 4 categories (she/her, he/him, they/them, and other/no response), with he/him as the reference group, and age in 4 categories, with 13-14 as the reference group (N=13,421). The findings indicated that compared with male-identified adolescents, female-identified adolescents were 32% more likely to be in the high engagement group (odds ratio [OR] 1.324, 95% CI 1.184-1.482; *P*<.001), gender-neutral adolescents were 39% more likely to be in the high engagement group (OR 1.39, 95% CI 1.199-1.61; *P*=.03), and those who did not indicate pronouns or chose other were 18% more likely to be in the high engagement group (OR 1.179, 95% CI 1.04-1.337; *P*=.01). In addition, adolescents in the youngest age group were more likely to be in the highest engagement group than the 3 older age groups (all *P*<.001).

Figure 1. Odds ratios and CIs for a logistic regression model testing the effects of pronouns and age on app engagement (total activities over 30 days; high vs low engagement).

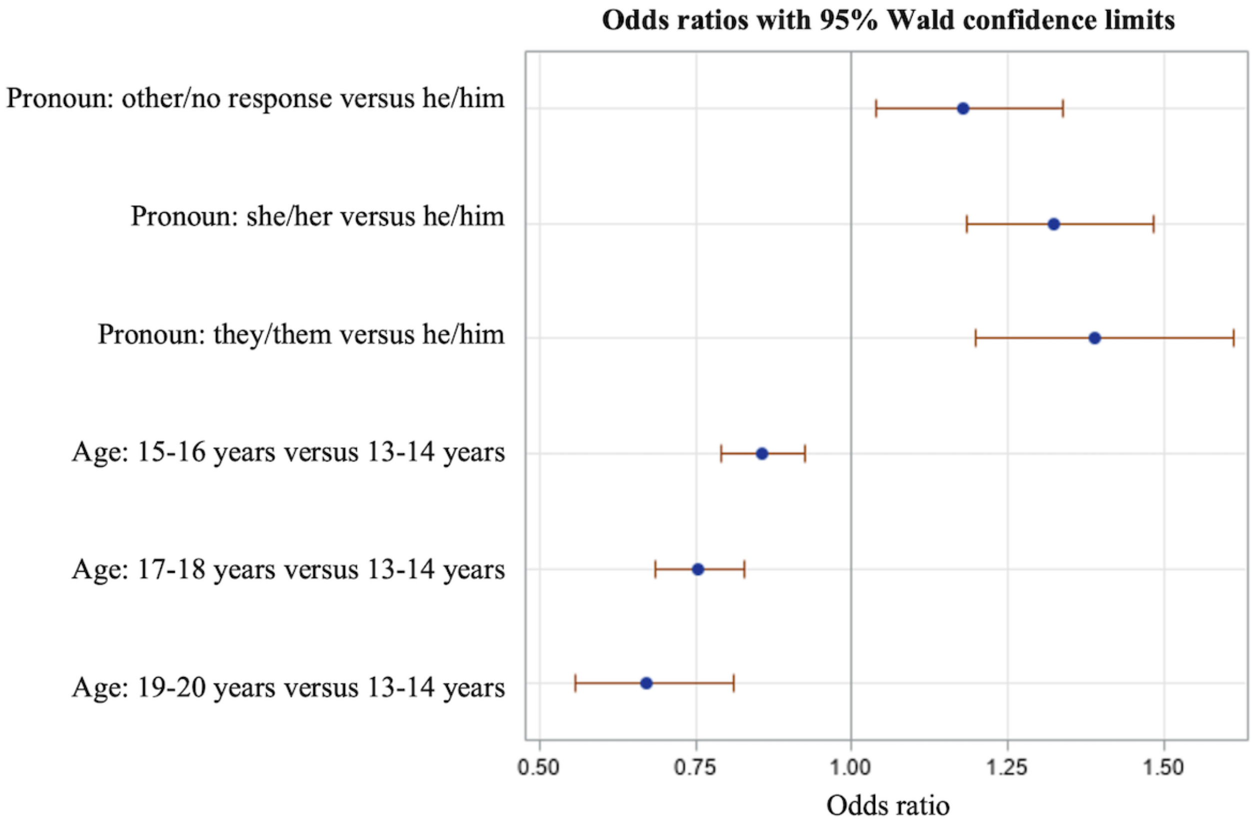


Table 3. Logistic regression model predicting high versus low app engagement (total activities completed; N=13,421).

Effect	Odds ratio (95% CI)	*P* value
Pronoun		
Other/no response vs he/him	1.179 (1.04-1.337)	.01
She/her vs he/him	1.324 (1.184-1.482)	<.001
They/them vs he/him	1.390 (1.199-1.61)	<.001
Age (years)		
15-16 vs 13-14	0.855 (0.792-0.924)	<.001
17-18 vs 13-14	0.754 (0.686-0.829)	<.001
19-20 vs 13-14	0.673 (0.558-0.811)	<.001

"Superusers," defined as engaging with BeMe at least 2 SDs above the mean of the full sample (engaged >7 out of 30 days), were more likely to be female (351/557, 63% vs 7265/12,880, 56.4%), less likely to be male (38/557, 6.8% vs 1432/12,880, 11.12%; $\chi^2_5$16.1; *P*=.007), and had lower PHQ-8 scores (15.04 vs 15.84; $F_{2,237}$=5.08; *P*=.03) compared with nonsuperusers. There were no significant differences in age, 7-item Generalized Anxiety Disorder Questionnaire, WHO-5, or 4-item Perceived Stress Scale scores according to the BeMe superuser status. BeMe superusers were more likely to engage with live coaching (306/557, 54.9% vs 2237/12,880, 17.37%; χ^2_5=491.1, *P*<.001), clinical resources (267/557, 47.9% vs 2149/12,880, 16.68%; χ^2_5=353.6, *P*<.001), and crisis support resources (322/557, 57.8% vs 2180/12,880, 16.93%; χ^2_5=588.9, *P*<.001).

Satisfaction and Helpfulness of Content and Features

Quick pulse surveys captured thumbs up or down ratings of different content and features of BeMe that were accessed by the participants. The surveys were launched at different times throughout the study period, resulting in varying sample sizes. Among the 8468 surveys, 5362 (63.32%) rated the BeMe activities as helpful for boosting mood. Among the 6072 surveys, 4408 (72.6%) rated the BeMe activities as helpful for coping with a big feeling. Completed by fewer participants owing to strategic placement of pulse surveys to prevent survey exhaustion and preserve the user experience, 91.66% (1044/1139) of adolescents planned to use a skill they learned from BeMe when coping with a stressor; 89.8% (158/176) rated the BeMe content as helpful; 86.2% (493/572) learned something about themselves from BeMe; 85.8% (235/274) felt relaxed after using BeMe; 84.3% (369/438) gained help with self-esteem; 83.4% (586/703) felt more hopeful; and 82% (46/56) felt less alone.

Satisfaction and Helpfulness of Coaching

On average, the adolescents rated the BeMe coaching sessions 4.2 out of 5 stars (SD 1.2; n=893). Over four-fifths of the responses indicated that the sessions were helpful (747/894, 83.6%) and provided content that an adolescent would use (747/894, 83.6%; Table 4).

Table 4. Satisfaction and impact of digital activities and coaching.

Feature type	Values, N	Yes, n (%)
Content		
Helpfulness	176	158 (89.77)
Self-efficacy	1139	1044 (91.66)
Hope	703	586 (83.36)
Self-identity	572	493 (86.19)
Self-esteem	438	369 (84.25)
Social connection	56	46 (82.14)
Relaxation	274	235 (85.77)
Interactive activities		
Helpfulness	8468	5362 (63.32)
Useful in coping with a big feeling	6072	4408 (72.6)
Coaching		
Helpfulness	894	747 (83.56)
Self-efficacy	894	747 (83.56)
Overall rating		
5 stars	893	532 (59.57)
4 stars	893	163 (18.25)
3 stars	893	84 (9.41)
2 stars	893	34 (3.81)
1 star	893	78 (8.73)

Predictors of Satisfaction and Helpfulness

Logistic regression models were run to examine the predictors of BeMe satisfaction and the perceived helpfulness of content and interactive activities. Modeled outcomes were yes versus no or no response on surveys of satisfaction and perceived helpfulness. An initial model with inclusion of WHO-5 Well-being Index scores was nonsignificant, and we show a second model without these scores (Table 5). The tested predictor variables were gender identity in 4 categories (she/her as the reference group), age in 4 categories (13-14 years as the reference group), and high versus low total activities (<7 vs >7 activities). The only predictor that was significant was total activities, with those completing ≥7 activities being almost 3.9 times as likely to have indicated that they were satisfied with and found help from BeMe content or an activity compared with those with low total activities (P<.001; Figure 2; Table 5).

Figure 2. Odds ratios and CIs for a logistic regression model testing the effects of pronouns, age, and app engagement (high vs low activities over 30 days) on satisfaction (positive endorsement of at least 1 survey after a piece of content or interactive activity).

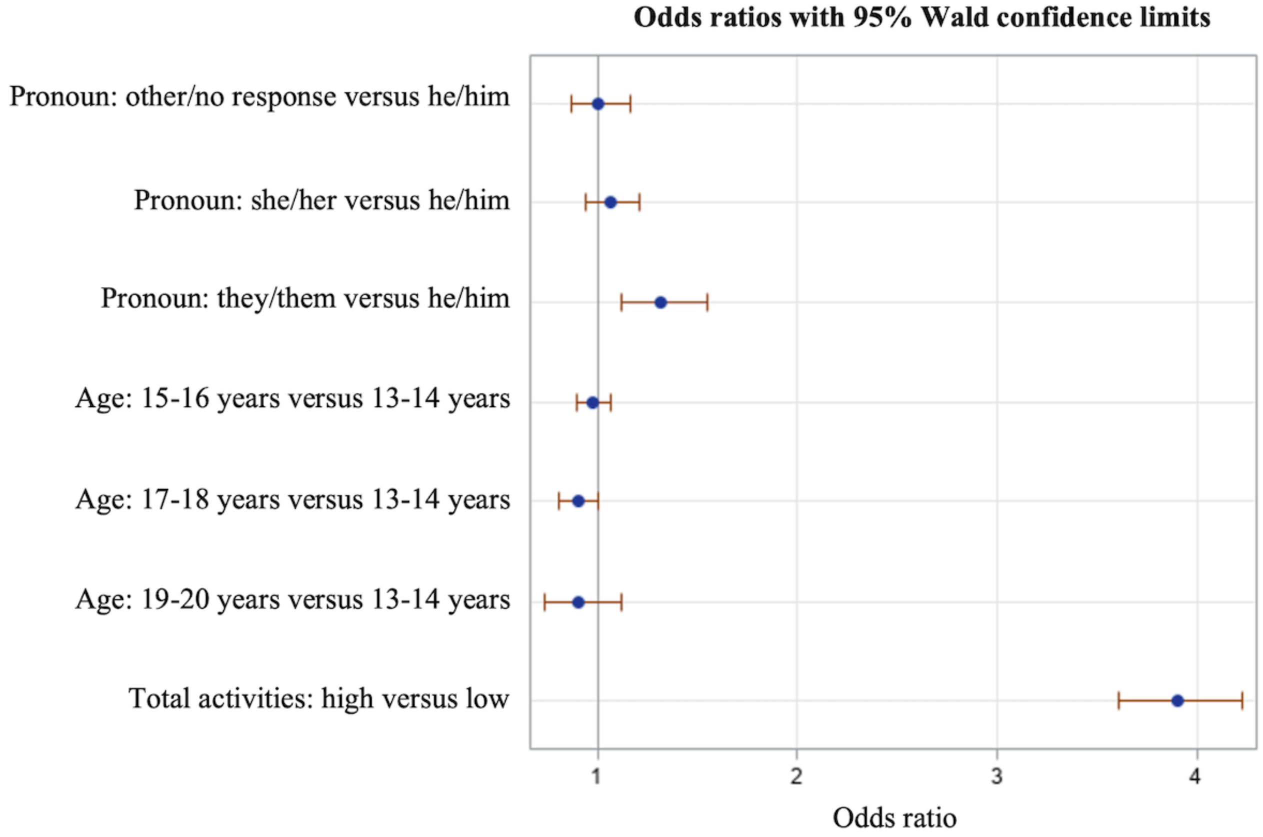


Table 5. Logistic regression model predicting at least 1 positive indicator of satisfaction or impact (N=13,421).

Effect	Odds ratio (95% CI)	*P* value
Pronoun		
Other/no response vs he/him	1.004 (0.87-1.159)	.96
She/her vs he/him	1.066 (0.938-1.212)	.33
They/them vs he/him	1.315 (1.116-1.549)	.001
Age (years)		
15-16 vs 13-14	0.977 (0.896-1.065)	.59
17-18 vs 13-14	0.900 (0.809-1.002)	.06
19-20 vs 13-14	0.905 (0.730-1.120)	.36
Total activities: high vs low	3.903 (3.607-4.224)	<.001

Discussion

Principal Findings

In the first large-scale evaluation of the acceptability and utility of the BeMe app, designed by and for adolescents, >6 times as many adolescents were enrolled than initially sought for the study. In 8 weeks, >13,000 adolescents accessed the app without the enticement of any financial incentive for using the app or completing within-app surveys. The adolescents reported interest in mood boosting, stress management, help with relationships, and discovering their identity, and most participants (11,005/13,421, 82%) identified multiple areas of interest. The great response suggests the demand for multifocal and multifunctional digital mental health programs among adolescents.

Characteristic of other mental health and digital program evaluations, BeMe attracted more adolescents who identified as female than male and notably attracted a sizable proportion of adolescents identifying as gender neutral (1391/13,421, 10.36%) or another gender (671/13,421, 5%). According to the literature, adolescent girls tend to be more willing to seek mental health treatment than boys [50] and are more likely to be attracted to wellness and mental health apps [51]. BeMe also attracted and engaged younger adolescents. Early adolescence is a major transitional period physically (puberty), environmentally (eg, transition to high school), and socially, when many new stressors arise [52] and when support may be particularly welcomed.

Across a 30-day evaluation of the BeMe app, participants' use averaged >2 days, in over 8 sessions, engaging in >11 activities, with a great deal of variation, indicating diversity in adolescents' needs and preferences. Notably, there was no in-app guidance toward a specific pathway; therefore, adolescents were able to gravitate toward the features that were most relevant to them. This was a conscious decision among the app's designers and the BeMe Teen Advisory Board who worked together to create an experience that would focus on adolescent choice rather than explicit guidance. A better understanding of adolescent-guided pathways will allow for enhanced personalization over time and the ability to guide adolescents toward journeys that support their presentation (eg, clinical functioning) and their needs and preferences.

The observed level of engagement is satisfactory for a beta version of an app that does not ask or require participants to move in any particular pathway or journey through the app experience. For example, prior work has similarly found that users engage with certain app features just a couple of times, with 1 study observing that only 15.6% of unique app resources were engaged with at least once [53,54]. Another prior study of engagement with a mental health app that did not involve user prompts found that participants engaged in a total of 6 sessions on average, similar to the total of 7.94 sessions observed in this study [55].

Although the BeMe platform was initially designed to encourage exploratory self-navigation by teens (rather than a forced pathway), its developers were cautious about incorporating features in the initial design that could encourage excessive daily use, thereby avoiding overuse of the intervention. Social media platforms that also include content and live human connections (eg, messaging) have saturated the adolescent market and are being used in a way that prevents interactions with other areas of life for some teens (eg, school, in-person friend connections, and family relationships). This has likely contributed to a time-sink problem, whereby some teens report they are on social media "almost constantly" (35% in a 2022 Pew survey [16]) and that it would be difficult to give up social media (36% in the same survey). In contrast, the beta version of the BeMe app avoided some of the common features of social media platforms that promote constant use, such as algorithms designed to promote long-term content engagement [56] and social features that encourage social feedback and almost constant connections. There was an expectation that adolescents would use BeMe at variable rates, which was indeed found in this study. The findings of this study indicate that encouraging further engagement (eg, guidance toward assessment completion, personalized content pathways, and purposeful promotion of coaching sessions) could drive even further toward meaningful engagement.

A preliminary analysis of BeMe's "superusers" confirmed active use among female adolescents and engagement in the live coaching, clinical service linkage, and crisis support features among those who used BeMe for >7 days in a month. A more detailed analysis of these users could help identify features they returned to use within the app to inform further app development, recruitment, onboarding, and impact, in line with previous work with adults [57]. A future examination of app use as it relates to changes in clinical symptoms and cognitive mediators of symptom incidence and course (eg, hope and self-esteem) will further inform targeted app development. Superusers also showed slightly lower PHQ-8 scores than those who used the app less often. Given that assessments could be completed at any time in a user's journey, this could indicate either a positive impact of BeMe use on depression symptoms among superusers or less depression among superusers at the times they started to use the app. A more targeted journey for adolescents with depression will ensure that the experience supports them.

App engagement is greater among younger than older adolescents and among female and gender-neutral adolescents than male adolescents. BeMe's constellation of services and support (eg, coaching) may be of greater interest to female adolescents and those who seek support for managing mood in particular. More work is needed to understand how best to reach and support male adolescents. Engagement patterns also indicate that there is potential for expanding app design, content, activities, and coaching features that resonate with older adolescents. BeMe could benefit from features that support developmental milestones of the older adolescent years (eg, graduating from high school, transitioning toward increasing independence in living situation, social interactions, work, and college).

Anxiety was the most common topic raised in coaching sessions, followed by relationships of multiple types, school, self-esteem, and depression. Developmental milestones of adolescence have, in many ways, been interrupted for this pandemic-stalled generation. Coaching may support adolescents to get back on track. The preponderance of sessions coded with the "other" topic suggests that the initial topic list should be expanded. Post hoc examination of sessions coded as other suggests that additional topics could include exploring adolescent identity or selfhood, attention or motivation, and social or conversational skills. Notably, the adolescents who engaged in the coaching sessions rated them highly, on average, 4.2 out of 5 stars. High ratings contribute to the promise of the paraprofessional model in delivering coping skill support digitally to adolescents and young adults [22].

Although a minority of the sample (1 in 10 or fewer) completed the well-being, anxiety, and depression assessments, a majority of those who self-selected to complete these measures indicated clinically concerning levels of distress, which supports the multiple offerings of BeMe, including individual counseling, safety planning, and clinical resources. BeMe can also be useful as a platform for tracking changes in anxiety and depressive symptoms over time using clinically validated measures with adolescents. Mood data from digital platforms such as BeMe can help to add nuance to the understanding of adolescent emotional experiences. Mood rating data from BeMe's interactive mood rating feature indicate that most moods expressed while using the platform are lower arousal moods (eg, low, chill, and bored), and least endorsed moods are higher arousal moods (eg, excited and furious). This is consistent with prior literature showing that adolescents experience low-intensity emotions (both positive and negative) more frequently than high-intensity emotions, regardless of age [58]. A future investigation could develop a more nuanced understanding of the relationships among mood rating completion, adolescent individual characteristics (eg, age and gender), and clinical functioning (eg, depression and anxiety).

The single-item response measures throughout the intervention platform provide a unique way for adolescents to share feedback and provide a pulse on satisfaction and helpfulness. The responses were very positive for cognitive precursors aimed at reducing depression and anxiety such as hopefulness, perceived helpfulness, and the ability to cope with a big feeling. The likelihood of implementing these skills is high and warrants further study.

About one-fifth of the adolescents on the BeMe platform accessed each of the live supports, including coaching sessions, clinical resources, and crisis support tools. Digital support through a platform such as BeMe may be particularly appealing to a generation that exhibits decreased stigma regarding mental health challenges, while also displaying a growing mistrust of conventional mental health assistance in comparison with previous generations [59]. Similar rates of connection to coaching and crisis support as to traditional mental health services suggest that a platform like BeMe may be able to address challenges some adolescents have with accessing traditional clinical interventions (eg, need for caregiver consent and stigma) and foster greater trust and satisfaction with accessible alternatives. Crisis support use and high completion (1129/2502, 45.12%) of digital safety plans among those who access crisis resources highlight the utility of these features for a generation that is struggling with high and increasing suicidality [60], emergency department visits for suicidal behavior [61], and suicide completions [62]. Future investigations should examine the pathways between access to clinical and crisis support services through a digital platform such as BeMe, linkage to such services, and subsequent clinical functioning.

The multiple modalities that make up BeMe's platform were designed to support across the acuity spectrum and offer options for engagement in multiple live features that are adolescent led (eg, coaching and crisis support). The beta version allowed unlimited engagement with 24/7 live crisis support and live coaching for 14 hours per day, but this pattern might need adaption as adolescents' use of the live service and its impact on clinical functioning are assessed over time. The dissemination of a multimodal platform such as BeMe is best supported by organizations that support health at the population level (eg, health plans) or invest in the well-being of teens and their families (eg, employers of teens' parents). BeMe's Teen Advisory Board and other adolescents can inform ways to best iterate upon the delivery of live features (eg, coaching) for operational efficiency and dissemination at scale.

Limitations

A study limitation is the incomplete data on measures of interest, given the self-driven nature of the BeMe app. Assessment completion was opportunistic, in that adolescents had to find the mood and well-being assessments or access content to trigger a pulse survey. The variety of information collected on participant characteristics was also limited to minimize burden in this initial evaluation of app use.

Conclusions

Overall, BeMe is a promising example of the combination of digital and live interactive support in practice. This study contributes to the growing body of work demonstrating the utility and impact of the combination of digital and live human support in digital interventions [21,63], and demonstrated acceptability and utility in a large sample of adolescents. The positive responses to BeMe's content, activities, and coaching service are encouraging. The next stage of evaluation will measure the changes in clinical functioning and well-being associated with BeMe use over time.

Acknowledgments

This study was funded by the California Health Care Foundation (grant G-32490).

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JJP, DER, and NPC designed the study and acquired funding. DER and NPC chose and designed the study assessments and the intervention content. DER and NPC led the data accrual and management. AC led institutional approvals. JJP, AC, and YW had access to the deidentified study data downloaded from Box. YW performed data analyses. JJP, MAB, and AC drafted the manuscript and incorporated feedback from the coauthors.

Conflicts of Interest

DER and NPC are employees of BeMe Health. All other authors declare no conflicts of interest related to this study.

Multimedia Appendix 1
BeMe terms of service.

[PDF File (Adobe PDF File), 124 KB - mhealth_v11i1e47183_app1.pdf]

Multimedia Appendix 2
BeMe samples.
[PNG File , 3948 KB - mhealth_v11i1e47183_app2.png]

References

1. Ivey-Stephenson AZ, Demissie Z, Crosby AE, Stone DM, Gaylor E, Wilkins N, et al. Suicidal ideation and behaviors among high school students - youth risk behavior survey, United States, 2019. MMWR Suppl 2020 Aug 21;69(1):47-55 [FREE Full text] [doi: 10.15585/mmwr.su6901a6] [Medline: 32817610]
2. Youth risk behavior survey data summary and trends report: 2009-2019. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBSDataSummaryTrendsReport2019-508.pdf [accessed 2023-03-15]
3. Coronavirus disease 2019. Centers for Disease Control and Prevention. 2022 Sep 09. URL: https://www.cdc.gov/media/releases/2022/p0331-youth-mental-health-covid-19.html [accessed 2022-12-01]
4. Radhakrishnan L, Leeb RT, Bitsko RH, Carey K, Gates A, Holland KM, et al. Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic - United States, January 2019-January 2022. MMWR Morb Mortal Wkly Rep 2022 Feb 25;71(8):319-324 [FREE Full text] [doi: 10.15585/mmwr.mm7108e2] [Medline: 35202358]
5. Murthy VH. Protecting youth mental health: the U.S. surgeon general's advisory. Department of Health and Human Services. 2021. URL: https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf [accessed 2023-03-15]
6. Hagan JFJ, Shaw JS, Duncan PM. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents. 4th edition. Washington, DC: American Academy of Pediatrics; 2017.
7. Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. J Clin Child Adolesc Psychol 2016;45(2):91-113. [doi: 10.1080/15374416.2015.1046177] [Medline: 26087438]
8. Hogue A, Henderson CE, Ozechowski TJ, Robbins MS. Evidence base on outpatient behavioral treatments for adolescent substance use: updates and recommendations 2007-2013. J Clin Child Adolesc Psychol 2014;43(5):695-720. [doi: 10.1080/15374416.2014.915550] [Medline: 24926870]
9. Weersing VR, Jeffreys M, Do MT, Schwartz KT, Bolano C. Evidence base update of psychosocial treatments for child and adolescent depression. J Clin Child Adolesc Psychol 2017 Jan;46(1):11-43 [FREE Full text] [doi: 10.1080/15374416.2016.1220310] [Medline: 27870579]
10. Lu W, Todhunter-Reid A, Mitsdarffer ML, Muñoz-Laboy M, Yoon AS, Xu L. Barriers and facilitators for mental health service use among racial/ethnic minority adolescents: a systematic review of literature. Front Public Health 2021 Mar;9:641605 [FREE Full text] [doi: 10.3389/fpubh.2021.641605] [Medline: 33763401]
11. Planey AM, Smith SM, Moore S, Walker TD. Barriers and facilitators to mental health help-seeking among African American youth and their families: a systematic review study. Child Youth Serv Rev 2019 Jun;101:190-200 [FREE Full text] [doi: 10.1016/j.childyouth.2019.04.001]
12. Brown A, Rice SM, Rickwood DJ, Parker AG. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia Pac Psychiatry 2016 Mar;8(1):3-22. [doi: 10.1111/appy.12199] [Medline: 26238088]
13. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019 Apr 01;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
14. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
15. Ceci L. Healthcare apps available Apple app store 2022. Statista. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2023-02-09]
16. Vogels EA, Gelles-Watnick R, Massarat N. Teens, social media and technology 2022. Pew Research Center. 2022 Aug 10. URL: https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/ [accessed 2022-09-26]
17. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
18. Conley CS, Raposa EB, Bartolotta K, Broner SE, Hareli M, Forbes N, et al. The impact of mobile technology-delivered interventions on youth well-being: systematic review and 3-level meta-analysis. JMIR Ment Health 2022 Jul 29;9(7):e34254 [FREE Full text] [doi: 10.2196/34254] [Medline: 35904845]
19. Rizvi SL, Finkelstein J, Wacha-Montes A, Yeager AL, Ruork AK, Yin Q, et al. Randomized clinical trial of a brief, scalable intervention for mental health sequelae in college students during the COVID-19 pandemic. Behav Res Ther 2022 Feb;149:104015 [FREE Full text] [doi: 10.1016/j.brat.2021.104015] [Medline: 34958980]

20. Bruehlman-Senecal E, Hook CJ, Pfeifer JH, FitzGerald C, Davis B, Delucchi KL, et al. Smartphone app to address loneliness among college students: pilot randomized controlled trial. JMIR Ment Health 2020 Oct 20;7(10):e21496 [FREE Full text] [doi: 10.2196/21496] [Medline: 33079071]
21. Thrul J, Meacham MC, Tice C, Kelly O, Ramo DE. Live counselor contact in a Facebook intervention predicts smoking cessation outcomes. Psychol Addict Behav 2020 Mar;34(2):360-369 [FREE Full text] [doi: 10.1037/adb0000541] [Medline: 31829664]
22. Rosenberg BM, Kodish T, Cohen ZD, Gong-Guy E, Craske MG. A novel peer-to-peer coaching program to support digital mental health: design and implementation. JMIR Ment Health 2022 Jan 26;9(1):e32430 [FREE Full text] [doi: 10.2196/32430] [Medline: 35080504]
23. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009 Feb 23;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
24. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: a national survey. Media Commun 2016;4(3):13-23 [FREE Full text] [doi: 10.17645/mac.v4i3.515]
25. Ribanszki R, Saez Fonseca JA, Barnby JM, Jano K, Osmani F, Almasi S, et al. Preferences for digital smartphone mental health apps among adolescents: qualitative interview study. JMIR Form Res 2021 Aug 27;5(8):e14004 [FREE Full text] [doi: 10.2196/14004] [Medline: 34128814]
26. Byron P. 'Apps are cool but generally pretty pointless': LGBTIQ+ young people's mental health app ambivalence. Media International Australia 2019 May;171(1):51-65 [FREE Full text] [doi: 10.1177/1329878x19844034]
27. Dickinson D, Burrus B, White L, Winston L, Burns O. Beyond the PDF: developing interactive digital health education products-designed with teens, for teens. In: Proceedings of the APHA 2021 Annual Meeting and Expo. 2021 Presented at: APHA' 21; October 24-27, 2021; Denver, CO URL: https://apha.confex.com/apha/2021/meetingapp.cgi/Paper/510365
28. Schleider JL. The fundamental need for lived experience perspectives in developing and evaluating psychotherapies. J Consult Clin Psychol 2023 Mar;91(3):119-121. [doi: 10.1037/ccp0000798] [Medline: 36795434]
29. Aschbrenner KA, Naslund JA, Tomlinson EF, Kinney A, Pratt SI, Brunette MF. Adolescents' use of digital technologies and preferences for mobile health coaching in public mental health settings. Front Public Health 2019 Jul 02;7:178 [FREE Full text] [doi: 10.3389/fpubh.2019.00178] [Medline: 31312629]
30. Turk L, Montenegro M, Krastev S. How to design mental health tools for gen Z. The Decision Lab. URL: https://thedecisionlab.com/insights/health/how-to-design-mental-health-tools-for-gen-z [accessed 2023-02-16]
31. Oud M, de Winter L, Vermeulen-Smit E, Bodden D, Nauta M, Stone L, et al. Effectiveness of CBT for children and adolescents with depression: a systematic review and meta-regression analysis. Eur Psychiatry 2019 Apr;57:33-45. [doi: 10.1016/j.eurpsy.2018.12.008] [Medline: 30658278]
32. Kothgassner OD, Goreis A, Robinson K, Huscsava MM, Schmahl C, Plener PL. Efficacy of dialectical behavior therapy for adolescent self-harm and suicidal ideation: a systematic review and meta-analysis. Psychol Med 2021 May;51(7):1057-1067 [FREE Full text] [doi: 10.1017/S0033291721001355] [Medline: 33875025]
33. Coyne LW, McHugh L, Martinez ER. Acceptance and commitment therapy (ACT): advances and applications with children, adolescents, and families. Child Adolesc Psychiatr Clin N Am 2011 Apr;20(2):379-399. [doi: 10.1016/j.chc.2011.01.010] [Medline: 21440862]
34. Bluth K, Blanton PW. Mindfulness and self-compassion: exploring pathways to adolescent emotional well-being. J Child Fam Stud 2014 Oct 01;23(7):1298-1309 [FREE Full text] [doi: 10.1007/s10826-013-9830-2] [Medline: 25328369]
35. Baños RM, Etchemendy E, Mira A, Riva G, Gaggioli A, Botella C. Online positive interventions to promote well-being and resilience in the adolescent population: a narrative review. Front Psychiatry 2017 Jan 30;8:10 [FREE Full text] [doi: 10.3389/fpsyt.2017.00010] [Medline: 28194117]
36. Naar-King S. Motivational interviewing in adolescent treatment. Can J Psychiatry 2011 Nov;56(11):651-657. [doi: 10.1177/070674371105601103] [Medline: 22114919]
37. Home page. BeMeHealth. URL: https://beme.com/ [accessed 2023-08-18]
38. Afkinich JL, Blachman-Demner DR. Providing incentives to youth participants in research: a literature review. J Empir Res Hum Res Ethics 2020 Jul 12;15(3):202-215. [doi: 10.1177/1556264619892707] [Medline: 31829071]
39. Fisher CB. Adolescent and parent perspectives on ethical issues in youth drug use and suicide survey research. Ethics Behav 2003;13(4):303-332. [doi: 10.1207/S15327019EB1304_1] [Medline: 15000093]
40. Grady C. Payment of clinical research subjects. J Clin Invest 2005 Jul;115(7):1681-1687 [FREE Full text] [doi: 10.1172/JCI25694] [Medline: 16007244]
41. Stanley B, Brown G, Brent DA, Wells K, Poling K, Curry J, et al. Cognitive-behavioral therapy for suicide prevention (CBT-SP): treatment model, feasibility, and acceptability. J Am Acad Child Adolesc Psychiatry 2009 Oct;48(10):1005-1013 [FREE Full text] [doi: 10.1097/CHI.0b013e3181b5dbfe] [Medline: 19730273]
42. Mossman SA, Luft MJ, Schroeder HK, Varney ST, Fleck DE, Barzman DH, et al. The generalized anxiety disorder 7-item scale in adolescents with generalized anxiety disorder: signal detection and validation. Ann Clin Psychiatry 2017 Nov;29(4):227-34A [FREE Full text] [Medline: 29069107]

43. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health 2002 Mar;30(3):196-204. [doi: 10.1016/s1054-139x(01)00333-0] [Medline: 11869927]
44. Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the short form Perceived Stress Scale (PSS-4). J Health Psychol 2013 Dec;18(12):1617-1628. [doi: 10.1177/1359105313508346] [Medline: 24155195]
45. Blom EH, Bech P, Högberg G, Larsson JO, Serlachius E. Screening for depressed mood in an adolescent psychiatric context by brief self-assessment scales--testing psychometric validity of WHO-5 and BDI-6 indices by latent trait analyses. Health Qual Life Outcomes 2012 Dec 11;10(1):149 [FREE Full text] [doi: 10.1186/1477-7525-10-149] [Medline: 23227908]
46. de Wit M, Pouwer F, Gemke RJ, Delemarre-van de Waal HA, Snoek FJ. Validation of the WHO-5 Well-Being Index in adolescents with type 1 diabetes. Diabetes Care 2007 Aug;30(8):2003-2006. [doi: 10.2337/dc07-0447] [Medline: 17475940]
47. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
48. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
49. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer; 2001.
50. Chandra A, Minkovitz CS. Stigma starts early: gender differences in teen willingness to use mental health services. J Adolesc Health 2006 Jun;38(6):754.e1-754.e8. [doi: 10.1016/j.jadohealth.2005.08.011] [Medline: 16730608]
51. Antezana G, Venning A, Smith D, Bidargaddi N. Do young men and women differ in well-being apps usage? Findings from a randomised trial. Health Informatics J 2022 Jan;28(1):14604582211064825 [FREE Full text] [doi: 10.1177/14604582211064825] [Medline: 35128952]
52. Bonnie RJ, Backes EP. The Promise of Adolescence: Realizing Opportunity for All Youth. Washington, DC: National Academies Press; 2019.
53. Shi J, Lo B, Wong HW, Hollenberg E, Sanches M, Abi-Jaoudé A, et al. Assessing the usability and user engagement of thought spot - a digital mental health help-seeking solution for transition-aged youth. Internet Interv 2021 Apr 26;24:100386 [FREE Full text] [doi: 10.1016/j.invent.2021.100386] [Medline: 33936952]
54. Nardi W, Roy A, Dunsiger S, Brewer J. Analyzing the impact of mobile app engagement on mental health outcomes: secondary analysis of the unwinding anxiety program. J Med Internet Res 2022 Aug 15;24(8):e33696 [FREE Full text] [doi: 10.2196/33696] [Medline: 35969440]
55. Silva Almodovar A, Surve S, Axon DR, Cooper D, Nahata MC. Self-directed engagement with a mobile app (Sinasprite) and its effects on confidence in coping skills, depression, and anxiety: retrospective longitudinal study. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e64 [FREE Full text] [doi: 10.2196/mhealth.9612] [Medline: 29549066]
56. McClucskey M. How addictive social media algorithms could finally face a reckoning in 2022. Time. 2022 Jan 04. URL: https://time.com/6127981/addictive-algorithms-2022-facebook-instagram/ [accessed 2023-03-15]
57. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
58. Larson R, Lampman-Petraitis C. Daily emotional states as reported by children and adolescents. Child Dev 1989 Oct;60(5):1250-1260 [FREE Full text] [doi: 10.2307/1130798]
59. Addressing Gen Z mental health challenges. McKinsey. 2022 Jan 14. URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/addressing-the-unprecedented-behavioral-health-challenges-facing-generation-z [accessed 2022-09-09]
60. Teen girls experiencing increased sadness and violence. Centers for Disease Control and Prevention. 2023 Feb 13. URL: https://www.cdc.gov/media/releases/2023/p0213-yrbs.html [accessed 2023-02-24]
61. Yard E, Radhakrishnan L, Ballesteros MF, Sheppard M, Gates A, Stein Z, et al. Emergency department visits for suspected suicide attempts among persons aged 12-25 years before and during the COVID-19 pandemic - United States, January 2019-May 2021. MMWR Morb Mortal Wkly Rep 2021 Jun 18;70(24):888-894. [doi: 10.15585/mmwr.mm7024e1] [Medline: 34138833]
62. Curtin SC, Ahmad MF, Ahmad FB. Provisional numbers and rates of suicide by month and demographic characteristics: United States, 2021. Centers for Disease Control and Prevention. 2022 Sep. URL: https://stacks.cdc.gov/view/cdc/120830#:~:text=The%20provisional%20age%2Dadjusted%20suicide,higher%20in%20all%20other%20months [accessed 2023-03-15]
63. Kunkle S, Yip M, Hunt J, Ξ W, Udall D, Arean P, et al. Association between care utilization and anxiety outcomes in an on-demand mental health system: retrospective observational study. JMIR Form Res 2021 Jan 26;5(1):e24662 [FREE Full text] [doi: 10.2196/24662] [Medline: 33496679]

Abbreviations

IV: independent variable
LGBTQIA: lesbian, gay, bisexual, transgender, queer, intersex, asexual, and similar minority
OR: odds ratio
PHQ-8: 8-item Patient Health Questionnaire
WHO-5: World Health Organization–Five Well-being Index

Edited by L Buis; submitted 10.03.23; peer-reviewed by A Knapp; comments to author 12.04.23; revised version received 26.04.23; accepted 03.08.23; published 28.08.23.

Please cite as:
Prochaska JJ, Wang Y, Bowdring MA, Chieng A, Chaudhary NP, Ramo DE
Acceptability and Utility of a Smartphone App to Support Adolescent Mental Health (BeMe): Program Evaluation Study
JMIR Mhealth Uhealth 2023;11:e47183
URL: https://mhealth.jmir.org/2023/1/e47183
doi:10.2196/47183
PMID:37639293

Review

Design Guidelines of Mobile Apps for Older Adults: Systematic Review and Thematic Analysis

Miguel Gomez-Hernandez[1,2], MSc; Xavier Ferre[1,3], PhD; Cristian Moral[1], PhD; Elena Villalba-Mora[1,4], PhD

[1]Center for Biomedical Technology, Universidad Politécnica de Madrid, Pozuelo de Alarcón, Spain

[2]Emerging Technologies Research Lab, Department of Human-Centred Computing, Monash University, Melbourne, Australia

[3]School of Software Engineering, Tongji University, Shanghai, China

[4]Biomedical Research Networking Center in Bioengineering Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain

Corresponding Author:
Xavier Ferre, PhD
Center for Biomedical Technology
Universidad Politécnica de Madrid
Campus Montegancedo
Pozuelo de Alarcón, 28223
Spain
Phone: 34 91067925
Email: xavier.ferre@upm.es

Abstract

Background: Mobile apps are fundamental tools in today's society for practical and social endeavors. However, these technologies are often not usable for older users. Given the increased use of mobile apps by this group of users and the impact that certain services may have on their quality of life, such as mobile health, personal finance, or online administrative procedures, a clear set of guidelines for mobile app designers is needed. Existing recommendations for older adults focus on investigations with certain groups of older adults or have not been extracted from experimental results.

Objective: In this research work, we systematically reviewed the scientific literature that provided recommendations for the design of mobile apps based on usability testing with older adults and organized such recommendations into a meaningful set of design guidelines.

Methods: We conducted a systematic literature review of journal and conference articles from 2010 to 2021. We included articles that carried out usability tests with populations aged >60 years and presented transferable guidelines on mobile software design, resulting in a final set of 40 articles. We then carried out a thematic analysis with 3 rounds of analysis to provide meaning to an otherwise diverse set of recommendations. At this stage, we discarded recommendations that were made by just 1 article, were based on a specific mobile app and were therefore nontransferrable, were based on other authors' literature (as opposed to recommendations based on the results of usability tests), or were not sufficiently argued. With the remaining recommendations, we identified commonalities, wrote a faithful statement for each guideline, used a common language for the entire set, and organized the guidelines into categories, thereby giving shape to an otherwise diverse set of recommendations.

Results: Among the 27 resulting guidelines, the rules *Simplify* and *Increase the size and distance between interactive controls* were transversal and of the greatest significance. The rest of the guidelines were divided into 5 categories (*Help & Training*, *Navigation*, *Visual Design*, *Cognitive Load*, and *Interaction*) and consequent subcategories in *Visual Design* (*Layout*, *Icons*, and *Appearance*) and *Interaction* (*Input* and *Output*). The recommendations were structured, explained in detail, and illustrated with applied examples extracted from the selected studies, where appropriate. We discussed the design implications of applying these guidelines, contextualized with relevant studies. We also discussed the limitations of the approach followed, stressing the need for further experimentation to gain a better understanding of how older adults use mobile apps and how to better design such apps with these users in mind.

Conclusions: The compiled guidelines support the design of mobile apps that cater to the needs of older adults because they are based on the results of actual usability tests with users aged >60 years.

(JMIR Mhealth Uhealth 2023;11:e43186) doi:10.2196/43186

KEYWORDS

tablet; smartphone; older user; design recommendations; usability testing; user experience design; UX design; design; mobile app; tool; quality of life; software; training; visual design; older adults; mobile phone

Introduction

Background

Mobile apps are becoming increasingly prevalent in the lives of older adults. The Pew Internet Research Center reported that 42% of older Americans (aged >65 years) had a smartphone and 32% owned a tablet in 2017, compared with 18% and 27% in 2013, respectively [1,2]. The importance of mobile apps for older adults became more apparent during recent events such as the COVID-19 pandemic to mitigate the effects of undesired self-isolation. Simultaneously, the older population is growing globally [3]. Because the interplay between mobile apps and older populations is gaining relevance, this paper pays special attention to both.

Touchscreen interfaces allow intuitive and direct manipulation interactions that depict real-world metaphors [4]. However, mobile devices still present substantial difficulties for older people with their nonconventional input methods and the limited size of their displays [5,6]. Other potential challenges include unexpected sensitivity of the touch surface, nonintuitive multifinger gestures, and a conceptual model that differs from desktop computers [4]. Knowledge about recommendations to design mobile apps tailored to address the limitations that older users experience can lead to a better adoption of mobile technologies by this population.

Some of the reasons older adults use mobile apps are to remain independent and active in society [7], monitor and improve their health condition through mobile health (mHealth) [8-10], or remember important information [11]. However, when using technology, older people encounter physical (visual, auditory, and motor changes and dexterity) and cognitive (decline in memory and attention) disadvantages associated with the aging process [12]. In relation to visual perception, declines in contrast sensitivity, acuity, and the ability to discriminate colors can affect symbol and character identification, button-striking accuracy, and reading rates [13]. The lack of motivation, experience, knowledge, access, understanding, and usability also challenge the adoption of technologies [14]. Therefore, mobile design efforts must address the needs and expectations of older adults.

The design of mobile apps needs to consider the user experience (UX) of older users and base design decisions on the results of usability tests with this group of users. The International Organization for Standardization defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [15]. However, the design of mobile apps for older adults has neglected some usability aspects. Design guidelines have focused extensively on visual and haptic issues (eg, high contrast, button type, and button size), whereas textual interface elements have been disregarded (eg, ease of text entry, button feedback, and font type) [16]. Therefore, this paper aimed to analyze all aspects of mobile app interaction design based on usability testing results.

When dealing with usability, we need to consider the overall design approach beyond user interface (UI) design, taking a user-centered design focus. The significance of involving older users in the design of mobile apps must be recognized if they constitute all or part of the target user population [17]. In addition to involving older adults (ie, end users) in testing the usability of mobile apps, user-centered evaluation can include experts [18]. Petrovčič et al [16] pointed out that good evaluations include both end users and experts. Dickinson et al [19] discussed that the challenge of including older users in technology design is that they demand additional technical, organizational, and managerial resources, compared with evaluations with experts. For example, one of the difficulties found in the segment of older people is that they are a heterogeneous group that modifies, uses, and interacts with technology in diverse ways [20].

Designing for older people is not a new realm. Within the human-computer interaction (HCI) field, interest in the topic of aging in connection with technology has grown, as have research-derived guidelines for the design of mobile apps that target older adults. However, these guidelines can be confusing, contradictory, complex, or have become obsolete [21]. Nurgalieva et al [21] and Petrovčič et al [16] also pointed out the limited number of validations, repeatability, and reproducibility of guideline-related studies. Therefore, this research work attempts to extract usability guidelines that are based on experimentation with older adults.

When referring to older adults, different authors have used different definitions. The definition of older adults is context dependent, and the generalizability of the findings for older people may distort the characteristics of this population. Vines et al [22] adopted a critical approach to the existing research that presented older adults as a homogeneous group. Older adults from technologically advanced countries might not experience mobile apps similar to older adults from regions where technology is less prevalent. There is also a difference in the use of the technology by those aged 60-74 years versus those aged >85 years. In this paper, older adults were defined as those aged ≥60 years.

Objective

In summary, we aimed to create a set of design guidelines for mobile apps for older adults that stemmed from published usability testing results with these user cohorts. Thus, our research question was as follows: *What are the demonstrated heuristics to carry out the design of mobile apps for older adults?* We described similar studies; the methodology used; and the results obtained, in particular, the proposed set of design recommendations extracted from the selected studies and discussed the results in the context of related literature.

Previous Work

Numerous mobile app design guidelines for older people have been developed through end-user evaluations (such as those included in this review) and heuristic evaluations [5,6,23]. Others have reviewed some of the literature (not systematically) on the design of mobile apps to propose a few guidelines [24-27]. In addition, Iancu and Iancu [28] provided a theoretical overview of the subject.

Some scoping reviews of the literature analyzed mHealth solutions for older adults [8,9,29,30]. Nimmanterdwong et al [31] reviewed the literature to illustrate the challenges and opportunities of applying human-centered design methodologies in the creation of mHealth solutions for older adults. Furthermore, Nurgalieva et al [21] systematically reviewed the trends and gaps in touchscreen design guidelines for older adults to systematize their knowledge of abilities and design categories. They focused on the characteristics of the older population, the quality of methods, and efforts to catalog the guidelines. In their review, they included studies grounded in secondary data (eg, literature reviews) and expert evaluations, which were excluded in our review. Another difference with our review stems from their definition of older adults (people aged >55 years), and their inclusion of guidelines specifically centered on certain pathologies (eg, Alzheimer disease). Petrovčič et al [16] also presented a systematic review of guidelines from 9 expert evaluations. Other systematic reviews on mobile apps for older users focused on specific health conditions such as cognitive decline [32] or older users' cognitive, visual, and psychomotor challenges with mobile apps [33].

To our knowledge, no systematic review has extracted recommendations for mobile app design from primary data collected through evaluation activities, such as usability tests with older adults aged >60 years.

Methods

Overview

We performed a 2-step study starting with a systematic review and subsequently conducted a thematic analysis. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [34] recommendations for conducting systematic literature reviews. The following subsections detail the search strategies, eligibility criteria, data extraction, quality assessment, and analysis methods. For the review, we collected and evaluated the publications collaboratively using Parsifal, an online tool that supports the performance of systematic reviews. We chose the systematic literature review method because it offers a comprehensive and clear overview of the guidelines written to date and identifies gaps and further research topics.

Search Strategy

First, we derived search terms from our research question: "older people," "mobile," and "usability." The term "guideline" was not considered for 3 reasons: our endeavor was to unpack guidelines based on primary data resulting from usability evaluations; oftentimes, the term "guideline" is not used explicitly; and sometimes, recommendations are offered implicitly throughout the article. We selected the most relevant databases in the field of this review (Web of Science, IEEE, and Scopus) to broadly cover the literature on usability studies that address aging and mobile technology. Then, we tested different strings in the 3 databases to identify various spellings and synonyms. The final search string was as follows: "(older* OR elder* OR ageing OR aging OR senior*) AND (mobile* OR smartphone* OR tablet*) AND (usability* OR UX OR ''user experience'' OR acceptability OR acceptance)."

To keep our search broad, we applied our search string to the fields of "title," "abstract," and "keywords" in each database. We searched for conference papers and academic articles written in English and published in peer-reviewed scientific journals and proceedings within the last 12 years (from 2010 to early 2021). The search strategy was discussed by all authors and conducted by the first author. The initial search yielded 4168 articles.

Eligibility Criteria

The second step of the process consisted of refining the search by filtering the previous results to retain only the articles that met our eligibility criteria. First, we discarded duplicate articles (n=889). Then, we applied our inclusion and exclusion criteria to the remaining articles and filtered them by title, abstract, and keywords. The first inclusion or exclusion of articles was carried out by the first author who held weekly meetings with the rest of the authors to specify which exclusion criteria were applied and to check how the review was progressing. The exclusion criteria are shown in Textbox 1.

Textbox 1. Exclusion criteria.

- Articles that were not written in English.
- Articles that were ≤4 pages long. In our experience, articles shorter than 4 pages usually did not have robust discussions and details that support the design recommendations offered.
- Articles not dealing with mobile apps. We defined mobile apps as those that run on tablets, smartphones, or smartwatches.
- Articles without an explicit usability test. We excluded qualitative research articles, heuristic evaluations, surveys of attitudes, expert reviews, and literature reviews that did not perform a usability test with older adults (aged >60 years). We considered a usability test to be a test with end users who use a system or prototype.
- Articles not providing specific results for older adults aged ≥60 years. Thus, we excluded articles that did not offer results distinguishing between older adults and other user groups, such as caregivers or experts who were not necessarily old. We also excluded articles that did not explicitly match the results within a range of years, for example, articles in which only the average or median age was mentioned.
- Articles exclusively focusing on the contribution to hardware development.
- Full-text articles were inaccessible.
- To validate the eligibility criteria, we peer reviewed 12 articles each. Therefore, we decided to integrate the following exclusion criterion:
  - Articles in which the results were not transferable either because there are no recommendations about the design of mobile apps for older adults or because the article only provided usability tests that informed the results of its own app, but no design recommendations could be extracted.

Data Extraction

The 41 articles that met the eligibility criteria were individually reviewed by 2 authors to extract relevant information to later complete the quality assessment stage. All the collected information was stored in a data extraction spreadsheet.

After individual analyses, 6 consensus meetings were held to discuss the results obtained and reach an agreement on the extraction of qualitative data, such as errors identified through usability testing, design recommendations, or publication venues. In each of the 6 meetings, we peer reviewed 10 to 12 articles and documented the results in a collaborative spreadsheet on Microsoft Teams (Microsoft Corporation). During the analysis, 1 author used the snowballing method, which eventually allowed us to include 5 more articles. The different aspects that were analyzed in each article are presented in Textbox 2.

Textbox 2. Aspects considered during article analyses.

- *Title and digital object identifier* of the article.
- *First impressions and a summary* of the article to point out the features that were considered important but were not covered in the following fields.
- *Number of older participants* involved in the usability testing. For cases in which >1 group of users participated in the evaluation, we considered only those groups that included users aged ≥60 years. If different iterations of usability testing were performed, we analyzed whether the participants were the same in each iteration. If the participants remained throughout the iterations, all were considered. In contrast, if the participants differed in each iteration, we considered only 1 group (the largest group of participants). This item allowed us to perform the following quality assessment: the higher the number of participants, the more robust the results were considered to be, and therefore, more relevance was assigned to the article.
- Article's venue and quality based on its relation to human-computer interaction (HCI; assessed considering the aims and scope of the venue) and the position of the venue in quality rankings, using sources such as the Journal Citation Report from Clarivate for journals or the Computing Research & Education ranking of conferences. We considered articles published in an HCI venue to have undergone a stricter methodological scrutiny for the usability testing. This item was added to perform the following quality assessment: the more the venue is related to HCI and the better its quality ranking, the more relevance was assigned to the article.
- *The number of times the participants used the system* was evaluated. "One-cut study" stands for evaluations in which participants used the system only once, as compared with longitudinal studies in which the time was expressed in weeks. Time was relevant to the quality assessment because it reflected the reliability and robustness of the results presented in the article. In other words, a longitudinal study can offer more robust results than a one-cut study; therefore, it would be assessed as more relevant.
- *Year of publication*. If the article had both online and printed versions, the earliest date of publication was considered. We believe that newer results should be better aligned with the current technology and technological abilities of the older population. This is essential because these aspects evolve, and the results become obsolete very quickly.
- *Methodological soundness*. We focused on whether the article provided sufficient information about the design of the evaluation, which included recruitment, eligibility criteria, test protocol and procedure, and participants' demographics. The existence of this information increases the external validity of the results and affects the relevance of the article in the quality assessment stage.
- *Number of tasks defined in the usability test*. We considered a usability test that prompted the user to perform several tasks to provide more reliable results than a usability test with fewer tasks.
- *Venue of the usability test*. Although this is not a central issue in assessing the articles, this item offered us complementary information on the characteristics of usability tests performed with older adults in the literature.
- *The existence of approval from an ethical committee*. We decided that such approval was not essential to assess the articles because obtaining approval to perform a usability study was not required by all journal venues. However, it provided complementary information about the articles.
- *Design recommendations*. We extracted the explicit design recommendations for mobile app design. Accordingly, we excluded participants' wishes and preferences, as they were not tested. If the study used mixed methods, we divided the recommendations according to each method. We included the recommendations if they were based on usability tests or a mixture of methods (including a usability test). However, we excluded the recommendations based exclusively on methods other than usability tests as well as recommendations from participants who were not older adults.
- *Errors* in the system identified during usability testing. These were compiled to explain the sources of design recommendations. Errors were extracted according to method used.
- *The level of agreement between reviewers* (0%-100%) and comments on the authors' discrepancies when reviewing the article. This item was added to register the discussions that the reviewers had during data extraction, as this information may be useful for the subsequent data analysis.

Quality Assessment

The same 2 authors who performed the data extraction in the previous stage used 3 qualitative values (high, medium, and low) based on some of the data extracted to assess the quality of each article considering 4 dimensions.

Dimension 1: Is the Quality of the Venue Sufficient?

If the venue was prestigious in the HCI field, it was rated as high. If the venue was not highly ranked but was related to HCI or if the venue was highly ranked in other fields but not related to HCI, the venue was rated as medium. Finally, if the venue was not prestigious in any field and was not related to HCI, it was rated as low. For this assessment, we used the Journal Citation Report and Computing Research & Education rankings as well as the authors' knowledge.

Dimension 2: Do the Authors Use and Describe a Proper Methodology for Evaluation in Their Article?

We considered the testing protocol as the most relevant information, rather than recruitment, eligibility criteria, and participants' demographics. If the methodology contained all the items, the rating was high. If the methodology contained all items except the testing protocol or procedure or if the methodology contained the testing protocol or procedure but not the other items, the rating was medium. Otherwise, the rating for this dimension was low.

Dimension 3: Are the Number of Tasks Tested and the Number of Users Involved Sufficient to Obtain Robust and Reliable Knowledge?

We weighted the number of users more than the tasks used because some studies did not necessarily involve tasks in their usability testing; therefore, the number of users was more important to us. If the number of users and tasks were higher than 10 and 5, respectively, the article was rated as high. If 5 to 10 users participated in the evaluation and the number of tasks was ≤5, a medium rating was given. If the number of users and tasks were <5, we rated it as low.

Dimension 4: Can We Extract Guidelines From the Usability Tests?

Even though the existence of transferable results was already an exclusion criterion, the authors evaluated the extent of this transferability based on their experience and previous knowledge, considering not only the number of results but also the extent to which the findings could be generalized.

Procedure

The results of the quality assessment were then transferred to Parsifal using the following scheme: for each of the 4 questions, we assigned 4.0, 2.0, and 0.0 points if qualitative ratings were high, medium, and low, respectively. The total score was the sum of the 4 question scores. The cutoff total score to discard low-quality articles was ≤2.0 points (all values were low and only 1 medium, at most). Consequently, 6 articles were excluded because of their low quality.

Thematic Analysis

The 2 authors who were not involved in the data extraction and quality assessment stages performed the data analysis using qualitative thematic analysis. The purpose of this division of labor was to ensure that the data analysis was not biased by the performance of the previous stages.

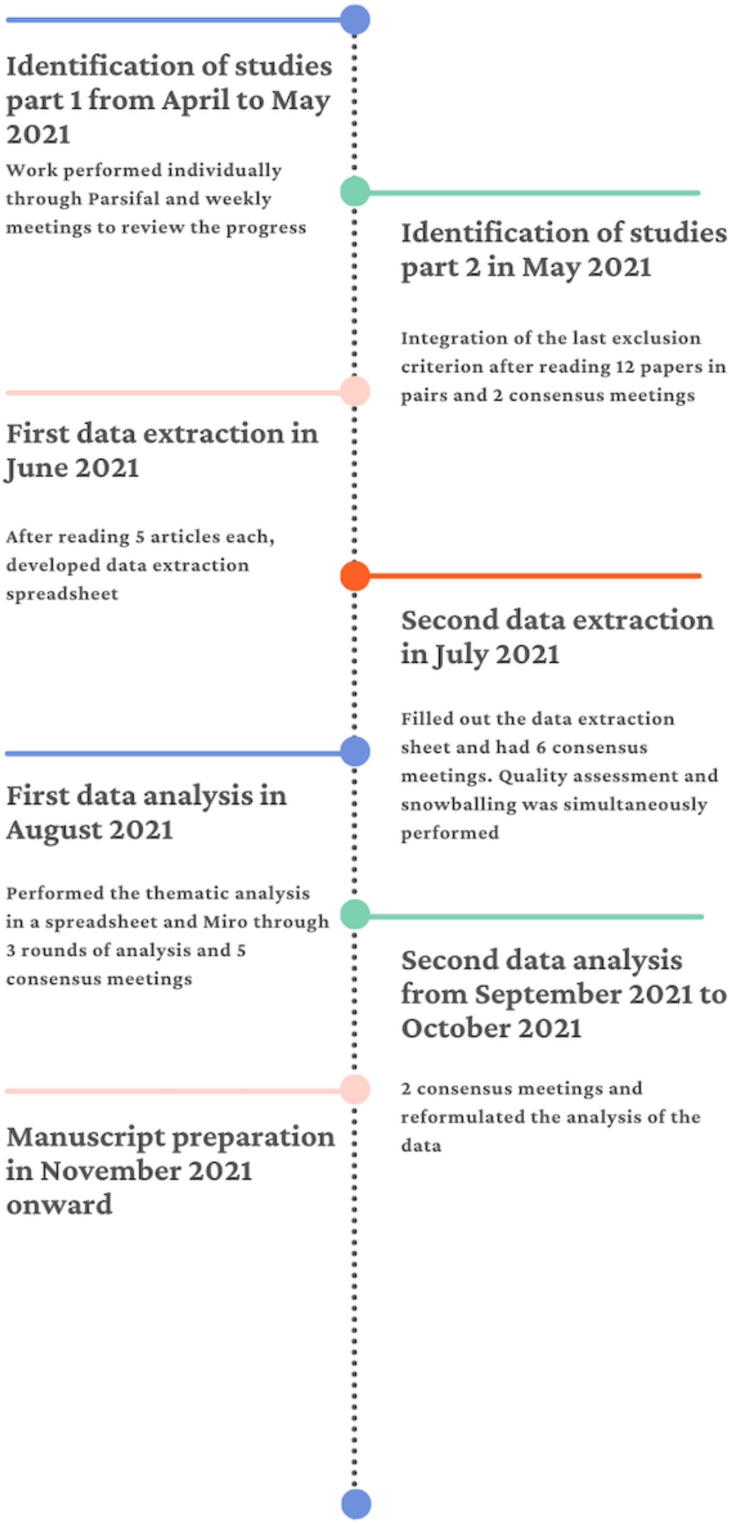
Our goal was to identify, analyze, and interpret patterns of meaning within a set of design recommendations. We performed 3 rounds of analyses. The first round consisted of grouping guidelines according to the terms' commonalities. Then, we organized the concrete guidelines into broader recommendations. In the final round, we grouped the recommendations into themes using a holistic approach. The resulting themes and categorization of the guidelines were first analyzed individually by all the authors and then discussed at 3 consensus meetings to agree on the results. We paid special attention to the consistency of the results and clarity of the terms used in the categorization. During this process, we used a shared spreadsheet to organize the data and Miro (Miro Corporation) to visualize the data. We also used a top-down representation of recommendations ranging from general to concrete. On the basis of our individual analysis of the results and discussions during the consensus meetings, we decided to exclude the following recommendations:

1. *Guidelines based on other's literature*, that is, guidelines that are not tested with users. Some articles pointed out design recommendations that were extracted from literature references and were not directly tested (eg, Pereira et al [35]).
2. *Guidelines that applied to concrete mobile apps and cannot be transferred*. Some articles made recommendations that we considered too specific to be generically applied. For example, specific display measures such as "button sizes should be at least 200 mm ^{2}" [36] or banning specific UI elements such as "do not use the picker" [37]. We considered that this type of guideline required further experimentation.
3. *Guidelines that were not sufficiently argued in the article*. Some articles pinpointed recommendations without explanation; therefore, they were excluded. For example, the recommendation "allow the user to select the icons preferred" [38] was not sufficiently argued and contradicted with other guidelines that did not allow such flexibility.
4. *Guidelines that were supported in only 1 article*. If the recommendation appeared in a single article, it was ignored because it was not considered generalizable, for example, "Maintain link underlined" [39].

Workflow

From the search strategy through the thematic analysis stage, we worked individually and held consensus meetings along the way to agree on certain criteria, strategies, tools, and so on. First, we identified the studies; second, we extracted data from the selected articles; and finally, we analyzed the data thematically. We performed the review for 6 months before manuscript preparation. The workflow is illustrated in Figure 1.

Figure 1. Workflow of the study.



Results

Systematic Review Results

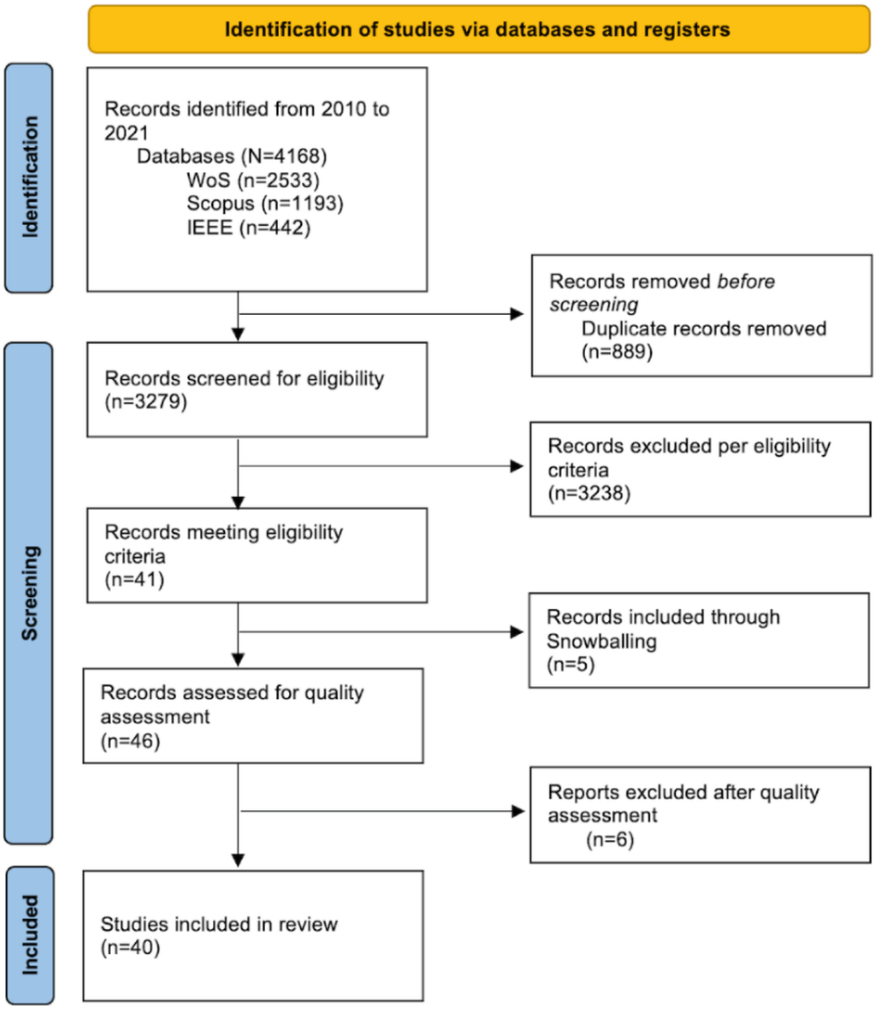
Overview

A total of 40 primary studies were retrieved. The initial search yielded 4168 articles (n=2533, 60.77% from Web of Science; n=1193, 28.62% from Scopus; and n=442, 10.6% from IEEE). Removal of duplicates decreased the number of articles to 3279. Through the screening of abstracts, titles, and keywords, 3238 articles were excluded. Subsequently, we assessed the full text of these articles, resulting in a final number of 40 studies. Figure 2 illustrates the flowchart of the review process.

On the basis of the exclusion criteria, the number of articles excluded (n=3238) for each criterion was as follows: the technologies at play were not mobile (n=343, 10.59%), articles had <4 pages (n=30, 92.65%), the researchers were not able to access the full article (n=1, 0.03%), articles were not written in

English (n=6, 18.53%), the methodology did not include a usability test (n=1165, 35.98%), the results could not be transferred (n=94, 2.9%), there were no specific recommendations for older people (n=1583, 48.89%), and articles did not deal with software technology (n=16, 0.49%).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for study selection. WoS: Web of Science.



Study Characteristics

Of the 40 studies, 16 compared features of mobile touchscreens and were not related to a concrete domain; 17 studies were related to the health care domain; 5 studies addressed social engagement; 1 study concerned entertainment; and 1 was related to housing.

With respect to the number of participants involved in the usability tests, 10 studies involved 3 to 9 participants, 11 studies performed tests with 10 to 19 participants, 14 with 20 to 29 participants, and 5 with >30 participants.

Regarding the duration for which the participants tested the apps, 35 studies encompassed only 1 test session, and 5 investigations were longitudinal. Among the latter, 2 tested the design for ≥1 month and 3 between 1 and 4 weeks. Although the number of tasks used in the usability tests was unspecified in 9 studies, 26 studies performed 1 to 9 tasks, and in 5 studies, 10 tasks were completed. The venue of the test was unspecified in 18 studies, whereas 10 studies performed the usability tests in a laboratory, 4 in day-care centers, and 8 in the participants' homes. Of the 40 studies, 12 (30%) received ethics approval to perform their tests.

Thematic Analysis Results: Guidelines

Golden Rules

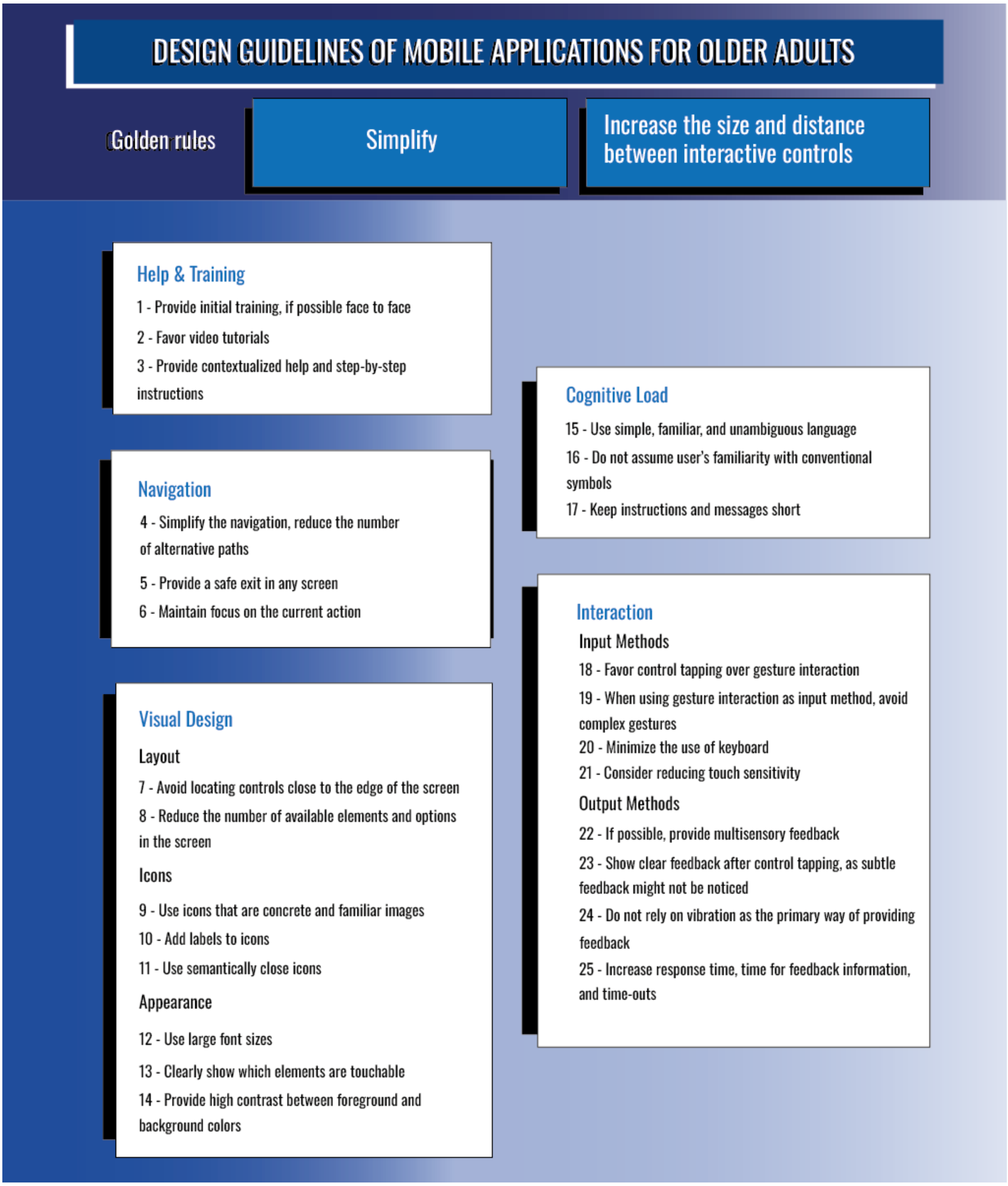
There are 2 guidelines of special significance, as both have been mentioned in 15 studies:

- *Simplify* [36,37,39-51]. There were different recommendations in the 15 studies, but they can be summarized as the need to simplify the design. The cognitive difficulties experienced by these users call for extra effort to simplify the product concept and any element that the users need to understand to successfully operate the mobile app.
- *Increase the size and distance between interactive controls* [36,39,41,43-45,49,52-59]. The size of interactive controls (eg, buttons or form entries) should be augmented to facilitate older people's interactions with them. Spacing should also be large to avoid accidental tapping because older users with motor limitations may have less precision

in interacting with controls. If possible, a touch area that exceeds the visual component should be defined.

Because these 2 recommendations were recurrent and of greater significance than the rest of the guidelines, we have chosen to represent them separately as golden rules that should be followed in the design of any mobile app to be used by older users. The rest of the recommendations gathered have been expressed as guidelines, grouped into 5 categories, as detailed in the following subsections and shown in Figure 3.

Figure 3. Summary of the guidelines.



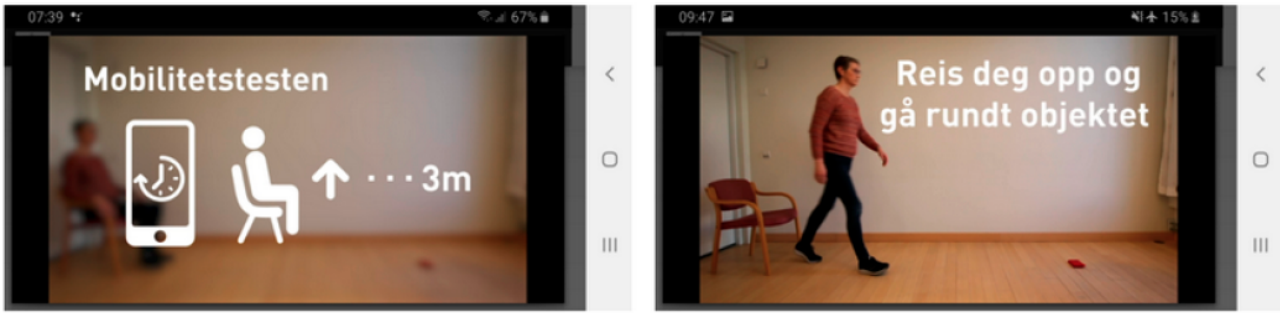
Help & Training

This category includes guidelines 1-3:

- *Provide initial training, if possible face to face* [52,60]. Considering the difficulty in providing initial training, a face-to-face demonstration of the system before its first use should be provided, as this population requires special support. The lack of familiarity with technology may affect the ability of a significant number of older users to benefit from written documentation. Face-to-face training would be most helpful for critical apps (health, personal finance, etc).

- *Favor video tutorials* [61-63]. When an older person is learning how to use an app, it should offer help through a video rather than written instructions. Written instructions can complement the videos; however, they should not stand alone. For example, Bergquist et al [62] provided instructional videos explaining how to perform the tests included in their app (Figure 4).
- *Provide contextualized help and step-by-step instructions* [41,62,65]. Having to search the help subsystem for how to solve a specific problem can be very time-consuming for older users and sometimes unsuccessful. Guidance on how to use the system, especially for complex tasks, should be structured and accessible. Designers should provide contextual help by focusing on the visible interface context.

Figure 4. Example application of guideline 2, “Favor video tutorials” (from Bergquist et al [62]).

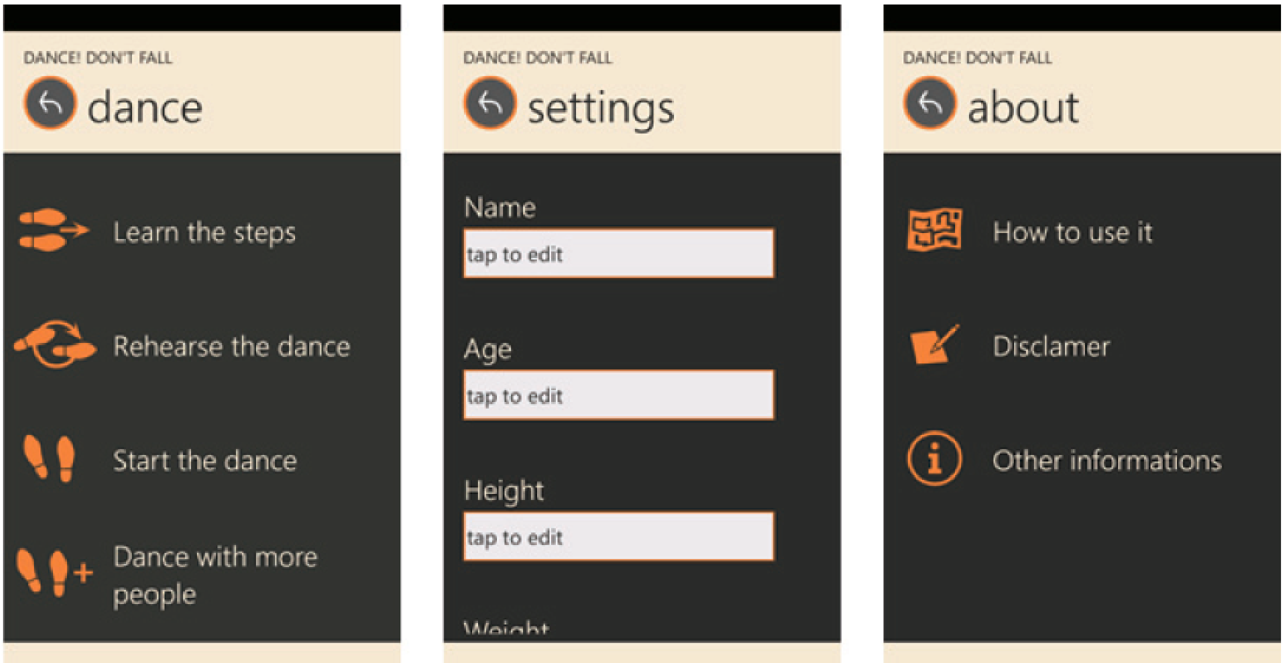


Navigation

This category includes guidelines 4-6:

- *Simplify the navigation, reduce the number of alternative paths* [40-42]. Navigating through a mobile app can pose a significant challenge for users with cognitive difficulties because they may be lost if they cannot remember all the steps to perform a task. We need to provide a navigation that is simple and uses logic involving very few rules, works everywhere, and reduces the number of alternatives. For simpler navigation, the complexity of the number of optimal paths and optimal path length should be reduced.
- *Provide a safe exit in any screen* [39,53]. Make sure that every screen includes an apparent exit on the interface, so that older adults can avoid anxiety when they do not know what to do in the app. By safe exit, we mean any way to return to a previous safe state, such as a return function; the back button (as in the app designed by Barros et al [53], displayed in Figure 5); or a cancel option.
- *Maintain focus on the current action* [39,66]. Because older users will have more difficulties maintaining concentration, we need to help them focus on the current action. Do not display secondary functions. Instead, attract users’ attention to the most important or typical button to tap in the next step. Thus, older users will be able to proceed more easily through the app and avoid navigation difficulties.

Figure 5. Example application of guideline 5, “Provide a safe exit in any screen” (from Barros et al [53]).



Visual Design

Layout

This category includes guidelines 7 and 8:

- *Avoid locating controls close to the edge of the screen* [53,67]. Position interactive elements far from the edge of the screen, so that older adults will avoid involuntary interactions when using the mobile app.
- *Reduce the number of available elements and options in the screen* [43,54,64]. Simplify the layout, even at the cost of reducing the set of available functionalities. Older adults tend to have a better UX with an app when layouts are simple.

Icons

This category includes guidelines 9-11:

- *Use icons that are concrete and familiar images* [60,68,69]. The use of abstract icons should be avoided, and icons should depict real-world representations. For example, add graphical content to labels such as medication package pictures that are meaningful to older users.
- *Add labels to icons* [39,44,45,68-70]. To improve understandability and user performance with a system, designers should add textual support to icons and buttons. The textual support should represent the purpose of the icon.
- *Use semantically close icons* [45,68,70]. "Semantic distance refers to the closeness of relationship between the icon and the function it represents" [71]. A small semantic distance will have a positive impact on icon recognition by older people.

Appearance

This category includes guidelines 12-14:

- *Use large font sizes* [36,41,43,72]. Visual acuity diminishes with age; therefore, a large font size will help older users read the text. Ensure that the font size is sufficiently large to be visible to older people.
- *Clearly show which elements are touchable* [45,66,73]. Users experiencing cognitive strain will have more difficulty distinguishing interactive elements from noninteractive ones. In addition, older users may not be familiar with conventional affordances in mobile apps. Therefore, enhance the difference between touchable and nontouchable elements at the interface through clear boundaries and avoid ambiguous elements. For instance, enable cues for interaction so that the older person knows whether an element is selectable or draggable.
- *Provide high contrast between foreground and background color* [39,41,44,72]. Visual acuity diminishes with age, so a strong contrast between the text color and the background color will help older users read the text. Even if it mostly benefits users with visual impairments, this recommendation can eventually benefit a broad range of users, for example, when using a screen with low brightness or when fatigued.

Cognitive Load

This category includes guidelines 15-17:

- *Use simple, familiar, and unambiguous language* [36,53]. Users who lack familiarity with technology will find it challenging to interpret technical terms and common symbols used in the UI. They will have added difficulty coping with ambiguous language because they lack the heuristics for interpreting the UIs that technology-proficient users possess. Use simple terms and clear feedback in mobile apps. In this way, older users, regardless of their cultural background, can understand it, and the technology does not cause as much anxiety. Harte et al [36] use an example of simple, familiar, and unambiguous language as shown in Figure 6.
- *Do not assume users' familiarity with conventional symbols* [44,60,74]. Designers should not take for granted that older users will understand usual conventions, such as "?" being a help button or "→" being a send button. They should create symbols that are understandable and adapt to the cultural context of the person, regardless of their familiarity with technology.
- *Keep instructions and messages short* [37,63]. Instructions and text on how to use a system should be short to avoid overwhelming older users with the cognitive effort of reading extensive messages.

Figure 6. Example application of guideline 15, “Use simple, familiar, and unambiguous language” (from Harte et al [36]).

Interaction

Input Methods

This category includes guidelines 18-21:

- *Favor control tapping over gesture interactions* [47,55,75]. Favor direct manipulation on the screen (control tapping or single tap) over gesture interaction. The latter requires advanced motor skills that may be difficult for older users and can hinder good UX with a system. By control tapping, we mean requiring the user to place the finger over a specific control appearing on the screen, as opposed to making a gesture such as pinch or swipe. For example, Barbosa Neves et al [75] designed an app whose only input interaction was single control tapping.
- *When using gesture interaction as input method, avoid complex gestures* [39,50,76]. We cannot rely on users remembering gestures because there is no hint in the UI that helps the users to recall the set of available gestures. Due to skin aging, wrinkling, or hand tremors, older users may lose contact with the screen, and the gesture may not be correctly interpreted by the system. This adds to the problem of lack of familiarity with technology, as gesture-based interaction is an advanced feature, aside from perhaps a very common gesture such as pinch to zoom.
- *Minimize the use of keyboard* [53,54,73]. Virtual keyboards require fine motor abilities, which are difficult for older users with hand tremors or arthritis. As a possible alternative, the use of voice input could be explored and usability tested.
- *Consider reducing touch sensitivity* [64,74]. A high sensitivity to touch produces involuntary taps on the screen by a certain number of older users. Designers should consider how high control sensitivity is and consider reducing it if there is a risk of involuntary taps by older users. Thus, these users will be able to move their hands over the screen with less fear of accidentally tapping on the controls.

Output Methods

This category includes guidelines 22-25:

- *If possible, provide multisensory feedback* [4,39,49,58,60]. Because older users may experience perception limitations, multisensory feedback will increase the probability that messages will get to users correctly. In this manner, we provide multiple options to users who have limitations in hearing or vision.
- *Show clear feedback after control tapping, as subtle feedback might not be noticed* [56,74]. Limitations in perception may lead the user to miss subtle feedback; therefore, feedback should be clear and always provided as a response to an explicit user action, such as control tapping. Older users may not notice subtle changes in the color of a pressed button, and they have a higher risk of tapping outside the target. Therefore, provide bolder interaction feedback anytime the tap has occurred so that the user is aware of having tapped a control.
- *Do not rely on vibration as the primary way of providing feedback* [49,67]. Designers should not consider vibration and tactile feedback as the only means of conveying information because older users may not notice it. Current mobile phones provide weak vibration motors, but this could change in the future.
- *Increase response time, time for feedback information, and time-outs* [4,39,57,61]. Long time-outs in input interaction modes allow users time to interpret the screen and decide on their next action. In this regard, the time for feedback information on the screen should be long enough for users to process, as, for example, in the case of pop-up messages.

Discussion

Principal Findings

The guidelines obtained address various issues at diverse abstraction levels. This is due to the method used to obtain the guidelines, as different studies made recommendations at different abstraction levels. We omitted solutions that were applied only to a specific design and were not easily transferrable to other problems and domains, as mentioned in the *Thematic Analysis* section.

Overall, we believe that designers should choose design options that would benefit older users without the need to create a different version of the mobile app. Thus, the same mobile app could be used by a wide range of users, regardless of their age. This approach aligns with the Universal Design perspective, which embraces "the design of products, environments, programmes, and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" [77]. We acknowledge that it may be necessary to either make adaptations for users with severe limitations or create a simplified mode for older users when complex operations are necessary for other users. However, we believe that it is wrong to disregard older users from the beginning. Designers should adopt the philosophy of trying to choose the design options that can best accommodate older users without losing other users.

The golden rule of *Simplify* is at the highest level of abstraction. Simplicity is related to higher UX. People love designs that make their lives simpler [78], so it is a guideline not just for older users, who not only prefer a simpler system but also require it. Cognitive limitations associated with age increase the challenges of understanding and remembering how to use a complex system. The aim of reducing complexity for older users aligns with the recommendations of Fisher et al [79].

Eyesight and motor limitations in older users necessitate the golden rule of *Increasing the size and distance between interactive controls*. Reduced motor skills cause older users to have more trouble when tapping small controls or controls that are too close together. Web design guidelines for older users, such as those developed by Kurniawan and Zaphiris [80], identified the need to provide larger targets for these users. Touchscreens already require larger touch targets than desktop or web systems accessed with a mouse, because fingers are much bigger than a pointer and have much less precision in target selection compared with mouse clicking [81]. For older users, this problem can be aggravated by their lack of dexterity and motor skills.

The *Help & Training* guidelines provide specific advice on how to provide proactive help [82]. Given that a certain number of older users will not be familiar with technology, learning through exploration is not a good strategy to favor this type of users. The survey by Leung et al [83] confirmed that preference for trial-and-error strategies decreases with age when learning to use mobile apps. The survey also found that older users were more interested in demonstrations followed by an opportunity to replicate steps to obtain feedback, compared with internet information or contact with help-desk staff.

The recommendation to favor video tutorials over written instructions matches our own experience in performing usability tests on eHealth apps for geriatric patients, wherein we observed users tiring easily from lengthy textual instructions [84,85]. Although the study by Ahmad [86] focused on younger tech-proficient adults (aged >50 years), it showed the user preference for video tutorials. In addition, designers should be aware that a significant proportion of the older population has a low literacy level [19]. In our experience, this is especially relevant in usability testing with older users in a medium- to low-income neighborhood [84,85]. Relatedly, Androutsou et al [87] identified the difficulty that older users experience when reading text in notifications. These findings also relate to the guideline that recommends keeping instructions and messages short.

The heuristic on help and documentation by Nielsen [88] advises listing concrete steps to be completed. Older users will better understand a set of instructions if they are provided step-by-step as they proceed with their task with the app. The study by Leung et al [68] showed that older users place more significance on learning task steps than on gaining a general understanding when learning to use mobile apps.

By reducing the number of possible alternatives in the app, we offer less freedom to the user; however, we can reduce the length of the optimal path. For procedural tasks, such as health-monitoring activities (as the ones described by Villalba-Mora et al [84] and Moral et al [85]), it is preferable to use a wizard navigation style in which users are only offered 2 options: go to the next step or exit. Such a simple navigation scheme helps most older users avoid the cognitive burden of deciding where to tap next to continue with their tasks.

The suggestion to maintain focus on the current action follows these same principles. Secondary functions, even if valuable to a minority group of users, can distract a significant number of older users. It is better to attract the attention of the user to the action that we expect to be taken in the most typical scenario, thus preventing user errors [80].

Regarding the need to use larger fonts, Harte et al [36] recommended that for simple interface elements, text sizes should be at least 10 points (Didot system); however, Morey et al [41] recommended using font sizes that are at least 30pts for critical text and at least 20 points for secondary text. Because there is no consensus on the exact minimum font size, we opted to state that the font size should be large to direct the attention of the designer to this issue. This topic needs further research through usability testing.

Reduced motor skills affect the dexterity necessary to properly hold the mobile device and to avoid tapping on an app control on the screen when it is too close to the screen edge. Both golden rules lead to a screen design with less clutter, thereby reducing the number of elements available at a given time. Guideline 8, namely, *Reduce the number of available elements and options in the screen*, explicitly states this corollary to both golden rules. As a result, there is an additional limitation on possible design solutions, which requires extra effort to complete the design work. Every element placed on a screen needs to have a clear purpose and be relevant to most users.

Icons have received substantial attention in the analyzed studies. Although it is a very specific interaction element, icons are relevant in mobile apps and pose a significant difficulty in the design endeavor. According to the study by Leung et al [68], older users identify icon objects and interpret icon meanings less accurately than younger users. The suggestion to use icons that are concrete and familiar images aligns with the recommendation by Petrovčič et al [16] to use meaningful icons, though they refer to the design of mobile phones and operating systems rather than individual mobile apps.

Regarding the recommendation to minimize the use of the keyboard, in the study by Soares Guedes et al [89], almost all study participants (aged 60-76 years) had difficulty using the virtual keyboard because the keys were small and required precision.

In the usability testing of our eHealth apps for older patients, we observed users tapping twice or thrice on the same area of the screen inadvertently when trying to tap just once, due to their lack of soft-movement skills [84,85]. The recommendation to consider decreasing touchscreen sensitivity addresses this problem. Nevertheless, reducing too much sensitivity could lead to the interaction being unnatural for users who are used to a fast response. This topic needs further research by means of usability testing with different types of users.

The effectiveness of vibration feedback for older users remains unclear and open for discussion. Huppert [90] states that in users who are aged >50 years, the ability to perceive vibrations is diminished. Therefore, vibration feedback is not as effective as visual or auditory feedback, supporting the guideline of not using vibration as the primary method of feedback. Because 2 of the studies considered in our review (de Almeida et al [39] and Leitão and Silva [58]) recommended using vibration as one possible multisensory feedback, we have not ruled out its use as such. Nevertheless, further research on this topic is needed.

Huppert [90] also identified the age-related decline in the speed at which information is processed as the reason why older users have difficulties with tasks in which information is presented for very brief periods or rapid responses are required. Therefore, the last of the compiled set of recommended guidelines is to *Increase response time, time for feedback information, and time-outs*.

Limitations

One limitation of this study is that potential errors could have occurred during the revision and synthesis of the articles. To avoid this, we performed these tasks in pairs and held periodic meetings along the way. Two reviewers extracted the data, and 2 different authors performed the thematic analysis. The screening was performed by 1 author, because reviewing 4174 articles was best done alone rather than in pairs. Furthermore, to prevent errors, we performed the snowballing review after the screening to include additional articles that matched our inclusion criteria.

Regarding the age group, we set the minimum age cutoff at 60 years. Different choices for the user group age range could lead to different results. Considering a younger age would include additional studies whose user characteristics differ from those of the target users. By contrast, a higher minimum age cutoff could severely restrict the number of studies included, thus affecting the quality and practical applicability of the results.

Moreover, we extracted studies that performed usability tests and excluded studies that used other methods, such as qualitative inquiries, focus groups, and ethnographic observations. We decided to follow this strategy not only to narrow down the search but also because usability tests are the gold standard for assessing usability. HCI remains our theoretical tenet. Nevertheless, we acknowledge that other valuable guidelines could be proposed in studies that use different methods.

We excluded gray literature and unpublished articles, so the sources were limited to articles in journals and conferences with data that have gone through a thorough review process. Due to the considerable number of articles screened, we believed that the number of articles to review was sufficient. We used 3 databases that could have limited our screening phase; however, we believe that Web of Science, Scopus, and IEEE are the broadest and most relevant to the HCI field. Moreover, we trialed preliminary keywords in different databases because the search strategy could have missed relevant articles.

Conclusions

There is sufficient experimental evidence in the published literature about design features in mobile apps that improve usability for older adults. We systemically extracted design recommendations stemming from this experimental evidence, where usability tests were conducted with actual older users. Through a thematic analysis, we organized these findings into a set of 27 recommendations, including 2 golden rules and 25 design guidelines classified into 5 categories (Help & Training, Navigation, Visual Design, Cognitive Load, and Interaction).

Design guidelines do not ensure usability or a good UX on their own, but we hope that these design guidelines will support the design of better mobile apps through a user-centered design approach, catering to the needs and characteristics of older users and helping bridge the age-related digital divide.

We plan to apply these guidelines to real mobile development projects in which older users are either part of or the entire user target to assess their applicability. Further research is also necessary to gain a deeper understanding of how the different guidelines are nuanced based on the diverse characteristics of older users in terms of age, previous experience with mobile technologies, or physical and cognitive limitations, among others.

Acknowledgments

This research was partially funded by the Spanish Ministry of Science and Innovation research grant PID2019-108408RB-C21—Active aging with Unobtrusive Personalized monitoring project and the European Project POSITIVE (Maintaining and Improving the Intrinsic Capacity Involving Primary Care and Caregivers; reference 20683) funded by the European Institute of Innovation & Technology Health. The authors' home institution has been cofinanced by the *Fondo Europeo de Desarrollo Regional* (European Regional Development Fund).

The authors would like to thank the reviewers of the first draft of the article for their thorough review, providing insight into the relevant aspects. Thanks to their review efforts, the authors have been able to greatly improve the presentation of their work. They would also like to thank Diego López Suárez for the graphic design of the figure that summarizes the guidelines.

Authors' Contributions

XF presented the idea. Consequently, all the authors designed the query and criteria to select the articles. MG-H retrieved and screened the articles until the final number of articles was reached. CM and MG-H performed data extraction and quality assessment. Then, XF and EV-M analyzed the data and conducted the thematic analysis. MG-H and XF prepared the initial manuscript, and all the authors reviewed and approved the final version. Throughout this process, all authors held periodic meetings to review the progress, give feedback, and agree on changes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 41 KB - mhealth_v11i1e43186_app1.docx]

References

1. Anderson M, Perrin A. Technology use among seniors. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2021-11-02]
2. Smith A. Older adults and technology use. Pew Research Center. 2014 Apr 03. URL: https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/ [accessed 2021-11-03]
3. World Population Ageing 2019: Highlights. New York, NY: United Nations; 2019:2020-2092.
4. Harada S, Sato D, Takagi H, Asakawa C. Characteristics of elderly user behavior on mobile multi-touch devices. In: Proceedings of the 14th IFIP TC 13 International Conference on Human-Computer Interaction. 2013 Presented at: INTERACT '13; September 2-6, 2013; Cape Town, South Africa p. 323-341 URL: https://link.springer.com/chapter/10.1007/978-3-642-40498-6_25 [doi: 10.1007/978-3-642-40498-6_25]
5. Silva PA, Holden K, Jordan P. Towards a list of heuristics to evaluate smartphone apps targeted at older adults: a study with apps that aim at promoting health and well-being. In: Proceedings of the 48th Hawaii International Conference on System Sciences. 2015 Presented at: HICSS '15; January 5-8, 2015; Kauai, HI p. 3237-3246 URL: https://ieeexplore.ieee.org/document/7070205 [doi: 10.1109/hicss.2015.390]
6. Salman HM, Wan Ahmad WF, Sulaiman S. Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective. IEEE Access 2018;6:22578-22591 [FREE Full text] [doi: 10.1109/access.2018.2827358]
7. Kurniawan S. Older people and mobile phones: a multi-method investigation. Int J Hum Comput Stud 2008 Dec;66(12):889-901 [FREE Full text] [doi: 10.1016/j.ijhcs.2008.03.002]
8. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
9. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
10. Kascak LR, Rébola CB, Sanford JA. Integrating Universal Design (UD) principles and mobile design guidelines to improve design of mobile health applications for older adults. In: Proceedings of the 2014 International Conference on Healthcare Informatics. 2014 Presented at: ICHI '14; September 15-17, 2014; Verona, Italy p. 343-348 URL: https://ieeexplore.ieee.org/document/7052509 [doi: 10.1109/ichi.2014.54]
11. Inglis EA, Szymkowiak A, Gregor P, Newell AF, Hine N, Shah P, et al. Issues surrounding the user-centred development of a new interactive memory aid. Univers Access Inf Soc 2003 Aug 06;2(3):226-234 [FREE Full text] [doi: 10.1007/s10209-003-0057-7]
12. Czaja SJ, Boot WR, Charness N, Rogers WA. Designing for Older Adults: Principles and Creative Human Factors Approaches. 3rd edition. Boca Raton, FL: CRC Press; 2019.

13. Echt KV, Burridge AB. Predictors of reported internet use in older adults with high and low health literacy: the role of socio-demographics and visual and cognitive function. Phys Occup Ther Geriatr 2011 Mar 02;29(1):23-43 [FREE Full text] [doi: 10.3109/02703181.2010.547657]
14. Czaja SJ, Sharit J. Age differences in attitudes toward computers. J Gerontol B Psychol Sci Soc Sci 1998 Sep;53(5):P329-P340. [doi: 10.1093/geronb/53b.5.p329] [Medline: 9750571]
15. ISO 9241-210:2019: Ergonomics of Human-System Interaction — Part 210: Human-Centred Design For Interactive Systems. Geneva, Switzerland: International Organization for Standardization; Jul 2019.
16. Petrovčič A, Taipale S, Rogelj A, Dolničar V. Design of mobile phones for older adults: an empirical analysis of design guidelines and checklists for feature phones and smartphones. Int J Hum Comput Interact 2018;34(3):251-264 [FREE Full text] [doi: 10.1080/10447318.2017.1345142]
17. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
18. Norman D. The Design of Everyday Things: Revised and Expanded Edition. New York, NY: Basic Books; 2013.
19. Dickinson A, Arnott J, Prior S. Methods for human – computer interaction research with older people. Behav Inf Technol 2007;26(4):343-352 [FREE Full text] [doi: 10.1080/01449290601176948]
20. Grates MG, Heming AC, Vukoman M, Schabsky P, Sorgalla J. New perspectives on user participation in technology design processes: an interdisciplinary approach. Gerontologist 2019 Jan 09;59(1):45-57. [doi: 10.1093/geront/gny112] [Medline: 30260407]
21. Nurgalieva L, Jara Laconich JJ, Baez M, Casati F, Marchese M. A systematic literature review of research-derived touchscreen design guidelines for older adults. IEEE Access 2019;7:22035-22058 [FREE Full text] [doi: 10.1109/access.2019.2898467]
22. Vines J, Pritchard GW, Wright PC, Olivier P, Brittain K. An age-old problem: examining the discourses of ageing in HCI and strategies for future research. ACM Trans Comput Hum Interact 2015 Feb 17;22(1):1-27 [FREE Full text] [doi: 10.1145/2696867]
23. Mi N, Cavuoto LA, Benson K, Smith-Jackson T, Nussbaum MA. A heuristic checklist for an accessible smartphone interface design. Univ Access Inf Soc 2013 Oct 20;13(4):351-365 [FREE Full text] [doi: 10.1007/s10209-013-0321-4]
24. Alnanih R. Usability issues and design guidelines for user interfaces for elderly users. Int J Adv Sci Eng Inf Technol 2019;28(13):138-148 [FREE Full text]
25. Al-Razgan M, Al-Khalifa H, Al-Shahrani M, AlAjmi H. Touch-based mobile phone interface guidelines and design recommendations for elderly people: a survey of the literature. In: Proceedings of the 19th International Conference on Neural Information Processing. 2019 Presented at: ICONIP '12; November 12-15, 2012; Doha, Qatar p. 568-574 URL: https://link.springer.com/chapter/10.1007/978-3-642-34478-7_69 [doi: 10.1007/978-3-642-34478-7_69]
26. Loureiro B, Rodrigues R. Design guidelines and design recommendations of multi-touch interfaces for elders. In: Proceedings of the 7th International Conference on Advances in Computer-Human Interactions, ACHI. 2014 Presented at: ACHI '14; March 23-27, 2014; Barcelona, Spain p. 1-15 URL: https://pdfs.semanticscholar.org/4e5c/a76ed1ee52dc7f7c32a254d8cccaad43bdb5.pdf
27. Motti L, Nadine V, Gorce P. Interaction techniques for older adults using touchscreen devices: a literature review from 2000 to 2013. J Interact Pers-Syst 2013;3(2):123:1-12326 [FREE Full text] [doi: 10.46298/jips.3180]
28. Iancu I, Iancu B. Designing mobile technology for elderly. A theoretical overview. Technol Forecast Soc Change 2020 Jun;155:119977. [doi: 10.1016/j.techfore.2020.119977]
29. Pattison M, Stedmon A. Inclusive design and human factors: designing mobile phones for older users. PsychNology J 2006;4(3):267-284 [FREE Full text]
30. Paiva JO, Andrade RM, de Oliveira PA, Duarte P, Santos IS, Evangelista AL, et al. Mobile applications for elderly healthcare: a systematic mapping. PLoS One 2020 Jul 30;15(7):e0236091 [FREE Full text] [doi: 10.1371/journal.pone.0236091] [Medline: 32730266]
31. Nimmanterdwong Z, Boonviriya S, Tangkijvanich P. Human-centered design of mobile health apps for older adults: systematic review and narrative synthesis. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e29512 [FREE Full text] [doi: 10.2196/29512] [Medline: 35029535]
32. Chirayus K, Nanthaamornphong A. A systematic mapping review: mobile user interface design guidelines for the elderly with cognitive impairments. In: Proceedings of the 23rd International Computer Science and Engineering Conference. 2019 Presented at: ICSEC '19; October 30-November 01, 2019; Phuket, Thailand p. 35-42 URL: https://ieeexplore.ieee.org/document/8974698 [doi: 10.1109/icsec47112.2019.8974698]
33. Elguera PL, Zapata DR. Elderly users and their main challenges usability with mobile applications: a systematic review. In: Proceedings of the 8th International Conference on Design, User Experience, and Usability. 2019 Presented at: DUXU '19; July 26–31, 2019; Orlando, FL p. 423-438 URL: https://link.springer.com/chapter/10.1007/978-3-030-23570-3_31 [doi: 10.1007/978-3-030-23570-3_31]
34. Sarkis-Onofre R, Catalá-López F, Aromataris E, Lockwood C. How to properly use the PRISMA statement. Syst Rev 2021 Apr 19;10(1):117 [FREE Full text] [doi: 10.1186/s13643-021-01671-z] [Medline: 33875004]

35. Pereira A, Silva F, Ribeiro J, Marcelino I, Barroso J. Smart remote control design for seniors. In: Proceedings of the 9th International Conference on Universal Access in Human-Computer Interaction. 2015 Presented at: UAHCI '15; August 2-7, 2015; Los Angeles, CA p. 484-495 URL: https://link.springer.com/chapter/10.1007/978-3-319-20681-3_46 [doi: 10.1007/978-3-319-20681-3_46]
36. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
37. Sani Z, Petrie H. User evaluation of an app for liquid monitoring by older adults. In: Proceedings of the 11th International Conference on Universal Access in Human–Computer Interaction. 2017 Presented at: UAHCI '17; July 9–14, 2017; Vancouver, BC p. 86-97 URL: https://link.springer.com/chapter/10.1007/978-3-319-58700-4_8 [doi: 10.1007/978-3-319-58700-4_8]
38. Vasconcelos A, Silva PA, Caseiro JM, Nunes F, Teixeira LF. Designing tablet-based games for seniors: the example of CogniPlay, a cognitive gaming platform. In: Proceedings of the 4th International Conference on Fun and Games. 2012 Presented at: FnG '12; September 4-6, 2012; Toulouse, France URL: https://dl.acm.org/doi/10.1145/2367616.2367617 [doi: 10.1145/2367616.2367617]
39. de Almeida RX, Ferreira SB, Soares HP. Recommendations for the development of web interfaces on tablets/iPads with emphasis on elderly users. Procedia Comput Sci 2015;67:140-149 [FREE Full text] [doi: 10.1016/j.procs.2015.09.258]
40. Morey SA, Barg-Walkow LH, Rogers WA. Managing heart failure on the go: usability issues with mHealth apps for older adults. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):1-5 [FREE Full text] [doi: 10.1177/1541931213601496]
41. Morey SA, Stuck RE, Chong AW, Barg-Walkow LH, Mitzner TL, Rogers WA. Mobile health apps: improving usability for older adult users. Ergon Des 2019 Oct;27(4):4-13 [FREE Full text] [doi: 10.1177/1064804619840731]
42. Petrovčič A, Šetinc M, Burnik T, Dolničar V. A comparison of the usability of a standard and an age-friendly smartphone launcher: experimental evidence from usability testing with older adults. Int J Rehabil Res 2018 Dec;41(4):337-342. [doi: 10.1097/MRR.0000000000000307] [Medline: 30085996]
43. Abdullah N, Abdul Hamid NF. Interface design features of mobile application for senior citizens. Indones J Electr Eng Comput Sci 2019;14(1):436 [FREE Full text] [doi: 10.11591/ijeecs.v14.i1.pp436-442]
44. Silva T, Mota M, Silva C. Development and UI/UX testing of an iTV companion application for seniors. In: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and E-Health. 2019 Presented at: ICT4AWE '19; May 2-4, 2019; Crete, Greece p. 124-134 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0007744601240134 [doi: 10.5220/0007744601240134]
45. Li Q, Luximon Y. Older adults' use of mobile device: usability challenges while navigating various interfaces. Behav Inf Technol 2020;39(8):837-861 [FREE Full text] [doi: 10.1080/0144929x.2019.1622786]
46. Batu SY, Kim YH, Cheng H. Senior-friendly icon design for the mobile phone. In: Proceedings of the 6th International Conference on Digital Content, Multimedia Technology and its Applications. 2010 Presented at: IDCCMTA '10; August 16-18, 2010; Seoul, South Korea p. 103-108 URL: https://ieeexplore.ieee.org/document/5568719
47. Tsai TH, Tseng KC, Chang YS. Testing the usability of smartphone surface gestures on different sizes of smartphones by different age groups of users. Comput Human Behav 2017 Oct;75:103-116 [FREE Full text] [doi: 10.1016/j.chb.2017.05.013]
48. Siek KA, Ross SE, Khan DU, Haverhals LM, Cali SR, Meyers J. Colorado care tablet: the design of an interoperable personal health application to help older adults with multimorbidity manage their medications. J Biomed Inform 2010 Oct;43(5 Suppl):S22-S26 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.007] [Medline: 20937480]
49. Hwangbo H, Yoon SH, Jin BS, Han YS, Ji YG. A study of pointing performance of elderly users on smartphones. Int J Hum Comput Interact 2013;29(9):604-618 [FREE Full text] [doi: 10.1080/10447318.2012.729996]
50. Murano P, Ali RM. A preliminary investigation into the usability of force touch interaction in a mobile context. Int J Interact Mob Technol 2018;12(5):169-181 [FREE Full text] [doi: 10.3991/ijim.v12i5.9208]
51. Cornet VP, Daley CN, Srinivas P, Holden RJ. User-centered evaluations with older adults: testing the usability of a mobile health system for heart failure self-management. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):6-10 [FREE Full text] [doi: 10.1177/1541931213601497] [Medline: 30930610]
52. Holden RJ, Campbell NL, Abebe E, Clark DO, Ferguson D, Bodke K, Brain Health Patient Safety Laboratory. Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Res Social Adm Pharm 2020 Jan;16(1):54-61 [FREE Full text] [doi: 10.1016/j.sapharm.2019.02.011] [Medline: 30853507]
53. Barros AC, Leitão R, Ribeiro J. Design and evaluation of a mobile user interface for older adults: navigation, interaction and visual design recommendations. Procedia Comput Sci 2014;27:369-378 [FREE Full text] [doi: 10.1016/j.procs.2014.02.041]
54. Panagopoulos C, Kalatha E, Tsanakas P, Maglogiannis I. Evaluation of a mobile home care platform. In: Proceedings of the 12th European Conference on Ambient Intelligence. 2015 Presented at: AmI '15; November 11-13, 2015; Athens, Greece p. 328-343 URL: https://link.springer.com/chapter/10.1007/978-3-319-26005-1_22 [doi: 10.1007/978-3-319-26005-1_22]
55. Menghi R, Ceccacci S, Gullà F, Cavalieri L, Germani M, Bevilacqua R. How older people who have never used touchscreen technology interact with a tablet. In: Proceedings of the 16th IFIP TC 13 International Conference on Human-Computer

Interaction. 2017 Presented at: INTERACT '17; September 25–29, 2017; Mumbai, India p. 117-131 URL: https://link.springer.com/chapter/10.1007/978-3-319-67744-6_8 [doi: 10.1007/978-3-319-67744-6_8]
56. Kobayashi M, Hiyama A, Miura T, Asakawa C, Hirose M, Ifukube T. Elderly user evaluation of mobile touchscreen interactions. In: Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction. 2011 Presented at: INTERACT '11; September 5-9, 2011; Lisbon, Portugal p. 83-99 URL: https://link.springer.com/chapter/10.1007/978-3-642-23774-4_9 [doi: 10.1007/978-3-642-23774-4_9]
57. Burrows M, Dutton R, Schulze E, Zirk A. PIA - practical support for everyday life results of the user tests. In: Proceedings of the 2014 conference on eChallenges. 2014 Presented at: eChallenges '14; October 29-30, 2014; Belfast, UK p. 1-7 URL: https://ieeexplore.ieee.org/document/7058157
58. Leitão R, Silva P. Target and spacing sizes for smartphone user interfaces for older adults: design patterns based on an evaluation with users. In: Proceedings of the 19th Conference on Pattern Languages of Programs. 2012 Presented at: PLoP '12; October 19-21, 2012; Tucson, Arizona p. 1-13 URL: https://dl.acm.org/doi/10.5555/2821679.2831275
59. Nicolau H, Jorge J. Elderly text-entry performance on touchscreens. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility. 2012 Presented at: ASSETS '12; October 22-24, 2012; New York, NY p. 127-134 URL: https://dl.acm.org/doi/10.1145/2384916.2384939 [doi: 10.1145/2384916.2384939]
60. Neves BB, Franz R, Munteanu C, Baecker RM, Ngo M. "My hand doesn't listen to me!": adoption and evaluation of a communication technology for the 'oldest old'. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul Republic of Korea p. 1593-1602 URL: https://dl.acm.org/doi/10.1145/2702123.2702430 [doi: 10.1145/2702123.2702430]
61. Hill NL, Mogle J, Wion R, Kitt-Lewis E, Hannan J, Dick R, et al. App-based attention training: Incorporating older adults' feedback to facilitate home-based use. Int J Older People Nurs 2018 Mar;13(1). [doi: 10.1111/opn.12163] [Medline: 28752612]
62. Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K. App-based self-administrable clinical tests of physical function: development and usability study. JMIR Mhealth Uhealth 2020 Apr 27;8(4):e16507 [FREE Full text] [doi: 10.2196/16507] [Medline: 32338616]
63. Philips B, Smits C, Govaerts PJ, Doorn I, Vanpoucke F. Empowering senior cochlear implant users at home via a tablet computer application. Am J Audiol 2018 Nov 19;27(3S):417-430 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0014] [Medline: 30452746]
64. Antonelli HL, Watanabe WM, de Mattos Fortes RP. Adapting web menus to mobile devices for elderly interactions. In: Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2018 Presented at: DSAI '18; June 20-22, 2018; Thessaloniki, Greece p. 47-54 URL: https://dl.acm.org/doi/10.1145/3218585.3218593 [doi: 10.1145/3218585.3218593]
65. Conte S, Munteanu C. Help!: I'm stuck, and there's no F1 key on my tablet!. In: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. 2019 Presented at: MobileHCI '19; October 1-4, 2019; Taipei, Taiwan p. 1-11 URL: https://dl.acm.org/doi/10.1145/3338286.3340121 [doi: 10.1145/3338286.3340121]
66. Lee N, Seaborn K, Hiyama A, Inami M, Hirose M. Evaluating a smartphone-based social participation app for the elderly. In: Proceedings of the 4th International Conference on Human Aspects of IT for the Aged Population. 2018 Presented at: ITAP '18; July 15-20, 2018; Las Vegas, NV p. 505-517 URL: https://link.springer.com/chapter/10.1007/978-3-319-92034-4_38 [doi: 10.1007/978-3-319-92034-4_38]
67. Hölzl M, Schaffer C. An adaptive and book-oriented mobile touch screen user interface concept for novice senior users. In: Proceedings of the 2013 International Conference on Advances in Mobile Computing & Multimedia. 2013 Presented at: MoMM '13; December 2-4, 2013; Vienna, Austria p. 576-584 URL: https://dl.acm.org/doi/abs/10.1145/2536853.2536869 [doi: 10.1145/2536853.2536869]
68. Leung R, McGrenere J, Graf P. Age-related differences in the initial usability of mobile device icons. Behav Inf Technol 2011;30(5):629-642 [FREE Full text] [doi: 10.1080/01449290903171308]
69. Pino M, Granata C, Legouverneur G, Rigaud A. Assessing design features of a graphical user interface for a social assistive robot for older adults with cognitive impairment. In: Proceedings of the 29th International Symposium on Automation and Robotics in Construction. 2012 Presented at: ISARC '12; June 26 - 29, 2012; Eindhoven, the Netherlands URL: https://www.iaarc.org/publications/proceedings_of_the_29th_isarc/assessing_design_features_of_a_graphical_user_interface_for_a_social_assistive_robot_for_older_adults_with_cognitive_impairment.html [doi: 10.22260/isarc2012/0085]
70. Dosso C, Chevalier A. How do older adults process icons during a navigation task? Effects of aging, semantic distance, and text label. Educ Gerontol 2021;47(3):132-147 [FREE Full text] [doi: 10.1080/03601277.2021.1886634]
71. Isherwood SJ, McDougall SJ, Curry MB. Icon identification in context: the changing role of icon characteristics with user experience. Hum Factors 2007 Jun;49(3):465-476. [doi: 10.1518/001872007X200102] [Medline: 17552310]
72. Jonsson O, Haak M, Tomsone S, Iwarsson S, Schmidt SM, Martensson K, et al. Cross-national usability study of a housing accessibility app: findings from the European innovage project. J Usability Stud 2016 Nov 01;12(1):26-49 [FREE Full text] [doi: 10.1093/geroni/igx004.4867]

73. Piper A, Campbell R, Hollan J. Exploring the accessibility and appeal of surface computing for older adult health care support. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, GA p. 907-916 URL: https://dl.acm.org/doi/10.1145/1753326.1753461 [doi: 10.1145/1753326.1753461]
74. Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, et al. Enhancing home health mobile phone app usability through general smartphone training: usability and learnability case study. JMIR Hum Factors 2018 Apr 26;5(2):e18 [FREE Full text] [doi: 10.2196/humanfactors.7718] [Medline: 29699969]
75. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
76. Fang LS, Hong WM. High-aged Using Smart Phone Interface Gesture Operation Research. Adv Transdiscipl Eng 2018;7(40):231-238. [doi: 10.3233/978-1-61499-898-3-231]
77. Article 2 -definitions. Convention on the Rights of Persons with Disabilities (CRPD). New York, NY: United Nations; 2006 Dec 13. URL: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-2-definitions.html [accessed 2023-07-12]
78. Maeda J. The Laws of Simplicity. Cambridge, MA: MIT Press; 2007.
79. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
80. Kurniawan S, Zaphiris P. Research-derived web design guidelines for older people. In: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility. 2005 Presented at: Assets '05; October 9-12, 2005; Baltimore, MD p. 129-135 URL: https://dl.acm.org/doi/10.1145/1090785.1090810 [doi: 10.1145/1090785.1090810]
81. Harley A. Touch targets on touchscreens. Nielsen Norman Group. 2019. URL: https://www.nngroup.com/articles/touch-target-size/ [accessed 2022-01-19]
82. Joyce A. Help and documentation: the 10th usability heuristic. Nielsen Norman Group. 2020 Dec 13. URL: https://www.nngroup.com/articles/help-and-documentation/ [accessed 2022-01-19]
83. Leung R, Tang C, Haddad S, Mcgrenere JL, Graf P, Ingriany V. How older adults learn to use mobile devices: survey and field investigations. ACM Trans Access Comput 2012 Dec;4(3):1-33 [FREE Full text] [doi: 10.1145/2399193.2399195]
84. Villalba-Mora E, Ferre X, Pérez-Rodríguez R, Moral C, Valdés-Aragonés M, Sánchez-Sánchez A, et al. Home monitoring system for comprehensive geriatric assessment in patient's dwelling: system design and UX evaluation. Front Digit Health 2021 May 7;3:659940 [FREE Full text] [doi: 10.3389/fdgth.2021.659940] [Medline: 34713133]
85. Moral C, Pérez-Rodríguez R, Villalba-Mora E, Barrio-Cortes J, Ferre X, Rodríguez-Mañas L. Integrated health system to assess and manage frailty in community dwelling: co-design and usability evaluation. Digit Health 2023 Jun 20;9:20552076231181229 [FREE Full text] [doi: 10.1177/20552076231181229] [Medline: 37361432]
86. Ahmad B. Usability recommendations for designers of smartphone applications for older adults: an empirical study. In: Castro LM, Cabrero D, Heimgärtner R, editors. Software Usability. London, UK: IntechOpen; 2022.
87. Androutsou T, Kouris I, Anastasiou A, Pavlopoulos S, Mostajeran F, Bamiou DE, et al. A smartphone application designed to engage the elderly in home-based rehabilitation. Front Digit Health 2020 Sep 8;2:15 [FREE Full text] [doi: 10.3389/fdgth.2020.00015] [Medline: 34713028]
88. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Proceedings of the 1994 SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: CHI '94; April 24-28, 1994; Boston, MA p. 152-158 URL: https://dl.acm.org/doi/10.1145/191666.191729 [doi: 10.1145/191666.191729]
89. Soares Guedes L, Ribeiro CC, Ounkhir S. How can we improve the interaction of older users with devices? In: Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2020 Presented at: DSAI '20; December 2-4, 2020; Virtual Event p. 158-162 URL: https://dl.acm.org/doi/10.1145/3439231.3440611 [doi: 10.1145/3439231.3440611]
90. Huppert F. Designing for older users. In: Clarkson J, Keates S, Coleman R, Lebbon C, editors. Inclusive Design: Design for the Whole Population. London, UK: Springer; 2003:30-49.

Abbreviations

HCI: human-computer interaction
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UI: user interface
UX: user experience

Edited by L Buis; submitted 03.10.22; peer-reviewed by J Andrews, SB Ho; comments to author 10.04.23; revised version received 20.06.23; accepted 11.07.23; published 21.09.23.

Please cite as:
Gomez-Hernandez M, Ferre X, Moral C, Villalba-Mora E
Design Guidelines of Mobile Apps for Older Adults: Systematic Review and Thematic Analysis
JMIR Mhealth Uhealth 2023;11:e43186
URL: https://mhealth.jmir.org/2023/1/e43186
doi:10.2196/43186
PMID:37733401

Original Paper

Designing a Mobile e-Coaching App for Immigrant Informal Caregivers: Qualitative Study Using the Persuasive System Design Model

Shweta Premanandan[1], MBA; Awais Ahmad[2], MSc; Åsa Cajander[2], PhD; Pär Ågerfalk[1], PhD; Michal Dolezel[3], PhD; Lisette van Gemert-Pijnen[4], PhD

[1]Department of Informatics and Media, Uppsala University, Uppsala, Sweden

[2]Division of Visual Information and Interaction, Department of Information Technology, Uppsala University, Uppsala, Sweden

[3]Department of Information Technologies, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

[4]Department of Psychology, Health, and Technology, Faculty of Behavioral, Management and Social Sciences, University of Twente, Enschede, Netherlands

Corresponding Author:
Shweta Premanandan, MBA
Department of Informatics and Media
Uppsala University
Ekonomikum (Plan 3)
Kyrkogårdsgatan 10
Uppsala, 751 20
Sweden
Phone: 46 707689331
Email: shweta.premanandan@im.uu.se

Abstract

Background: Informal caregivers are vital in caring for their family and friends at home who may have illnesses or disabilities. In particular, the demands for caregiving can be even more challenging for those with limited resources, support systems, and language barriers, such as immigrant informal caregivers. They face complex challenges in providing care for their relatives. These challenges can be related to sociocultural diversity, language barriers, and health care system navigation. Acknowledging the global context of the increasing number of immigrants is essential in designing inclusive mobile health apps.

Objective: This study aims to investigate the needs of immigrant informal caregivers in Sweden and discuss the application of the Persuasive System Design Model (PSDM) to develop an e-coaching prototype. By addressing the unique challenges faced by immigrant informal caregivers, this study will contribute to the development of more effective and inclusive mobile health apps.

Methods: The participants were considered immigrants and included in the study if they and their parents were born outside of Sweden. Through various channels, such as the National Association of Relatives, rehabilitation departments at municipalities, and immigrant groups, we recruited 13 immigrant informal caregivers. These immigrant informal caregivers were primarily women aged 18 to 40 years. Most participants belonged to the Middle Eastern region whereas some were from North Africa. However, all of them spoke Arabic. We used semistructured interviews to gather data from the participants in Arabic, which were translated into English. Data were analyzed using thematic analysis and discussed in relation to the extended PSDM. The needs of the caregivers were compared with the description of persuasive design principles, and a design principle was chosen based on the match. The PSDM was extended if the need description did not match any principles. Several brainstorming and prototyping sessions were conducted to design the mobile e-coaching app.

Results: Immigrant informal caregivers have various needs in their caregiving role. They reported a need for training on the illness and future caregiving needs, assistance with understanding the Swedish language and culture, and help with accessing internet-based information and services. They also required recognition and appreciation for their efforts, additional informal support, and easy access to health care services, which can be important for their mental health. The PSDM was adapted to the informal caregiving context by adding "facilitating conditions" and "verbal encouragement" as additional persuasive design principles. This study also presents the subsequent mobile e-coaching app for immigrant informal caregivers in Sweden.

Conclusions: This study revealed important immigrant informal caregivers' needs based on which design suggestions for a mobile e-coaching app were presented. We also proposed an adapted PSDM, for the informal caregiving context. The adapted PSDM can be further used to design digital interventions for caregiving.

(JMIR Mhealth Uhealth 2023;11:e50038) doi:10.2196/50038

KEYWORDS

e-coaching; mobile health; mHealth; immigrant informal caregivers; designing app; persuasive system design; user needs; caregiver; app; design; users; aging; development; diversity; language barrier; inclusion; training; mental health; mobile phone

Introduction

Background

Informal caregivers provide care for their relatives and friends in the long term, often without proper training or experience [1]. Consequently, informal caregiving can take a toll on the well-being of informal caregivers (hereafter referred to as caregivers) and impact their ability to continue providing quality care [2]. Approximately 80% of long-term care in Europe is believed to be provided by informal caregivers, with women accounting for approximately two-thirds of this care provision [3]. In Sweden, 1 in 5 individuals is an informal caregiver [4], which is expected to increase owing to Europe's aging population. Hence, informal caregiving is crucial for traditional health care. It encompasses a broad range of support, including practical assistance such as domestic work and personal care, emotional support, and administrative tasks such as coordinating care and interacting with public authorities [5]. The impact of caregiving extends beyond the patients themselves, affecting the physical and psychological well-being of caregivers [6]. Many caregivers face a range of psychological challenges, including depression, anxiety, and posttraumatic stress [7]. Caregiving frequently brings about physical, emotional, and financial stress, impacting the caregiver's well-being and ability to provide quality care [8]. Caregiving demands can be especially challenging for those with limited resources or support systems, such as immigrant informal caregivers.

Several studies have highlighted the prevalence of informal caregiving in Sweden and its impact on caregivers' well-being, emphasizing the need for better support systems. Families are still the major providers of care for older adults in Sweden, suggesting that it is common to be an informal caregiver [9]. The study by Jegermalm [10] indicated that relatively few caregivers in Sweden have any kind of support aimed directly at them as caregivers. This lack of support is a critical issue that must be addressed to improve the well-being of informal caregivers. The research by Berglund et al [11] further supports this point, emphasizing that caregivers have worse perceptions of self-rated health and psychological well-being than noncaregivers in Sweden. Hence, it is important to recognize and support informal caregivers in Sweden. Caregivers in Sweden need recognition and support from the health care system and are interested in being considered as care partners [12]. Moreover, there are significant gender-related differences in the number of male and female caregivers based on different caregiving tasks [10]. In particular, studies have highlighted that women are more likely to be engaged in intensive caregiving, including personal care and other caregiving tasks. Simultaneously, males were more likely to provide practical assistance to mothers, neighbors, and friends. These gender-related differences highlight the pressing need to address caregivers' support needs, particularly women, who are often engaged in intensive caregiving tasks. Overall, these findings suggest that better support systems for informal caregivers in Sweden are essential to ensure that the social care system is better equipped to meet the needs of both caregivers and care recipients.

There is a growing immigrant population in Europe. In Sweden alone, with a total population of approximately 10 million, 102,000 people immigrated in 2022 [13]. Owing to a tougher sociopolitical climate, immigrant caregivers were forced to care for themselves to a greater extent with limited support from authorities and other public support systems. Immigrant informal caregivers often face complex challenges in providing care for their relatives. These challenges can be related to sociocultural diversity, language barriers, and navigating the health care system [14]. Galiana-Gómez de Cádiz et al [15] highlighted the heavy burden of care and limited respite from caregiving responsibilities that immigrant informal caregivers often face. Informal caregivers, particularly those from racial or ethnic minority groups, often provide high-intensity care without help from formal caregivers, experience unmet needs, and may benefit from culturally sensitive programs and policies [16]. Immigrant informal caregivers often face challenges in accessing support, services, and resources because of the gendered nature of care work and their immigrant social locations [17]. Moreover, health care and community service providers face difficulties when providing culturally appropriate care to immigrant families. These obstacles include language and communication barriers, as well as differences in the understanding of disability [18]. In Sweden, the welfare system is reported to perceive asylum seekers as individuals whose grievances need not be given due consideration and who are seen as subject to deportation. This perception has unfortunate consequences, as it hinders their ability to access formal care systems and creates a significant demand for informal care [5]. Immigrant informal caregivers often face competing priorities and beliefs, feel out of control, and need education and culturally tailored support systems [19].

Digital technologies have become an increasingly important resource for supporting informal caregivers in traditional health care. Mobile health (mHealth) refers to the use of mobile devices such as smartphones, tablets, and wearable technology in health care. It encompasses various apps and services designed to improve health care delivery, patient monitoring, and access to health-related information through mobile technology. Most mHealth apps for caregivers provide educational resources and information to caregivers. Some telecare technology apps may also use remote monitoring tools and devices, such as wearable sensors and video cameras, to support caregivers in monitoring the well-being of their loved ones [20]. This technology has been shown to improve the quality of life of caregivers and care recipients and reduce the burden on the health care system. mHealth apps have also been used to provide social support for caregivers. For example, internet-based forums and social media groups have been created to connect caregivers, allowing them to share their experiences and offer support and advice to one another [21]. Care coordination platforms provide a centralized location for care information, allowing real-time communication and collaboration among caregivers, health care providers, and other support individuals [22,23]. By streamlining care coordination and reducing the administrative burden, care coordination platforms can help reduce caregiving's overall stress and strain. This streamlining of care coordination can be particularly beneficial for caregivers responsible for coordinating care for multiple individuals or for juggling multiple care-related tasks. In addition, mobile apps provide various resources and tools to help manage caregiving duties such as scheduling and tracking medications, tracking care recipient health metrics, and managing finances. Mobile apps can help reduce the overall burden of caregiving by providing care-related information and resources at the caregiver's fingertips. Mobile apps can also improve the quality of care, allowing caregivers to access information and resources to help make care decisions.

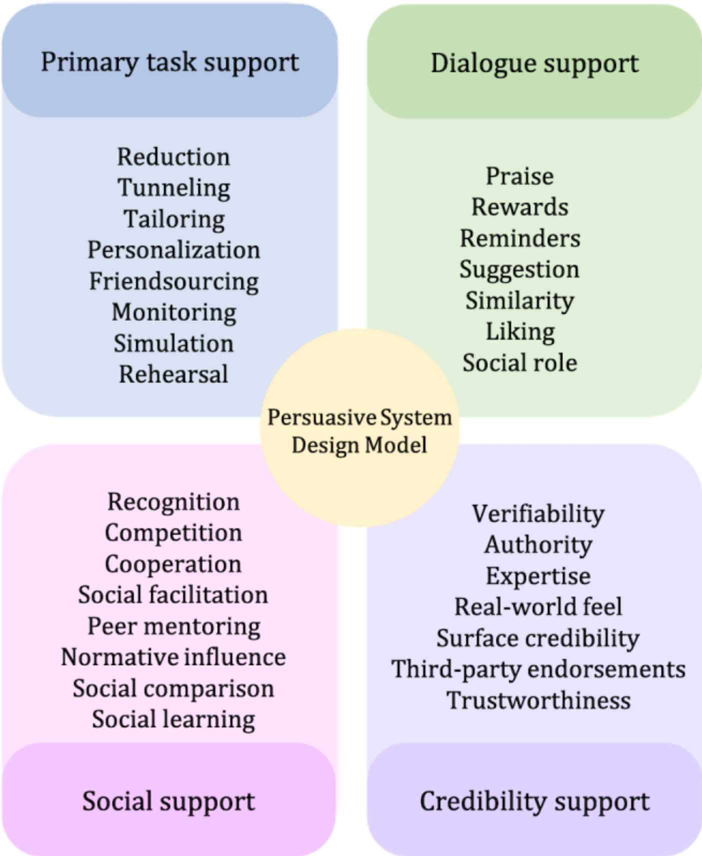
e-Coaching apps are a kind of mHealth app that provides users with coaching, support, and feedback through various methods, such as text-based conversations, audio and video calls, and interactive modules [24]. e-Coaching aims to help individuals improve their well-being, skills, and behaviors in a specific area or in general. e-Coaching apps can be used by individuals who want to achieve personal or professional goals or by organizations who want to support their employees. They offer the flexibility and convenience of being accessible anywhere and anytime, provided they have an internet connection. They also provide personalized support by adjusting the coaching content based on the user's needs and progress [25].

In addition, e-coaching apps can be cost-effective compared with traditional coaching methods and can reach a larger audience. These apps have been used in health, wellness, and personal and professional development. In the health care field, e-coaching apps have been used to support individuals with chronic conditions such as diabetes or to provide mental health support to individuals struggling with stress and anxiety [26]. In informal caregiving, e-coaching apps can provide support, resources, and education to caregivers looking after older adult family members, friends, or loved ones [27]. These apps can help caregivers manage stress, provide caregiving tips, and connect them with support networks and resources. By leveraging technology, e-coaching apps can offer scalable and accessible support to caregivers, helping them care for their loved ones while also improving their well-being.

Designing mHealth apps with a persuasive design approach can be highly beneficial in encouraging positive behavior change and promoting health in immigrant informal caregivers. Designing mHealth apps using a persuasive design approach can be achieved through several approaches, including the Persuasive System Design Model (PSDM). The PSDM was introduced in 2009 to provide a comprehensive method for developing apps that use persuasive design principles. By using theories of social psychology, persuasive design aims to inform, persuade, and convince people to adopt new behaviors [12]. In health care, a persuasive design can be used to encourage healthy behavior and potentially prevent or manage illnesses. The PSDM offers a systematic approach to designing engaging and applicable interventions that involves analyzing major aspects of persuasive systems, understanding the context, and designing system qualities. However, the existing literature suggests that PSDM needs to be adapted to incorporate the use and user context [28]. In a recent study, the PSDM was adapted to the context of informal caregivers [27]. As part of this adaptation, the "self-monitoring" principle was reformulated, and 2 additional principles, "friendsourcing" and "peer mentoring," were introduced to the PSDM. In this study, building upon this adapted model, we further extended it to incorporate the experiences of immigrant informal caregivers.

The adapted PSDM [27] proposes 30 design principles or strategies that are categorized into 4 dimensions: primary task support, dialogue support, credibility support, and social support. These design principles are grouped into four dimensions, as shown in Figure 1 [25], which is an adaptation from the study by Oinas-Kukkonen and Harjumaa [29]: (1) primary task support that helps users perform their target behaviors; (2) dialogue support that uses design principles that motivate users through feedback and interaction with the app; (3) credibility support using techniques that make potentially prevent or manage illnesses look and feel trustworthy to users; and (4) social support, which uses techniques that leverage social influence.

Figure 1. The adapted Persuasive System Design Model for informal caregiving.



Research Objective and Questions

Despite the growing immigrant population in Sweden, there is limited knowledge about the challenges immigrant informal caregivers face [30]. These caregivers may face unique stressors such as language barriers, cultural differences, limited support networks, and financial difficulties [30]. Navigating the health care system, understanding medical instructions, and accessing information on the web can be particularly challenging, leading to increased stress and confusion in caring for their relatives. Immigrant informal caregivers may also struggle with cultural differences in their caregiving approach and navigating cultural norms and values in a new country [30]. These stressors combined with limited social networks and financial insecurity can lead to feelings of isolation and loneliness. The burden of caregiving can be further compounded by the need to balance work and caregiving responsibilities, making it a complex and challenging experience for immigrant informal caregivers [8]. Hence, this study explores the needs and challenges faced by immigrant informal caregivers and presents the design of a persuasive mobile e-coaching app. This paper addressed the following research question:

- How can a persuasive design approach be used to design a mobile e-coaching app, addressing the needs and challenges that immigrant caregivers face in Sweden?

The research question was divided into 3 subquestions:

- What are the needs and challenges immigrant informal caregivers in Sweden face?
- How do the persuasive design model principles address these needs and challenges?
- How can persuasive design principles be incorporated into the design of an effective mobile e-coaching app for immigrant informal caregivers in Sweden?

This study explores the use of the adapted PSDM in the designing of a mobile e-coaching app to pursue the aforementioned research questions. To begin this process, we conducted interviews with Swedish immigrant informal caregivers to discern their specific needs. We then formulated design recommendations rooted in adapted PSDM principles. These insights serve as the foundation for proposing modifications to the PSDM, tailored to the distinctive needs of immigrant informal caregivers in Sweden. Finally, we have presented the design of the mobile e-coaching app. Ultimately, our objective is to tackle the multifaceted needs and challenges faced by this group by advocating for the development of an e-coaching app that seamlessly integrates persuasive systems design principles.

Methods

Overview

Our research adopted a qualitative approach, beginning with the exploration of caregivers' needs for a mobile e-coaching app. This research adopts an interpretive, evolutionary, and complementary ontological stance. Our research has been informed by a reflective and hermeneutic epistemological stance [31].

We gathered information on their caregiving experiences and context through semistructured interviews. This section outlines the participant recruitment process and data collection and analytical procedures. In addition, we describe the process of adapting the PSDM to the context of informal caregiving. We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Multimedia Appendix 1) to report this study's method and qualitative findings [32].

Study Design

Figure 2 illustrates a stepwise process of the study design described in this section.

Figure 2. Study design. PSDM: Persuasive System Design Model.

Participants and Recruitment

The study was conducted in Uppsala, Sweden, over 8 months from April 2022. The participants were considered immigrants and included in the study if they and their parents were born outside of Sweden. They were recruited through the National Association of Relatives (Anhörigas Riksförbund), rehabilitation departments at municipalities, advertisements on Facebook groups, and public libraries in areas with a higher concentration of immigrants. We also recruited participants through Selmagruppen, a nonprofit association of immigrant women in Sweden that aims to help immigrant women integrate with and understand Swedish society.

Information leaflets were distributed in these caregiver associations, with links to the study's page on the university's web page. Caregivers were given the option to either use the web-based registration form to input their information or contact the research team via phone or email. In addition, the snowballing technique was used to reach out to certain caregivers, where they were contacted through referrals from the participants who had already been interviewed [33].

We received interest from 16 immigrant informal caregivers willing to participate in the study. However, only 13 of them could be interviewed. Three of these caregivers could not participate in the study because of other commitments, despite their initial interest. Caregivers who were interested in participating in the study were from the Uppsala and Stockholm regions. The demographic data of the participants and their care recipients are presented in Table 1. We used the American Medical Association's age classification to determine the age of the caregivers and care recipients [34].

The research team did not have any relationship with the participants before the commencement of the study.

Table 1. Participants' information[a].

Age group of caregivers (years)	Sex of caregiver	Condition	Age group of care recipient (years)	Country
18-40	Female	Type 1 diabetes	65-90	Syria
18-40	Female	Autism	0-12	Morocco
18-40	Female	Autism	0-12	Iraq
18-40	Female	Autism and ADHD[b]	0-12	Yemen
18-40	Female	Autism and ADHD	0-12	Morocco
41-64	Female	ADHD and mental development delay	18-40	Egypt
18-40	Female	Glanzmann thrombasthenia	0-12	Syria
18-40	Female	Bone issues due to aging	65-90	Egypt
65-90	Male	Diabetes	65-90	Palestine
41-64	Female	Schizophrenia and hemiplegia	41-64	Syria
41-64	Female	Diabetes	41-64	Iran
18-40	Male	Diabetes	65-90	Syria
18-40	Female	Vestibular vertigo	41-64	Syria

[a]The care recipients are able to perform basic activities of daily living independently but are dependent on instrumental activities of daily living.

[b]ADHD: attention-deficit/hyperactivity disorder.

Procedure and Measures

A semistructured interview guide was used with open-ended questions [35] (Multimedia Appendix 2). SP, PÅ, and ÅC collaborated to create an interview guide approved by the Swedish Ethical Review Authority. The guide contained questions about informal caregiving, including caregiving tasks, experience, assistance from family and friends, respite care, and formal health care. It also included questions about the potential help and support needed by caregivers.

The interviews were conducted in Arabic by a research assistant employed at the university. The research assistant has had prior experience in conducting interviews for research projects. Before the interviews, SP had several mock interview sessions with the research assistant. Field notes were taken during the interviews by the research assistant, which were discussed with SP after the interviews to understand the participants better.

The interviews were conducted one-on-one and audio-recorded. After the interviews were completed, the research assistant transcribed and translated them into English for analysis. Although some informal caregivers preferred web-based videoconferencing [36] for convenience and flexibility, others preferred to do so in person at their homes. The interviews lasted between 60 and 75 minutes. The transcripts of the 13 interviews were reviewed by SP and ÅC, who thoroughly examined them and engaged in discussions regarding data saturation. On the basis of their analysis, it was concluded that data saturation had been sufficiently reached.

Interview Data Analysis

Data were pseudonymized and entered into qualitative data analysis software. Data were analyzed using the thematic analysis described by Braun and Clarke [35]. Following each interview, SP engaged in discussions with the research assistants to review the interview proceedings. During these discussions, SP took quick notes regarding the participant's context as well as any noteworthy and interesting thoughts associated with it. SP reviewed the data repeatedly to gain familiarity with it while taking notes of ideas. To examine the needs and challenges encountered by immigrant informal caregivers, an initial set of codes was systematically developed by SP. Subsequently, SP and ÅC analyzed these codes and looked for broader themes, gathering the most relevant data for potential themes. The broader themes were then reviewed and refined by SP and ÅC to ensure their importance to the research question. Quotes were linked to relevant themes, and the most important themes were selected and defined to explore the needs of caregivers for a mobile e-coaching app. Finally, the themes along with the researchers' interpretations and illustrative quotes have been described in the Results section. Data were stored and analyzed using the qualitative data analysis software MaxQDA. An inductive approach was used to explore caregivers' needs for the e-coaching app.

Elicitation of Design Suggestions and Adaptation of the PSDM

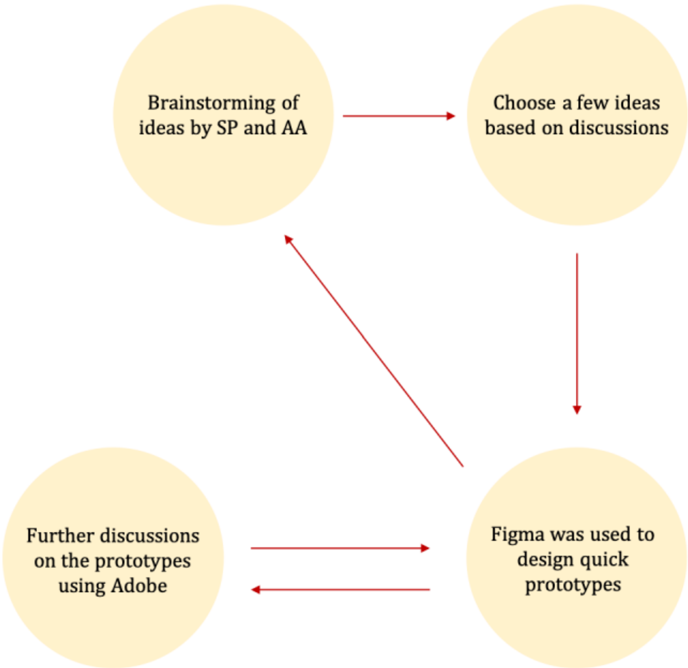
The caregivers' needs were mapped using the adapted PSDM. The description of the need was compared with the description of the persuasive design principles from the PSDM, and a design principle was chosen based on this match.

If the description of the needs did not map with any potential design principles from the PSDM, we described the design suggestion in detail and checked whether any design principles were close to it. Otherwise, we suggested a new design principle for the PSDM. Once we described the new design principle, we checked for any overlap with an existing design principle from the PSDM. Once no overlap was found, we formally added it as a design principle in the extended model.

Prototype Design

During the brainstorming sessions, SP and AA generated multiple ideas for the prototype, as illustrated in Figure 3. Careful attention was paid to the description of the informal caregivers' needs and adherence to the adopted persuasive system design principles to ensure that the resulting prototype would be responsive to users' requirements. To visualize and discuss the initial ideas, Figma, a prototyping tool [37], was used. However, as the prototype progressed, Adobe Illustrator was used to create esthetically appealing prototypes [38]. In the Results section, the main screenshots of the prototype that address the needs of informal caregivers are presented. These visuals demonstrate how the prototype meets specific requirements and showcase its features and user interfaces.

Figure 3. Prototype process.



Ethical Considerations

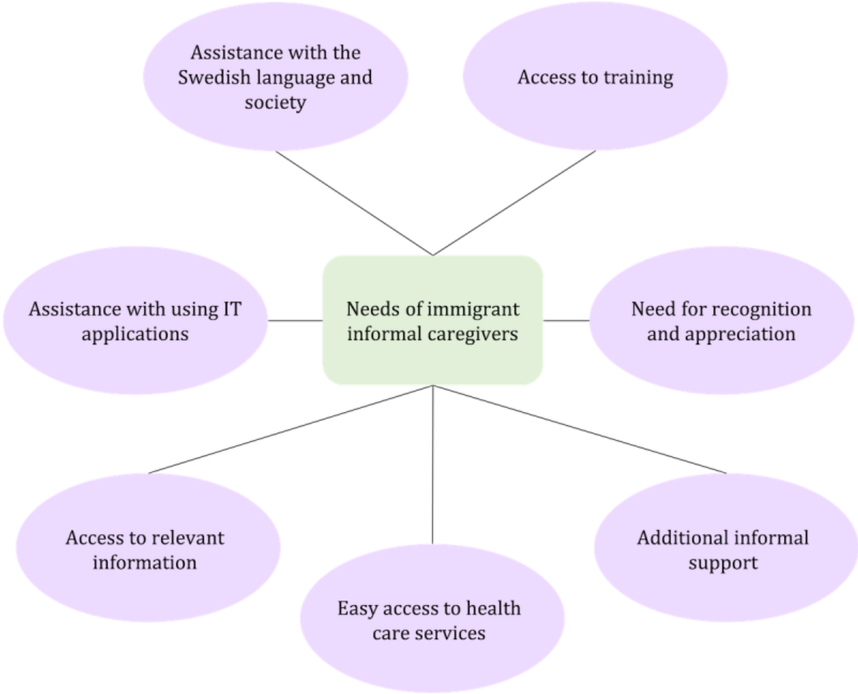
The Swedish Ethical Review Authority approved this research project (Dnr: 2021-03656), which was conducted with verbal and written information provided to participants. Written informed consent was obtained before the interviews. At least 2 weeks separated participants' receipt of information about the study and their consent to participate—enough time for them to consider their involvement. No relationship was found between the authors and the informal caregivers. Some participants experienced emotional distress or discomfort during this process. If this occurred, the interview was stopped to contact the informal caregivers' association for assistance.

Results

Overview

In this section, we reveal the findings from our empirical data and answer the research questions related to the design of the mobile e-coaching app. The first research question focused on identifying caregivers' needs through semistructured interviews. In this section, we describe the needs of immigrant informal caregivers. Figure 4 summarizes these needs. The second research question pertains to the design suggestions that can address the needs of immigrant informal caregivers identified in the first research question. We proposed design suggestions using the adapted PSDM for each requirement. These suggestions are based on our interpretation of the needs and understanding of the persuasive system design principles. Finally, to answer the third research question, we present the prototype and a visual representation of the proposed design.

Figure 4. Needs of immigrant informal caregivers.



Description of Needs

Access to Training

Most caregivers reported that they had taken some formative training courses on the illness, especially those who cared for a patient with autism. They reported that these intensive courses were aimed at teaching them to provide care to their care recipients, especially in the early years after diagnosis. In addition, some caregivers also expressed that they would like to have training on what to expect a few years later as the condition or the age of the care recipient or caregiver advances. They felt that such training could help them be better prepared in their caregiving process.

Some other caregivers (especially those caring for patients without autism) were not aware of such training courses but emphasized that access to such courses early on would certainly be helpful and enable them to be better prepared. Some caregivers felt that the formal health care system, at different stages, needs to anticipate challenges that caregivers may face and design courses that could help them:

> *I went to ["rehabilitation for autism"] to take care of autistic children and took training courses once every two weeks, meaning twice a month. I attended training courses for six weeks to learn more about autism and how to deal with my child to improve and develop.* [Morocco, aged 37 years, son with attention-deficit/hyperactivity disorder (ADHD) and autism]

> *At the moment, I need to cover his free hours and also I mean, I also want training courses that teach me how to deal with him as a teenager. /.../ When he turns eighteen how his relationship with the opposite sex will be and how I will explain to him what it means to deal with... (she is referring to girls with her gestures).* [Yemen, female, aged 44 years, son with ADHD and autism]

Assistance With the Swedish Language and Society

Most participant caregivers stated that they needed help with understanding the Swedish language. As most of the information and services from the counties are provided in Swedish, it becomes even more difficult for immigrant informal caregivers to find and understand these information and services. This creates an additional challenge for these caregivers who may already be dealing with the demands of caregiving while adjusting to a new country and culture. They feel that if they could learn the language, they would be able to search for information related to caregiving, which would ultimately help them better serve their relatives.

In addition, they feel that Swedish culture is quite different from their native culture in many aspects, especially when interacting with government agencies and traditional health care systems. Differences in communication styles, values, and expectations may lead to misunderstandings or challenges in navigating these systems and situations. Cultural differences in communication may also lead to difficulties in understanding and following medical instructions or in communicating with medical officials. Hence, they expressed that understanding the nuances of living in a society so different from their native society is challenging, especially when they are also caregivers:

> *I wish there were a program that would help me learn the language so that I could care and learn at the same time. To learn how to search for anything I want for my son and to learn the language to complete my studies in the same field because I chose to work in the same field and help people with special needs.* [Egypt, female, aged 48 years, son with ADHD and mental development delay]

Assistance With Using IT Applications

Most immigrant informal caregivers expressed that they had limitations in accessing web-based content and using web-based applications. Their internet use was limited to Google and YouTube, and they felt that this was insufficient, as most information and services in Sweden were hosted on the internet. They highlighted the need for assistance in using such IT applications. This assistance may include guidance on how to install apps from app stores and how to use them effectively. For example, caregivers may need assistance using health care apps to book appointments or access medical records.

Moreover, some of them have not used IT systems much. Hence, they would need support throughout the process of using an IT application, from the initial installation to achieving a desired outcome. This may involve technical support for troubleshooting and resolving any issues that arise while using the app. Most caregivers reported that they do not have experience using IT applications and would need help while using them to accomplish their goals:

> *I research diabetes and others, [using] my computer. [I] only use Google and not any other page. I honestly haven't tried any mobile applications [apps] or websites. I have no experience in it and have not tried to learn it. It takes time and I don't have so much time to learn this.* [Syria, male, aged 30-40 years, father with diabetes]

Need for Recognition and Appreciation

Most immigrant informal caregivers felt that they lacked a general acknowledgment of their efforts from immediate or extended family members. According to the interviewed female caregivers, they felt this was due to the hierarchical man-woman relationships in their culture. They felt that some degree of understanding, recognition, and appreciation would go a long way to boost their morale and help them continue to provide care to their relatives. Some also felt that such motivation could be good for their mental health, as they spend a lot of time wondering if they are doing things right by their relatives.

Some caregivers spoke about the sacrifices they had to make to care for their relatives, for example, missing out on certain job opportunities and ignoring their physical limitations. They felt that an acknowledgment of them from the care recipient and their extended family members could instill a renewed sense of purpose and motivation. They felt that this could also strengthen their familial bonds, especially in the extended family:

> *He is a lazy man at home! Because of, as I said, the men who come from the Middle East, they are like served in the home, and he has been like that from the beginning [this statement is the participant's own views and is not a generalized statement reflecting the Middle East]... I have to fix, clean, do a lot. Not just like putting out breakfast or food, it's cleaning, washing up, and laundry. So maybe I will prepare lunch for like half past one, twelve-thirty. And also that- he complains a lot. /... / he will sit and eat lunch with medication; he has special medications to take with food. He goes to bed, the TV is on, the laptop, mobile phone, Facebook, and so. But now it has become a little like I'm thinking about myself. I myself need some space, and that's what helps me, I have a lot of girlfriends who tell me I am doing good. It's like… you need that sometimes to... It helps to keep going.* [Iran, female, aged 63 years, husband with type 1 diabetes]

Access to Relevant Information

Most immigrant informal caregivers stated that it would be helpful to access relevant information in their native language. They have had difficulty accessing information in languages other than their own, particularly Swedish, when researching topics such as autism. They mentioned that they often read books about topics such as autism, hyperactivity, and disability but wished this information was more readily available in other languages.

They were also worried about the trustworthiness of the information. They expressed concern about the reliability of the information they found on the internet and read multiple sources to confirm its accuracy, as they fear encountering false or incorrect information. They stated that most of the time they spent was to confirm the accuracy of the information by visiting multiple websites or talking to other caregivers:

> *And it's also another thing that I don't know how it is to have information in all other languages. For example, if I want to find out information about autism it's difficult. Sometimes I read books on my own about hyperactivity or disability, in general, but I wish this information were available in other languages in Sweden, for example, it would have made many things easier… I write the question, and when I open it, several sites appear to me. Then I read the first three until I find that the speech applies to all of them. Because sometimes I'm afraid to take false information. Or wrong answers, for example. But I still can't tell if the information is 100% true.* [Egypt, female, aged 48 years, son with ADHD and mental development delay]

Additional Informal Support

Most immigrant informal caregivers felt alone in their caregiving journey in terms of feeling that they were the only ones providing care for their care recipients. This was particularly felt by female caregivers who balanced caregiving responsibilities with other activities such as studying or working. As a result, many female caregivers emphasized the need for support from other family members or friends when they have other activities to focus on, such as their education or a job.

The caregivers interviewed stressed the need for a joint coaching session with family members and health care providers to support the primary caregiver. This could involve bringing together family members, health care providers, and the primary caregiver to discuss the needs and challenges associated with caregiving. By working together, these stakeholders could develop a plan to support the primary caregiver and ensure that their relative's needs are met:

> *[If my friends have to look after him] The person who stays with him must be able to understand his character. Someone who [can] expect what my child can do. For example, if he is sitting at the table, I know he will fall and injure himself, and I know how impulsive he is. And he can cause a problem because they will say that he is a normal child, and he can cause a problem.* [Morocco, female, aged 34 years, son with autism and ADHD]

Easy Access to Health Care Services

Most immigrant informal caregivers experienced difficulty in accessing doctors or having medical assistance on time. They felt that they had to wait for long hours and would only be able to see a nurse for a short period, leading to another waiting time for a physician's appointment. This made them feel that their care recipient's health was not taken seriously.

In addition, some caregivers struggled to understand the Swedish health care and insurance system, making it difficult to navigate the system and access appropriate services. This lack of understanding may be compounded by comparisons with health care systems in their home countries, which may have been more familiar and relevant to them. There were frequent comparisons to the health care system, the benefits or aid they received back home, and how that was more relevant. Immigrant informal caregivers stressed that by providing timely and efficient health care services, caregivers may be better able to manage their care recipient's health needs and feel more supported in their caregiving role:

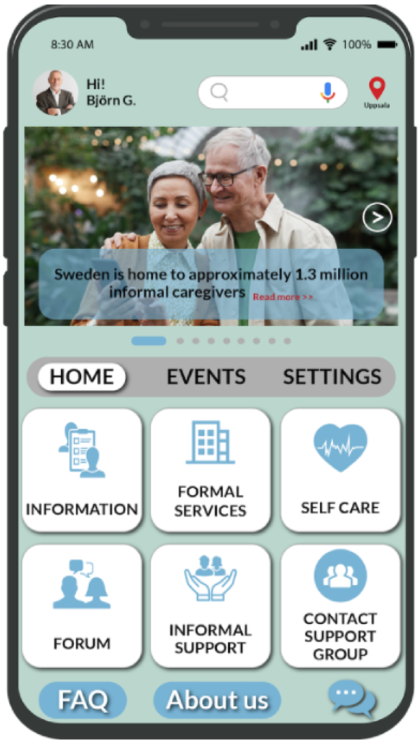
> *I would like something that can help get easy access to the doctors and the medical assistance because it is very hard to get that on time here in Sweden. It would also be nice if there is a file with all patient information and that is easy to access or a link or a website that helps the caregiver know what to do in an emergency.* [Morocco, female, aged 37 years, son with autism and ADHD]

Mobile e-Coaching App Design

Overview

In this subsection, we first present the design suggestions for the mobile e-coaching app using the adapted PSDM. Finally, we present an overview of the various functions of the mobile e-coaching app, outlining how each function is tailored to address the caregivers' needs. The main page of the app encompasses a range of functions, including "information," "formal services," "self-care," "forum," "informal support," and "contact support group," as depicted in Figure 5. We also delve into the practical implementations, demonstrating how caregivers' needs and PSDM design suggestions have been practically integrated into the aforementioned functions.

Figure 5. Home page of the app.



Access to Training

The "tailoring" design principle of PSDM can be applied to address this specific need, which is to provide caregivers with information on available training courses on the internet that are relevant to the illness and age of the patient, the relationship with the patient, the caregiver's age, and the caregiving stage that they are in. The app can personalize the information presented to the caregiver based on their location, care recipient's condition, caregiver's age, and stage of the caregiving process, considering their unique situation.

To further enhance the app's effectiveness, it is essential to anticipate caregivers' problems and provide them with relevant training sessions that will equip them to address future challenges. By doing so, caregivers can be adequately prepared to handle different scenarios in their caregiving journey, such as managing the patient's changing health conditions or coping with emotional and mental stress (Figures 5-7).

Figure 6. Information page.

Figure 7. Training courses.

On the main page of the mobile e-coaching app, we introduced an "Information" option (Figure 5). This option provides access to all relevant information about caregiving that can be useful for caregivers (as depicted in Figure 6). Within this "Information" page, among other options, caregivers can view information on the training courses available to them. We present information about 2 distinct types of courses: one facilitated by the state and the other by private organizations (as depicted in Figure 7). Here the caregiver is providing care for "autism" and hence can readily access courses tailored to their needs, such as "Autism Awareness for Parents and Caregivers." This is an implementation of the "tailoring" design principle.

Assistance With the Swedish Language and Society

The PSDM's "suggestion" design principle can significantly enhance the effectiveness of an app by providing caregivers with various options and resources to assist them in their tasks. To this effect, the app can offer a variety of resources, such as links to useful websites and books, tips and advice, and a glossary of Swedish medical terms and their meanings. This can help caregivers stay informed and be equipped to effectively manage their caregiving responsibilities.

Furthermore, the app can also incorporate the "social learning" design principle, which can facilitate interaction and community building between Swedish native and immigrant informal caregivers. This can be achieved by providing a web-based platform or space where caregivers can connect, share their experiences and knowledge, and learn from each other. This feature can be especially beneficial for immigrant informal caregivers who are unfamiliar with the Swedish language and culture. By engaging in conversations with native Swedish caregivers, they can gain insight and knowledge about the country's society and language, making it easier for them to integrate into the community. The provision of this web-based platform or space in the app can also provide a platform for caregivers to connect and network. Caregiving can be an isolating experience and having a community of like-minded individuals can be a resource for support and advice.

We designed a dedicated section to assist the caregivers in the Swedish language and society (Figure 8). They can access this section from the "Information" page (Figure 6). This section encompasses resources that include medical terms, information on Swedish language courses, and an array of learning materials. Notably, the app also includes an internet-based "Forum" (depicted in Figure 9), fostering connections among caregivers, especially with Swedish caregivers. This platform provides an opportunity for collaboration, networking, and cultural exchange and facilitates a smoother integration process.

Figure 8. Swedish language page.

Figure 9. Forum page.

Assistance With Using IT Applications

Immigrant caregivers have shown a limited understanding of using IT applications and therefore need assistance for a complete user journey, from installing the app from the App Store to obtaining some potential benefits from the app. Such support could include web-based tutorials, help files, and technical assistance. By providing such assistance, immigrant informal caregivers may be better able to access important services and information on the internet, ultimately assisting them in providing care and support to their relatives.

To assist caregivers better in understanding and using the app, an advanced level of support can be provided. Here, we introduce a new design principle to the PSDM called "facilitating conditions." It provides support structures that help users effectively navigate and use the system. The implementation of this design principle aims at making the app easier to use.

This feature can be especially beneficial for caregivers with limited technology experience or who require additional assistance in navigating the app's features. The advanced level of support can also aid caregivers in understanding the app's functionalities, ensuring that they can effectively use the system to manage their caregiving responsibilities.

Within the app, we implemented 2 particular functions: frequently asked question (FAQ) and "menu-based" or "decision tree" chat. These functions were intentionally designed for optimal accessibility and usability, remaining accessible from any page or section of the app (as illustrated in Figures 10 and 11). The FAQ page serves as a repository of general information about the app's functionalities. For instance, users can find guidance on tasks, such as adjusting language preferences within the app. The menu-based chat provides users with helpful instructions and assistance based on predefined options. These 2 functionalities combine to create a user-friendly support system, ensuring that users can easily navigate the app and receive the help they need in a manner that suits their preferences.

Figure 10. Frequently asked question page.

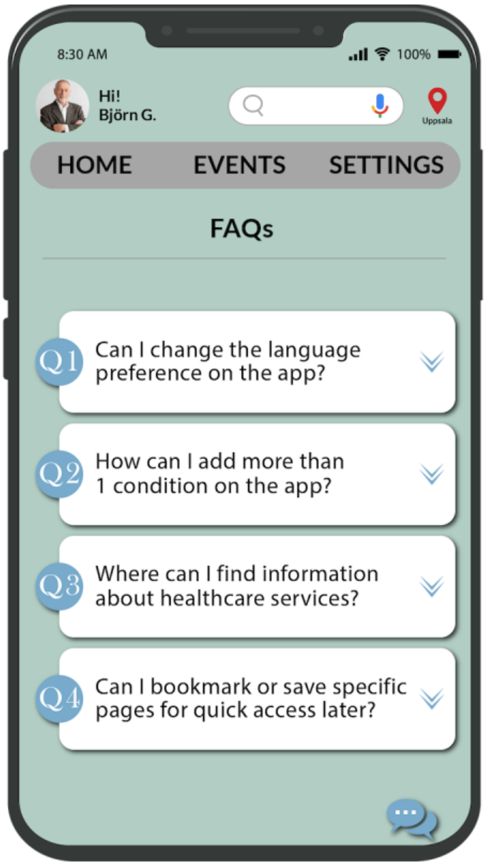
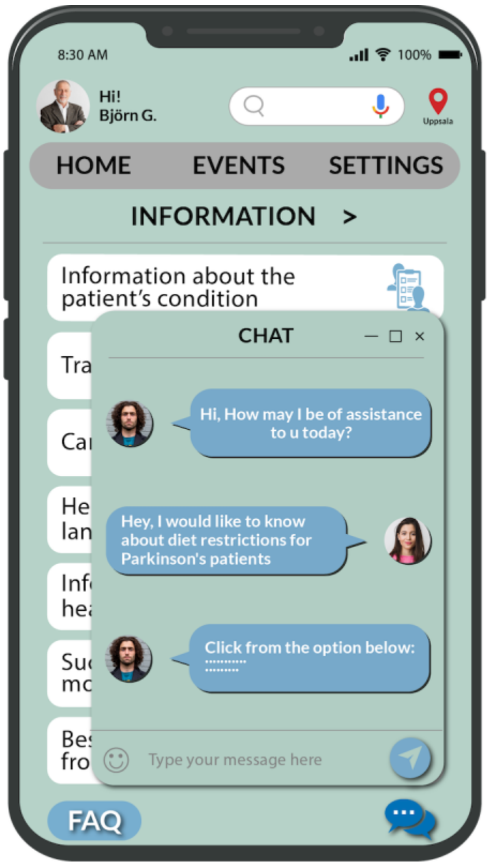


Figure 11. Chat page.



Need for Recognition and Appreciation

None of the 30 PSDM principles can be used to address this need of recognition and appreciation directly. Hence, we introduced a new design principle called "verbal encouragement." Its aim is to provide motivational messages and encouragement to continue in the users' tasks. Verbal encouragement can be implemented in the app by providing caregivers with inspirational and motivational experiences, best practices, or relevant tips from other caregivers. This could serve as a source of inspiration and validation for caregivers who feel overwhelmed or unappreciated (Figure 9).

In this e-coaching app, we developed a "Forum" space that offers caregivers a platform to share their experiences and situations. This forum, as illustrated in Figure 9, is not moderated and provides caregivers with a platform to share their experiences, ask questions, or simply vent their frustration. Caregivers can provide tips and advice based on individual caregivers' specific needs and challenges. Caregivers can also share anecdotes about what motivates them in their caregiving journey. They can use the "What motivates you?" tab. By sharing their stories and engaging in discussions, caregivers can gain a sense of validation and acknowledgment, which is an implementation of the "verbal encouragement" design principle. Ultimately, this may contribute positively toward their well-being and foster a sense of community through the app.

Access to Relevant Information

In the mobile e-coaching app, all information presented to caregivers will be sourced from credible and official web pages to ensure its accuracy and trustworthiness. Through this, the "trustworthiness" design principle of the PSDM will be implemented here to instill confidence in caregivers regarding the reliability of the information provided. The system will display the information source (as depicted in Figure 7), providing caregivers with transparency regarding where the information comes from, further enhancing the app's trustworthiness.

To make the information more relatable and engaging for caregivers, the "similarity" design principle of the PSDM will also be used. This principle involves presenting information in a way that imitates the user in a specific manner, creating a more personalized experience. Caregivers will be presented with information in their language and using texts and pictures that they can identify and relate to. By presenting relatable information, caregivers are more likely to engage with the app and feel more comfortable using it.

Additional Informal Support

In this e-coaching app, we provide a provision to schedule joint coaching sessions for family members and close friends with the health care professional. In the app, caregivers can also provide a quick summary of information that family and friends can use to pitch in. Caregivers can present a summary of the care recipient's condition and any notes the caregiver has entered related to the care recipient's illness, such as a quick fact sheet, on how to deal with unexpected situations in their absence.

We have also used the "Friendsourcing" principle [27] that motivates individuals such as family and friends in the caregiver's community or care volunteers to provide informal support to them. Caregivers can create a list of tasks in their account, and these care volunteers can choose the tasks they can assist with. This feature aims to unite caregivers and their care volunteers on a single platform to work together on tasks. On the basis of their previous contributions, the app can also recommend tasks to family and friends (Figures 12-14).

In this e-coaching app, we provide a provision to schedule joint coaching sessions for family members and close friends with the health care professional on the health care services page (Figure 15). Caregivers can choose to offer a summary of the care recipient's condition (such as a quick fact sheet) to their care volunteers. In addition, they can provide notes containing essential and pertinent details about caring for the individual and instructions on handling unforeseen situations when they are not around (Figure 14).

Caregivers also have the provision to share noncaregiving tasks with care volunteers when they need additional support (Figure 13). The tasks that the caregivers need repeated help with are stored in the app and provided as suggestions to begin with, as depicted in Figure 12. To access these lists of tasks, a caregiver's friends or family would need to log in to the app as a care volunteer through an invitation from the caregiver.

Figure 12. All tasks page.

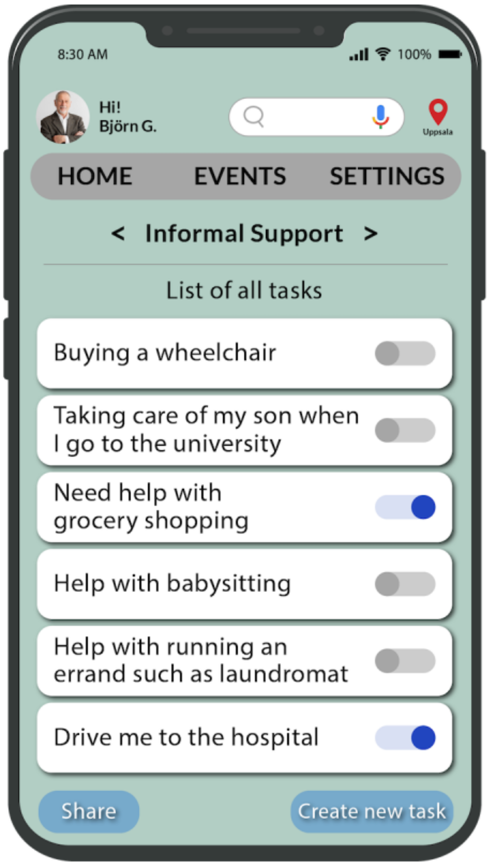


Figure 13. New task page.

Figure 14. Patient information page.

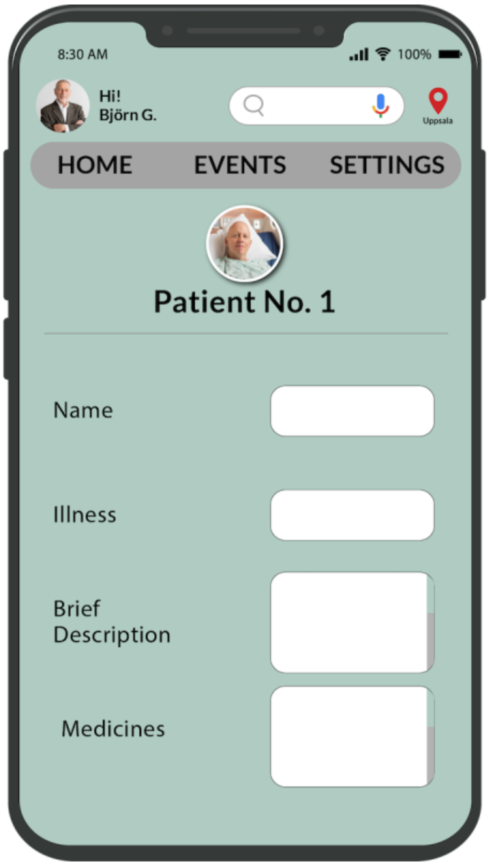


Figure 15. Health care services page.

Easy Access to Health Care Services

One way to apply the PSDM to address this need is to use the principle of "tailoring" to provide caregivers with personalized information about health care services and how to access them on the "Information" page. This could be done by gathering information about the caregiver's location, the care recipient's medical needs, and preferred language and then providing customized information about relevant health care services that meet those criteria.

Another persuasive system design principle that could address this need is "social learning." Caregivers could be provided with opportunities to connect with other caregivers through the e-coaching app on the "Forum" page, allowing them to share information and support each other as they navigate the health care system. This could help alleviate some of the stress and frustration associated with accessing health care services and could provide caregivers with a sense of community and belonging.

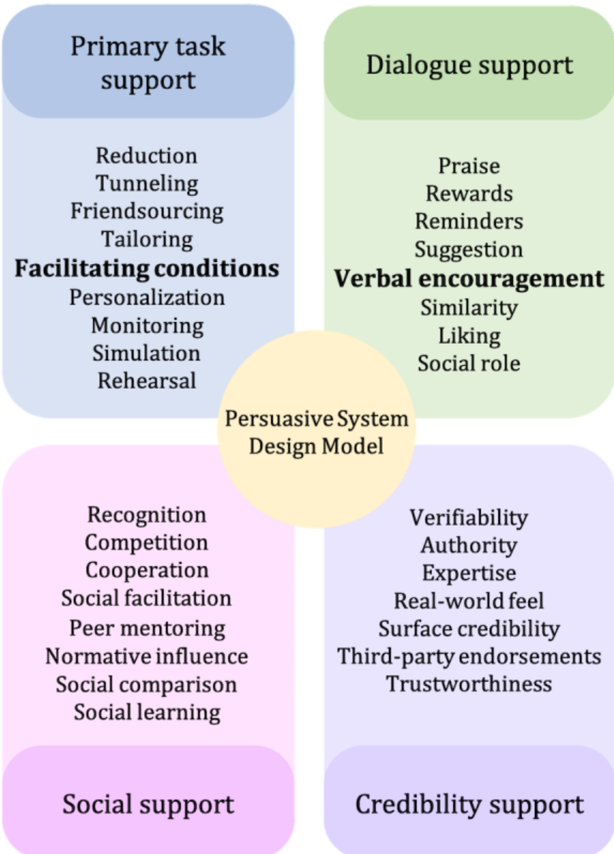
To ensure easy access to information about various health care services, we introduced a dedicated page titled "Healthcare Services" (as depicted in Figure 15). On this page, we offer information tailored to parameters such as care recipient's needs, caregiver's location, and care recipient's age. For instance, caregivers can find details on tasks, such as scheduling appointments with therapists, arranging personal assistance, and acquiring home care services.

Adaptation of the PSDM

Some of the needs, such as *assistance with using IT applications* and *the need for recognition and appreciation*, could not be mapped to any specific existing design principles in the PSD model. Hence, we extended the PSDM for informal caregivers, as illustrated in Figure 16. For the need *assistance with IT applications*, caregivers wanted to be able to receive assistance while using the IT application. This design principle of persuading the user by providing a support structure or help in the IT application has not been addressed in the existing PSDM. Hence, we introduced a design principle called "facilitating conditions." In "facilitating conditions," users can find help files or take the assistance of the menu-based chat to have help while using the IT application.

In addition, for the *need for recognition and appreciation*, caregivers felt the need to be recognized and appreciated for the work they were doing. This kind of persuasion technique is not addressed in the existing PSDM; hence, we introduce the design principle of "verbal encouragement."

Figure 16. Persuasive System Design Model for informal caregiving (PSD-I).



Discussion

Principal Findings

This study aimed to identify the needs of immigrant informal caregivers in Sweden and propose design suggestions and a prototype for a mobile e-coaching app to support these caregivers. These informal caregivers provide care for various medical conditions, such as autism, diabetes, ADHD; therefore, the aim is to discuss the common needs of caregivers. Hence, we have designed a generic mobile e-coaching app that may limit the specificity of information and coaching that can be provided compared with an e-coaching app focused on a specific condition. Through various channels, such as the National Association of Relatives, rehabilitation departments at the municipalities, and immigrant groups, we recruited 13 immigrant informal caregivers. These immigrant informal caregivers were mostly women and belonged to the age of 18 to 40 years. Most of the immigrant caregivers belonged to the Middle Eastern region, whereas some were from North African countries. However, all of them spoke Arabic. On the basis of the qualitative findings, 7 needs were identified: training access, language and societal assistance, IT application assistance, recognition and appreciation, access to relevant information, informal support, and easy access to health care services. The adapted PSDM [27] was used to propose design suggestions to address these needs. Although the PSDM is a useful tool for designing interventions, this study found that it needs to be adapted to fit the context of immigrant informal caregivers. As a result, the researchers added new design principles, namely facilitating conditions and verbal encouragement, to the PSDM, specifically for informal caregiving, PSD-I. The PSD-I model proposed in this study can guide information system researchers and designers in developing inclusive IT-based support interventions for informal caregivers. We do not claim to generalize statistically to “all immigrant caregivers” but rather apply our findings to theory [39]. Further studies are needed to generalize this result to a wider population.

This study also underscores the significance of early involvement and engagement of the ultimate end users, the caregivers, in the design and development process of mHealth apps. IT systems have shown considerable potential as efficient and cost-effective tools for delivering health care in the home environment. Nevertheless, the uncertainty surrounding their acceptance and adoption by end users calls for further investigation into the procedures and processes essential for the successful adoption of these systems [40-42]. Recognizing the well-being of end users and the genuine involvement of users in designing mHealth apps have emerged as pivotal factors in ensuring technology acceptance [42]. Within the scope of this study, we explored the circumstances and contexts of immigrant caregivers, valuing their unique needs and perspectives. This understanding serves as the foundation for designing and developing a mobile e-coaching app that is finely attuned to support these caregivers in their caregiving roles. As it aligns with existing research, we emphasize that gaining an understanding of caregivers and their needs is beneficial at the beginning of the design process. This contributes to the field of information system design.

Our findings indicate that immigrant informal caregivers value and need training for caregiving activities, which is in line with past research that has highlighted the importance of providing training and support to caregivers. For instance, a review by Sörensen et al [43] found that caregiver interventions, including psychoeducation and skills training, were associated with improved well-being, patient outcomes, and health care use. Similarly, a meta-analysis by Cheng et al [44] emphasized that caregiver interventions, including education and training, improve caregiver knowledge, self-efficacy, and mental health outcomes.

Moreover, studies have shown that tailored interventions may be more effective than generic interventions in improving caregiver outcomes than generic interventions [45]. Therefore, the application of the “tailoring” design principle suggested aligns with previous research, highlighting the importance of providing tailored interventions and support to caregivers. By providing personalized information and training sessions that address the unique needs and challenges of individual caregivers, the proposed app can improve caregiver outcomes and enhance preparedness in their caregiving journey.

On the basis of our findings, immigrant informal caregivers need assistance with the Swedish language and society. In countries with multicultural populations, extensive evidence indicates the importance of cultural and linguistic competence in health care. However, in Sweden, despite the growing multiculturality, only a limited number of studies have specifically addressed the challenges related to cross-cultural care [46]. Immigrant female family caregivers are known to avoid certain formal services for a variety of reasons, including the lack of cultural sensitivity and language issues [47]. On the other hand, health care providers also are often known to encounter language and communication issues when providing care to immigrant informal caregivers [18]. Moreover, immigrant informal caregivers face cultural differences between their native culture and Swedish culture, which further complicates their caregiving experiences. Interacting with government agencies and health care systems may be particularly challenging owing to differences in communication styles, values, and expectations [48]. These cultural differences can lead to misunderstandings, difficulties in following medical instructions, and ineffective communication with medical officials [49]. Immigrant informal caregivers highlighted the need to understand the nuances of living in a society that is different from their own, especially while fulfilling their caregiving responsibilities [50]. The findings suggest that health care and social service providers need to be aware of the unique challenges faced by immigrant informal caregivers and develop strategies to address their needs. Language interpretation services can empower immigrant informal caregivers to access information and services more effectively [18]. Culturally sensitive and inclusive approaches are essential for health care providers to bridge the communication and cultural gaps, ensuring effective collaboration and understanding between immigrant informal caregivers and health care professionals [51]. To improve the experiences of immigrant informal caregivers, the literature suggests the importance of tailored support services that consider the specific needs and challenges faced by this population [52]. Collaborative efforts between health care providers, community organizations, and policy makers are necessary to develop policies and programs that address the unique needs of immigrant informal caregivers and ensure equitable access to health care and social services [47]. In the e-coaching app, important links to web pages to learn the language and culture can be useful. This is addressed using the design principle of “suggestion.”

Our study and other related research also highlight that immigrant informal caregivers have limited knowledge and experience and little support for IT applications, and they need assistance in installing and effectively using them [53,54]. Their internet use is mainly limited to Google and YouTube, which they feel is insufficient as most services in Sweden are on the internet [55]. Limited experience with mHealth apps necessitates support throughout the process [54]. Immigrant informal caregivers' lack of familiarity with IT applications underscores the significance of assisting in helping them accomplish their goals. By facilitating a complete user journey from installing apps from the App Store, caregivers can better access essential services and information through the use of the e-coaching app. In the existing PSDM, no principle specifically addresses the provision of support structures [29]. Therefore, we extended the PSDM by adding a design principle, "facilitating conditions." Support functions such as FAQs, help files, and menu-based chats can help enhance the app's persuasiveness.

This study also highlighted that immigrant informal caregivers, especially women, often lack acknowledgment from family members, impacting their morale and mental health. Recognition and appreciation are crucial for their motivation and well-being as they navigate the challenges of caregiving [6]. Therefore, we introduced the "verbal encouragement" principle to address caregivers' needs, which was not present in the existing PSDM. It provides inspirational experiences, best practices, and tips for other caregivers. Caregiving might be emotionally and mentally draining, and the added stress of feeling unappreciated can exacerbate mental health issues [56,57]. Recognizing and appreciating caregivers can serve as a great source of motivation, especially when they face the demands and challenges of caregiving [6,58]. It is also perceived as beneficial for their mental well-being as they often question whether they adequately meet the needs of their relatives. In our app, we propose inspirational content, best practices, and a forum for caregivers to share their experiences and seek advice, addressing their need for acknowledgment and support.

Many immigrant informal caregivers expressed the need for access to information available in their native language, as they encountered difficulties when information was primarily provided in languages they were less proficient in, such as Swedish [50]. In addition to language barriers, immigrant informal caregivers also expressed concerns about the trustworthiness and reliability of the information they found [50,51]. Collaborative efforts among health care organizations, community support services, and culturally diverse caregiver networks can contribute to the development and dissemination of culturally sensitive information resources in multiple languages [59]. In our study, we use the "trustworthiness" design principle. Previous research has indicated that the trustworthiness of the information presented is an important concern for users [60]. Addressing this concern can increase the persuasiveness of the app.

One prominent issue identified in our study is the lack of informal support and the burden of sole responsibility carried by immigrant informal caregivers, particularly female caregivers. Many immigrant informal caregivers expressed the need for additional support from their family members or friends, especially when they have other obligations, such as studying at the university or working. This finding aligns with studies emphasizing the multiple roles and responsibilities that immigrant women often fulfill within their families [49,61]. These caregivers feel overwhelmed by the demands of caregiving, which can hinder their ability to engage in other important activities or maintain a healthy balance in life. To address this issue, most immigrant informal caregivers suggested implementing joint coaching sessions involving family members, health care providers, and the primary caregiver. These sessions aimed to create a supportive network that can share the caregiving responsibilities and provide respite for the primary caregiver. By actively involving family members in discussions about caregiving responsibilities, challenges, and needs, they can gain insights into the daily struggles faced by the primary caregiver. This increased awareness can promote shared responsibility and encourage family members to provide support and assistance when needed [49,61]. Health care providers play a crucial role in facilitating these joint coaching sessions. They can provide guidance, practical advice, and access to appropriate resources. It is important to note that these joint coaching sessions should be culturally sensitive and linguistically appropriate to ensure effective communication and participation among all stakeholders [49,50,52].

Our findings indicate that most immigrant informal caregivers experienced difficulty accessing doctors or having medical assistance on time. The difficulties immigrant informal caregivers face in accessing timely health care services emerge as a recurring theme across several studies [48,50,52]. The literature emphasizes the importance of providing timely and efficient health care services to caregivers and their care recipients [49,51]. By addressing the challenges immigrant informal caregivers face in accessing health care, health care systems can better support these caregivers in their caregiving role. Accessible and culturally sensitive services can contribute to their ability to manage their care recipient's health needs effectively and enhance their overall well-being. Another significant barrier to accessing appropriate health care services faced by immigrant informal caregivers is their limited understanding of their host countries' health care and social welfare systems in their host countries [48,61]. Navigating complex health care systems becomes particularly challenging for these caregivers, who may come from different cultural backgrounds with different health care practices and beliefs. This lack of understanding is further exacerbated by comparisons made with the health care systems in their home countries, which are often perceived as more familiar and relevant to their needs [62]. Health care professionals and policy makers should recognize and address the unique needs and challenges faced by immigrant informal caregivers. Culturally sensitive approaches, targeted information and support, interpreter services, and simplified explanations of the health care system and insurance policies can help alleviate the difficulties faced by immigrant informal caregivers in accessing and navigating health care services [59]. Existing research on persuasive design suggests that design principles such as "social comparison" and "competition" are commonly favored in similar apps [60]. However, considering that caregivers are vulnerable users, implementing these design principles may exacerbate their stress levels. In this regard, this study proposes an alternative approach, advocating for "social learning" as a means for caregivers to collaboratively share and learn from each other's caregiving experiences in their caregiving journey.

Limitations and Future Research

This study's findings should be considered in light of these limitations. First, we proposed an extended PSDM called PSD-I. A limitation of this study is that this extension is based on interviews with a group of immigrant caregivers that may not be fully representative, as we mostly received participants from the Middle Eastern region. More empirical evidence to explore the implications of PSD-I could contribute to the validity of this model in other contexts. In our study, we observed no discernible differences in caregiver needs when considering the age of the care recipient or the nature of the condition being cared for. Hence, we have put forth design recommendations for a general e-coaching app, albeit acknowledging that this approach may restrict the depth of information and coaching available compared with an e-coaching app tailored to a particular condition. As a result, future research efforts should concentrate on developing specialized e-coaching apps catering to specific caregiving scenarios, such as those involving young adult caregivers or care recipients. In addition, there are other caregiving contexts, such as stroke and cancer care, which we were unable to include in our study because of the unavailability of relevant caregiver data. Subsequent research endeavors could aim to improve the inclusivity of these underrepresented caregiving scenarios. Self-selection bias may also affect the sample's representativeness, as those who are more motivated to participate may have different experiences or needs than those who do not participate. To gain a more comprehensive understanding of the needs and experiences of immigrant informal caregivers in Sweden, further research with a larger sample size is needed. Further research is needed to determine the applicability of these results in other populations and settings.

Conclusions

In this study, we shed light on the significant needs of immigrant informal caregivers and provide insights for the design of mHealth apps in the caregiving domain. Our findings lead to several key conclusions regarding the needs of immigrant informal caregivers. First, it is apparent that these caregivers would greatly benefit from training programs designed to prepare them to care for their patients. This would help them develop the necessary skills and knowledge to perform their caregiving duties effectively. Second, the study found that many of these caregivers struggled with understanding the Swedish language and culture and therefore needed assistance. Third, the use of mHealth apps to access information and services is an area where these caregivers need support, as they may not be familiar with these systems. Fourth, the study revealed that many of these caregivers lacked recognition and appreciation for their work, which is a concern that needs to be addressed. Finally, access to relevant information about caring for their patients is essential for these caregivers, as it can help them provide better care and support. In addition, support is needed to help caregivers manage the stress and mental health impact of caring, as it can significantly impact their well-being and ability to perform their duties.

This study also proposes an extended and adapted PSDM for informal caregiving called PSD-I by including persuasive design principles of "facilitating conditions" and "verbal encouragement." This adapted model provides a structured approach for designing interventions that effectively support and empower informal caregivers in their caregiving roles. By applying the design principles of PSD-I, we present the design of a mobile e-coaching app that addresses immigrant informal caregivers' specific needs and challenges in Sweden. We also describe the implementation of these PSDM design principles for developing the e-coaching prototype. This prototype is a practical demonstration of how the PSD-I model can be effectively used to create tailored interventions for informal caregivers. PSD-I may serve as a resource for researchers and designers aiming to create impactful digital apps and services specifically tailored to the needs of the caregiving population.

Acknowledgments

The authors would like to thank the interview participants, Anhörigas Riksförbund and Selma Gruppen, for their time, input, and generous contributions. They also thank Kaoutar Kheir, the research intern, for conducting the interviews in Arabic and ensuring accurate transcription and translation into English. This work was funded by the European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska Curie grant 814 072. In addition, the Swedish Research Council and Swedish Cancer Society (ref 2019-01231; ref 20 1014 PjF) provided support for the study.

Authors' Contributions

SP is the principal author of this paper. SP designed the study, analyzed and interpreted the interviews, designed the prototype, and wrote the manuscript. AA was involved in discussions on the design of the prototype, wrote parts of the manuscript, and edited the manuscript draft. ÅC contributed to the study design, was involved in discussions on the design of the prototype, wrote parts of the paper, and critically revised the manuscript draft. PÅ contributed to the study design, was involved in discussions regarding the design of the prototype, and critically revised the manuscript. MD and LvG-P revised the manuscript draft. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOC File , 73 KB - mhealth_v11i1e50038_app1.doc]

Multimedia Appendix 2
Interview guide.
[DOCX File , 16 KB - mhealth_v11i1e50038_app2.docx]

References

1. Biliunaite I, Kazlauskas E, Sanderman R, Andersson G. Informal caregiver support needs and burden: a survey in Lithuania. BMJ Open 2022 Jan 04;12(1):e054607 [FREE Full text] [doi: 10.1136/bmjopen-2021-054607] [Medline: 34983768]
2. Ghahramani F, Wang J. Intention to adopt mHealth apps among informal caregivers: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 17;9(3):e24755 [FREE Full text] [doi: 10.2196/24755] [Medline: 33729166]
3. Zigante V. Informal care in Europe: exploring formalisation, availability and quality. European Commission. URL: https://op.europa.eu/en/publication-detail/-/publication/96d27995-6dee-11e8-9483-01aa75ed71a1 [accessed 2023-10-24]
4. 'Ageing in place' policy has led to a gradual downsizing of institutional. Eurocarers. URL: https://eurocarers.org/country-profiles/sweden/ [accessed 2023-02-16]
5. Herz M, Andreasson J, Andréasson F. The precariousness of asylum-seekers care and support: informal care within and because of the immigration process. Int J Care Caring 2023;7(1):50-66 [FREE Full text] [doi: 10.1332/239788221x16456304482483]
6. Ahmad A, Premanandan S, Langegård U, Cajander Å, Johansson B, Carlsson M, et al. Positive design framework for carer eSupport: qualitative study to support informal caregivers of patients with head and neck cancer in Sweden. JMIR Cancer 2023 May 30;9:e45748 [FREE Full text] [doi: 10.2196/45748] [Medline: 37252765]
7. Slev VN, Mistiaen P, Pasman HR, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
8. Sörensen S, Duberstein P, Gill D, Pinquart M. Dementia care: mental health effects, intervention strategies, and clinical implications. Lancet Neurol 2006 Nov;5(11):961-973. [doi: 10.1016/S1474-4422(06)70599-3] [Medline: 17052663]
9. Johansson L, Long H, Parker MG. Informal caregiving for elders in Sweden: an analysis of current policy developments. J Aging Soc Policy 2011 Oct;23(4):335-353. [doi: 10.1080/08959420.2011.605630] [Medline: 21985063]
10. Jegermalm M. Informal care and support for carers in Sweden: patterns of service receipt among informal caregivers and care recipients. Eur J Soc Work 2004;7(1):7-24 [FREE Full text] [doi: 10.1080/1369145042000217465]
11. Berglund E, Lytsy P, Westerling R. Health and wellbeing in informal caregivers and non-caregivers: a comparative cross-sectional study of the Swedish general population. Health Qual Life Outcomes 2015 Jul 29;13(1):109 [FREE Full text] [doi: 10.1186/s12955-015-0309-2] [Medline: 26216099]
12. Blanck E, Fors A, Ali L, Brännström M, Ekman I. Informal carers in Sweden - striving for partnership. Int J Qual Stud Health Well-being 2021 Dec;16(1):1994804 [FREE Full text] [doi: 10.1080/17482631.2021.1994804] [Medline: 34720063]
13. Immigration and emigration 1970–2022 and projection 2023–2070. Statistiska Centralbyrån. 2023. URL: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-projections/population-projections/pong/tables-and-graphs/immigration-and-emigration-by-sex-and-country-of-birth-and-projection/ [accessed 2023-06-09]
14. Aranda MP, Knight BG. The influence of ethnicity and culture on the caregiver stress and coping process: a sociocultural review and analysis. Gerontologist 1997 Jun;37(3):342-354. [doi: 10.1093/geront/37.3.342] [Medline: 9203758]
15. Galiana-Gómez de Cádiz MJ, de la Cuesta-Benjumea C, Donet-Montagut T. [Immigrant caregivers: characteristics of the care provided to dependent elders]. Enferm Clin 2008 Sep;18(5):269-272. [doi: 10.1016/s1130-8621(08)72386-6] [Medline: 18840336]
16. Navaie-Waliser M, Feldman PH, Gould DA, Levine C, Kuerbis AN, Donelan K. The experiences and challenges of informal caregivers: common themes and differences among whites, blacks, and hispanics. Gerontologist 2001 Dec;41(6):733-741. [doi: 10.1093/geront/41.6.733] [Medline: 11723341]
17. Flagler-George J. Squeezed in: the intersecting paradoxes of care for immigrant informal caregivers. UWSpace. 2014. URL: https://uwspace.uwaterloo.ca/handle/10012/8839 [accessed 2023-04-24]
18. Lindsay S, King G, Klassen AF, Esses V, Stachel M. Working with immigrant families raising a child with a disability: challenges and recommendations for healthcare and community service providers. Disabil Rehabil 2012;34(23):2007-2017. [doi: 10.3109/09638288.2012.667192] [Medline: 22455458]
19. Han HR, Choi YJ, Kim MT, Lee JE, Kim KB. Experiences and challenges of informal caregiving for Korean immigrants. J Adv Nurs 2008 Sep;63(5):517-526 [FREE Full text] [doi: 10.1111/j.1365-2648.2008.04746.x] [Medline: 18727754]

20. Sharma N, Brinke JK, Gemert-Pijnen JE, Braakman-Jansen LM. Implementation of unobtrusive sensing systems for older adult care: scoping review. JMIR Aging 2021 Oct 06;4(4):e27862 [FREE Full text] [doi: 10.2196/27862] [Medline: 34612822]
21. Yüce YK, Gülkesen KH. Development of a social support intervention with a network of caregivers to find wandering Alzheimer's patients as soon as possible: a social computing application in healthcare. In: Proceedings of the 8th International Symposium on Health Informatics and Bioinformatics. 2013 Presented at: ISHIB '13; September 25-27, 2013; Ankara, Turkey p. 1-8 URL: https://ieeexplore.ieee.org/document/6661678 [doi: 10.1109/hibit.2013.6661678]
22. Kondylakis H, Katehakis DG, Kouroubali A, Marias K, Flouris G, Patkos T, et al. CareKeeper: A platform for intelligent care coordination. In: Proceedings of the 21st International Conference on Bioinformatics and Bioengineering. 2021 Presented at: BIBE '21; October 25-27, 2021; Kragujevac, Serbia p. 1-4 URL: https://ieeexplore.ieee.org/document/9635445 [doi: 10.1109/bibe52308.2021.9635445]
23. Lin ZS, McKinstry B, Anderson S. Coordinating caregiving using smartphone technology: a collaborative software prototype approach. In: Proceedings of the 12th International Conference on Informatics, Management, and Technology in Healthcare. 2014 Presented at: ICIMTH '14; July 10-13, 2014; Athens, Greece p. 237-240 URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84904210072&doi=10.3233%2f978-1-61499-423-7-237&partnerID=40&md5=cd82109033e647cc4cc2e2147c97dd68
24. Kamphorst BA. E-coaching systems: what they are, and what they aren’t. Pers Ubiquit Comput 2017 May 23;21(4):625-632. [doi: 10.1007/s00779-017-1020-6]
25. Chatterjee A, Gerdes M, Prinz A, Martinez S. Human coaching methodologies for automatic electronic coaching (eCoaching) as behavioral interventions with information and communication technology: systematic review. J Med Internet Res 2021 Mar 24;23(3):e23533 [FREE Full text] [doi: 10.2196/23533] [Medline: 33759793]
26. Chatterjee A, Prinz A. Personalized recommendations for physical activity e-coaching (OntoRecoModel): ontological modeling. JMIR Med Inform 2022 Jun 23;10(6):e33847 [FREE Full text] [doi: 10.2196/33847] [Medline: 35737439]
27. Premanandan S, Ahmad A, Cajander Å, Ågerfalk P, van Gemert-Pijnen L. Design suggestions for a persuasive e-coaching application: a study on informal caregivers' needs. Digital Health 2023 May 30;9:20552076231177129 [FREE Full text] [doi: 10.1177/20552076231177129] [Medline: 37284014]
28. Asbjørnsen RA, Hjelmesæth J, Smedsrød ML, Wentzel J, Ollivier M, Clark MM, et al. Combining persuasive system design principles and behavior change techniques in digital interventions supporting long-term weight loss maintenance: design and development of eCHANGE. JMIR Hum Factors 2022 May 27;9(2):e37372 [FREE Full text] [doi: 10.2196/37372] [Medline: 35622394]
29. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:485-500 [FREE Full text] [doi: 10.17705/1CAIS.02428]
30. Gilliver SC, Sundquist J, Li X, Sundquist K. Recent research on the mental health of immigrants to Sweden: a literature review. Eur J Public Health 2014 Aug 07;24 Suppl 1(suppl 1):72-79. [doi: 10.1093/eurpub/cku101] [Medline: 25108001]
31. Purao S. Design research in the technology of information systems: truth or dare. GSU Department of CIS Working Paper 34. 2002. URL: https://www.researchgate.net/profile/Sandeep-Purao/publication/228606379_Design_research_in_the_technology_of_information_systems_Truth_or_dare/links/0fcfd5115fdfda361ad000000/Design-research-in-the-technology-of-information-systems-Truth-or-dare.pdf [accessed 2023-10-24]
32. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. Naderifar M, Goli H, Ghaljaie F. Snowball sampling: a purposeful method of sampling in qualitative research. Strides Dev Med Educ 2017 Sep 30;14(3):e67670 [FREE Full text] [doi: 10.5812/sdme.67670]
34. Age. National Institutes of Health. URL: https://www.nih.gov/nih-style-guide/age [accessed 2023-04-19]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
36. One platform to connect. Zoom. URL: https://zoom.us/ [accessed 2023-04-19]
37. The collaborative interface design tool. Figma. URL: https://www.figma.com/ [accessed 2023-02-24]
38. Industry-leading vector graphics software. Adobe Illustrator. URL: https://www.adobe.com/products/illustrator.html [accessed 2023-06-13]
39. Lee AS, Baskerville RL. Generalizing generalizability in information systems research. Inf Syst Res 2003 Sep;14(3):221-243 [FREE Full text] [doi: 10.1287/isre.14.3.221.16560]
40. Ahmad A, Mozelius P. Human-computer interaction for older adults - a literature review on technology acceptance of eHealth systems. J Eng Res Sci 2022 Apr;1(4):119-126 [FREE Full text] [doi: 10.55708/js0104014]
41. Freund M, Carey M, Dilworth S, Waller A, Mansfield E, Rose A, et al. Effectiveness of information and communications technology interventions for stroke survivors and their support people: a systematic review. Disabil Rehabil 2022 Aug;44(17):4563-4578. [doi: 10.1080/09638288.2021.1913245] [Medline: 33905279]

42. Guisado-Fernández E, Giunti G, Mackey LM, Blake C, Caulfield BM. Factors influencing the adoption of smart health technologies for people with dementia and their informal caregivers: scoping review and design framework. JMIR Aging 2019 Apr 30;2(1):e12192 [FREE Full text] [doi: 10.2196/12192] [Medline: 31518262]
43. Sörensen S, Pinquart M, Duberstein P. How effective are interventions with caregivers? An updated meta-analysis. Gerontologist 2002 Jun 01;42(3):356-372. [doi: 10.1093/geront/42.3.356] [Medline: 12040138]
44. Cheng ST, Li KK, Losada A, Zhang F, Au A, Thompson LW, et al. The effectiveness of nonpharmacological interventions for informal dementia caregivers: an updated systematic review and meta-analysis. Psychol Aging 2020 Feb;35(1):55-77. [doi: 10.1037/pag0000401] [Medline: 31985249]
45. Vandepitte S, Van Den Noortgate N, Putman K, Verhaeghe S, Faes K, Annemans L. Effectiveness of supporting informal caregivers of people with dementia: a systematic review of randomized and non-randomized controlled trials. J Alzheimers Dis 2016 Apr 08;52(3):929-965. [doi: 10.3233/JAD-151011] [Medline: 27079704]
46. Nkulu Kalengayi FK, Hurtig AK, Ahlm C, Ahlberg BM. "It is a challenge to do it the right way": an interpretive description of caregivers' experiences in caring for migrant patients in Northern Sweden. BMC Health Serv Res 2012 Nov 29;12(1):433 [FREE Full text] [doi: 10.1186/1472-6963-12-433] [Medline: 23194441]
47. Stewart MJ, Neufeld A, Harrison MJ, Spitzer D, Hughes K, Makwarimba E. Immigrant women family caregivers in Canada: implications for policies and programmes in health and social sectors. Health Soc Care Community 2006 Jul;14(4):329-340. [doi: 10.1111/j.1365-2524.2006.00627.x] [Medline: 16787484]
48. Lillekroken D, Halvorsrud L, Gulestø R, Bjørge H. Family caregivers' experiences of providing care for family members from minority ethnic groups living with dementia: a qualitative systematic review. J Clin Nurs 2023 May 16;32(9-10):1625-1641. [doi: 10.1111/jocn.16127] [Medline: 34786789]
49. Sagbakken M, Spilker RS, Ingebretsen R. Dementia and migration: family care patterns merging with public care services. Qual Health Res 2018 Jan;28(1):16-29. [doi: 10.1177/1049732317730818] [Medline: 28918700]
50. Stenberg J, Hjelm K. Migrant informal caregiver perceptions and experiences of caring for a family member with dementia: a systematic review and thematic synthesis. J Clin Nurs 2023 May;32(9-10):2178-2192. [doi: 10.1111/jocn.16390] [Medline: 35655377]
51. Monsees J, Schmachtenberg T, Hoffmann W, Kind A, Gilmore-Bykovskyi A, Kim AJ, et al. Dementia in people with a Turkish migration background: experiences and utilization of healthcare services. J Alzheimers Dis 2020;77(2):865-875 [FREE Full text] [doi: 10.3233/JAD-200184] [Medline: 32741814]
52. Berdai Chaouni S, De Donder L. Invisible realities: caring for older Moroccan migrants with dementia in Belgium. Dementia (London) 2019;18(7-8):3113-3129. [doi: 10.1177/1471301218768923] [Medline: 29649908]
53. Mollenkopf H, Kloé U, Obermann E, Klumpp G. The potential of ICT in supporting domiciliary care in Germany. Joint Research Centre (Seville site). 2010. URL: https://econpapers.repec.org/paper/iptiptwpa/jrc56575.htm [accessed 2023-06-13]
54. Koru G, Alhuwail D, Topaz M, Norcio AF, Mills ME. Investigating the challenges and opportunities in home care to facilitate effective information technology adoption. J Am Med Dir Assoc 2016 Jan;17(1):53-58. [doi: 10.1016/j.jamda.2015.10.008] [Medline: 26612483]
55. Rasi S. Impact of language barriers on access to healthcare services by immigrant patients: a systematic review. Asia Pac J Health Manag 2020;15(1):35-48 [FREE Full text] [doi: 10.24083/apjhm.v15i1.271]
56. Köhle N, Drossaert CH, Oosterik S, Schreurs KM, Hagedoorn M, van Uden-Kraan CF, et al. Needs and preferences of partners of cancer patients regarding a web-based psychological intervention: a qualitative study. JMIR Cancer 2015 Dec 29;1(2):e13 [FREE Full text] [doi: 10.2196/cancer.4631] [Medline: 28410157]
57. Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. J Soc Work Educ 2017 Jan 12;44(sup3):105-113 [FREE Full text] [doi: 10.5175/jswe.2008.773247702]
58. Feeney BC, Collins NL. Motivations for caregiving in adult intimate relationships: influences on caregiving behavior and relationship functioning. Pers Soc Psychol Bull 2003 Aug;29(8):950-968. [doi: 10.1177/0146167203252807] [Medline: 15189615]
59. Linton JM, Choi R, Mendoza F. Caring for children in immigrant families: vulnerabilities, resilience, and opportunities. Pediatr Clin North Am 2016 Feb;63(1):115-130. [doi: 10.1016/j.pcl.2015.08.006] [Medline: 26613692]
60. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
61. van Wezel N, van der Heide I, Devillé WL, Duran G, Hoopman R, Blom MM, et al. The Turkish version of the SPPIC validated among informal caregivers with a Turkish immigrant background. BMC Geriatr 2021 Apr 29;21(1):284 [FREE Full text] [doi: 10.1186/s12877-021-02161-6] [Medline: 33926385]
62. Ahmad M, van den Broeke J, Saharso S, Tonkens E. Persons with a migration background caring for a family member with dementia: challenges to shared care. Gerontologist 2020 Feb 24;60(2):340-349 [FREE Full text] [doi: 10.1093/geront/gnz161] [Medline: 31786594]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder

COREQ: Consolidated Criteria for Reporting Qualitative Research
FAQ: frequently asked question
PSDM: Persuasive System Design Model

Edited by L Buis; submitted 20.06.23; peer-reviewed by R Daynes-Kearney, S Watland; comments to author 31.07.23; revised version received 21.09.23; accepted 27.09.23; published 09.11.23.

Please cite as:
Premanandan S, Ahmad A, Cajander Å, Ågerfalk P, Dolezel M, van Gemert-Pijnen L
Designing a Mobile e-Coaching App for Immigrant Informal Caregivers: Qualitative Study Using the Persuasive System Design Model
JMIR Mhealth Uhealth 2023;11:e50038
URL: https://mhealth.jmir.org/2023/1/e50038
doi:10.2196/50038
PMID:37943598

Original Paper

Predictive Dispatch of Volunteer First Responders: Algorithm Development and Validation

Michael Khalemsky[1], PhD; Anna Khalemsky[1], PhD; Stephen Lankenau[2], PhD; Janna Ataiants[2], PhD; Alexis Roth[2], PhD; Gabriela Marcu[3], PhD; David G Schwartz[4], PhD

[1]Department of Management, Hadassah Academic College, Jerusalem, Israel

[2]School of Public Health, Drexel University, Philadelphia, PA, United States

[3]School of Information, University of Michigan, Ann Arbor, MI, United States

[4]The Graduate School of Business Administration, Bar-Ilan University, Ramat Gan, Israel

Corresponding Author:
Michael Khalemsky, PhD
Department of Management
Hadassah Academic College
ha-Neviim 37
Jerusalem, 9101001
Israel
Phone: 972 26291941
Email: michaelkh@hac.ac.il

Abstract

Background: Smartphone-based emergency response apps are increasingly being used to identify and dispatch volunteer first responders (VFRs) to medical emergencies to provide faster first aid, which is associated with better prognoses. Volunteers' availability and willingness to respond are uncertain, leading in recent studies to response rates of 17% to 47%. Dispatch algorithms that select volunteers based on their estimated time of arrival (ETA) without considering the likelihood of response may be suboptimal due to a large percentage of alerts *wasted* on VFRs with shorter ETA but a low likelihood of response, resulting in delays until a volunteer who will actually respond can be dispatched.

Objective: This study aims to improve the decision-making process of human emergency medical services dispatchers and autonomous dispatch algorithms by presenting a novel approach for predicting whether a VFR will respond to or ignore a given alert.

Methods: We developed and compared 4 analytical models to predict VFRs' response behaviors based on emergency event characteristics, volunteers' demographic data and previous experience, and condition-specific parameters. We tested these 4 models using 4 different algorithms applied on actual demographic and response data from a 12-month study of 112 VFRs who received 993 alerts to respond to 188 opioid overdose emergencies. Model 4 used an additional dynamically updated synthetic dichotomous variable, *frequent responder*, which reflects the responder's previous behavior.

Results: The highest accuracy (260/329, 79.1%) of prediction that a VFR will ignore an alert was achieved by 2 models that used events data, VFRs' demographic data, and their previous response experience, with slightly better overall accuracy (248/329, 75.4%) for model 4, which used the *frequent responder* indicator. Another model that used events data and VFRs' previous experience but did not use demographic data provided a high-accuracy prediction (277/329, 84.2%) of ignored alerts but a low-accuracy prediction (153/329, 46.5%) of responded alerts. The accuracy of the model that used events data only was unacceptably low. The J48 decision tree algorithm provided the best accuracy.

Conclusions: VFR dispatch has evolved in the last decades, thanks to technological advances and a better understanding of VFR management. The dispatch of substitute responders is a common approach in VFR systems. Predicting the response behavior of candidate responders in advance of dispatch can allow any VFR system to choose the best possible response candidates based not only on ETA but also on the probability of actual response. The integration of the probability to respond into the dispatch algorithm constitutes a new generation of individual dispatch, making this one of the first studies to harness the power of predictive analytics for VFR dispatch. Our findings can help VFR network administrators in their continual efforts to improve the response times of their networks and to save lives.

(JMIR Mhealth Uhealth 2023;11:e41551) doi:10.2196/41551

KEYWORDS

volunteer; emergency; dispatch; responder; smartphone; emergency response; smartphone-based apps; mobile phone apps; first responders; medical emergency; dispatch algorithms; dispatch decisions; dispatch prediction; smartphone app; decision-making; algorithm; mobile health; mHealth intervention; mobile phone

Introduction

Background

Emergency response apps, commonly smartphone based, are increasingly being used to identify and dispatch volunteer first responders (VFRs) to the location of a medical emergency [1]. Automated dispatch algorithms generally rely on a simple estimated time of arrival (ETA) calculation based on the locations of the VFRs and the incident as well as the known modes of transport. A key aspect lacking in these algorithms is a consideration of the likelihood of response; for instance, given a set of potential VFRs with equivalent ETAs, which subset should be alerted to maximize the likelihood of response? The automated dispatch of VFRs to medical emergencies is suboptimal owing to a large percentage of alerts *wasted* on VFRs with shorter ETA but a low likelihood of response. This results in delays until a volunteer who will actually respond can be identified and dispatched. Using actual demographic and response data taken from a 12-month study of 112 VFRs alerted to respond to opioid overdose emergencies, we applied a series of analytical methods and advanced classification models to learn and predict volunteer response behaviors. Our findings can be used to improve dispatch algorithms in VFR networks to optimize dispatch decisions and increase the likelihood of timely emergency responses.

Medical Emergencies

A medical emergency is an acute injury or illness that can result in death or long-term health complications [2]. Some common medical emergencies include out-of-hospital cardiac arrest (OHCA), severe trauma, opioid overdose, and anaphylaxis. OHCA is a leading cause of death worldwide [3], with a poor survival rate (only 5.6% in adults) [4]. Major trauma is the sixth leading cause of death worldwide [5,6]. Opioid overdose is a severe public health problem that has been consistently rising for the past 20 years and in the United States is the leading cause of accidental death [7]. The incidence of anaphylaxis ranges from 1.5 to 7.9 per 100,000 population per year in Europe [8,9].

Networks of VFRs

The immediate provision of first aid is crucial in lowering mortality and improving long-term prognosis, particularly in regard to OHCA [10-12] and opioid overdose events [13]. Emergency medical services (EMS) are the primary first aid provider [14,15], but EMS response times vary significantly among countries and geographies [16,17]. Interventions to achieve faster response times include the deployment of automatic external defibrillators (AEDs) in public places [18-21] and the establishment of local networks of VFRs [22-30]. Recently, there was a concerted effort to use smartphone apps for faster emergency response, such as PulsePoint, HelpAround, Heartrunner, and UnityPhilly. An extensive review of emergency response apps can be found in the study by Gaziel-Yablowitz and Schwartz [31].

An emergency response community (ERC) [32], a subtype of a VFR network, is a social network of patients who are prescribed to carry life-saving medication for themselves and can potentially help other patients who are without their medication in a medical emergency. Two projects that apply the ERC approach are the subjects of recent field studies: EPIMADA, which focused on patients at risk of anaphylaxis and their parents [33]; and UnityPhilly, which focuses on people who have experienced an opioid overdose [34].

Willingness to Respond, Barriers, and Facilitators

Once a person becomes a volunteer, they are expected to respond if available when a relevant event occurs. However, the actual rates of response to emergency alerts are far from 100%. Brooks et al [35] reported a response rate of 23% among PulsePoint volunteers. In a recent study, the willingness of cardiopulmonary resuscitation (CPR)–trained bystanders to respond to an OHCA event was 46.6% [36]. Another study analyzed barriers to receiving notifications and reported that 32% of the responders who were sent notifications did not receive the notification because, for example, they were away from their device (21%), their device was switched off (8%), or their device was out of network range (4%) [35]. Stress levels among responders varied for different medical conditions, different locations, and different demographic groups [37]. Younger age, higher education level, shorter time since the last CPR training, and cardiac arrest event in a public location were good predictors of bystanders' greater willingness to perform CPR. The main reasons for not performing CPR were panic, the perception of bystanders that they are not able to perform CPR correctly, and a fear of hurting the patient [38]. Familial experiences of receiving CPR were associated with an increase in responders' willingness to perform CPR [39]. The UnityPhilly study, which established a network of volunteers to provide naloxone to those experiencing an opioid overdose, reported that 17% of the alerted volunteers accepted the alert, and 11.9% of the alerted volunteers arrived at the scene [34].

Dispatch Algorithms and Decision-Making

Complexity of VFR Dispatch and Decision-Making

The complexity of VFR dispatch stems from 2 sources: unknown resource location and uncertain response. Emergency response services that try to optimize their own resources to maximize their effectiveness can determine the allocation of their resources, such as ambulance dispatch stations or police patrol districts, subject to constraints (eg, budgets) [40-42]. The administrators of a VFR network are unable to plan and control the location of their resources because VFRs perform their daily activities until called to action: they can be anywhere, enter and exit the area that the network covers, switch on and off their mobile phones, and so on. In addition, although ambulance staff or a police patrol are expected to respond to any event that they are dispatched to, VFRs decide for themselves whether to respond to a specific event.

Usual Location–Based Dispatch Approach Using Pagers and SMS Text Messages

In a typical location-based approach, VFRs are alerted based on their usual location (eg, home or work address) and not their actual location at the moment of the alert. VFRs may not provide any feedback to the system regarding their availability to respond to the specific event and just show up on the scene if they can; for example, this approach was used by Zijlstra et al [43] who sent SMS text messages to volunteers living within a 1000-meter radius of an OHCA event.

Current Location–Based Dispatch Approach

A current location–based dispatch approach is based on a smartphone app that continuously sends VFRs' locations (eg, geospatial coordinates) to a central server. When an emergency event is registered in the system, the dispatch algorithm selects volunteers based on their distance from the scene or, in a more advanced version, based on their ETA [44]. Such apps can also allow VFRs to set their availability status to control for their commitment, which was found to be an important factor of VFRs' willingness to volunteer [45]. Location-based dispatch is widely used in VFR networks [34,36,46,47]. Usually, location-based algorithms dispatch >1 volunteer, if available, but still limit the number of volunteers who are dispatched to prevent burnout and a decrease in self-efficacy. Sending a large number of responders to each event can lead to the "diffusion of responsibility" phenomenon and reduce willingness to respond [48].

Autonomous Dispatch Versus EMS-Mediated Dispatch

Some VFR networks are managed by EMS and are integrated into their business processes. In this case, the dispatch of VFRs is at the discretion of a human dispatcher, and the VFR system serves as a decision support system that provides the dispatcher with the necessary information, such as location and ETA, of volunteers that can be compared with the location and ETA of an ambulance. Once alerts are sent, the system constantly updates its recommendations based on the feedback from the alerted volunteers. This approach is used by the Life Guardians project managed by Israeli National EMS [46] and in several AED and CPR projects [36].

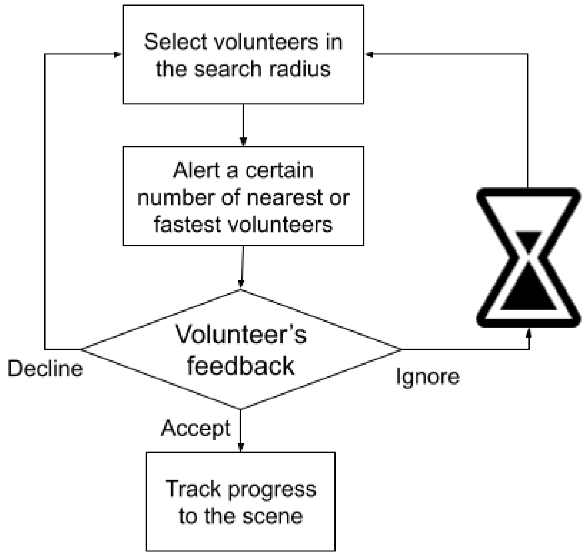
An alternative approach is autonomous dispatch, where VFRs are selected and alerted by an autonomous system according to a predefined business logic. The system can dispatch additional volunteers if the alerted volunteers ignore the alert, refuse to respond, or linger on the way. This approach was used by the UnityPhilly project [34] and the PulsePoint project [35].

Both approaches can be either registered (usual or expected) location based or current (dynamic) location based; for example, UnityPhilly uses a current location–based autonomous dispatch approach.

Integration of Volunteers' Feedback Into Dispatch Algorithms

Many smartphone apps for VFR networks allow alerted responders to accept or decline the alert. Such feedback lowers the uncertainty regarding the dispatcher, and, if a volunteer declines an alert, the dispatch algorithm can reconsider the selection of responders and send additional alerts to substitute volunteers (ie, to volunteers who were not initially selected by the algorithm [eg, because they had a longer ETA] but who, in the event that ≥1 of the initially selected volunteers decline or ignore the alert, can be dispatched to achieve the target number of responders). If an alerted volunteer ignores the alert and does not provide any feedback, the system waits for a set period of time and then considers the nonresponse a "no" and acts accordingly. Figure 1 depicts this process.

Figure 1. The dispatch process and feedback from alerted volunteers.



Profiling

Profiling is "the process of generating profiles from obtained data, associated to one or multiple subjects" [49]. Profiling of people is widely used in several areas, such as targeted advertising [50], donation solicitation [51], and volunteer recruitment [52]. Elsner et al [22,49] proposed to use the profiling of volunteers in dispatch algorithms to enhance the prediction of the volunteers' position, trajectory, and constraints. In this study, we used classification techniques to generate different behavioral profiles of volunteers that serve as independent variables for predicting responses to alerts.

The Purpose of the Study

The challenge of improving volunteer dispatch speed and response rates is recognized in fields ranging from food rescue operations [53] to OHCA response in which the optimization of the responder network is now taking center stage [54,55]. Studies such as the one by Gregers et al [56] have attempted to determine the optimal number of responders to dispatch, yet such studies base response viability solely on current ETA with no consideration of responder history or other characteristics that could improve responsiveness. Currently used dispatch algorithms that select volunteers based on their ETA without considering the likelihood of response may be suboptimal owing to a large percentage of alerts *wasted* on VFRs with shorter ETA but a low likelihood of response. We build on prior work on VFR optimization by presenting a novel approach for predicting whether a VFR will respond to, or ignore, a given alert. As such, the enhanced algorithm reduces the time that the system unnecessarily spends waiting for a response from volunteers who are likely to ignore the alert. The amount of time wasted depends on the specific dispatch algorithm; for example, in UnityPhilly trials, the system waited 2 minutes before dispatching a substitute volunteer. A faster dispatch of substitute volunteers has the potential to reduce the response time of the VFR network as a whole and improve its effectiveness. However, overdispatch of more VFRs than necessary to secure an effective emergency response can have a negative impact on future willingness to respond [48].

Methods

Data

We used data from the UnityPhilly study that piloted a smartphone-based app for requesting and providing ERC assistance to those suspected of experiencing an opioid overdose in the neighborhood of Kensington, PA, over 12 months from March 1, 2019, to February 28, 2020. Kensington has Philadelphia's highest concentration of overdose deaths and is also home to Prevention Point Philadelphia, which is a city-sanctioned syringe exchange program that also distributes naloxone and provides naloxone training. Recruitment occurred via face-to-face screening at Prevention Point's drop-in center, Prevention Point's substance use disorder treatment van, street intercepts, and chain referrals from enrolled participants. The inclusion criteria for participants were that they lived, worked, or used drugs within 4 zip codes around the Kensington neighborhood (19122, 19125, 19133, and 19134); possessed a smartphone with a data plan; were willing to have location and movements tracked via an app; were willing to carry naloxone; and were aged ≥18 years. Sampling purposefully targeted a mix of members of the Kensington community who used opioids nonmedically in the past 30 days and those who reported no nonmedical opioid use in the past 30 days. The study recruited 112 volunteers who were almost equally divided between people who reported opioid use in the past 30 days at baseline (n=57, 50.9%) and community members, that is, people who reported no opioid use at baseline (n=55, 49.1%).

At a research storefront in Kensington, the study enrollment procedure included obtaining written informed consent, the recording of contact information, structured baseline interviews, app installation and training, and naloxone distribution and training. During the informed consent procedure, participants agreed to participate in a baseline interview, monthly follow-up interviews, and brief surveys after overdose incidents. Project staff installed the app on the participant's smartphone and provided app training, which included watching an animated training video explaining app use and practicing using the app to send and receive alerts with project staff. Naloxone training included recognizing the signs of opioid overdose, practicing rescue breathing on a CPR dummy, and demonstrating how to administer intranasal naloxone. All participants received a kit containing 2 doses of intranasal naloxone. The UnityPhilly app enabled them to report opioid overdose events and to receive notifications about opioid overdose events reported by other members in their proximity. Participants received US $25 in cash for the baseline interview and US $5 for each completed follow-up monthly interview or incident survey. No compensation was offered or given for the use of the app to signal or respond to overdose incidents. More details about the study are available in prior publications [34].

The data used for this analysis consist of 4 components (Textbox 1 and Figure 2).

Of the 112 volunteers recruited to UnityPhilly, 27 (24.1%) were completely inactive as either signaler or responder (ie, they did not send or respond to a single alert). Of the remaining 85 volunteers, 80 (94%) received at least 1 alert and were defined as *responders*, and 52 (61%) who signaled at least 1 event were defined as *signalers* (many volunteers served in both roles). Figure 3 presents the distribution of responders and signalers.

Events that were canceled by the signaler for any reason were considered false alarms. For this analysis, we excluded these events because we were not able to distinguish between alarms ignored by the responder and alarms that were canceled before the responder had a chance to respond. Figure 4 describes the sample.

We used *alerts* as a unit of analysis.

Textbox 1. The 4 components of the data used for analysis.

Event

- This refers to an opioid overdose event. An event's characteristics are true or false alarm, signaler, weekday or weekend, and day or night.

Signaler

- This refers to a UnityPhilly user who witnesses an event and reports it to the system using the UnityPhilly app. A signaler's characteristics are age, gender, housing status, employment status, naloxone carriage adherence before joining the UnityPhilly community, opioid overdose witnessing experience before joining the UnityPhilly community, and experience in administering naloxone to a person experiencing an overdose before joining the UnityPhilly community.

Responder

- This refers to a UnityPhilly member who is selected by the UnityPhilly system (based on their location and estimated time of arrival) and notified in their UnityPhilly app about an event. The responder's characteristics are the same as those of the signaler.

Alert

- This refers to a notification sent to a specific responder about a specific event. The UnityPhilly app enables the responder to accept or decline an alert. However, many alerts are ignored, that is, neither accepted nor declined. An alert's characteristics are distance between the potential responder and the event scene at the moment of the alert, the number of previous alerts received by the responder since joining UnityPhilly, the number of previous false alerts received by the responder since joining, the number of previous alerts received by the responder since joining that were initiated by the same signaler, the number of previous false alerts received by the responder since joining that were initiated by the same signaler, the number of previous responses by the responder since joining, the number of previous responses to false alerts received by the responder since joining, and the number of previous responses to false alerts initiated by the same signaler that were received by the responder since joining.

Figure 2. Entities in the UnityPhilly data set. M: many.

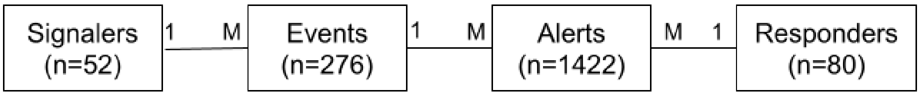


Figure 3. Distribution of responders and signalers in the UnityPhilly data set.

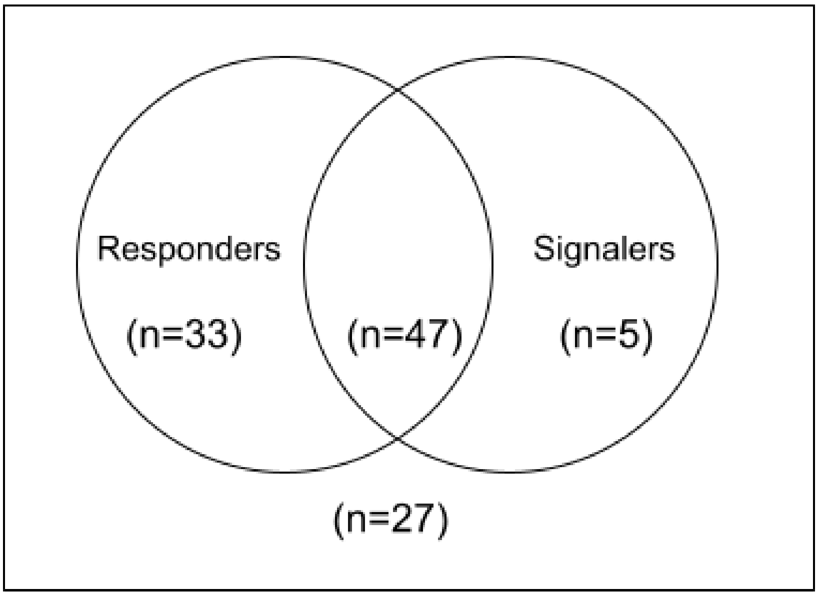
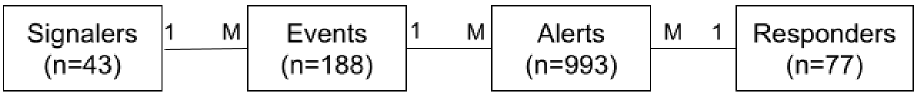


Figure 4. Sample used for this study. M: many.

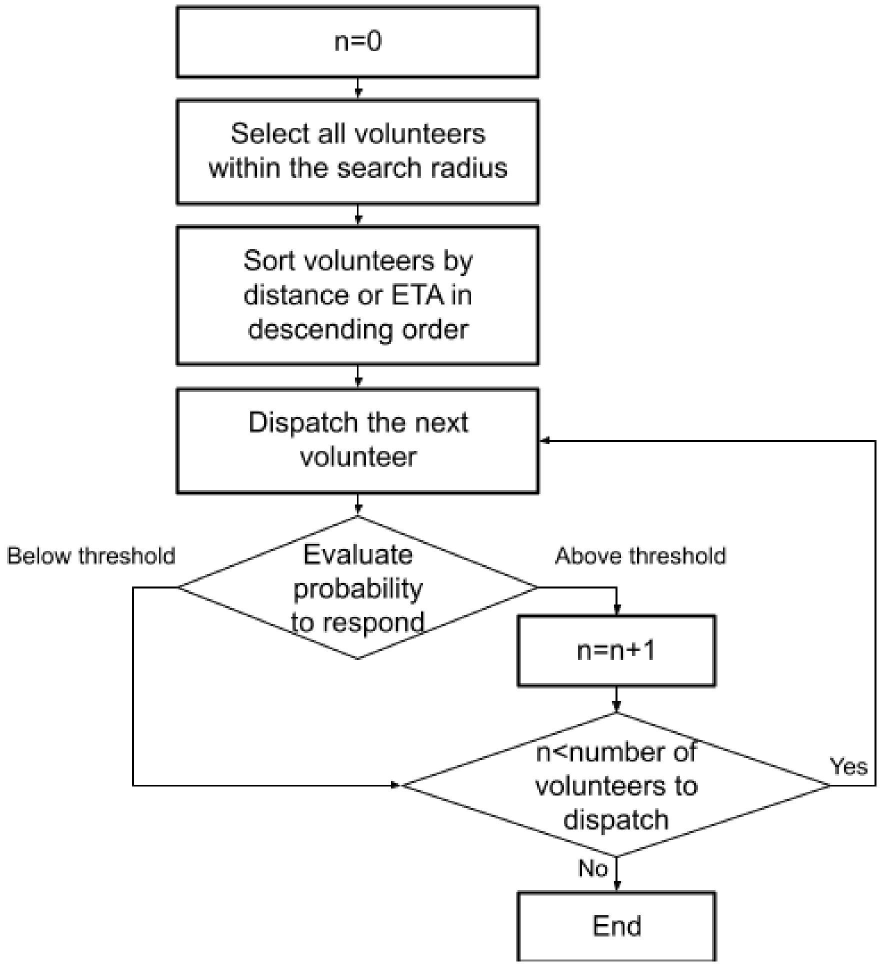


Analytical Approach

We used multiple analytical methods to classify the behavior of each volunteer identified as being in the proximity of an overdose event. We integrated data on specific volunteers and events into the dispatch algorithm in such a way that for each dispatched volunteer who is most likely to ignore the alert, an additional volunteer is dispatched right away (if available), until the maximum number of volunteers to be dispatched is reached, or no more volunteers are available. Volunteers for whom the algorithm predicts a low probability of response are still dispatched and thus are given the chance to respond. Figure 5 depicts this process.

We tested 4 models, based on different configurations of variables, to predict whether a given responder is likely to respond to a given event (Textbox 2 and Table 1).

Figure 5. Integration of the probability to respond into the dispatch algorithm. ETA: estimated time of arrival.



Textbox 2. The 4 models tested in this study.

Model 1

- This model is based solely on historic events and alerts data, incorporating no other data related to the potential responders.

Model 2

- This model is based on the events and alerts data, but it also integrates data on the responders' patterns of behavior through their previous experience in the volunteer first responder network, including previous alerts and false alerts, and previous responses, including responses to false alerts.

Model 3

- This model is based on the events and alerts data, as well as respondents' personal and demographic data, and ignores their previous experience in the network.

Model 4

- This model is based on the events and alerts data, as well as respondents' personal and demographic data, and dynamically calculates the

  frequent responder

  indicator that represents the responder's experience in the community before a specific alert. This indicator was calculated as follows:
  - <6 alerts: no
  - 6-10 alerts and response rate ≥50%: yes
  - 11-20 alerts and response rate ≥40%: yes
  - 21-30 alerts and response rate ≥30%: yes
  - ≥31 alerts and response rate ≥25%: yes
  - Otherwise: no

Table 1. Data used in each model.

Data	Model			
	1	2	3	4
Events and alerts data (weekday or weekend, day or night, and distance [m])	✓	✓	✓	✓
Responder's previous experience in UnityPhilly (previous alerts, previous false alerts, previous alerts by the same signaler, previous false alerts by the same signaler, previous responses, previous responses to false alerts, and previous responses to false alerts by the same signaler)		✓		✓
Responders' demographic data (age, gender, housing status, and employment status)			✓	✓
Responders' condition-specific characteristics (naloxone carriage adherence, history of witnessing opioid overdoses before joining UnityPhilly, and history of administering naloxone before joining UnityPhilly)			✓	✓
Frequent responder indicator (recalculated after each alarm)				✓

Classification

The classification analysis for all models was conducted using four classification algorithms suitable for binary classification: (1) the J48 decision tree algorithm, which is an extension of the C4.5 algorithm, implemented in Weka software (University of Waikato) used in the research; (2) random forest; (3) neural network (multilayer perceptron); and (4) logistic regression. The J48 algorithm creates univariate decision trees for classification and provides effective alternatives to other classification methods. The choice of the *best* classification model is based on the combination of different evaluation metrics. The main interest was to identify the model that succeeds in correctly classifying *any answer* class.

We used 4 evaluation metrics: accuracy, *F*-score, precision, and recall. Accuracy is the overall percentage of correctly classified instances. The *F*-score is the harmonic mean of the recall and precision metrics and can take values ranging between 0 (none of the instances were correctly classified) and 1 (all instances were correctly classified). Precision is the percentage of true positively classified instances out of all positively classified instances. Recall is the percentage of positively classified instances out of all positive instances. The best way to explain the trends found in this analysis is to explain the differences in the recall metric among the different classification algorithms and among the different classes.

Because of the relatively small overall number of cases in the data set, we did not use a percentage split for the training and test sets for models 1 to 3; instead, we used a cross-validation option with 10 folds. Model 4 includes the additional synthetic dichotomous variable called *frequent responder* that reflects the previous behavior of the responder. The variable is dynamically updated; therefore, a responder can change their behavior several times throughout the research period—from being active to inactive or vice versa. As the *frequent responder* variable cannot be treated as an independent sequence of values and behavioral patterns that must be preserved, there is no option to use cross-validation for classification analysis. For this reason, we split the data set into a training set with 66.9% (664/993) of the data and a test set with 33.1% (329/993) of the data.

All 4 algorithms were used for a binary classification task in a baseline analysis that included only the events and alerts data (model 1 in Table 1). The obtained results provide the baseline for comparison with the additional data related to the responder's previous experience data (models 2, 3, and 4 in Table 1). We claim that building a model that considers the responders' behavioral characteristics can improve the use of the dispatch algorithm. In this kind of analysis, precision in predicting nonresponse is more important than precision in predicting response because in the former case a mistake will delay the dispatch of a substitute responder, whereas in the latter case a mistake will result in the dispatch of too many volunteers.

The comparison between all classification techniques and all evaluation metrics for the 4 models is presented in Multimedia Appendix 1.

Ethical Considerations

All study procedures were approved by the Drexel University Institutional Review Board and registered with ClinicalTrials.gov (NCT03305497). Study enrollment included written informed consent. All data used for this research were deidentified. Participants received US $25 in cash for the baseline interview and US $5 for each completed follow-up monthly interview or incident survey. No compensation was offered or given for use of the app to signal or respond to overdose incidents.

Results

The results of this study are derived from an analysis of emergency events, volunteer participants' demographics, and behavior patterns.

Description of the Sample

Table 2 presents the characteristics of overdose events.

Table 3 presents the distribution and correlation of the responders' characteristics. Cramér *V* was used for categorical variables, and Spearman ρ was used for ordinal variables. ANOVA tests for age differences among the different subgroups of categorical or ordinal variables did not reveal any significant differences at the 5% significance level.

Table 2. Description of overdose event characteristics (n=188).

Variables	Values
Weekdays and weekends, n (%)[a]	
Weekday	136 (72.3)
Weekend	52 (27.7)
Days and nights, n (%)[a]	
Day	133 (70.7)
Night	55 (29.3)
Distance (meters; n=162[b]), mean (SD); median (IQR)[c]	3326 (2784); 2595 (955.09-5567.75)

[a]Cramér correlation between weekday/weekend and day/night is 0.006.

[b]For 26 (13.8%) of the 188 overdose events, distance data were not available.

[c]Distance during weekdays: mean 3611 (SD 2871) meters; distance during weekends: mean 2537 (SD 2384) meters; *P*=.03; distance during the day: mean 3507 (SD 2724) meters; distance during the night: mean 2870 (2910) meters; *P*=.19.

Table 3. Distribution and correlation of responders' characteristics (n=80).

Variable	Values, n (%)	Gender	Naloxone carriage adherence	Homelessness	Employment	History of witnessing an opioid overdose	History of administering naloxone	Age
Age[a]								
r	—[b]	0.07	0.18	0.13	0.14	−0.08	−0.07	1
P value	—	.54	.12	.26	.22	0.52	.54	—
Gender								
r	—	1	0.25	0.42	0.18	0.19	0.14	0.07
P value	—	—	.27	<.001	.25	.35	.58	.54
Male	35 (44)	—	—	—	—	—	—	—
Female	44 (55)	—	—	—	—	—	—	—
Intersex	1 (1)	—	—	—	—	—	—	—
Naloxone carriage adherence								
r	—	0.25	1	0.37	0.27	−0.18	−0.21	0.18
P value	—	.27	—	.02	.16	.05	.05	.12
All the time	36 (45)	—	—	—	—	—	—	—
Often	22 (28)	—	—	—	—	—	—	—
Sometimes	10 (13)	—	—	—	—	—	—	—
Seldom	2 (3)	—	—	—	—	—	—	—
Never	10 (13)	—	—	—	—	—	—	—
Homelessness								
r	—	0.42	0.37	1	0.37	0.22	0.09	0.13
P value	—	<.001	.02	—	.004	.16	.46	.26
Homeless	22 (28)	—	—	—	—	—	—	—
Not homeless	58 (73)	—	—	—	—	—	—	—
Employment								
r	—	0.18	0.27	0.37	1	0.15	0.14	0.14
P value	—	.25	.16	.004	—	.16	.58	.22
Part time	11 (14)	—	—	—	—	—	—	—
Full time	18 (23)	—	—	—	—	—	—	—
Unemployed	51 (64)	—	—	—	—	—	—	—
History of witnessing an opioid overdose (number of times)								
r	—	0.19	−0.18	0.22	0.15	1	0.63	−0.08
P value	—	.35	.05	.16	.16	—	<.001	.52
≤20	48 (60)	—	—	—	—	—	—	—
21-40	20 (25)	—	—	—	—	—	—	—
>40	12 (15)	—	—	—	—	—	—	—
History of administering naloxone (number of times)								
r	—	0.14	−0.21	0.09	0.14	0.63	1	−0.07
P value	—	.58	.05	.46	.58	<.001	—	.54
≤20	61 (81)	—	—	—	—	—	—	—
21-40	7 (9)	—	—	—	—	—	—	—
>40	7 (9)	—	—	—	—	—	—	—

[a]Age (y): mean 40.31 (SD 10.41); median 39.5 (IQR 32-47.75).

[b]Not applicable.

Significant correlations were found between gender and homelessness (P<.001) as well as between history of witnessing an opioid overdose and history of administering naloxone (P<.001).

Response Patterns

Textbox 3 and Figure 6 present how the alerted volunteers responded (true alarms only; n=993). Responders could change their decision.

Textbox 3. Volunteers' response patterns.

No answer

- Responder ignored the alert. This was the final status in 60.3% (599/993) of the alerts.

No go

- Responder notified the system that they are not able to respond. This was the final status in 23% (228/993) of the alerts.

En route

- Responder notified the system that they are on the way to the scene. This was the final status in 5.1% (51/993) of the alerts.

On scene

- Responder notified the system that they are on the scene. This status can be set automatically by the system (based on the responder's location) or manually by the responder. This was the final status in 2.6% (26/993) of the alerts.

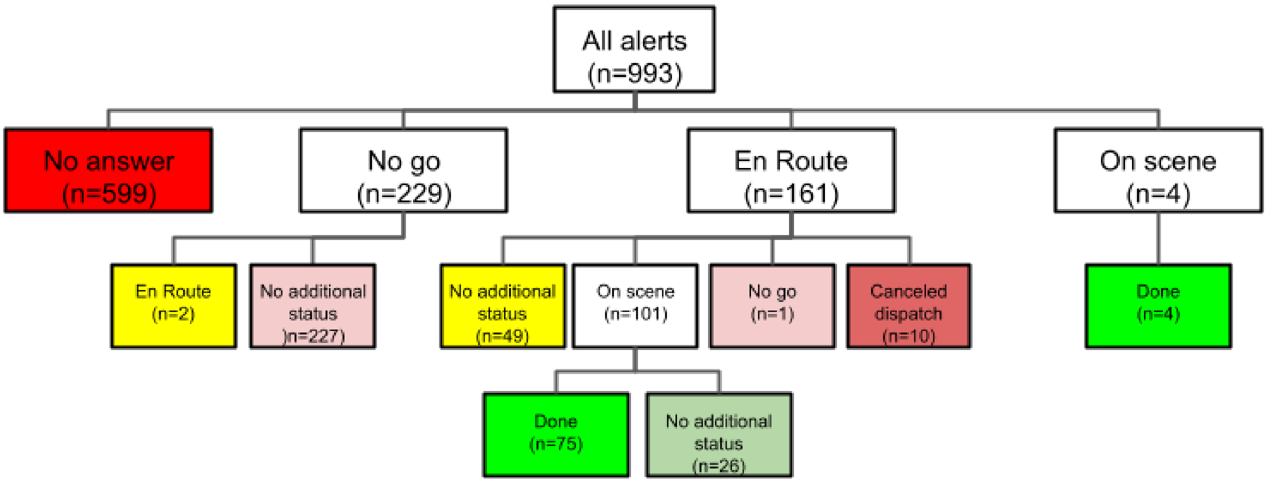
Done

- Responder performed the treatment. This was the final status in 7.9% (79/993) of the alerts.

Canceled dispatch

- This was the final status in 1% (10/993) of the alerts.

Figure 6. Response patterns (refer to Textbox 3 for an explanation of the terms used in this figure).



Classification Analysis of Response Patterns

Figure 7 presents the ability of each model to predict the responder's behavior. To compare model 4 with the other models, all models were tested using the test set of alerts (n=329).

For the test set, model 4 provided the best classification accuracy both overall and for ignored alerts. Model 3 provided the same classification accuracy for ignored alerts, slightly lower accuracy overall, and lower accuracy for answered alerts. Model 2 provided the best classification accuracy for ignored alerts; however, its accuracy was lower overall and significantly lower for answered alerts. Model 1's classification accuracy was the lowest.

Figure 8 presents the ability of models 1 to 3 to classify the responder's behavior, using the full data set (n=993).

For the full set, model 3 provided the best classification accuracy. Model 2 had similar accuracy for ignored events and lower accuracy both overall and for answered events. Model 1's classification accuracy was the lowest. Model 4 was not tested with the full set because the construction of the *frequent responder* variable requires training.

Figure 9 presents the J48 decision tree for model 4 for the test set.

Figure 7. Classification accuracy of models 1 to 4 using the test set (n=329).

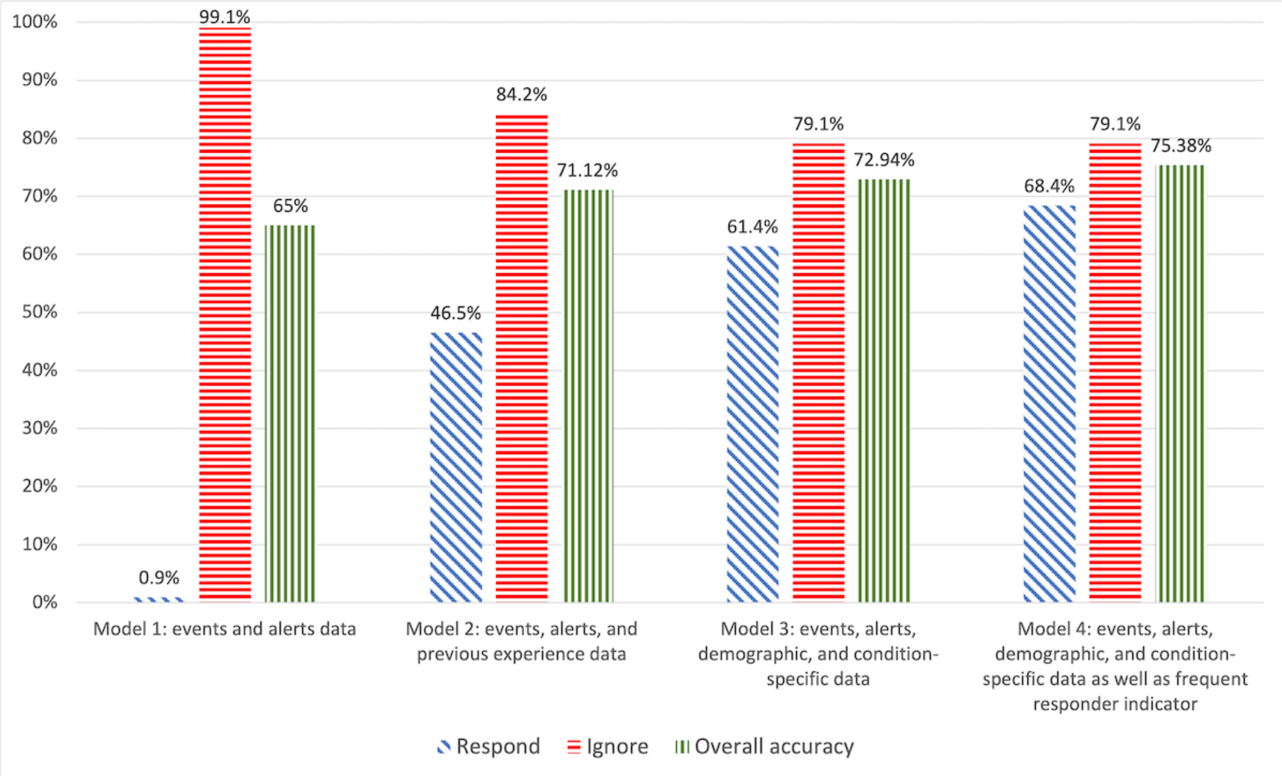


Figure 8. Classification accuracy of models 1 to 3 using the full set (n=993).

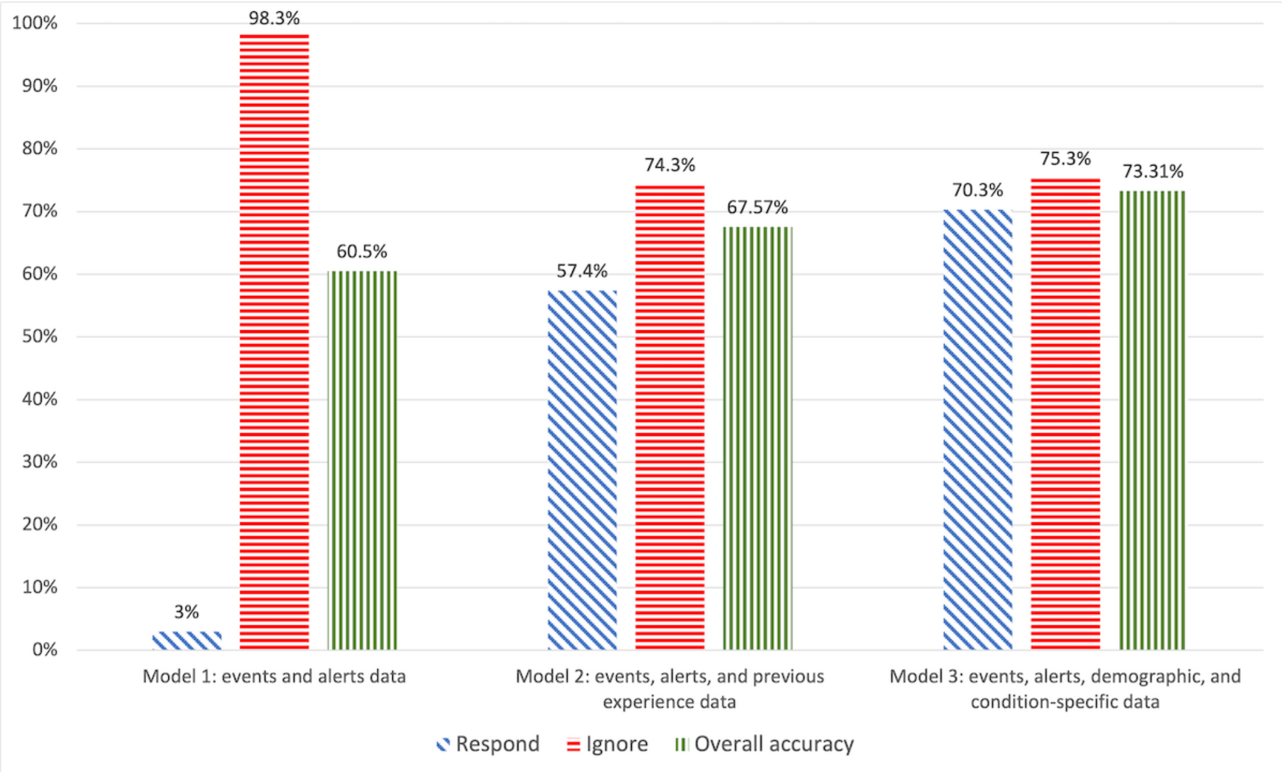
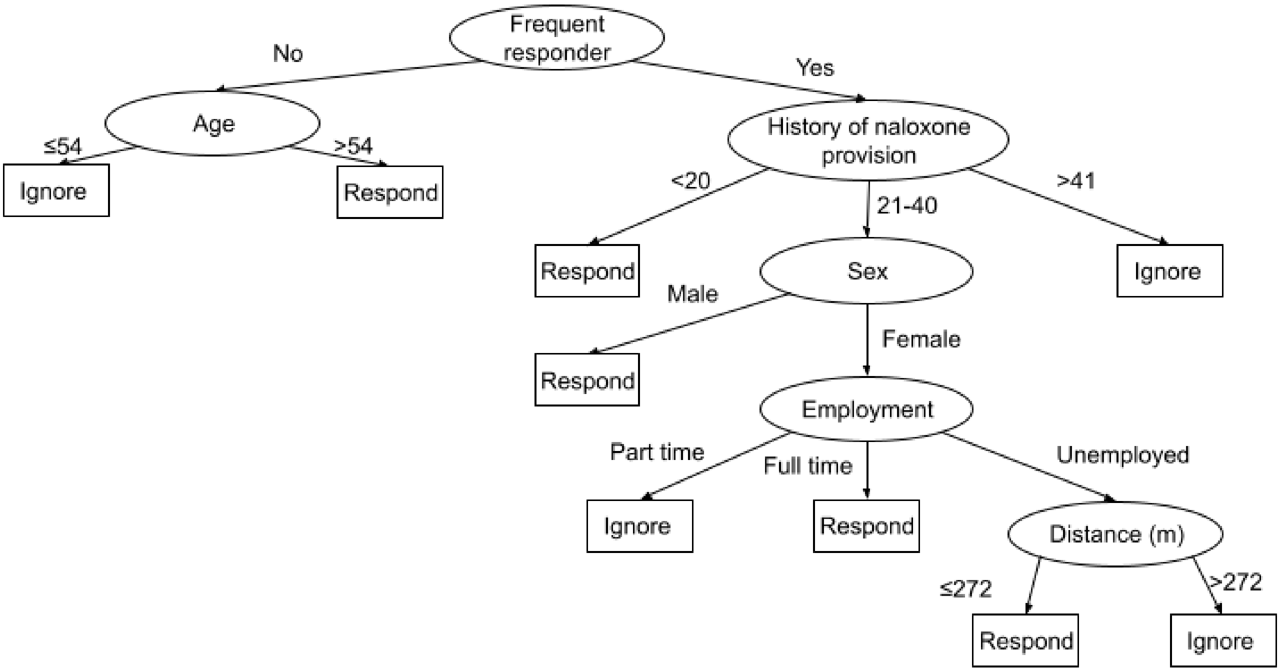


Figure 9. J48 decision tree for model 4 for the test set.



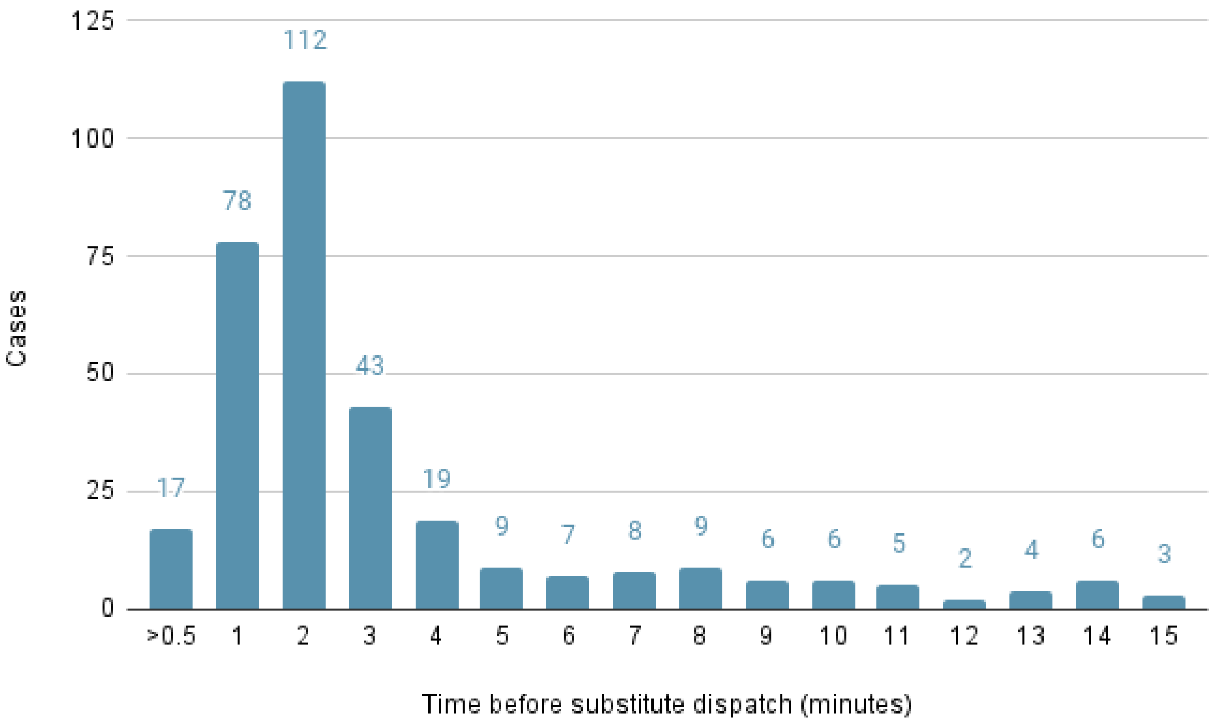
The analysis of the classification tree reveals 5 possible routes to the *response* result: infrequent responders aged >54 years, frequent responders who administered naloxone <20 times, male frequent responders who administered naloxone 21 to 40 times, fully employed female frequent responders who administered naloxone 21 to 40 times, and unemployed female frequent responders who administered naloxone 21 to 40 times in situations where the distance to the scene was <272 meters.

We have to remember that the overall accuracy is not very high and that there are false-positive and false-negative statistical errors in the classification output. A false-positive error occurs when the ignored alert is classified as a responded alert, and a false-negative error occurs when the responded alert is classified as an ignored alert.

Potential Time Savings

Substitute responders (responders who were not initially selected by the algorithm) were used in 73.4% (138/188) of the events. Substitute responders received 33.6% (334/993) of the alerts. Figure 10 presents the lengths of the delays (in min) before substitute responders were dispatched.

Figure 10. Time before substitute dispatch (n=334).



Factors Affecting Willingness to Respond to an Opioid Overdose Event

Table 4 presents the analysis of differences between alerts that were ignored and alerts that resulted in some responses (*en route*, *no go*, or *on scene*).

Significant differences between responded alerts and ignored alerts were found for the following variables: gender (higher response rate by male volunteers; *P*=.05), naloxone carriage adherence (*P*<.001), employment (higher response rate by volunteers who were unemployed; *P*<.001), age (slightly higher average age among volunteers who responded; *P*=.003), the number of previous alerts (higher among volunteers who responded; *P*=.003), previous false alerts (higher among volunteers who responded; *P*=.003), previous false alerts by the same signaler (lower among volunteers who responded; *P*=.02), previous responses (higher among volunteers who responded; *P*<.001), and previous responses to false alerts (higher among volunteers who responded; *P*<.001).

Table 4. Differences between responded alerts and ignored alerts (n=993).

Variable	Responded alerts (n=394)	Ignored alerts (n=599)	*P* value
Weekdays and weekends, n (%)			.49[a]
Weekday	289 (73.4)	451 (75.3)	
Weekend	105 (26.6)	148 (24.7)	
Days and nights, n (%)			.44[a]
Day	282 (71.6)	415 (69.3)	
Night	112 (28.4)	184 (30.7)	
Sex, n (%)			.05[a]
Male	182 (46.2)	239 (39.9)	
Female	212 (53.8)	360 (60.1)	
Naloxone carriage adherence, n (%)			<.001[a,b]
All the time	115 (29.2)	241 (40.2)	
Most of the time	174 (44.2)	150 (25)	
Sometimes	29 (7.4)	59 (9.8)	
Seldom	4 (1)	17 (2.8)	
Never	72 (18.3)	132 (22)	
Homelessness, n (%)			.18[a]
Yes	68 (17.3)	124 (20.7)	
No	326 (82.7)	475 (79.3)	
Employment, n (%)			<.001[a]
Part time	18 (4.6)	70 (11.7)	
Full time	70 (17.8)	110 (18.4)	
Unemployed	306 (77.7)	419 (69.9)	
Age (y), mean (SD)	42.91 (11.86)	40.47 (13.11)	.003[c]
Previous alerts, mean (SD)	25.00 (20.96)	21.09 (20.31)	.003[c]
Previous alerts by the same signaler, mean (SD)	3.35 (5.12)	3.52 (5.58)	.63[c]
Previous false alerts, mean (SD)	7.70 (7.00)	6.39 (6.46)	.003[c]
Previous false alerts by the same signaler, mean (SD)	0.70 (1.35)	0.96 (2.03)	.018[c]
Previous responses, mean (SD)	14.27 (14.01)	6.74 (10.40)	<.001[c]
Previous responses to false alerts, mean (SD)	1.39 (1.85)	0.82 (1.59)	<.001[c]
Previous false alerts by the same signaler, mean (SD)	0.13 (0.42)	0.13 (0.50)	.89[c]
Distance, mean (SD)	1947.24 (2290.24)	1726.98 (2127.02)	.16[c]
History of witnessed overdoses (number of times), n (%)			.54[a]
≤20	247 (62.7)	382 (63.8)	
21-40	80 (20.3)	106 (17.7)	
>40	67 (17)	111 (18.5)	
History of naloxone administration (number of times), n (%)			.07[a]
≤20	299 (86.4)	453 (80.6)	
21-40	14 (4)	38 (6.8)	
>40	33 (9.5)	71 (12.6)	

[a]P value for the chi-square test for the test of independence.
[b]P value for the Kendall τ test for ordinal variables.
[c]P value for the 2-tailed independent samples t test.

Discussion

Principal Findings

To the best of our knowledge, this is the first study that integrates the predictions of VFRs' response behavior into dispatch algorithms. We found that volunteers' past response behavior is the most influential predictor of future response behavior. Our findings suggest that the behavior-based approach can be applied to VFR dispatch to achieve a better response rate of the network as a whole.

Profiling of Responders and Personalization of VFR Dispatch

Model 1 used events and alerts data only and completely ignored any volunteer-related data. The ability of this model to predict volunteers' behavior was extremely low—only approximately 0.9% (3/329) of the responded alerts were classified correctly. This model would lead to the dispatch of all available volunteers, resulting in burnout, a "diffusion of responsibility," and low willingness to respond. We conclude that this model is unacceptable.

Model 2 assumed that the volunteers are completely anonymous and that the algorithm knows only their previous behavior in the VFR network. Using these data as well as events and alerts data, model 2 correctly predicted 84.2% (277/329) of the cases in the test set in which responders ignored an alert. Lower prediction accuracy for the full data set (738/993, 74.3%) can be explained by the inclusion of early events for which the model did not have enough data about previous behavior. The ability of model 2 to predict that a volunteer will respond to an event was lower (153/329, 46.5% for the test set and 570/993, 57.4% for the full data set). On the one hand, this model is expected to improve the response rate of the network, but, on the other hand, in half of the cases in which a volunteer responds, another volunteer would be dispatched unnecessarily. We conclude that this model should be used only if the responders are completely anonymous.

Model 3 used events and alerts data, responders' demographics, their prior experience of witnessing an opioid overdose, their prior experience in administering naloxone, and their naloxone carriage adherence *before* joining the VFR network. This model's ability to predict that a volunteer will ignore an alert was similar to that of model 2, but its ability to predict that a volunteer will respond to an event was higher (202/329, 61.4% in the test set and 698/993, 70.3% in the full data set). A closer look at the decision tree of this model reveals that the most influential variables were related to the volunteer's experience before joining the network: naloxone carriage adherence and the provision of naloxone to those experiencing an overdose. We conclude that this model can be used if data about a volunteer's app use behavior in the network are not currently available (eg, during the period between recruiting the volunteer and until they receive enough alerts).

Model 4 used all available data, including the frequency of events and alerts, volunteers' demographics, their prior overdose witnessing and naloxone provision experience before joining the VFR network, and their response behavior in the VFR network (according to the *frequent responder* indicator recalculated after each event). The ability of this model to predict that a volunteer will ignore an alert was similar to that of model 3, but its ability to predict that a volunteer will respond to an event was higher (225/329, 68.4% in the test set). The decision tree presented in Figure 9 reveals that the *frequent responder* indicator was the most influential variable. We conclude that model 4 achieves the best prediction accuracy and should be preferred whenever the necessary data are available.

Generalizability of the Proposed Approach

The dispatch of substitute responders is relevant whenever the initial subset of closest volunteers do not provide a response and is a common approach in VFR systems. However, valuable time is lost until nonresponse is identified. Although existing algorithms are based on technical variables such as distance from the scene and the ETA, this study introduces a completely new variable: volunteers' behavior and their probability to respond to a specific alert. The demonstrated importance of considering multiple factors in volunteer demographics and behavioral characteristics and the insights from the models we have tested are applicable wherever volunteer dispatch optimization is important. Such challenges are found in areas as diverse as food rescue operations [53], OHCA response [54,56], and mass casualty events [57]. Following our approach, these domains and more may find value in testing different sets of demographic and behavioral factors. Predicting the response behavior of candidate responders in advance of dispatch can allow any VFR system to choose the best possible response candidates based not only on ETA or location but also on the probability of actual response. The potential time savings depend on the network-specific period of time until a nonresponsive volunteer is considered unavailable, and a substitute responder is dispatched. The longer this time period, the greater the potential savings provided by a predictive dispatch algorithm.

The data used in our algorithm can be divided into four categories (Table 1): (1) event characteristics, (2) past responder behavior, (3) demographics, and (4) certain parameters specific to a medical condition relevant to the VFR network. The first 3 categories are directly generalizable because most responder mobilization apps collect and store these data, which, based on our findings, can be harvested for improved dispatch algorithms. The fourth data category includes factors that may differ depending on the medical condition relevant to the VFR network.

Factors Affecting Volunteers' Decision to Respond

Herein, we provide a brief discussion of the factors that affect volunteers' decisions to respond. A full analysis of these factors is beyond the scope of this research and should be pursued using a larger sample that may provide generalizability.

Experience, including the experience gained both before and after joining the VFR community, was found to be the most influential factor in volunteers' willingness to respond. In model 3, naloxone carriage adherence and experience in the provision of naloxone were the most influential factors, whereas in model 4, the *frequent responder* indicator was the most influential factor. Significant differences were found between responded alerts and ignored alerts for the following variables: naloxone carriage adherence, previous alerts, previous false alerts, previous false alerts by the same signaler, previous responses, and previous responses to false alerts.

Part-time employment led to lower willingness to respond.

The average age of volunteers who responded to alerts was a little higher than the average age of volunteers who ignored alerts. Model 4 revealed that age is an important factor for volunteers who are not *frequent responders*: volunteers aged >54 years are expected to respond.

Male volunteers had a higher willingness to respond, but this difference had borderline significance (P=.05). The results of model 4 were consistent with this difference.

Comparison With Prior Work

VFR dispatch has evolved in the last decades, thanks to technological advances and a better understanding of VFR network management. In a pretechnology era, VFRs (eg, volunteer firefighters) were alerted by sirens or other means rather than individually dispatched. Once pagers and SMS text messaging technology became available, the first generation of individual dispatch based on usual location was implemented (eg, the study by Zijlstra et al [43]). Further technological advances, including smartphone apps and GPS, enabled the second generation of individual dispatch based on current location [34,36,46,47] and the integration of VFRs' feedback into the algorithm [34]. The integration of the probability to respond based on event characteristics as well as VFRs' demographic data and previous behavior into the dispatch algorithm constitutes the third generation of individual dispatch, making this one of the first studies to harness the power of predictive analytics for VFR dispatch.

Limitations

A relatively small sample for a specific condition (opioid overdose) and a specific emergency intervention (the provision of naloxone) was used. The sample has specific socioeconomic characteristics: it included a significant proportion of people experiencing homelessness and those who were unemployed (volunteers may have lower motivation to help owing to these destabilizing factors), as well as a significant proportion of people dependent on drugs (volunteers may have lower response rates when intoxicated). The setting was very specific: a large number of outdoor opioid overdoses within a relatively small geographic area. The responders were aware that there were many trained bystanders nearby, and this could have led to the "diffusion of responsibility" phenomenon and reduced the willingness to respond. No randomization or control group was used.

Future Research

The proposed approach should be tested with a larger sample and for different conditions and interventions. A randomized study comparing the outcomes of the proposed dispatch algorithm with those of a regular location-based dispatch algorithm should be considered.

Machine learning techniques should be considered to calculate the *frequent responder* indicator. Future studies should examine whether the probability that a specific responder will respond to a specific event can be used instead of a binary indicator.

Further research is necessary on whether the proposed approach may have implications for multisided networks dispatching nonemergency services, such as ride sharing and package delivery.

Conclusions

In this research, we proposed a way to improve dispatch algorithms in VFR networks based on the individual characteristics of the volunteers and their behavior. We have shown that even in a relatively small sample, a classification model can predict with fair accuracy whether a specific volunteer will respond to a specific event or ignore it. Such prediction may improve the dispatchers' decision-making process and enable the dispatch of substitute responders without delay.

Our findings can help VFR network administrators in their continual efforts to improve the response rates and response times of their networks and to save lives.

Acknowledgments

This research was supported by a grant from the National Institute on Drug Abuse (R34DA044758).

Data Availability

The data sets analyzed during this study are not publicly available owing to Health Insurance Portability and Accountability Act (HIPAA) guidelines regarding patient privacy and the sensitivity of raw location data. Requests for data can be sent to the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of different algorithms for models 1 to 4.

[DOCX File , 30 KB - mhealth_v11i1e41551_app1.docx]

References

1. Matinrad N, Granberg TA, Vogel NE, Angelakis V. Optimal dispatch of volunteers to out-of-hospital cardiac arrest patients. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. 2019 Presented at: HICSS '19; January 8-11, 2019; Grand Wailea, HI p. 4088-4097 URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/1c1e1ead-4a8d-4f19-a4d2-65d2aee5b105/content [doi: 10.24251/hicss.2019.496]
2. Ramanayake RP, Ranasingha S, Lakmini S. Management of emergencies in general practice: role of general practitioners. J Family Med Prim Care 2014;3(4):305-308 [FREE Full text] [doi: 10.4103/2249-4863.148089] [Medline: 25657933]
3. Myat A, Song KJ, Rea T. Out-of-hospital cardiac arrest: current concepts. Lancet 2018 Mar 10;391(10124):970-979. [doi: 10.1016/S0140-6736(18)30472-0] [Medline: 29536861]
4. Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. Resuscitation 2010 Nov;81(11):1479-1487 [FREE Full text] [doi: 10.1016/j.resuscitation.2010.08.006] [Medline: 20828914]
5. Alberdi F, García I, Atutxa L, Zabarte M, TraumaNeurointensive Care Work Group of the SEMICYUC. Epidemiología del trauma grave. Med Intensiva 2014 Dec;38(9):580-588 [FREE Full text] [doi: 10.1016/j.medin.2014.06.012] [Medline: 25241267]
6. Minei JP, Schmicker RH, Kerby JD, Stiell IG, Schreiber MA, Bulger E, Resuscitation Outcome Consortium Investigators. Severe traumatic injury: regional variation in incidence and outcome. Ann Surg 2010 Jul;252(1):149-157 [FREE Full text] [doi: 10.1097/SLA.0b013e3181df0401] [Medline: 20531005]
7. Drug overdose deaths. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/drugoverdose/data/statedeaths.html [accessed 2020-12-20]
8. Panesar SS, Javad S, de Silva D, Nwaru BI, Hickstein L, Muraro A, EAACI Food Allergy and Anaphylaxis Group. The epidemiology of anaphylaxis in Europe: a systematic review. Allergy 2013 Nov;68(11):1353-1361 [FREE Full text] [doi: 10.1111/all.12272] [Medline: 24117770]
9. Turner PJ, Jerschow E, Umasunthar T, Lin R, Campbell DE, Boyle RJ. Fatal anaphylaxis: mortality rate and risk factors. J Allergy Clin Immunol Pract 2017 Sep;5(5):1169-1178 [FREE Full text] [doi: 10.1016/j.jaip.2017.06.031] [Medline: 28888247]
10. Hasselqvist-Ax I, Riva G, Herlitz J, Rosenqvist M, Hollenberg J, Nordberg P, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med 2015 Jun 11;372(24):2307-2315. [doi: 10.1056/NEJMoa1405796] [Medline: 26061835]
11. Malta Hansen C, Kragholm K, Pearson DA, Tyson C, Monk L, Myers B, et al. Association of bystander and first-responder intervention with survival after out-of-hospital cardiac arrest in North Carolina, 2010-2013. JAMA 2015 Jul 21;314(3):255-264. [doi: 10.1001/jama.2015.7938] [Medline: 26197186]
12. Sulovic LS, Pavlovic AP, Zivkovic JB, Zivkovic ZN, Filipovic-Danic SS, Trpkovic SV. Accidental drowning: the importance of early measures of resuscitation for a successful outcome. Case Rep Emerg Med 2018;2018:7525313 [FREE Full text] [doi: 10.1155/2018/7525313] [Medline: 29974001]
13. Walley AY, Xuan Z, Hackman HH, Quinn E, Doe-Simkins M, Sorensen-Alawad A, et al. Opioid overdose rates and implementation of overdose education and nasal naloxone distribution in Massachusetts: interrupted time series analysis. BMJ 2013 Jan 30;346:f174 [FREE Full text] [doi: 10.1136/bmj.f174] [Medline: 23372174]
14. Hanfling D, Altevogt BM, Viswanathan K, Gostin LO. Crisis Standards of Care: A Systems Framework for Catastrophic Disaster Response. Washington, DC: National Academies Press; 2012.
15. Razzak JA, Kellermann AL. Emergency medical care in developing countries: is it worthwhile? Bull World Health Organ 2002;80(11):900-905 [FREE Full text] [Medline: 12481213]
16. Chanta S, Mayorga ME, McLay LA. Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. Ann Oper Res 2011 Sep 27;221(1):133-159 [FREE Full text] [doi: 10.1007/s10479-011-0972-6]
17. Roudsari BS, Nathens AB, Arreola-Risa C, Cameron P, Civil I, Grigoriou G, et al. Emergency Medical Service (EMS) systems in developed and developing countries. Injury 2007 Sep;38(9):1001-1013 [FREE Full text] [doi: 10.1016/j.injury.2007.04.008] [Medline: 17583709]
18. Kitamura T, Kiyohara K, Sakai T, Matsuyama T, Hatakeyama T, Shimamoto T, et al. Public-access defibrillation and out-of-hospital cardiac arrest in Japan. N Engl J Med 2016 Oct 27;375(17):1649-1659. [doi: 10.1056/NEJMsa1600011] [Medline: 27783922]

19. Kiyohara K, Kitamura T, Sakai T, Nishiyama C, Nishiuchi T, Hayashi Y, et al. Public-access AED pad application and outcomes for out-of-hospital cardiac arrests in Osaka, Japan. Resuscitation 2016 Sep;106:70-75. [doi: 10.1016/j.resuscitation.2016.06.025] [Medline: 27373223]
20. Murakami Y, Iwami T, Kitamura T, Nishiyama C, Nishiuchi T, Hayashi Y, Utstein Osaka Project. Outcomes of out-of-hospital cardiac arrest by public location in the public-access defibrillation era. J Am Heart Assoc 2014 Apr 22;3(2):e000533 [FREE Full text] [doi: 10.1161/JAHA.113.000533] [Medline: 24755149]
21. Nakahara S, Tomio J, Ichikawa M, Nakamura F, Nishida M, Takahashi H, et al. Association of bystander interventions with neurologically intact survival among patients with bystander-witnessed out-of-hospital cardiac arrest in Japan. JAMA 2015 Jul 21;314(3):247-254. [doi: 10.1001/jama.2015.8068] [Medline: 26197185]
22. Elsner J, Meisen P, Thelen S, Schilberg D, Jeschke S. EMuRgency - a basic concept for an AI driven volunteer notification system for integrating laypersons into emergency medical services. Int J Adv Life Sci 2013;5(3 & 4):223-236. [doi: 10.1007/978-3-319-08816-7_46]
23. Berglund E, Claesson A, Nordberg P, Djärv T, Lundgren P, Folke F, et al. A smartphone application for dispatch of lay responders to out-of-hospital cardiac arrests. Resuscitation 2018 May;126:160-165 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.01.039] [Medline: 29408717]
24. Folke F, Lippert FK, Nielsen SL, Gislason GH, Hansen ML, Schramm TK, et al. Location of cardiac arrest in a city center: strategic placement of automated external defibrillators in public locations. Circulation 2009 Aug 11;120(6):510-517. [doi: 10.1161/CIRCULATIONAHA.108.843755] [Medline: 19635969]
25. Marshall AH, Cairns KJ, Kee F, Moore MJ, Hamilton AJ, Adgey AA. A Monte Carlo simulation model to assess volunteer response times in a public access defibrillation scheme in Northern Ireland. In: Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems. 2006 Presented at: CBMS '06; June 22-23, 2006; Salt Lake City, UT p. 783-788 URL: https://ieeexplore.ieee.org/document/1647666 [doi: 10.1109/cbms.2006.19]
26. Pijls RW, Nelemans PJ, Rahel BM, Gorgels AP. A text message alert system for trained volunteers improves out-of-hospital cardiac arrest survival. Resuscitation 2016 Aug;105:182-187 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.06.006] [Medline: 27327230]
27. Rea T, Blackwood J, Damon S, Phelps R, Eisenberg M. A link between emergency dispatch and public access AEDs: potential implications for early defibrillation. Resuscitation 2011 Aug;82(8):995-998. [doi: 10.1016/j.resuscitation.2011.04.011] [Medline: 21570169]
28. Sakai T, Iwami T, Kitamura T, Nishiyama C, Kawamura T, Kajino K, et al. Effectiveness of the new 'Mobile AED Map' to find and retrieve an AED: a randomised controlled trial. Resuscitation 2011 Jan;82(1):69-73. [doi: 10.1016/j.resuscitation.2010.09.466] [Medline: 21051130]
29. Smith CM, Wilson MH, Ghorbangholi A, Hartley-Sharpe C, Gwinnutt C, Dicker B, et al. The use of trained volunteers in the response to out-of-hospital cardiac arrest - the GoodSAM experience. Resuscitation 2017 Dec;121:123-126 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.020] [Medline: 29079507]
30. Takei Y, Kamikura T, Nishi T, Maeda T, Sakagami S, Kubo M, et al. Recruitments of trained citizen volunteering for conventional cardiopulmonary resuscitation are necessary to improve the outcome after out-of-hospital cardiac arrests in remote time-distance area: a nationwide population-based study. Resuscitation 2016 Aug;105:100-108 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.05.021] [Medline: 27263486]
31. Gaziel-Yablowitz M, Schwartz DG. A review and assessment framework for mobile-based emergency intervention apps. ACM Comput Surv 2018 Jan 10;51(1):1-32 [FREE Full text] [doi: 10.1145/3145846]
32. Schwartz DG, Bellou A, Garcia-Castrillo L, Muraro A, Papadopoulos N. Exploring mHealth participation for emergency response communities. Health Inf Sci Syst 2017;21:1378 [FREE Full text] [doi: 10.3127/ajis.v21i0.1378]
33. Khalemsky M, Schwartz DG, Silberg T, Khalemsky A, Jaffe E, Herbst R. Childrens' and parents' willingness to join a smartphone-based emergency response community for anaphylaxis: survey. JMIR Mhealth Uhealth 2019 Aug 27;7(8):e13892 [FREE Full text] [doi: 10.2196/13892] [Medline: 31456582]
34. Schwartz DG, Ataiants J, Roth A, Marcu G, Yahav I, Cocchiaro B, et al. Layperson reversal of opioid overdose supported by smartphone alert: a prospective observational cohort study. EClinicalMedicine 2020 Aug;25:100474 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100474] [Medline: 32954238]
35. Brooks SC, Simmons G, Worthington H, Bobrow BJ, Morrison LJ. The PulsePoint Respond mobile device application to crowdsource basic life support for patients with out-of-hospital cardiac arrest: challenges for optimal implementation. Resuscitation 2016 Jan;98:20-26 [FREE Full text] [doi: 10.1016/j.resuscitation.2015.09.392] [Medline: 26475397]
36. Andelius L, Malta Hansen C, Lippert FK, Karlsson L, Torp-Pedersen C, Kjær Ersbøll A, et al. Smartphone activation of citizen responders to facilitate defibrillation in out-of-hospital cardiac arrest. J Am Coll Cardiol 2020 Jul 07;76(1):43-53 [FREE Full text] [doi: 10.1016/j.jacc.2020.04.073] [Medline: 32616162]
37. Riegel B, Mosesso VN, Birnbaum A, Bosken L, Evans LM, Feeny D, et al. Stress reactions and perceived difficulties of lay responders to a medical emergency. Resuscitation 2006 Jul;70(1):98-106. [doi: 10.1016/j.resuscitation.2005.10.029] [Medline: 16753251]

38. Swor R, Khan I, Domeier R, Honeycutt L, Chu K, Compton S. CPR training and CPR performance: do CPR-trained bystanders perform CPR? Acad Emerg Med 2006 Jun;13(6):596-601 [FREE Full text] [doi: 10.1197/j.aem.2005.12.021] [Medline: 16614455]
39. Chew KS, Ahmad Razali S, Wong SS, Azizul A, Ismail NF, Robert SJ, et al. The influence of past experiences on future willingness to perform bystander cardiopulmonary resuscitation. Int J Emerg Med 2019 Dec 12;12(1):40 [FREE Full text] [doi: 10.1186/s12245-019-0256-5] [Medline: 31830912]
40. Zaffar MA, Rajagopalan HK, Saydam C, Mayorga M, Sharer E. Coverage, survivability or response time: a comparative study of performance statistics used in ambulance location models via simulation–optimization. Oper Res Health Care 2016 Dec;11:1-12 [FREE Full text] [doi: 10.1016/j.orhc.2016.08.001]
41. Başar A, Çatay B, Ünlüyurt T. A taxonomy for emergency service station location problem. Optim Lett 2011 Aug 9;6(6):1147-1160 [FREE Full text] [doi: 10.1007/s11590-011-0376-1]
42. Camacho-Collados M, Liberatore F. A decision support system for predictive police patrolling. Decis Support Syst 2015 Jul;75:25-37 [FREE Full text] [doi: 10.1016/j.dss.2015.04.012]
43. Zijlstra JA, Stieglis R, Riedijk F, Smeekes M, van der Worp WE, Koster RW. Local lay rescuers with AEDs, alerted by text messages, contribute to early defibrillation in a Dutch out-of-hospital cardiac arrest dispatch system. Resuscitation 2014 Nov;85(11):1444-1449 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.07.020] [Medline: 25132473]
44. Rao G, Choudhury S, Lingras P, Savage D, Mago V. SURF: identifying and allocating resources during out-of-hospital cardiac arrest. BMC Med Inform Decis Mak 2020 Dec 30;20(Suppl 11):313 [FREE Full text] [doi: 10.1186/s12911-020-01334-4] [Medline: 33380330]
45. Timmons S, Vernon-Evans A. Why do people volunteer for community first responder groups? Emerg Med J 2013 Mar;30(3):e13 [FREE Full text] [doi: 10.1136/emermed-2011-200990] [Medline: 22508195]
46. Jaffe E, Dadon Z, Alpert EA. Wisdom of the crowd in saving lives: the life guardians app. Prehosp Disaster Med 2018 Oct;33(5):550-552. [doi: 10.1017/S1049023X18000754] [Medline: 30220268]
47. Metelmann C, Metelmann B, Kohnen D, Brinkrolf P, Andelius L, Böttiger BW, et al. Smartphone-based dispatch of community first responders to out-of-hospital cardiac arrest - statements from an international consensus conference. Scand J Trauma Resusc Emerg Med 2021 Feb 01;29(1):29 [FREE Full text] [doi: 10.1186/s13049-021-00841-1] [Medline: 33526058]
48. Van de Velde S, Roex A, Vangronsveld K, Niezink L, Van Praet K, Heselmans A, et al. Can training improve laypersons helping behaviour in first aid? A randomised controlled deception trial. Emerg Med J 2013 Apr;30(4):292-297 [FREE Full text] [doi: 10.1136/emermed-2012-201128] [Medline: 22562070]
49. Elsner J, Meisen P, Ewert D, Schilberg D, Jeschke S. Prescient profiling – AI driven volunteer selection within a volunteer notification system. In: Jeschke S, Isenhardt I, Hees F, Henning K, editors. Automation, Communication and Cybernetics in Science and Engineering. Cham, Switzerland: Springer; 2014:597-607.
50. van Dam JW, van de Velden M. Online profiling and clustering of Facebook users. Decis Support Syst 2015 Feb;70:60-72 [FREE Full text] [doi: 10.1016/j.dss.2014.12.001]
51. Schetgen L, Bogaert M, Van den Poel D. Predicting donation behavior: acquisition modeling in the nonprofit sector using Facebook data. Decis Support Syst 2021 Feb;141:113446 [FREE Full text] [doi: 10.1016/j.dss.2020.113446]
52. Ward AM, McKillop DG. Profiling: a strategy for successful volunteer recruitment in credit unions. Financ Account Manag 2010;26(4):367-391 [FREE Full text] [doi: 10.1111/j.1468-0408.2010.00506.x]
53. Shi ZR, Yuan Y, Lo K, Lizarondo L, Fang F. Improving efficiency of volunteer-based food rescue operations. Proc AAAI Conf Artif Intell 2020 Apr 03;34(08):13369-13375 [FREE Full text] [doi: 10.1609/aaai.v34i08.7051]
54. Bray JE, Smith CM, Nehme Z. Community volunteer responder programs in cardiac arrest: the horse has bolted, it's time to optimize. J Am Coll Cardiol 2023 Jul 18;82(3):211-213. [doi: 10.1016/j.jacc.2023.05.018] [Medline: 37438007]
55. Jonsson M, Berglund E, Baldi E, Caputo ML, Auricchio A, Blom MT, ESCAPE-NET Investigators. Dispatch of volunteer responders to out-of-hospital cardiac arrests. J Am Coll Cardiol 2023 Jul 18;82(3):200-210. [doi: 10.1016/j.jacc.2023.05.017] [Medline: 37438006]
56. Gregers MC, Andelius L, Kjoelbye JS, Juul Grabmayr A, Jakobsen LK, Bo Christensen N, et al. Association between number of volunteer responders and interventions before ambulance arrival for cardiac arrest. J Am Coll Cardiol 2023 Mar 21;81(7):668-680 [FREE Full text] [doi: 10.1016/j.jacc.2022.11.047] [Medline: 36792282]
57. Yafe E, Walker BB, Amram O, Schuurman N, Randall E, Friger M, et al. Volunteer first responders for optimizing management of mass casualty incidents. Disaster Med Public Health Prep 2019 Apr 4;13(2):287-294. [doi: 10.1017/dmp.2018.56] [Medline: 29860960]

Abbreviations

AED: automatic external defibrillator
CPR: cardiopulmonary resuscitation
EMS: emergency medical services
ERC: emergency response community

ETA: estimated time of arrival
OHCA: out-of-hospital cardiac arrest
VFR: volunteer first responder

Edited by L Buis; submitted 30.07.22; peer-reviewed by A Mavragani, L Wu, X Chen; comments to author 12.04.23; revised version received 03.06.23; accepted 17.10.23; published 28.11.23.

Please cite as:
Khalemsky M, Khalemsky A, Lankenau S, Ataiants J, Roth A, Marcu G, Schwartz DG
Predictive Dispatch of Volunteer First Responders: Algorithm Development and Validation
JMIR Mhealth Uhealth 2023;11:e41551
URL: https://mhealth.jmir.org/2023/1/e41551
doi:10.2196/41551
PMID:38015602

Personal Information Protection and Privacy Policy Compliance of Health Code Apps in China: Scale Development and Content Analysis

Jiayi Jiang[*], PhD; Zexing Zheng[*], PhD

Law School, Central South University, Changsha, China

[*]all authors contributed equally

Corresponding Author:
Zexing Zheng, PhD

Abstract

Background: Digital technologies, especially contact tracing apps, have been crucial in monitoring and tracing the transmission of COVID-19 worldwide. China developed health code apps as an emergency response to the pandemic with plans to use them for broader public health services. However, potential problems within privacy policies may compromise personal information (PI) protection.

Objective: We aimed to evaluate the compliance of the privacy policies of 30 health code apps in the mainland of China with the Personal Information Protection Law (PIPL) and related specifications.

Methods: We reviewed and assessed the privacy policies of 30 health code apps between August 26 and September 6, 2023. We used a 3-level indicator scale based on the information life cycle as provided in the PIPL and related specifications. The scale comprised 7 level-1 indicators, 26 level-2 indicators, and 71 level-3 indicators.

Results: The mean compliance score of the 30 health code apps was 59.9% (SD 22.6%). A total of 13 (43.3%) apps scored below this average, and 6 apps scored below 40%. Level-1 indicator scores included the following: general attributes (mean 85.6%, SD 23.3%); PI collection and use (mean 66.2%, SD 22.7%); PI storage and protection (mean 63.3%, SD 30.8%); PI sharing, transfer, disclosure, and transmission (mean 57.2%, SD 27.3%); PI deletion (mean 52.2%, SD 29.4%); individual rights (mean 59.3%, SD 25.7%); and PI processor duties (mean 43.7%, SD 23.8%). Sensitive PI protection compliance (mean 51.4%, SD 26.0%) lagged behind general PI protection (mean 83.3%, SD 24.3%), with only 1 app requiring separate consent for sensitive PI processing. Additionally, 46.7% (n=14) of the apps needed separate consent for subcontracting activities, while fewer disclosed PI recipient information (n=13, 43.3%), safety precautions (n=11, 36.7%), and rules of PI transfer during specific events (n=10, 33.3%). Most privacy policies specified the PI retention period (n=23, 76.7%) and postperiod deletion or anonymization (n=22, 73.3%), but only 6.7% (n=2) were committed to prompt third-party PI deletion. Most apps delineated various individual rights: the right to inquire (n=25, 83.3%), correct (n=24, 80%), and delete PI (n=24, 80%); cancel their account (n=21, 70%); withdraw consent (n=20, 60%); and request privacy policy explanations (n=24, 80%). Only a fraction addressed the rights to obtain copies (n=4, 13.3%) or refuse advertisement of automated decision-making (n=1, 3.3%). The mean compliance rate of PI processor duties was only 43.7% (SD 23.8%), with significant deficiencies in impact assessments (mean 5.0%, SD 19.8%), PI protection officer appointment (mean 6.7%, SD 24.9%), regular compliance audits (mean 6.7%, SD 24.9%), and complaint management (mean 37.8%, SD 39.2%).

Conclusions: Our analysis revealed both strengths and significant shortcomings in the compliance of privacy policies of health code apps with the PIPL and related specifications considering the information life cycle. As China contemplates the future extended use of health code apps, it should articulate the legitimacy of the apps' normalization and ensure that users provide informed consent. Meanwhile, China should raise the compliance level of relevant privacy policies and fortify its enforcement mechanisms.

(JMIR Mhealth Uhealth 2023;11:e48714) doi:10.2196/48714

KEYWORDS

contact tracing; privacy policy; personal information protection; compliance; content analysis; COVID-19

Introduction

Background

The COVID-19 pandemic has posed significant public health challenges worldwide, prompting many countries to adopt various digital strategies, including contact-tracing apps to monitor the transmission of the virus [1,2]. These technological innovations have allowed for an unprecedented level of information collection, aggregation, analysis, and dissemination [3], and have brought significant benefits, such as the ability to track cases, identify potential outbreaks, and inform public health interventions [4,5]. However, the extensive use and accelerated development of contact-tracing apps have raised concerns regarding individual privacy rights and personal information (PI) [6-9], with lasting and profound impacts on data governance and PI protection [10]. To address these concerns, proponents of these apps, including governments and developers, should follow legal guidelines concerning privacy principles and policy content [11-13].

Although the short-term use of contact-tracing apps may be justified given the public emergencies the pandemic caused, the legitimacy of long-term use after the pandemic should be assessed to enhance PI protection and improve data governance capabilities [14]. As the global pandemic situation has stabilized, some countries have evaluated the necessity of using contact-tracing apps and decided to shut down such services. In the United Kingdom, the National Health Service COVID-19 app closed down on April 27, 2023 [15], after preventing around 1 million cases, 44,000 hospitalizations, and 9600 deaths in its first year alone [16]. In India, the contact-tracing feature of the Aarogya Setu app has been disabled, and the contact-tracing data collected has been deleted [17]. Singapore's government has been progressively rolling back its TraceTogether and SafeEntry platforms as the global pandemic situation stabilized, with all identifiable data collected via the two platforms wiped from their servers and databases [18].

In China, health code apps have been implemented as an essential tool for classifying citizens based on different transmission risk levels, quickly locating people who are potentially infected, and implementing control measures promptly [19-21]. These apps are designed to be a dynamic health certification that allows government agencies, employers, and communities to identify personal health risk levels and grants individuals the qualifications for mobility and work resumption [22]. Health code apps collect various types of sensitive PI. According to the Personal Health Information Code-Data Format (GB/T 38962-2020), PI collected to generate health codes includes the user's identity (eg, name, contact information, and medical history), health (eg, body temperature and current symptoms), travel (eg, residence and geographical location), and health certificate (eg, health risk level, evaluation, and medical examination results), among other information. Sensitive PI collected to generate health codes flows from users to digital platforms, medical institutions, and governments in existing practices [23]. The collection of PI is susceptible to invasion, which can lead to discrimination and harm [24]. Therefore, the increasing risk of leaks and abuse of PI due to the aggregated storage of data collected to generate health codes has raised concerns about privacy violations from public powers [25,26].

Instead of shutting down health code apps in the postpandemic era, China has promoted the use of health code apps for a broader public health service scope [27]. Each citizen will have a dynamic electronic health file and a multifunctional health code by 2025 [28], which will act as a major index of personal health information in disease prevention, medical care, rehabilitation, and health management [29]. Meanwhile, health code apps will serve as strategic health and medical service platforms. This strategy will turn individual health data, such as medical records and biometrics, into critical assets to reinforce government administration and social management [30]. Such an extension has raised concerns regarding the normalization of health code apps in the postpandemic society and the routinization of expanded government power [31,32].

China has realized the importance and urgency of PI protection and established a regulatory framework. Some national voluntary standards have come into effect, in particular the Information Security Technology–Personal Information Specification (GB/T 35273-2020; PI Specifications), which laid out granular guidelines for how PI should be collected, used, and shared to operate health code apps [33]. In addition, the Personal Information Protection Law (PIPL), which came into force on November 11, 2021, guarantees the rights of individuals and places constraints on PI processors. PIPL is regarded as a milestone for regulating PI protection specifically [34].

The existing application and potential normalization of health code apps have presented significant challenges for PI protection [14,35,36]. However, the legal compliance of health code apps remains unclear. Experts, authorities, and users need to assess the risks of PI protection and determine the future of health code apps. Notably, several potential problems within the privacy policies of health code apps may compromise the effectiveness of legal protections for PI, including the readability of the privacy policies, extensive PI collection, multiple processing purposes, indeterminate storage duration, and ambiguous privacy policy content [12,13,37,38].

Objective

In this study, we aimed to collect the privacy policies of health code apps developed by the provincial administrative regions in the mainland of China and assess the compliance of these privacy policies with the PIPL and PI Specifications from the information life cycle. We hope this study can contribute to the global discussion on balanced policies for PI protection in digital health initiatives in the postpandemic era, providing insights for policy makers, health code developers, and users across different countries while highlighting the importance of improving legal compliance and strengthening enforcement.

Methods

We conducted a content analysis of the privacy policies of health code apps developed in 31 provincial administrative regions in the mainland of China and evaluated their compliance with the PIPL and PI Specifications.

Ethical Considerations

Ethical review does not apply to our research because no experiments on human participants were completed.

Apps Access and Privacy Policies Collection

We searched for health code apps developed by provincial administrative regions on August 24, 2023. We accessed the health code apps of various provincial administrative regions through the National Government Service Platform, a national digitally integrated platform of government services available on the WeChat mini-program (operated by Tencent) and Alipay (operated by Alibaba). We obtained and reviewed the full text of corresponding privacy policies as text files or screenshots from the WeChat mini-program, Alipay, and Baidu, a well-known Chinese search engine, between August 25 and September 6, 2023.

Scale Development and Scoring

We used level-1 evaluation indicators based on the information life cycle as provided in the PI Specifications and the PIPL. These indicators encompassed the following stages: PI collection and use; PI storage and protection; PI sharing, transfer, disclosure, and transmission; PI deletion; general attributes; individual rights; and PI processor duties. We further elaborated these categories into 26 level-2 indicators and 71 level-3 indicators, each aligned with the specific provisions of the PIPL and PI Specifications. We provided brief explanations, example sentences, and corresponding references to provisions of the PI Specifications and the PIPL in Multimedia Appendix 1.

We assigned a score of 1 for each level-3 indicator if the privacy policy complied with the specific indicator and a score of 0 if it did not. Each level-3 indicator's compliance rate was determined as the proportion of policies that scored "1" from the sample of 30 apps. The scoring rate of each level-2 indicator was the arithmetic mean of the scoring rates of all associated level-3 indicators. Likewise, the compliance rate of level-1 indicators was the mean of its corresponding level-2 rates, thus representing the overall compliance of each app at specific information life cycle stages. For each privacy policy, the aggregate of all level-3 indicators was calculated as a total score and converted into a percentage system as a final score to denote the overall compliance of a given policy. Two independent raters (JJ and ZZ) collaboratively assessed all 30 privacy policies between August 25 and September 6, 2023.

Results

Sample Collection

We accessed the health code apps of all 31 provincial administrative regions in the mainland of China and obtained the full text of 30 privacy policies, including 23 from WeChat, 3 from Alipay, and 4 sourced manually from Baidu. The privacy policy of the health code app of Chongqing City was unavailable on the referenced platforms or search engine. Notably, the health code apps for Heilongjiang Province and Qinghai Province lacked distinct privacy policies. Heilongjiang's approach involved a tick box where users ensured the accuracy of information for COVID-19 prevention and control, while Qinghai integrated its privacy provisions within the user agreement. In addition, the health code apps of Ningxia Hui Autonomous Region and Tibet Autonomous Region used a common privacy policy template, differing only in the basic PI processor information.

Compliance Evaluation

The overall compliance landscape among the 30 assessed privacy policies of health code apps presented a mixed picture. The mean compliance score of the 30 privacy policies was 59.9% (SD 22.6%). A total of 17 (56.7%) apps surpassed the mean score, while 13 (43.3%) apps fell below it.

The evaluation results on the privacy policies' level-1 and level-2 indicators are listed in Table 1. The level-1 indicators were ranked from highest to lowest scores as follows: general attributes (mean 85.6%, SD 23.3%); PI collection and use (mean 66.2%, SD 22.7%); PI storage and protection (mean 63.3%, SD 30.8%); individual rights (mean 59.3%, SD 25.7%); PI sharing, transfer, disclosure, and transmission (mean 57.2%, SD 27.3%); PI deletion (mean 52.2%, SD 29.4%); and PI processors duties (mean 43.7%, SD 23.8%). The names and evaluation results of each app are listed in Multimedia Appendix 2.

Table . Compliance evaluation rates for level-1 and level-2 indicators in privacy policies.

Evaluation results on level-1 and level-2 indicators		Compliance rate (%), mean (SD)
General attributes		85.6 (23.3)
	PI[a] processors and service	93.3 (21.3)
	Policy transparency	95.6 (18.7)
	Policy maintenance	74.2 (34.5)
PI collection and use		66.2 (22.7)
	Collection and use of general PI in service functions	83.3 (24.3)
	Collection and use of sensitive PI in service functions	51.4 (26.0)
PI storage and protection		63.3 (30.8)
	Storage security	65.8 (32.6)
	Security incidents	60.0 (35.9)
PI sharing, transfer, disclosure, and transmission		57.2 (27.3)
	Subcontracting of PI processing	46.7 (28.7)
	PI sharing and transfer	53.3 (32.4)
	Public disclosure	85.0 (32.0)
	Cross-border transmission	53.3 (38.6)
PI deletion		52.2 (29.4)
	Retention period	76.7 (42.3)
	Deletion and cessation	40.0 (27.1)
Individual rights		59.3 (25.7)
	Inquiry of PI	83.3 (37.3)
	Obtain copies of PI	13.3 (34.0)
	Correction of PI	80.0 (40.0)
	Deletion of PI	80.0 (40.0)
	Explanation regarding PI processing	80.0 (40.0)
	Consent withdrawal	31.7 (27.3)
	Deregistration	70.0 (45.8)
	Consent exception scenarios	63.3 (48.2)
PI processor duties		43.7 (23.8)
	PI protection officer disclosure	6.7 (24.9)
	Compliance audits	6.7 (24.9)
	Impact assessment procedures	5.0 (19.8)
	Request management	60.7 (30.3)
	Complaint management	37.8 (39.2)

[a]PI: personal information.

The privacy policies' general attributes (mean 85.6%, SD 23.3%) scored high, which reflected their transparency and maintenance. Some level-2 indicators scored notably high, including PI processors and service (mean 93.3%, SD 21.3%) and policy transparency (mean 95.6%, SD 18.7%). These high scores indicated that most privacy policies identified who was responsible for processing PI and providing services. Policy maintenance was another strong area, with a score of 74.2% (SD 34.5%). Of the 30 apps, most appeared to be diligent in updating their policies, with 25 (83.3%) indicating they did so occasionally. Additionally, 80% (n=24) of the apps notified users about updates through various methods (eg, email or pop-up alerts). More than half of the privacy policies exceeded requirements by obtaining separate consent for specific policy changes (n=20, 66.7%) or by clearly marking the effective or updated dates (n=20, 66.7%). Specifically, 16 (53.3%) apps updated their privacy policies after the PIPL came into force. However, concerningly, 10 (33.3%) apps failed to mention either the effective date or the updated date of their policies, while 4 apps updated their privacy policies before the PIPL came into effect.

The scoring rate of collection and use of general PI (mean 83.3%, SD 24.3%) was high, which indicated that a majority of health code apps were attentive to articulating the collection and use of general PI in their privacy policies. The privacy policies of almost all apps delineated the purpose (n=29, 96.7%) and methods (n=29, 96.7%) of collecting and using general PI. A large number (n=28, 93.3%) of apps listed the types of PI collected and used, while 25 (83%) elaborated on the specific service functions that require such information. This suggested that users were generally well-informed about how their PI would be collected and used. However, a modest majority of the privacy policies differentiated between necessary PI and nonessential PI (n=19, 63.3%) and explained the consequences of failing to provide certain types of PI (n=20, 66.7%). This left room for improvement in ensuring that individuals fully understood which information was mandatory versus optional and the consequences of not providing it. By contrast, the scoring rate of collection and use of sensitive PI (mean 51.4%, SD 26.0%) was relatively low. Most privacy policies communicated the implications of processing sensitive PI (n=24, 80%) and required consent for collecting PI from minors (n=24, 80%). More than half of the privacy policies outlined protective measures (n=19, 63.3%) and specified the purposes (n=18, 60%) for collecting and using sensitive PI. However, only about one-third of health code apps highlighted what constituted sensitive PI (n=11, 36.7%) or sufficiently described the necessity for processing such sensitive PI (n=11, 36.7%). Only 1 app explicitly required separate consent for processing sensitive PI.

In the PI storage and protection stage (mean 63.3%, SD 30.8%), the scoring rate of level-2 indicators varied slightly. The mean compliance rate of storage security was 65.8% (SD 32.6%). Most of the 30 apps outlined the level of technical security measures satisfactorily (n=27, 90%) and informed users about potential security risks (n=25, 83.3%). Moreover, around half of the apps (n=16, 53.3%) extended their security explanations to include organizational management measures, while only 11 (36.7%) policies discussed the PI security agreements or certifications obtained. As for security incidents (mean 60.0%, SD 35.9%), although a significant portion of apps were committed to notifying (n=23, 76.7%) and reporting security incidents (n=23, 76.7%), only 26.7% (n=8) of PI processors were committed to assuming legal responsibilities in the event of such incidents.

In the stage of PI sharing, transfer, disclosure, and transmission (mean 57.2%, SD 27.3%), the scoring rate of level-2 indicators varied substantially. As for public disclosure (mean 85.0%, SD 32.0%) in the privacy policies, we observed high compliance in specifying conditions for potential public PI disclosure (n=25, 83.3%) and requiring separate consent for such practices (n=26, 86.7%). These rates indicated a high degree of transparency and respect for user consent for public disclosure. The compliance rate of PI sharing and transfer (mean 53.3%, SD 32.4%) indicated a mixed landscape in terms of transparency and user consent in these practices. Most apps required separate consent for sharing or transferring PI (n=27, 90%), whereas only half of the apps described the methods of PI transfer (n=16, 53.3%) and types of PI involved in the transfer (n=15, 50%). Less than half of the policies disclosed the basic information about PI recipients (n=13, 43.3%) and safety precautions (n=11, 36.7%). Rules governing PI transfer during specific events were missing in 66.7% (n=20) of privacy policies. As for cross-border transmission (mean 53.3%, SD 38.6%), most apps specified storage locations of PI (n=22, 73.3%), while only one-third of apps mentioned compliance with relevant cross-border transmission laws (n=10, 33.3%). As for subcontracting PI processing (mean 46.7%, SD 28.7%), less than half of the apps required separate consent for these activities (n=14, 46.7%) or ensured supervision (n=14, 46.7%), mainly via signed agreements.

In the stage of PI deletion (mean 52.2%, SD 29.4%), most privacy policies of the 30 apps stated the PI retention period (n=23, 76.7%), with 2 mentioning that PI would be retained for more than 6 months. By contrast, the mean scoring rate of deletion and cessation was low (mean 40.0%, SD 27.1). Although most apps committed to PI deletion or anonymization after the retention period (n=22, 73.3%), only 2 (n=2, 6.7%) apps claimed to notify third parties to delete or cease processing PI.

Concerning individual rights (mean 59.3%, SD 25.7%), most of the 30 apps explained individuals' various rights effectively, including the right to inquire (n=25, 83.3%), correct (n=24, 80%), and delete PI (n=24, 80%); cancel the account (n=21, 70%); withdraw consent (n=18, 60%); and request an explanation of the privacy policy (n=24, 80%). Providing this information empowered individuals to exercise their rights. However, only a few apps (n=4, 13.3%) recognized the right to obtain copies, and only 1 app explained the right to refuse business marketing using automated decision-making. In addition, a majority of the apps (n=21, 70%) listed exceptions for obtaining consent as provided by applicable laws or administrative regulations.

Concerning PI processor duties, we found a mean compliance rate of 43.7% (SD 23.8%). While many of the 30 apps provided methods for individuals to inquire (n=25, 83.3%), correct (n=24, 80%), and delete PI (n=24, 80%) as well as cancel their account (n=21, 70%) and withdraw consent (n=18, 60%), there was a significant shortfall in institutional oversight and risk management. Specifically, only a small percentage of PI processors appointed a PI protection officer (n=2, 6.7%) or conducted a PI protection assessment (n=2, 6.7%), with only 1 case assessing the purpose, method, impact, and protective measures for processing PI. Regular compliance audits were almost nonexistent (n=2, 6.7%). Furthermore, although many apps provided avenues for inquiries and complaints by disclosing contact information for requests (n=25, 83.3%) and means for complaints (n=15, 50%), only a minority of them committed to addressing these within 30 days or a legal time limit (n=19, 63.3% for requests and n=9, 30% for complaints). Only 3 apps explained the limitations of the use of automated decision-making in the information system.

Discussion

Principal Findings

In this study, we reviewed 30 privacy policies of health code apps in the mainland of China and assessed the compliance of these privacy policies with the PIPL and PI Specifications. Bardus et al [13] presented a systematic review of COVID-19 contact-tracing apps used worldwide to analyze apps' approach to data protection and privacy. However, they only identified one health code app (ie, Alipay Health Code) on May 28, 2020, and found the privacy policy was not available, which excluded China from their scope. In addition, Ni et al [33] referred to the PI Specifications and developed a scale to evaluate the compliance of the privacy policies of China's chronic disease apps. However, their study was conducted before the PIPL came into force, so their scale could not reveal the regulatory development and the mandatory requirement of the PIPL. Therefore, it is necessary to re-evaluate the compliance of the privacy policies of health code apps based on regulatory developments in China.

Our findings illustrated a mixed landscape of compliance status, revealing both areas of commendable adherence and notable gaps in the alignment with the legal framework. While 13 of the 30 apps scored below the mean average, a concerning 20% (n=6) scored under 40%, signaling the urgent need for improvements and highlighting potential threats to PI. When examining compliance across the information life cycle, we found that the highest alignment was in the realm of general attributes with a mean compliance of 85.6% (SD 23.3%). This indicates a prevalent transparency among PI processors regarding their basic information, service range, and privacy policy content and updates. However, 14 (46.7%) apps did not mark any updates after the PIPL came into force, raising concerns regarding the timeliness of policy updates in alignment with regulatory changes.

In the realm of PI collection and use, the compliance rate of sensitive PI protection (mean 51.4%, SD 26.0%) was significantly lower than the rate of general PI protection (mean 83.3%, SD 24.3%). Such discrepancy contradicts the special protection for sensitive PI as provided in the PIPL (in particular, section 2, chapter 2 of the PIPL) and may reduce users' risk awareness. According to the PIPL, PI processors may process sensitive PI only when there is a specified purpose and sufficient necessity, and when stringent protective measures are adopted (article 28 of the PIPL). Meanwhile, PI processors should notify the users of "the necessity of processing such sensitive PI" and "the influence on the individual's rights and interests" (article 30 of the PIPL), and obtain separate consent (article 29 of the PIPL). A majority of the 30 apps elaborated on the specific purpose (n=18, 60%) and influence of processing sensitive PI (n=24, 80%), and the various strict protective measures (n=19, 63.3%). In addition, 80% (n=24) of the apps ensured explicit consent for collecting the PI of minors. However, only 1 app required separate consent for processing sensitive PI, while the other 29 apps only obtained general consent for processing all types of PI. In addition, although 60% of the apps described the specific purpose for processing sensitive PI, fewer apps underscored the exact sensitive PI collected for the health code apps (n=11, 36.7%) and explained the necessity of processing such sensitive PI (n=11, 36.7%). As a result, users may not fully understand the specific sensitive PI involved, why such processing becomes necessary, and the exact content of their consent, in particular, whether the specific consent required for processing sensitive PI is implied or mixed with general consent.

While the overall mean compliance of PI storage and protection stood at a relatively satisfactory level (mean 63.3%, SD 30.8%), the depth and commitment underpinning this compliance varied significantly. Although a robust 90% (n=27) of the 30 apps explained their technical security measures and over 80% (n=25) informed users of potential risks, only 36.7% (n=11) of the assessed apps discussed or provided evidence of PI security agreements or certifications. This shortfall underscores the potential vulnerability of PI, considering that such certifications often serve as benchmarks for best practices. Equally concerning is the lack of commitment to legal responsibility, which is not compliant with articles 66 to 70 of the PIPL. While 76.7% (n=23) of the assessed apps pledged to notify and report security incidents, only 26.7% (n=8) of PI processors were explicitly committed to bearing legal responsibilities during such breaches. This discrepancy raises pressing questions about accountability and reinforces the need for stronger mechanisms that can assure users of adequate protection and remediation in the face of potential PI infractions.

In the stage of PI sharing, transfer, disclosure, and transmission (mean 57.2%, SD 27.3%), most of the 30 apps required separate consent for PI transfer (n=27, 90%) and public disclosure (n=26, 86.7%). However, less than half of the evaluated apps required separate consent for subcontracting practices (n=14, 46.7%) and were committed to supervising such practices (n=14, 46.7%). This is notably lower than the expected standards set out in article 23 of the PIPL, which may reduce users' situational awareness concerning the flow and security of their PI. Even fewer apps disclosed the basic information of PI recipients (n=13, 43.3%), safety precautions (n=11, 36.7%), and rules of PI transfer during specific events (n=10, 33.3%). The resultant opacity diminishes users' ability to grasp the full trajectory of their PI, undermining their trust and inhibiting informed consent. Another note of concern is that only 2 apps notified third parties to promptly delete PI or cease processing due to user request or other circumstances. Not doing so can lead to prolonged PI retention beyond necessity, heightening the risk of data breaches or misuse.

In the stage of PI deletion, roughly one-quarter of the 30 apps (n=7, 23.3%) did not mention their PI retention period, which is the minimum period necessary for achieving the purpose of processing according to article 19 of the PIPL. While this extended retention raises concerns about the purpose and implications of such a practice, another deficiency is evident in the deletion and cessation protocols. Even though a significant percentage (n=22, 73.3%) of apps asserted the deletion or anonymization of PI after the defined retention span, only 6.7% (n=2) made a proactive commitment to ensure that third parties were notified to delete or halt the processing of PI promptly. This lack of third-party engagement presents potential vulnerabilities in the comprehensive safeguarding of PI, especially when the use of third-party services is common.

When evaluating individual rights concerning their PI, we found that a predominant number of the 30 apps efficiently elucidated the diverse rights of users to inquire (n=25, 83.3%), correct (n=24, 80%), and delete their PI (n=24, 80%); request an explanation (n=24, 80%); lodge a complaint; withdraw their consent (n=18, 60%); and cancel their account (n=21, 70%). Additionally, most apps (n=19, 63.3%) elaborated on exceptions for obtaining consent as provided by laws and regulations. This comprehensive coverage underscores a commendable effort in empowering individuals to exercise their rights with exemptions of informed consent provided by law. However, only 13.3% (n=4) acknowledged the right to obtain PI copies, while this right is explicitly provided in article 45 of the PIPL. More concerning is the fact that only 1 app addressed the right to refuse business marketing through automated decision-making, while article 24 of the PIPL calls for transparency, fairness, and the right to receive an explanation and to opt out of such marketing. These results reflect and further highlight a broad acknowledgment of individual rights by most apps. These crucial rights might inadvertently hinder users from fully realizing their entitlements under PI protection norms.

As for the duties of PI processors, our findings revealed a stark incongruity between policy statements and tangible practices for ensuring PI protection. Notably, the overall mean rate of compliance regarding the duties of PI processors was only 43.7% (SD 23.8%), with the 30 assessed health code apps particularly falling short in critical areas of impact assessment procedures (conduct PI protection assessment: n=2, 6.7%; assess the purpose, method, and impact of protective measures: n=1, 3.3%), PI protection officer appointment (n=2, 6.7%), regular compliance audits (n=2, 6.7%), and complaint management (convenient means to lodge complaints: n=15, 50%; commit to responding within legal time limits: n=9, 30%; dispute resolution involving external parties: n=10, 33.3%). These key components are outlined in articles 50, 52, 54, and 55 of the PIPL. The deficiencies in impact assessment, for instance, can result in unforeseen risks or breaches as changes in technology or external threats evolve. The absence of dedicated PI protection officers indicates there is no designated authority to ensure that PI is handled in strict accordance with the law. Without regular compliance audits, apps may drift from best practices over time, unknowingly exposing PI to risks. Lastly, inadequate complaint management mechanisms not only breach the PIPL but also degrade user trust, leading to potential withdrawals from the app or caution against sharing sensitive PI.

Recommendations

In 2020, during the phase termed "the people's war against the epidemic" [39], the right to life and health, constituting public health, was prioritized over the protection of PI. This only changed when the PIPL came into force on November 11, 2021. Health code apps essentially compromised individual rights for necessary prevention and control of the COVID-19 pandemic, which in turn required PI processors to properly protect the use of PI. The once temporary and mandatory use of health code apps could be seen as a "trade-off" to win the war. However, the legitimacy of processing sensitive PI and the further retention of such PI have come into question, especially when PIPL was enacted. Once China eased its stringent zero-COVID policy [40], green health codes were no longer required for movement and travel [41].

In light of the evolving situation, China's government faces a crossroads regarding the future of health code apps, where it must choose between shutting down its services or expanding their use. This choice calls for a balanced policy that considers both public health and PI protection.

First, it is essential to reconsider and clearly articulate the legitimacy of extending health code apps more broadly in public health services such that users are notified and their consent is separately obtained. Initially, the deployment of health code apps was an emergency response to prevent and control the pandemic by performing the statutory duties of government agencies. Because this rationale no longer holds, the continuation of intrusive surveillance through these apps should not be allowed without further evaluation [3]. It is inadequate to justify the routinization of health code apps based merely on general purposes such as providing convenience to citizens, protecting people's health, or incorporating big data technologies in public health. Furthermore, the benefits and resources associated with health code apps do not sufficiently justify processing sensitive PI in their expanded use. Government agencies should encourage more discussions on the necessity and purpose of the continued use of health code apps to address the intensified concerns about data privacy, data security, and data governance as a whole [42].

Second, as China looks toward the normalization of health code apps, it becomes paramount to not only uplift the compliance level of relevant privacy policies but also fortify their enforcement mechanisms. Areas of improvement, based on the information life cycle, encompass enhanced clarity during PI collection and use, improved storage protection measures, more transparent sharing and transfer protocols, clearer deletion guidelines, and a broader acknowledgment of individuals' rights accompanied by actionable exercising avenues. Moreover, PI processors should be diligent in their responsibilities and conduct regular audits and impact assessments. The COVID-19 pandemic not only necessitated and justified the intervention of health code apps during public health emergencies but also brought new challenges to PI protection in their broader application in the postepidemic era. Legal protections for PI seek to facilitate the processing of PI to achieve public benefits while still furnishing reasonable and sufficient protection [43]. Striking a nuanced balance between public interests and PI protection has become an important theme, requiring more effort and thought alongside the rapid development of big data technologies. Only through such concerted efforts can we ensure that these apps serve as not only functional tools but also vanguards of PI, fostering trust and confidence in their user base.

Limitations

One limitation of this study is the lack of empirical evaluation of protective measures for processing sensitive PI. Technological equipment and expertise are necessary to assess whether the protective measures adopted by health code apps are stringent enough to prevent abuse and leakage of sensitive PI. We hope that technological professionals can engage in health code assessment and provide insightful research. Another limitation is that this study's focus on the informed consent model does not address the gaps between the health code practices and best practices of contact-tracing apps. Further studies comparing both the existing and future practices of Chinese health code apps with other countries' practices can use this study as a starting point.

Conclusion

Health code apps are not only an innovation for monitoring and controlling public health emergencies such as the COVID-19 pandemic, but they can also act as a strategic health and medical service platform. Our analysis of 30 privacy policies sheds light on the multifaceted nature of compliance with the PIPL and related specifications. Although commendable strides have been made, significant gaps remain in pivotal areas of the information life cycle. These discrepancies not only pinpoint the exigence of robust PI protection measures but also underscore the importance of fostering trust among users. Only with sufficient PI protection can health code apps and other contact-tracing apps worldwide achieve the maximum value for both public and private interests.

Acknowledgments

The research is funded by the Postdoctoral Research Fund of Central South University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluation scale.
[XLSX File, 24 KB - mhealth_v11i1e48714_app1.xlsx]

Multimedia Appendix 2
Health code apps and scoring rates.
[XLSX File, 14 KB - mhealth_v11i1e48714_app2.xlsx]

References

1. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461. [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
2. Kolasa K, Mazzi F, Leszczuk-Czubkowska E, Zrubka Z, Péntek M. State of the art in adoption of contact tracing apps and recommendations regarding privacy protection and public health: systematic review. JMIR Mhealth Uhealth 2021 Jun 10;9(6):e23250. [doi: 10.2196/23250] [Medline: 34033581]
3. Rothstein MA. Public health and privacy in the pandemic. Am J Public Health 2020 Sep 1;110(9):1374-1375. [doi: 10.2105/AJPH.2020.305849] [Medline: 32614613]
4. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug 1;2(8):e435-e440. [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
5. Jalabneh R, Syed HZ, Pillai S, Jalabneh R, Syed HZ, Pillai S, et al. Use of mobile phone apps for contact tracing to control the COVID-19 pandemic: a literature review. In: Nandan Mohanty S, Saxena SK, Satpathy S, Chatterjee JM, editors. Applications of Artificial Intelligence in COVID-19: Springer; 2021:389-404. [doi: 10.1007/978-981-15-7317-0]
6. Fahey RA, Hino A. COVID-19, digital privacy, and the social limits on data-focused public health responses. Int J Inf Manage 2020 Dec 1;55:102181. [doi: 10.1016/j.ijinfomgt.2020.102181] [Medline: 32836638]
7. Akinbi A, Forshaw M, Blinkhorn V. Contact tracing apps for the COVID-19 pandemic: a systematic literature review of challenges and future directions for neo-liberal societies. Health Inf Sci Syst 2021 Apr 13;9(1):18. [doi: 10.1007/s13755-021-00147-7] [Medline: 33868671]
8. Seberger JS, Patil S. Post-COVID public health surveillance and privacy expectations in the United States: scenario-based interview study. JMIR Mhealth Uhealth 2021 Oct 5;9(10):e30871. [doi: 10.2196/30871] [Medline: 34519667]
9. Sharma T, Islam MM, Das A, Haque SMT, Ahmed SI. Privacy during pandemic: a global view of privacy practices around COVID-19 apps. Presented at: COMPASS '21; Jun 28 to Jul 2, 2021; Virtual Event, Australia p. 215-229. [doi: 10.1145/3460112.3471958]
10. Li VQT, Ma L, Wu X. COVID-19, policy change, and post-pandemic data governance: a case analysis of contact tracing applications in East Asia. Policy Soc 2022 Jan 22;41(1):129-142. [doi: 10.1093/polsoc/puab019]
11. Hatamian M, Wairimu S, Momen N, Fritsch L. A privacy and security analysis of early-deployed COVID-19 contact tracing Android apps. Empir Softw Eng 2021 Mar 19;26(3):36. [doi: 10.1007/s10664-020-09934-4] [Medline: 33776548]

12. Zhang M, Chow A, Smith H. COVID-19 contact-tracing apps: analysis of the readability of privacy policies. J Med Internet Res 2020 Dec 3;22(12):e21572. [doi: 10.2196/21572] [Medline: 33170798]
13. Bardus M, Al Daccache M, Maalouf N, Al Sarih R, Elhajj IH. Data management and privacy policy of COVID-19 contact-tracing apps: systematic review and content analysis. JMIR Mhealth Uhealth 2022 Jul 12;10(7):e35195. [doi: 10.2196/35195] [Medline: 35709334]
14. Wang H. Reflection and foresight on personal information protection and optimization in public health emergencies in China-from the perspective of personal information collection during the period of China's dynamic-zero COVID-19 prevention and control policy. Int J Environ Res Public Health 2023 Jan 11;20(2):1290. [doi: 10.3390/ijerph20021290] [Medline: 36674045]
15. NHS COVID-19 App. GOV.UK. 2023 Mar 28. URL: www.gov.uk/government/collections/nhs-covid-19-app#full-publication-update-history [accessed 2023-09-14]
16. Kendall M, Tsallis D, Wymant C, Di Francia A, Balogun Y, Didelot X, et al. Epidemiological impacts of the NHS COVID-19 app in England and Wales throughout its first year. Nat Commun 2023 Feb 22;14(1):858. [doi: 10.1038/s41467-023-36495-z] [Medline: 36813770]
17. Aryan A. Contact tracing feature of Aarogya Setu disabled, data deleted: govt. The Economic Times. 2023 Feb 9. URL: economictimes.indiatimes.com/tech/technology/contact-tracing-feature-of-aarogya-setu-disabled-data-deleted-govt/articleshow/97744445.cms [accessed 2023-11-03]
18. Yu E. Singapore officially deactivates contact tracing system, to ‘refurbish’ wearables. ZDNET. 2023 Feb 8. URL: www.zdnet.com/article/singapore-officially-deactivates-contact-tracing-system-to-refurbish-wearables/ [accessed 2023-09-14]
19. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8. [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]
20. Zhou SL, Jia X, Skinner SP, Yang W, Claude I. Lessons on mobile apps for COVID-19 from China. J Safety Sci Resilience 2021 Jun 1;2(2):40-49. [doi: 10.1016/j.jnlssr.2021.04.002]
21. Ye Q, Zhou J, Wu H. Using information technology to manage the COVID-19 pandemic: development of a technical framework based on practical experience in China. JMIR Med Inform 2020 Jun 8;8(6):e19515. [doi: 10.2196/19515] [Medline: 32479411]
22. Circular on succeeding the endeavour of prevention and control of COVID-19 epidemic in a legal, scientific and accurate way. National Health Commission of the People’s Republic of China. 2020 Feb 25. URL: www.nhc.gov.cn/jkj/s3577/202002/69b3fdcbb61f499ba50a25cdf1d5374e.shtml [accessed 2023-09-14]
23. Liang F. COVID-19 and health code: how digital platforms tackle the pandemic in China. Soc Media Soc 2020 Aug 11;6(3):2056305120947657. [doi: 10.1177/2056305120947657] [Medline: 34192023]
24. Wentao H. Conceptions of definition of sensitive personal information in China. China Leg Sci 2018(5):235-254. [doi: 10.14111/j.cnki.zgfx.2018.05.013]
25. Ning Y, Ren R, Nkengurutse G. China’s model to combat the COVID-19 epidemic: a public health emergency governance approach. Glob Health Res Policy 2020 Jul 14;5:34. [doi: 10.1186/s41256-020-00161-4] [Medline: 32685691]
26. Shen T, Wang C. Big data technology applications and the right to health in China during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Jul 8;18(14):7325. [doi: 10.3390/ijerph18147325] [Medline: 34299776]
27. Circular on deepening the service campaign for “Internet plus Healthcare” and "five ones”. Gov.CN. 2020 Dec 4. URL: www.gov.cn/zhengce/zhengceku/2020-12/10/content_5568777.htm [accessed 2023-09-14]
28. Circular on issuing the plan for Informatization in universal health during the 14th five-year plan period. National Health Commission of the People’s Republic of China. 2022 Nov 9. URL: www.nhc.gov.cn/guihuaxxs/s3585u/202211/49eb570ca79a42f688f9efac42e3c0f1.shtml [accessed 2023-09-14]
29. Guiding opinions on strengthening the work on health and hygiene statistic. National Health Commission of the People’s Republic of China. 2020 Aug 28. URL: www.nhc.gov.cn/guihuaxxs/s10743/202008/f4fce10ec10548fdbb1743ad29790024.shtml [accessed 2023-09-14]
30. Huang G, Hu A, Chen W. Privacy at risk? Understanding the perceived privacy protection of health code apps in China. Big Data Soc 2022 Oct 31;9(2):205395172211351. [doi: 10.1177/20539517221135132]
31. Cong W. From pandemic control to data-driven governance: the case of China’s health code. Front Polit Sci 2021 Apr 14;3. [doi: 10.3389/fpos.2021.627959]
32. Liu CC, Graham R. Making sense of algorithms: relational perception of contact tracing and risk assessment during COVID-19. Big Data Soc 2021 Feb 17;8(1):205395172199521. [doi: 10.1177/2053951721995218]
33. Ni Z, Wang Y, Qian Y. Privacy policy compliance of chronic disease management apps in China: scale development and content evaluation. JMIR Mhealth Uhealth 2021 Jan 28;9(1):e23409. [doi: 10.2196/23409] [Medline: 33507159]
34. Calzada I. Citizens’ data privacy in China: the state of the art of the Personal Information Protection Law (PIPL). Smart Cities 2022 Sep 8;5(3):1129-1150. [doi: 10.3390/smartcities5030057]
35. Kui S. The stumbling balance between public health and privacy amid the pandemic in China. Chin J Comp Law 2021 Feb 1;9(1):25-50. [doi: 10.1093/cjcl/cxaa035]

36. Zhang X. Decoding China's COVID-19 health code apps: the legal challenges. Healthcare (Basel) 2022 Aug 5;10(8):1479. [doi: 10.3390/healthcare10081479] [Medline: 36011136]
37. Xiong B, Lin F. How to balance governance efficiency and privacy protection? A textual analysis of the privacy policies of the COVID-19 contact-tracing app in China and Singapore. Int J Chin Comp Phil Med 2020 Jan 1;18:113-143. [doi: 10.24112/ijccpm.181691]
38. Krehling L, Essex A. A security and privacy scoring system for contact tracing apps. J Cybersecur Priv 2021 Oct 14;1(4):597-614. [doi: 10.3390/jcp1040030]
39. Gathering the strong power of the whole nation to fight the war against the epidemic - on resolutely winning the battle against COVID-19 prevention and control. Xinhua News. 2020 Feb 2. URL: www.xinhuanet.com/politics/2020-02/02/c_1125523580.htm [accessed 2023-09-14]
40. China to manage COVID-19 with measures against class B infectious diseases. Xinhua Net. 2022 Dec 27. URL: english.news.cn/20221227/0f5ffdc62cd841af8c274be6b16264e7/c.html [accessed 2023-09-14]
41. Joint Prevention and Control Mechanism for COVID-19 of the People's Republic of China. Circular on further optimizing the implementation of COVID-19 prevention and control measures. National Health Commission of the People's Republic of China. 2022 Dec 7. URL: www.nhc.gov.cn/xcs/gzzcwj/202212/8278e7a7aee34e5bb378f0e0fc94e0f0.shtml [accessed 2023-09-14]
42. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 May 4;46(7):427-431. [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
43. Gostin L, Wiley L. Public Health Law: Power, Duty, Restraint: University of California Press; 2016.

Abbreviations

PI: personal information
PIPL: Personal Information Protection Law

Edited by L Buis; submitted 04.05.23; peer-reviewed by L Krehling, W Shuang; revised version received 15.09.23; accepted 04.10.23; published 14.11.23.

Please cite as:
Jiang J, Zheng Z
Personal Information Protection and Privacy Policy Compliance of Health Code Apps in China: Scale Development and Content Analysis
JMIR Mhealth Uhealth 2023;11:e48714
URL: https://mhealth.jmir.org/2023/1/e48714
doi:10.2196/48714

Original Paper

Critical Criteria and Countermeasures for Mobile Health Developers to Ensure Mobile Health Privacy and Security: Mixed Methods Study

Rita Rezaee[1,2,3], MD, PhD; Mahboobeh Khashayar[4], MSc; Saeed Saeedinezhad[3], PhD; Mahdi Nasiri[3,5], PhD; Sahar Zare[6], PhD

[1]Department of Health Information Technology, Shiraz University of Medical Sciences, Shiraz, Iran
[2]Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[3]Health Human Resources Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[4]Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran
[5]Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran
[6]Health Information Management Research Center (HIMRC), Kashan University of Medical Sciences, Kashan, Iran

Corresponding Author:
Sahar Zare, PhD
Health Information Management Research Center (HIMRC)
Kashan University of Medical Sciences
5th of Qotb -e Ravandi Blvd Kashan
Kashan, 87159-73449
Iran
Phone: 98 31 55548883
Email: zare.sahar89@gmail.com

Abstract

Background: Despite the importance of the privacy and confidentiality of patients' information, mobile health (mHealth) apps can raise the risk of violating users' privacy and confidentiality. Research has shown that many apps provide an insecure infrastructure and that security is not a priority for developers.

Objective: This study aims to develop and validate a comprehensive tool to be considered by developers for assessing the security and privacy of mHealth apps.

Methods: A literature search was performed to identify papers on app development, and those papers reporting criteria for the security and privacy of mHealth were assessed. The criteria were extracted using content analysis and presented to experts. An expert panel was held for determining the categories and subcategories of the criteria according to meaning, repetition, and overlap; impact scores were also measured. Quantitative and qualitative methods were used for validating the criteria. The validity and reliability of the instrument were calculated to present an assessment instrument.

Results: The search strategy identified 8190 papers, of which 33 (0.4%) were deemed eligible. A total of 218 criteria were extracted based on the literature search; of these, 119 (54.6%) criteria were removed as duplicates and 10 (4.6%) were deemed irrelevant to the security or privacy of mHealth apps. The remaining 89 (40.8%) criteria were presented to the expert panel. After calculating impact scores, the content validity ratio (CVR), and the content validity index (CVI), 63 (70.8%) criteria were confirmed. The mean CVR and CVI of the instrument were 0.72 and 0.86, respectively. The criteria were grouped into 8 categories: authentication and authorization, access management, security, data storage, integrity, encryption and decryption, privacy, and privacy policy content.

Conclusions: The proposed comprehensive criteria can be used as a guide for app designers, developers, and even researchers. The criteria and the countermeasures presented in this study can be considered to improve the privacy and security of mHealth apps before releasing the apps into the market. Regulators are recommended to consider an established standard using such criteria for the accreditation process, since the available self-certification of developers is not reliable enough.

(JMIR Mhealth Uhealth 2023;11:e39055) doi:10.2196/39055

KEYWORDS

telemedicine; mobile apps; privacy; computer security, confidentiality; mHealth; mobile health

Introduction

More than 5.19 billion people now use mobile phones, which indicates that mobile phones form an important part of daily life worldwide [1]. Mobile phone features, including mobility, instantaneous availability, and direct communication, have changed the provision of health care services. These features introduce mobile health (mHealth). Of about 2 million smartphone apps available in app stores, 318,000 are health apps [2]. According to a World Health Organization report [3], the penetration of mHealth, with promising results, in low- and middle-income countries would be even more.

mHealth has improved the patient care status through the provision of health care anytime and anywhere [4]. Even in recent years, the integration of mHealth and wireless technologies has provided clinicians with an opportunity to collect real-time data via wearable sensors [5]. Health information is deemed sensitive, and its protection is of significance. Nevertheless, smartphones are vulnerable to a wide range of security threats [6]. Moreover, electronic transmission of information has brought about concerns about its privacy and security. A national survey showed that 1 of the common reasons for people not having downloaded health apps is concern about apps gathering their data [7,8]. The privacy and confidentiality of information, as a human right, have long been considered in law and regulations. Well-known examples are the Health Insurance Portability and Accountability Act (HIPAA) rules, the General Data Protection Regulation (GDPR), and the Common Rule [9-11]. The terms "security," "privacy" and "confidentiality" are all separate yet connected concepts that need to be addressed. The National Committee for Vital and Health Statistics [12] defines and distinguishes these concepts as follows:

> *Health information privacy is an individual's right to control the acquisition, uses, or disclosures of his or her identifiable health data. Confidentiality, which is closely related, refers to the obligations of those who receive information to respect the privacy interests of those to whom the data relate. Security is altogether different. It refers to physical, technological, or administrative safeguards or tools used to protect identifiable health data from unwarranted access or disclosure.*

Despite the importance of the privacy and confidentiality of patients' information, studies report that mHealth apps may share the information with third parties, which raises the risk of violating patients' privacy and confidentiality [13-15]. Dehling et al [16] evaluated the information security and privacy of 24,405 health-related apps and revealed that most apps request access to sensitive information. Robillard et al [17] reported that most of the apps do not include privacy policies and terms of the agreement. Moreover, it has been shown that many apps provide an insecure infrastructure and security is not a priority for the developers [18]. Similar studies emphasize assessing mHealth apps for the privacy, security, and confidentiality of information to minimize the associated risks [16,19,20].

Criteria have been proposed in previous studies for assessing mHealth apps. Benjumea et al [21] proposed a novel scale to assess the privacy policy of mHealth apps. However, the scale considers only specific items associated with the privacy policy content based on the GDPR rather than considering security and privacy in general. Another study [22] also proposed a heuristic evaluation approach to assessing the privacy of mHealth apps, but that is a time-consuming approach because heuristics require a close reading of the privacy policy. Another study proposed a security-testing method for Android mHealth apps designed based on a threat analysis, considering probable attack scenarios and vulnerabilities associated with the domain [18]. They assessed security using novel dynamic and static analysis testing methods that were expensive to perform. Benjumea et al [23] conducted a scoping review on studies exploring privacy issues in mHealth apps. Finding that most studies assess the apps based on heterogeneous criteria, Benjumea et al [23] emphasized the importance of developing a scale based on more objective criteria for evaluating privacy issues. In addition, the mHealth field faces a variety of legal and cultural differences over privacy between nations, so it needs a comprehensive tool for assessing both privacy and security issues [24]. Thus, developing a comprehensive tool assessing both privacy and security sounds necessary. This study aims to develop and validate a comprehensive tool to be considered by developers for assessing both the security and the privacy of mHealth apps targeting patients.

Methods

Study Design

This study was conducted to answer the following question: What security and privacy criteria should be considered when developing or assessing mHealth apps targeting patients based on 3 main phases: item generation, tool development, and tool evaluation? These main phases [25] were performed based on 4 steps: (1) identifying criteria associated with mHealth apps' security/privacy according to a literature search (item generation); (2) conducting an expert panel for determining the categories and subcategories according to meaning, repetition, and overlap (tool development); (3) testing the validity of the instrument (tool evaluation); and (4) testing the reliability of the instrument (tool evaluation).

Stage 1: Literature Review

An unstructured literature search was performed to identify papers on app development, assessment, security, or privacy that reported criteria for the security and privacy of mHealth. PubMed, Scopus, Web of Science, and Cochrane were searched for English language papers published until December 15, 2021, without a time limitation. The search strategy (Multimedia Appendix 1) included a combination of 4 keywords: ("mobile device" OR "mobile phone" OR smartphone OR "smart Phone" OR mHealth OR "mobile health") AND (App OR apps OR application*) AND (security OR privacy OR confidentiality OR cybersecurity) AND (guideline* OR standard* OR criteria OR risk* OR assess* OR evaluat* OR measure).

The HIPAA and GDPR websites were searched for relevant criteria. After removing duplicate papers, the titles and abstracts of the studies were screened for inclusion. The full text of potentially relevant papers was investigated based on study objectives. Studies substantially focusing on security or privacy, not just mentioning them in passing, and stating clear criteria for assessing the privacy/security of mHealth apps were included. Studies evaluating the privacy or security of mHealth apps were also included to specify the criteria used for evaluation. Papers proposing a secure architecture, investigating technical solutions for mHealth apps (eg, access control, authentication approaches, encryption methods), presenting technical solutions for connecting mHealth apps to cloud computing or the internet of things devices or conducted on wearable devices without connecting to a mobile device, and discussing mobile phone access to electronic health records were excluded. Papers focusing on mHealth apps targeting users other than patients, focusing on app quality or determining functional requirements, and examining user experiences were also excluded. The criteria were extracted using content analysis.

Stage 2: Expert Panel

The list of primary criteria extracted through the literature search was presented to a focus group including 2 health information technology (HIT) specialists, 2 medical informatics specialists, and 1 software and IT specialist. The focus group discussion consisted of 4 major steps: designing research, collecting data, analyzing, and reporting results through a moderated interaction [26]. The experts discussed and categorized the criteria and decided over their inclusion or exclusion based on the relevancy, clarity, importance, comprehensiveness, and overlap with other included criteria, and they determined subcategories based on meaning, repetition, and overlap. This method can have a high level of validity due to the interaction among experts that confirms, reinforces, or rejects the individual respondents' contributions. The criteria extracted through the focus group discussion were used in the next stage.

Stage 3: Testing the Validity of the Instrument

Quantitative and qualitative methods were used for validating the instrument. To validate the instrument based on the qualitative approach, face validity was checked through face-to-face interviews by 8 HIT specialists and 5 software and IT experts. The inclusion criteria for the experts included specialists in HIT, IT, or software, with a master's degree in science or higher, with at least 1-year work experience in software security, network security, health information security, or mobile app development. The criteria were modified based on the experts' comments.

To validate the instrument quantitatively, the impact score was calculated for each criterion. The impact score determines inappropriate criteria. Thus, the criteria were evaluated based on a 5-point Likert scale ranging from 5 (very important) to 1 (not at all important). The impact score for each criterion was calculated as follows:

$$\text{Impact score} = \text{Frequency (\%)} \times \text{Importance}$$

Content validity was evaluated by 16 other IT (n=8, 50%) and software (n=8, 50%) experts, of whom 3 (18.8%) experts did not participate. Thus, to make sure the most essential criteria for the study objective were chosen, the content validity ratio (CVR) was measured. The CVR was calculated based on the following formula:

According to the Lawshe table, if the number of experts in the panel is 13, the minimal acceptable CVR is 0.54.

In addition, to ensure the relevancy and clarity of each criterion, the content validity index (CVI) was measured. Thus, the 13 experts also completed a 4-point scale based on relevance, clarity, and simplicity for the criteria. The CVI was calculated using the following formula:

The criteria were included in the final assessment tool if the CVI was ≥0.79 [27,28]. If the CVI was between 0.70 and 0.79, it needed to be calculated after the criteria were revised by the experts. Criteria with a CVI of <0.70 were removed.

Stage 4: Testing Reliability

To assess the reliability of the final tool, the hypertensive self-care app developed in our previous study [29] was selected. The app needs to record a variety of personal information. In total, 30 experts in HIT, medical informatics, IT, and software assessed the reliability of the instrument. The instrument was distributed among these experts twice in a 2-month interval. They were asked to assess the privacy and security of the self-care app using the criteria provided in the checklist. After collecting expert opinions about the self-care app, the data were analyzed using the Cronbach α.

Ethical Considerations

The research was conducted according to the principles stated by the Vice-Chancellorship for Research Affairs of Shiraz University of Medical Science and approved by the Ethics Review Board of the Vice-Chancellorship for Research Affairs of Shiraz University of Medical Science (ethical code IR.SUMS.REC.1397.500).

Results

Study Selection

The search strategy retrieved 10,092 papers, of which 1902 (18.8%) were duplicates. Of the 8190 (81.2%) remaining papers, 8072 (98.6%) were irrelevant. To retrieve the greatest number of possible relevant papers, our search strategy included smartphone or mobile devices as a synonym for mHealth ("mobile device" OR "mobile phone" OR smartphone OR "smart Phone" OR mHealth OR "mobile health"); this resulted in retrieving papers basically irrelevant to the health discipline, in addition to those relevant to the health discipline—for example, studies associated with payment/banking/commercial apps were also retrieved in the primary result. In total, 33 (0.4%) studies were deemed eligible for inclusion in the research (Figure 1). The characteristics of the included studies [13,14,16,18-20,24,30-56] are presented in Multimedia Appendix 2.

A total of 218 criteria were extracted based on the literature search; of these, 119 (54.6%) were removed as duplicates (showing the same idea) and 10 (4.6%) were deemed irrelevant to the security or privacy of mHealth apps. The remaining 89 (40.8%) criteria were presented to the expert panel. As shown in Figure 2, 63 (70.8%) criteria were confirmed at last.

The mean CVR of the total instrument was 0.72, while the mean CVI was 0.86. Multimedia Appendix 3 shows the complete list of removed criteria in the different phases of the study.

Finally, to measure the reliability of the instrument, the experts were asked to assess the hypertensive self-care app using the instrument. When measuring the reliability of the instrument, 18 (28.6%) of the 63 criteria received the lowest and the highest score of the Likert spectrum (“not at all” and “completely”) equally. Since the variance of equal data was 0, these 18 criteria did not automatically enter for calculating the Cronbach α value. Thus, the test was performed with 45 (71.4%) criteria. The Cronbach α value was 0.89.

The 63 criteria were grouped into 8 categories: authentication and authorization (n=8, 12.7%), access management (n=6, 9.5%), security (n=13, 20.6%), data storage (n=4, 6.3%), integrity (n=2, 3.2%), encryption and decryption (n=5, 9.5%), privacy policy (n=15, 23.8%), and privacy policy content (n=10, 15.9%); see Textbox 1.

Figure 1. Flow diagram of study selection. EHR: electronic health record.

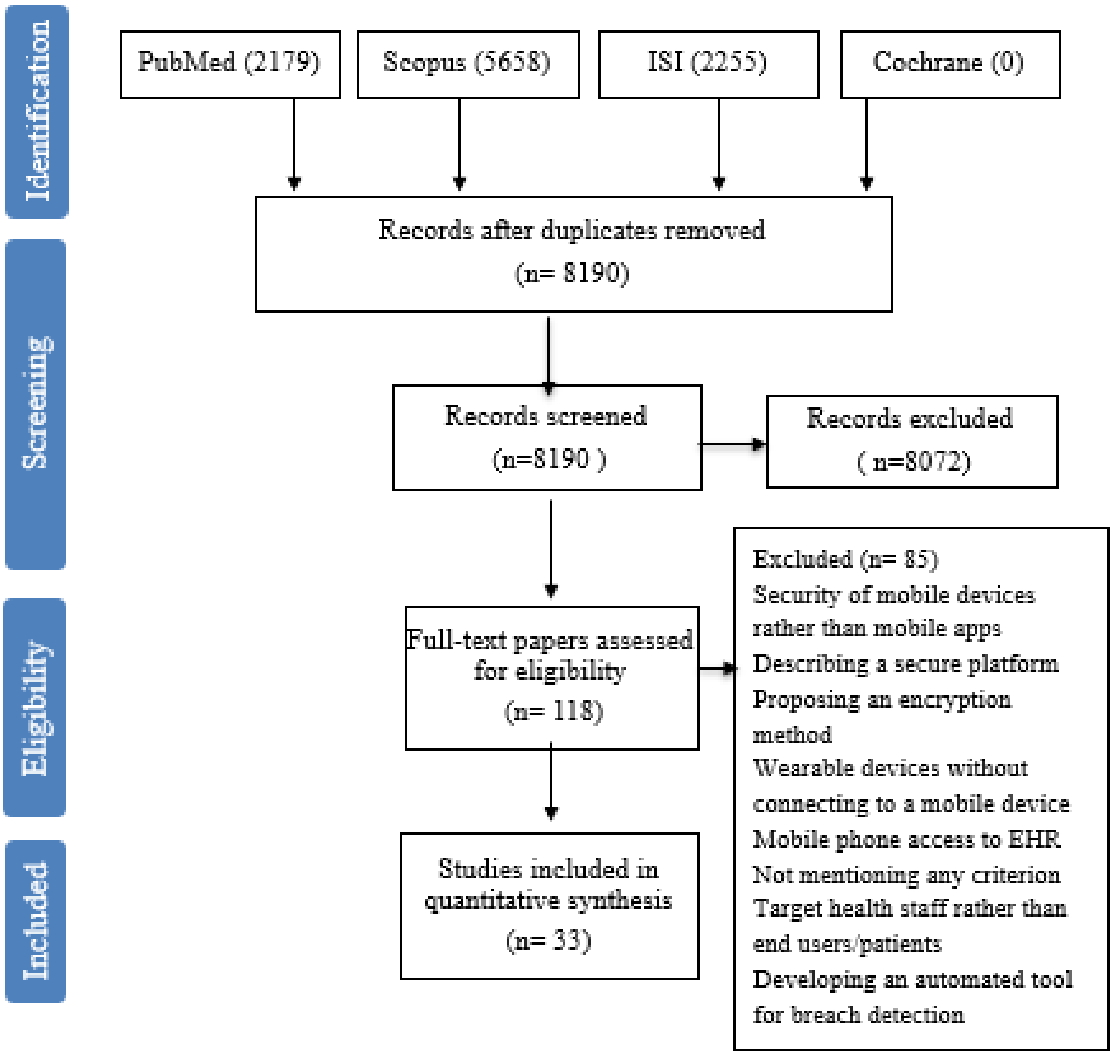
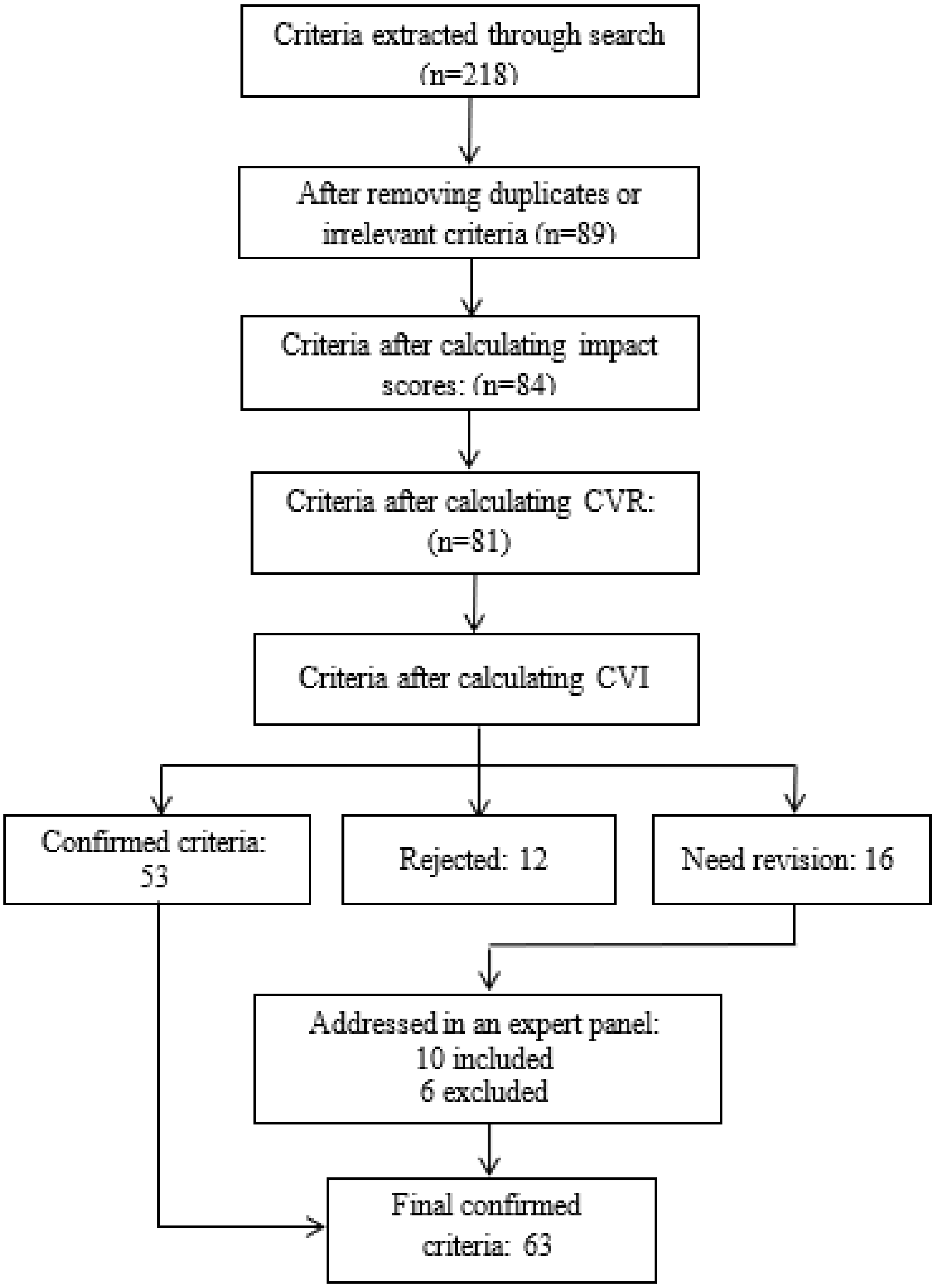


Figure 2. Flowchart of criteria determination. CVI: content validity index; CVR: content validity ratio.



Textbox 1. Final privacy and security assessment criteria.

1. Authentication and authorization

1.1. Is there any registration/log-in available in the app?

1.2. Does the app capture a unique username or “fixed device identifier” used as a user identifier (for both patient and health care provider)?

1.3. Are there procedures to verify that any person or entity claiming access to electronic protected health information complies with its claim?

1.4. Are there any ways to monitor the log and report errors?

1.5. Are there any steps to create, change, and protect the password?

1.6. Are the passwords complex enough (ie, of a minimum length, alphanumeric with upper- and lowercase letters and symbols)?

1.7. Are the passwords updated periodically?

1.8. Is the user’s account locked after a determined number of consecutive unsuccessful log-in attempts?

2. Access management

2.1. Is there patient-centric access control?

2.2. Are there measures taken to access the health information needed in an emergency?

2.3. Is the user allowed to access personal information and to participate in treatment?

2.4. Does the app facilitate the provision of an electronic copy of data?

2.5. Is the app capable of cutting off or blocking a person's access at any time?

2.6. Are users allowed to control the access level of their health information by third parties?

3. Security

3.1. Does the app use secure connections (Secure Socket Layer [SSL]/Transport Layer Security [TLS])?

3.2. Can the data be remotely controlled if the mobile phone is lost/stolen?

3.3. Does the app use a secure platform for transmitting health data?

3.4. Does the app protect network traffic by strong coding?

3.5. Are default measures present to protect against, identify, and report security incidents/malware?

3.6. Does the app use external devices?

3.7. Does the app use random number generators?

3.8. Are users able to change individual profiles according to the policy of the mobile health (mHealth) app?

3.9. Does the app require interaction with the user while performing a sensitive operation or communicating with an untrusted app?

3.10. Does the app use cookies?

3.11. Is the security policy transparent and easy to find?

3.12. Are there reminders for periodic system security updates?

3.13. Has anyone been appointed to assume security responsibility?

4. Data storage

4.1. Are data stored locally on the device? If no, are the users notified about using another platform for storing their data?

4.2. Are data centers in a secure condition?

4.3. Are data stored on the mobile phone or to the app company’s own servers?

4.4. Are there any steps to recover lost data or any backup?

5. Integrity

5.1. Are there electronic mechanisms to verify that health information is not unauthorized, altered, or destroyed (eg, check-sum verification or digital signatures)?

5.2. Are security measures in place to prevent the unauthorized destruction or tampering of health information that is being exchanged electronically?

6. Encryption and decryption

6.1. Does the app use a strong modern encryption/decryption mechanism?

6.2. Is a proper method of encryption selected and implemented (eg, use encryption through https rather than http)?

6.3. Are the data stored encrypted?

6.4. Are the data transmitted encrypted?
6.5. Is the username/password/keys encrypted?
7. Privacy
7.1. Is there a privacy policy on the app or a link to the full privacy policy?
7.2. Are there any restrictions on the use or disclosure of information contained in the app?
7.3. Are there restrictions on the collection of information?
7.4. Does the app have the ability to disclose information on social media by the user?
7.5. Has the principle of protecting the confidentiality of data been met?
7.6. Does the app state which regulation it complies with and which country the regulation belongs to?
7.7. Does the app ask normal permissions and provide justification for that?
7.8. Is identifiable information anonymized and de-identifiable? If anonymization is not possible, are users informed?
7.9. Have any measures been taken to notify the users of their privacy rights?
7.10. Will the user be informed of any leaks or breaches?
7.11. Does the app have the ability to manage alerts (eg, hide them from the lock screen)?
7.12. Is the privacy policy easy to find, clear, readable, and up to date?
7.13. Are users able to manipulate or completely delete personal profiles and any data archives?
7.14. Are users informed about any security or privacy measures?
7.15. Does the app prevent disclosure of data about the location or sensor type of the user?
8. Privacy policy content
8.1. Is there a time limit for data retention?
8.2. Is the content of the contract with third parties clearly stated?
8.3. Does the app mention the collection of user data and how they are being used?
8.4. Does the privacy policy describe the purpose and the type of information collected?
8.5. Is the data ownership specified?
8.6. Are the administrative details stated (identify data controller or responsible legal entity, legal jurisdiction governing policy, jurisdictions under which transmitted data will be processed, date of policy and next review)?
8.7. Is there an explanation about the retention policy for the health information?
8.8. Does the privacy policy explain the manipulation of data by the developer or third parties?
8.9. Does the privacy policy explain the complaints procedures?
8.10. Does the privacy policy explain the procedures for changing the terms of the policy?

Discussion

Principal Findings

In this study, we developed an instrument for assessing the security and privacy of mHealth apps. The criteria proposed in this tool were classified into 8 categories: authentication and authorization, access management, security, data storage, integrity, encryption and decryption, privacy, and privacy policy. These criteria can be considered by mHealth app developers to improve the privacy and security of their apps before releasing them into the market.

Authentication and Authorization

The criteria in the tool suggest implementing rigorous authentication and authorization techniques. More time and effort should be devoted to preventing unauthorized access to personal health information. The developers are asked to provide a unique master ID and a secret key identity for users to control role-based access and verify users' activities according to the defined identity and roles. Authentication via a fingerprint or a personal identification number is necessary for internal storage, internal cache, external storage, and databases [57]. Audit trails should be in place to track logs, protect data, and identify which user's health data was handled and by whom. Each user should be able to create, change, and protect their passwords. The developers should make sure the passwords are strong enough and are changed periodically, because there are tools that produce 10^{14} guesses in an hour to find the correct password [58]. There are some strategies to be used by developers to make sure passwords are secure; these include enforcing password complexity; making passwords unviewable, even to the app administrator; and locking a user's account after a determined number of consecutive unsuccessful log-in attempts. System-generated passwords can be strong, but they do not guarantee memorability. Using Optiwords8 passwords [59], based on the picture superiority effect on the mobile phone keyboards, guarantees the security of passwords, while keeping them usable and memorable as a result.

Access Management

mHealth app developers need to define access controls for their team members as well as users. For those apps providing health care provider–patient communication, granting access to specific app functions should be based on predetermined and confirmed roles and attributes. Patients should be users allowed to control the access level of their health information by third parties. Greene et al [60] proposed the ShareHealth framework, which provides cryptographically enforced access to data. The framework takes advantage of combining a robust cryptographic scheme, hash chains (to control access by data time), and attribute-based encryption (to control access by data type). Rectification, deleting, or blocking of data should be facilitated for users [53].

Security

Some mHealth apps use connections for several purposes, including fetching mail, sending analytics data, or checking for updates. To protect the authenticity, confidentiality, and integrity of the connection, developers are encouraged to use an up-to-date version of the Transport Layer Security protocol and its predecessor, the Secure Socket Layer (SSL) [54]. SSL protocols provide an encrypted link that connects a server and a client and makes sure the transmitted data remain impossible to read and are kept private; however, if the coding is not strong enough, hackers would be able to interpret health data during transmission [44]. There should be a functionality of remote control of data to securely transfer, retrieve, or completely erase health information if the mobile phone is stolen/lost [35]. However, it is safer to store data on users' own devices rather than on the app company's servers [13]. Some apps use external devices, such as cameras, sensors, or payment apps, to improve their functionality, but this endangers users' confidentiality through attacks, such as external-device misbonding [48]. Moreover, using cookies can jeopardize user privacy especially those used for data analysis by third parties [14]. Users should be able to manipulate their profile or delete it completely when they stop using an app [31].

Encryption and Decryption

Bhanot and Hans [61] compared various encryption algorithms based on different criteria, such as cryptography type, key management, keys number, and bit numbers used in a key. They found that elliptic-curve cryptography and blowfish encryption algorithms are the best, providing higher security levels as well as faster encryption speeds, which is required for mobile devices due to less power consumption [61]. Security measures, such as wired equivalent privacy, which is used to provide security to mobile devices, are vulnerable to hackers [62,63]. Thus, developers are required to perform a security risk analysis to determine vulnerabilities at each stage of design and implementation throughout testing and use. Arora et al [64] suggest using a "red team" for risk analysis. Red team experts are charged with hacking cyber systems in order to detect weaknesses.

Privacy

Papageorgiou et al [49] found that although many of the studied apps ask for dangerous permissions (eg, read/write external storage, access camera, location, and contacts), they do not follow well-known regulations, such as HIPAA. Developers are required to collect data as much as they need to provide their services, so they are required to provide reasons for permissions they ask for, the type of data they collect, and how the data will be used by them or third parties, including insurance companies, government institutions, or even research centers [18,38]. Third-party usage of health data can bring about privacy intrusions, such as loss of insurance coverage or higher insurance premiums [65]. Complying with regulations and which country these regulations belong to is also important because when enforcing privacy rights, the regulations may differ from the users' own country [13]. Users' records should be stored in incognito forms, which are anonymized and unidentifiable; if anonymization is not possible, users should be informed [40].

All mHealth apps need to provide a transparent, precise, and well-readable privacy policy statement or a link to the complete privacy policy. Procedures for refusing data sharing, consequences of not providing/sharing data, procedures for changing the terms of the policy, procedures for editing or deleting data held by developers/third parties, procedures for complaints, and procedures for handling data for vulnerable users are subsets of "user rights" a privacy policy should contain. In addition, a data retention policy, data ownership, date of the policy, and next reviews should be contained as "administrative details" of the privacy policy. Users' access to their health information is another right. A systematic review [66] indicated that patients' access to their health information has a positive impact. A similar study [21] proposed a 14-criteria scale for assessment of a privacy policy based on the GDPR. Although the items by proposed Benjumea et al [21] overlap our proposed criteria (some with different words but similar concepts), they include 5 items not included in our tool; 2 items are "legal basis for processing" and "legitimate interests from controller" that imply the bases for the processing determined by the GDPR. This may be similar to the criteria associated with permission/consent and how users' data will be processed/used, which are considered in our tool in general. Another item is "transfers to non-EU countries," which sounds similar to the "regulation the mHealth app comply with and the country (as general, not only European ones) that the regulation belongs to" also considered in our tool. The fourth item is "obligation to provide personal data," which can be considered as a subset of "user rights" [34] (existent among our criteria). As mentioned earlier, users need to be informed about the consequences of not providing their information. The last item is "existence of automated decision-making or profiling," which is not included in our tool. It also worth to note that the criteria proposed in our study are general criteria for assessing both privacy and security classified into 8 categories. We tried to determine a comprehensive list of criteria, but we also faced a restriction to limit our criteria to general important aspects of privacy and security, because including a large number of criteria makes it difficult for assessors to consider all of them and this may result in rejection of the tool. That is why we tried to use general concepts that cover more specific criteria (eg, user rights) or merge some criteria into a single one (eg, administrative details).

Limitations

In this study, a list of criteria was proposed using published papers. A limitation of this study is conducting an unstructured literature search, due to which we missed some related papers. However, to the best of our knowledge, many of the criteria included in our study overlap those that were not included. Another limitation is the large number of included criteria, which may make it difficult for assessors to consider all of them; however, we tried to limit our criteria to important ones to make them more applicable, and we also used general concepts that cover more specific criteria (eg, user rights) or merged some items into a single one (eg, administrative details). Another limitation is the difficulty in assessing some criteria—for example, app compliance with regulations may not be clearly stated in the app. It is recommended that future studies verify the proposed criteria using mobile apps. However, they should be considered in conjunction with other assessment strategies, such as risk analysis, data leakage detection, and continuous revision accordingly. Moreover, this study focused on the security and privacy challenges of mHealth apps, but there are other important challenges, such as interoperability. Thus, it is recommended that future studies combine both aspects to obtain not only a secure system but also an interoperable one, because mHealth apps communicate with a variety of sources.

Conclusion

With the evolution in the health field through smartphones and mHealth apps, privacy and security challenges need to be addressed. The proposed comprehensive criteria can be used as a quick guide for app designers, developers, regulators, and even researchers. The criteria and the countermeasures presented in this study can be considered to improve the privacy and security of an mHealth app before releasing it into the market. Regulators are recommended to consider an established standard using such criteria for the accreditation process, since the available self-certification of developers is not reliable enough.

Acknowledgments

We acknowledge all the participants who helped us conduct this research. This paper was extracted from MK's Master of Science thesis on health information management.

RR, MK, SS, MN, and SZ made substantial contributions to conception and design. SZ and MK made substantial contributions to data collection through a literature search. SZ and RR drafted the manuscript, and all the authors have read it to revise it critically for important intellectual content.

MN was affiliated with the Health Human Resources Research Center at Shiraz University of Medical Sciences at the time of the study and is currently affiliated with the Department of Computer Engineering and Information Technology at Shiraz University of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
This is the search strategy.
[DOCX File , 15 KB - mhealth_v11i1e39055_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies.
[DOCX File , 61 KB - mhealth_v11i1e39055_app2.docx]

Multimedia Appendix 3
The complete list of removed criteria.
[DOCX File , 15 KB - mhealth_v11i1e39055_app3.docx]

References

1. Kemp S. Digital 2020: global digital overview. DATAREPORTAL. 2020. URL: https://datareportal.com/reports/digital-2020-global-digital-overview [accessed 2023-02-09]
2. Szpunar MJ, Parry BL. A systematic review of cortisol, thyroid-stimulating hormone, and prolactin in peripartum women with major depression. Arch Womens Ment Health 2018 Apr;21(2):149-161 [FREE Full text] [doi: 10.1007/s00737-017-0787-9] [Medline: 29022126]
3. Buranarach M, Chalortham N, Thein YM, Supnithi T. Design and implementation of an ontology-based clinical reminder system to support chronic disease healthcare. IEICE Trans Inf Syst 2011;E94-D(3):432-439. [doi: 10.1587/transinf.e94.d.432]

4. Silva B, Rodrigues J, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
5. Munos B, Baker PC, Bot BM, Crouthamel M, de Vries G, Ferguson I, et al. Mobile health: the power of wearables, sensors, and apps to transform clinical trials. Ann N Y Acad Sci 2016 Jul;1375(1):3-18. [doi: 10.1111/nyas.13117] [Medline: 27384501]
6. Zubaydi F, Saleh A, Aloul F, Sagahyroon A. Security of mobile health (mHealth) systems. 2015 Presented at: IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE); November 2-4, 2015; Belgrade, Serbia p. 2-4. [doi: 10.1109/bibe.2015.7367689]
7. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
8. Özkan Ö, Aydin SY, Aydinoğlu AU. Security and privacy concerns regarding genetic data in mobile health record systems: an empirical study from Turkey. bioRxiv 2019. [doi: 10.1101/678912]
9. Edemekong P, Annamaraju P, Haydel M. Health Insurance Portability and Accountability Act. Treasure Island, FL: StatPearls Publishing; 2020.
10. European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46. Official J Eur Union 2016;59:1-88 [FREE Full text] [doi: 10.5593/sgemsocial2019v/1.1/s02.022]
11. Williams ED. Federal protection for human research subjects: an analysis of the common rule and its interactions with FDA regulations and the HIPAA Privacy Rule. Congressional Research Service, Library of Congress. 2005. URL: http://www.fas.org/sgp/crs/misc/RL32909.pdf [accessed 2023-02-09]
12. Cohn S. Privacy and confidentiality in the nationwide health information network. National Committee on Vital and Health Statistics. 2006. URL: https://library.ahima.org/doc?oid=75960#.Y-RqgnZBxD8 [accessed 2023-02-09]
13. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy issues in smartphone applications: an analysis of headache/migraine applications. Headache 2018 Jul;58(7):1014-1027 [FREE Full text] [doi: 10.1111/head.13341] [Medline: 29974470]
14. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
15. Knorr K, Aspinall D, Wolters M. On the privacy, security and safety of blood pressure and diabetes apps. In: Federrath H, Gollmann D, editors. ICT Systems Security and Privacy Protection. SEC 2015. IFIP Advances in Information and Communication Technology, Vol 455. Cham: Springer; 2015:571-584.
16. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
17. Robillard J, Feng T, Sporn A, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Sep;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
18. Knorr K, Aspinall D. Security testing for Android mHealth apps. 2015 Presented at: 2015 IEEE Eighth International Conference on Software Testing, Verification and Validation Workshops (ICSTW); April 13-17, 2015; Graz, Austria p. 13-17. [doi: 10.1109/icstw.2015.7107459]
19. Scott K, Richards D, Adhikari R. A review and comparative analysis of security risks and safety measures of mobile health apps. AJIS 2015 Nov 22;19:1-18. [doi: 10.3127/ajis.v19i0.1210]
20. Grindrod K, Boersema J, Waked K, Smith V, Yang J, Gebotys C. Locking it down: the privacy and security of mobile medication apps. Can Pharm J (Ott) 2017;150(1):60-66 [FREE Full text] [doi: 10.1177/1715163516680226] [Medline: 28286594]
21. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Assessment of the fairness of privacy policies of mobile health apps: scale development and evaluation in cancer apps. JMIR Mhealth Uhealth 2020 Jul 28;8(7):e17134 [FREE Full text] [doi: 10.2196/17134] [Medline: 32720913]
22. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
23. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
24. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
25. Hinkin TR. A review of scale development practices in the study of organizations. J Manag 2016 Jun 30;21(5):967-988. [doi: 10.1177/014920639502100509]
26. Morgan D. Focus Groups as Qualitative Research. New York, NY: SAGE Publications; 1996.

27. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci 2015 Jun;4(2):165-178 [FREE Full text] [doi: 10.15171/jcs.2015.017] [Medline: 26161370]
28. Rodrigues IB, Adachi JD, Beattie KA, MacDermid JC. Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. BMC Musculoskelet Disord 2017 Dec 19;18(1):540 [FREE Full text] [doi: 10.1186/s12891-017-1914-5] [Medline: 29258503]
29. Zare S, Rezaee R, Aslani A, Shirdeli M, Kojuri J. Moving toward community based telehealth services using mhealth for hypertensive patients. Int J Technol Assess Health Care 2019 Sep 24;35(5):379-383. [doi: 10.1017/s0266462319000655]
30. Jones J, Hook S, Park S, Scott L. Privacy, security and interoperability of mobile health applications. LNCS (subseries LN Artif Intell LN Bioinform) 2011;6767:46-55. [doi: 10.1007/978-3-642-21666-4_6]
31. Adhikari R, Richards D, Scott K. Security and privacy issues related to the use of mobile health apps. 2014 Presented at: Proceedings of the 25th Australasian Conference on Information Systems (ACIS); December 8-10, 2014; Auckland, New Zealand.
32. Rowan M, Dehlinger J. A privacy policy comparison of health and fitness related mobile applications. Procedia Comp Sci 2014;37:348-355. [doi: 10.1016/j.procs.2014.08.051]
33. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
34. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
35. Kramer GM, Kinn JT, Mishkind MC. Legal, regulatory, and risk management issues in the use of technology to deliver mental health care. Cogn Behav Pract 2015 Aug;22(3):258-268. [doi: 10.1016/j.cbpra.2014.04.008]
36. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
37. Plachkinova M, Andres S, Chatterjee S. A taxonomy of mHealth apps - security and privacy concerns. 2015 Presented at: Proceedings of the Annual Hawaii International Conference on System Sciences; January 5-8, 2015; Hawaii. [doi: 10.1109/hicss.2015.385]
38. Scott KM, Gome GA, Richards D, Caldwell PHY. How trustworthy are apps for maternal and child health? Health Technol 2015 Mar 12;4(4):329-336. [doi: 10.1007/s12553-015-0099-x]
39. Bruggemann T, Hansen J, Dehling T, Sunyaev A. An information privacy risk index for mHealth apps. 2016 Presented at: 4th Annual Privacy Forum (APF); September 7-8, 2016; Frankfurt, Germany. [doi: 10.1007/978-3-319-44760-5_12]
40. Chen C, Osman M, Zaaba Z, Talib A. Managing secure personal mobile health information. Adv Intell Syst Comput 2017:347-356. [doi: 10.1007/978-3-319-49073-1_38]
41. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. Prof Psychol: Res Pract 2016 Apr;47(2):155-162. [doi: 10.1037/pro0000069]
42. Loy JS, Ali EE, Yap KY. Quality assessment of medical apps that target medication-related problems. J Manag Care Spec Pharm 2016 Oct;22(10):1124-1140 [FREE Full text] [doi: 10.18553/jmcp.2016.22.10.1124] [Medline: 27668561]
43. Mense A, Steger S, Sulek M, Jukicsunaric D, Meszaros A. Analyzing privacy risks of mHealth applications. 2016 Presented at: Special Topic Conference (STC) of the European-Federation-for-Medical-Informatics (EFMI) on Transforming Healthcare with the Internet of Things; April 2016; Paris, France p. 17-19.
44. Morera EP, de la Torre Díez I, Garcia-Zapirain B, López-Coronado M, Arambarri J. Security recommendations for mHealth apps: elaboration of a developer's guide. J Med Syst 2016 Jun;40(6):152. [doi: 10.1007/s10916-016-0513-6] [Medline: 27147515]
45. Asaddok N, Ghazali M. Exploring the usability, security and privacy taxonomy for mobile health applications. 2017 Presented at: 5th International Conference on Research and Innovation in Information Systems (ICRIIS); July 16-17, 2017; Langkawi Island, Malaysia. [doi: 10.1109/icriis.2017.8002472]
46. Wu E, Torous J, Hardaway R, Gutheil T. Confidentiality and privacy for smartphone applications in child and adolescent psychiatry: unmet needs and practical solutions. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):117-124. [doi: 10.1016/j.chc.2016.07.006] [Medline: 27837937]
47. Gabel A, Schiering I, Müller S, Ertas F. mHealth applications for goal management training - privacy engineering in neuropsychological studies. IFIP Adv Inf Commun Technol 2018:330-345. [doi: 10.1007/978-3-319-92925-5_22]
48. Hussain M, Al-Haiqi A, Zaidan A, Zaidan B, Kiah M, Iqbal S, et al. A security framework for mHealth apps on Android platform. Comp Security 2018 Jun;75:191-217. [doi: 10.1016/j.cose.2018.02.003]
49. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
50. Scott KM, Richards D, Londos G. Assessment criteria for parents to determine the trustworthiness of maternal and child health apps: a pilot study. Health Technol 2018 Jan 25;8(1-2):63-70. [doi: 10.1007/s12553-018-0216-8]

51. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
52. Al-Sharo YM. Networking issues for security and privacy in mobile health apps. Int J Adv Comp Sci Appl 2019;10(2):186-191. [doi: 10.14569/ijacsa.2019.0100225]
53. Iwaya LH, Fischer-Hübner S, Åhlfeldt RM, Martucci LA. Mobile health systems for community-based primary care: identifying controls and mitigating privacy threats. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11642 [FREE Full text] [doi: 10.2196/11642] [Medline: 30892275]
54. Müthing J, Brüngel R, Friedrich CM. Server-focused security assessment of mobile health apps for popular mobile platforms. J Med Internet Res 2019 Jan 23;21(1):e9818 [FREE Full text] [doi: 10.2196/jmir.9818] [Medline: 30672738]
55. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mHealth applications: a scoping review. IEEE Access 2020;8:104247-104268. [doi: 10.1109/access.2020.2999934]
56. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33 [FREE Full text] [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
57. Srivastava M, Thamilarasu G. MSF: a comprehensive security framework for mHealth applications. 2019 Presented at: 7th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW); August 26-28, 2019; Istanbul, Turkey p. 26-28. [doi: 10.1109/ficloudw.2019.00026]
58. Dev J. Usage of botnets for high speed MD5 hash cracking. 2013 Presented at: Third International Conference on Innovative Computing Technology (INTECH); August 29-31, 2013; London, UK p. 29-31. [doi: 10.1109/intech.2013.6653658]
59. Guo Y, Zhang Z, Guo Y. Optiwords: a new password policy for creating memorable and strong passwords. Comp Security 2019 Aug;85:423-435 [FREE Full text] [doi: 10.1016/j.cose.2019.05.015]
60. Greene E, Proctor P, Kotz D. Secure sharing of mHealth data streams through cryptographically-enforced access control. Smart Health (Amst) 2019 Apr;12:49-65 [FREE Full text] [doi: 10.1016/j.smhl.2018.01.003] [Medline: 31598544]
61. Bhanot R, Hans R. A review and comparative analysis of various encryption algorithms. Int J Security Appl 2015 Apr 30;9(4):289-306. [doi: 10.14257/ijsia.2015.9.4.27]
62. Bajwa MI. mHealth security. Pak J Med Sci 2014 Jul 31;30(4):904-907 [FREE Full text] [doi: 10.12669/pjms.304.5210] [Medline: 25097542]
63. Choi CQ. Digital danger. Sci Am 2012 Nov 13;307(6):14-14. [doi: 10.1038/scientificamerican1212-14]
64. Arora S, Yttri J, Nilse W. Privacy and security in mobile health (mHealth) research. Alcohol Res 2014;36(1):143-151 [FREE Full text] [Medline: 26259009]
65. Braghin C, Cimato S, Libera A. Are mHealth apps secure? A case study. 2018 Presented at: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); July 23-27, 2018; Tokyo, Japan p. 23-27. [doi: 10.1109/compsac.2018.10253]
66. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]

Abbreviations

CVI: content validity index
CVR: content validity ratio
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
HIT: health information technology
mHealth: mobile health

Edited by L Buis; submitted 27.04.22; peer-reviewed by J Ropero, KL Mauco; comments to author 25.07.22; revised version received 16.08.22; accepted 21.09.22; published 02.03.23.

Please cite as:
Rezaee R, Khashayar M, Saeedinezhad S, Nasiri M, Zare S
Critical Criteria and Countermeasures for Mobile Health Developers to Ensure Mobile Health Privacy and Security: Mixed Methods Study
JMIR Mhealth Uhealth 2023;11:e39055
URL: https://mhealth.jmir.org/2023/1/e39055
doi:10.2196/39055
PMID:36862494

Content and Quality of Mobile Apps for the Monitoring of Musculoskeletal or Neuropathic Pain in Australia: Systematic Evaluation

Joshua Simmich[1,2,3], PhD; Megan Heather Ross[1,2,3], PhD; Nicole Emma Andrews[1,2,3,4,5], PhD; Atiyeh Vaezipour[1,2,3], PhD; Trevor Glen Russell[1,2,3], PhD

1

2

3

4

5

Corresponding Author:
Joshua Simmich, PhD

Abstract

Background: Mobile apps offer a potential mechanism for people with persistent pain to monitor pain levels conveniently within their own environment and for clinicians to remotely monitor their patients' pain. However, the quality of currently available apps and the usefulness of included features from a clinical perspective are not known.

Objective: The aim of this study was to examine the content and quality of currently available smartphone apps designed for monitoring the intensity or presence of musculoskeletal or neuropathic pain.

Methods: A systematic search was performed in the Australian Apple and Google Play stores. Apps were included if they were designed to monitor the intensity or presence of musculoskeletal or neuropathic pain and were available in the English language within the Australian app stores. Data pertaining to the intended use of the app and clinical population were extracted by using a custom-designed data extraction form, and app quality was assessed by using the 23-item Mobile App Rating Scale.

Results: Of the 2190 apps screened, 49 met the inclusion criteria. Apps were primarily designed for adult users (36/49, 73%) with nonspecific musculoskeletal or neuropathic pain conditions, arthritis, and joint pain. All apps monitored pain intensity, with almost half (23/49, 47%) also specifying pain location. Overall, the mean quality scores from the Mobile App Rating Scale ranged from 1.5 to 4.4 (out of 5.0). Between 20% (10/49) and 22% (11/49) of apps involved clinicians, consumers, or both in their development, and 20% (10/49) had published literature related to the development or use of the app in clinical scenarios. Although 71% (35/49) had data sharing features, only 5 apps enabled client-clinician communication through the app.

Conclusions: The overall quality of mobile apps that are currently available for monitoring pain intensity is acceptable. Presently, mobile apps for remote pain monitoring lack functionality for clinicians to view data between consults. Both users and clinicians should be aware of the limitations of these apps and make informed choices in using or recommending apps that best suit the clinical need.

(JMIR Mhealth Uhealth 2023;11:e46881) doi:10.2196/46881

KEYWORDS

pain; monitoring; digital health; mobile application; digital health; mobile app; pain management; pain level; chronic pain; smartphone; musculoskeletal pain; neuropathic pain; remote

Introduction

Background

Persistent or chronic pain has been recognized as a global public health priority [1]. It is estimated that 20% of the adult population experience pain globally [2]. Persistent pain is considered a stand-alone disease [3,4], with musculoskeletal pain being by far the most prevalent pain condition [5,6]. Persistent pain is linked to changes in neural signaling and reorganization of the brain's structure and function [7]. It is well established that persistent pain can have a devastating effect on individuals, interfering with relationships, mental health, and the ability to engage in meaningful and important activities [8-11]. The condition is a significant contributor to the opioid crisis [12] and incurs substantial costs to society through health system expenditures, decreased productivity, decreased quality of life, and the need for the provision of informal care [13,14].

Persistent pain management services often experience considerable health service strain, as evidenced in Australia, where high demand for such services has resulted in prolonged waiting times to access care [15]. This reality underscores the importance of innovation to improve service delivery, which could potentially be achieved through digital health technology [16-18]. Digital health technologies could empower patients to self-manage within their own environment, thereby decreasing the need for hospital or clinic visits. Digital health could also facilitate the remote monitoring of up-to-date personalized data to improve service delivery by assisting with patient triage, decreasing waiting times of those with urgent needs, and facilitating more timely treatment decisions [19-21]. By enhancing self-management, reducing barriers to accessing care, and addressing existing inefficiencies in pain management services, digital health innovations may lead to the more effective management of persistent pain.

Self-management refers to the “ability to monitor one’s condition and to effect the cognitive, behavioural and emotional responses necessary to maintain a satisfactory quality of life” [22]. Monitoring outcomes of behavior, such as pain intensity, is therefore a critical but underused component of pain self-management [23], as well as an established behavior change technique in its own right [24]. Self-monitoring could help people with persistent pain to become more aware of any patterns in their pain. Self-monitoring could also help individuals to identify behaviors or circumstances that may change their pain and help them to make changes in their behavior or lifestyle that may reduce their pain levels, thereby gaining an enhanced feeling of self-control [25]. In addition, monitoring has the potential to improve information exchange between people with pain and their clinicians. People with persistent pain can struggle to accurately recall pain intensity or fluctuations beyond the past several days [26], often leading to the overestimation of past pain intensity [27,28]. By regularly tracking their pain intensity, individuals can provide more accurate data to their clinicians, who would be better placed to provide feedback and support with pain management strategies. Additionally, receiving feedback from others, especially trusted others such as clinicians, is also a key behavior change technique for chronic disease management [24]. Overall, the monitoring of pain can be a valuable technique for people with persistent pain, helping them to better understand the nature of their pain and communicate this to clinicians.

Mobile health (mHealth) apps are being explored as potential tools for management and monitoring in people with persistent pain [29]. Although evidence suggests that mHealth apps may improve communication between health care professionals and patients [30], increase patients’ engagement with their health [31], and lead to health benefits [29,32,33], there remains a need for the systematic evaluation of the content and quality of available mHealth apps that focus primarily on pain monitoring. The Mobile App Rating Scale (MARS) is a widely used and reliable tool for assessing the quality of mobile phone apps [34] and has been used to appraise the quality of mHealth apps for low back pain [35-38], shoulder pain [39], and neck pain [40], as well as those for cancer [41] and arthritis [42,43]. However, these prior appraisals have largely focused on the use of apps for the broader concept of the self-management of pain, whereas the monitoring of pain intensity is only 1 component of self-management. A previous evaluation found that only half of commercially available pain management apps had a monitoring feature and that such apps instead more commonly gave instructions on new techniques (such as exercises or stretches), encouraged goal-setting, and provided education about the link between behaviors and pain [44]. To date, existing research has been limited to investigating the content, but not the quality, of general pain monitoring apps [45] or investigating both the content and the quality of apps for monitoring general pain and cancer pain [41]. However, musculoskeletal pain is by far the single largest category of persistent pain, with cancer-related pain constituting only a small minority of pain cases. Additionally, persistent musculoskeletal pain may overlap with neuropathic pain [46,47]. Therefore, there is a need for and a gap in the literature regarding the evaluation of apps for tracking and monitoring musculoskeletal or neuropathic pain symptoms.

Aim

The aim of this study was to systematically review and appraise the content and quality of currently available mobile phone apps that were primarily designed for monitoring musculoskeletal or neuropathic pain intensity over time.

Methods

Search Strategy and Selection

The Apple App Store and the Google Play Store were searched in Australia on February 2, 2022. The keywords used for the search were based on the most common pain conditions in Australia and Canada [48,49]. The following keywords were used: *pain*, *arthritis*, *headache*, *migraine*, *post surgery*, *fracture*, and *fibromyalgia*. Search results and app details (eg, developer, cost, and version date and number) were downloaded to a spreadsheet, using Python-based App Store and Play Store scrapers developed by the Digital Methods Initiative [50,51]. Search results were first manually screened, by 2 reviewers (JS and AV), for apps that appeared in both stores and were then uploaded to Covidence (Veritas Health Innovation Ltd) for initial screening based on the names and descriptions of the apps.

Apps were included if they were available for public use and if monitoring musculoskeletal or neuropathic pain over time was a primary focus of the app. Apps were excluded if they were for monitoring reproductive pain or cancer pain. Although *migraine* and *headache* were initially added as search terms, the large volume of apps designed specifically for monitoring migraine and headache symptoms led to the decision to consider these separately; therefore, such apps were also excluded. Apps were additionally excluded if they were considered generic health monitoring apps (ie, apps for monitoring many health symptoms, without pain as the primary focus) or were no longer publicly available in Australia at the time of data extraction or evaluation.

Four reviewers (JS, MHR, AV, and NEA) independently performed the initial screening based on the inclusion criteria.

Apps that met the selection criteria, according to 2 reviewers, were downloaded onto Apple or Android devices for a full review (Multimedia Appendix 1 provides device details). Disagreements were resolved through discussion or consultation with a third reviewer.

Data Extraction

Apps that met the inclusion criteria were purchased (if applicable) and downloaded. One reviewer extracted general information about the apps by using a custom-made data extraction spreadsheet, and the second reviewer checked the veracity of this information. Extracted information included details of consumer and clinician involvement in app development, target populations, app features (eg, gamification, symptoms monitored, and monitoring frequency), pain tracking features (eg, intensity, location, and description), additional app features (eg, mood, exercise, and physical activity tracking), and data sharing features.

Quality Appraisal

Each mobile app was independently rated by 2 reviewers from a panel of 4, ensuring that all reviewers participated in the process. Reviewers rated each app on the MARS, which consists of 23 items across categories, namely engagement, functionality, aesthetics, information quality, and subjective quality [34]. The MARS was scored on a 5-point scale (1: inadequate; 2: poor; 3: acceptable; 4: good; 5: excellent), as per standard instructions for use. Mean scores were calculated for the first 4 categories (engagement, functionality, aesthetics, and information quality), and an overall mean score for the MARS was obtained by averaging these 4 means. As recommended [34], all reviewers viewed the MARS training material. App quality ratings were piloted by having all 4 reviewers rate the same app initially. Subsequently, all reviewers met to discuss the ratings for this app to ensure that they had the same interpretation of the rating scale and process, before proceeding to rate the remaining apps.

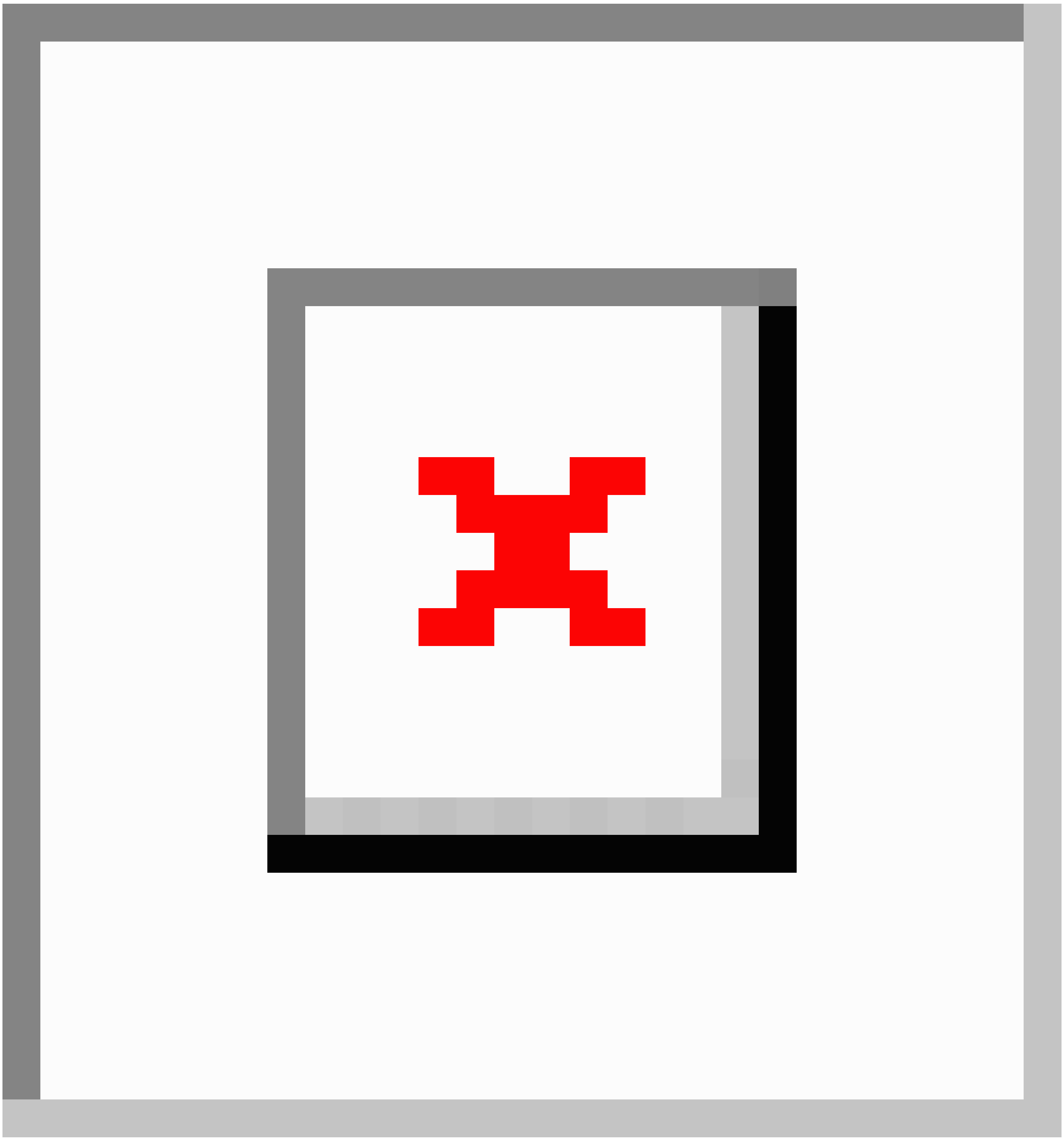
Interrater reliability for the MARS was calculated by using the 1-way random effects intraclass correlation (ICC; ICC [1,1]), under the assumption that the 2 reviewers rating each app were randomly selected from the larger population of 4 reviewers [52]. A score of greater than 0.7 on the ICC was considered to indicate acceptable reliability between reviewers, while a score of greater than 0.8 indicated good reliability and a score of greater than 0.9 indicated excellent reliability.

Results

Overview of Apps

Of the 2190 apps screened, 151 were downloaded for a full app review, and a total of 49 met the inclusion criteria and were screened by using the MARS. Mobile apps were primarily excluded for being designed for specific clinical studies or health facilities and not being designed primarily for pain monitoring (Figure 1). Details of the data extracted from the app stores are included in Multimedia Appendix 2, and the data extracted by reviewers for each mobile app are included in Multimedia Appendix 3.

Figure 1. Flowchart of the app selection process. MSK: musculoskeletal.



In total, 35% (17/49) of included apps were available on both the Apple App Store and the Google Play Store, with 49% (24/49) only available in the App Store and 16% (8/49) only available in the Play Store. The majority of apps (35/49, 71%) were completely free to download and use, while an additional 10% (5/49) offered conditional free use but asked for payments to eliminate restrictions or advertisements. For those that required payment, the purchase cost ranged between Aus $1.49 (US $0.96) and Aus $10.99 (US $7.10), with some requiring an ongoing subscription or in-app purchases (Multimedia Appendix 2). Of the 49 included apps, 10 (20%) had published literature supporting their development, efficacy, or both; 10 (20%) were developed in consultation with clinicians; and 11 (22%) involved consumers in the design (Multimedia Appendix 3). Apps were primarily designed for use by adults (36/49, 73%), with only 3 (6%) designed for adolescents and 1 (2%) exclusively for children. Further, 5 (10%) used some form of gamification, primarily in the form of awarding points and achieving targets (for logging pain and activities).

Pain Monitoring

The frequency at which users can enter pain ratings ranged from an unlimited number (ie, multiple entries per day) to the weekly logging of pain intensity, with 94% (46/49) of apps permitting at least daily recording. All apps recorded pain intensity (numeric rating scale: 24/49, 49%; visual analog scale: 19/49, 39%; Likert scale: 5/49, 10%; Wong-Baker Faces: 1/49, 2%) with a range of anchors and descriptors, including text, colors, and emojis (or faces). Around two-thirds (32/49, 65%) of apps graphically depicted change in pain symptoms over time with a chart or graph. Of the 49 apps, almost half (n=23, 47%) also recorded pain location, with 9 (18%) apps using a body chart to record pain location and the remaining 14 (29%) using text to describe the location of pain. Apps that recorded additional dimensions of pain were less frequent, with 15 (31%) recording pain type and quality and 10 (20%) recording duration and frequency. Reminders to enter data (ie, to monitor and track pain) were included in 18 (37%) of the apps, with the ability to customize the frequency of these reminders featured in 10 (20%) apps. Further details regarding the pain monitoring features of included apps are provided in Multimedia Appendix 3.

Additional Features

Many apps included features that enabled users to monitor and track other symptoms and events, including medication (28/49, 57%), mood (25/49, 51%), and customizable or free-text notes (20/49, 41%). More detailed additional features are provided in Multimedia Appendix 4. Further, of the 49 apps, 20 (41%) included educational information or resources within the app or links to external resources for pain education, additional support, and further condition-specific information. All apps lacked the capacity to individualize management based on pain data entered by the patient; for example, they did not automatically adjust exercises or provide advice to manage a pain flare-up.

Data Sharing

Of the 49 included apps, 35 (71%) had the ability to share data from the app (Multimedia Appendix 3). A wide range of file formats for exporting were used, with the most common being CSV and PDF formats (either as a file for the device or as an attachment in an email). Further, 5 apps had data sharing functionalities within the app platform itself, requiring clinicians to be a user of the app to receive shared data. No apps offered real-time data sharing with clinicians. Apps primarily shared raw data (34/49, 70%) and graphs (32/49, 65%), with fewer apps providing a summary of the data (18/49, 37%). Of the included apps, 32 had a privacy policy listed on the store or within the app, whereas 17 apps did not.

MARS Ratings

The ICC (1,1) for agreement on the MARS ratings between reviewers was 0.72 (95% CI 0.61-0.81), indicating good reliability, but the uncertainty ranged from less than acceptable (<0.7) to better than good (>0.8). The median overall MARS score for included mobile apps was 3.1 (range 1.5-4.4). Almost 60% (29/49, 59%) of apps scored ≥3.0 (acceptable) overall; the highest mean score was for the functionality (mean 3.5, SD 0.66) domain (Multimedia Appendix 5). Included apps, on average, had the lowest scores for the engagement domain, with a mean of 2.5 (SD 0.63), which is below *acceptable* on the MARS.

Discussion

Principal Results

This study systematically reviewed the content and appraised the quality of currently available mobile phone apps that were primarily designed for monitoring musculoskeletal or neuropathic pain intensity over time. In an era where digital health solutions are increasingly prevalent, providing a snapshot of the landscape of pain monitoring apps is critical. These findings have important implications for the potential use of these apps in clinical practice from the perspectives of both clinicians and people with persistent pain.

In conducting this evaluation, we identified a range of apps that had different methods for highlighting painful body parts and monitoring pain over time, ranging from basic methods to sophisticated methods. In accordance with existing literature on general pain monitoring apps [45] and pain management apps [53], only 31% (15/49) of the included apps allowed users to rate or describe other dimensions of their pain, such as its quality. It is likely that unidimensional ratings of pain intensity are used for their simplicity for users and app developers, but it is critical to consider that pain is an inherently multidimensional experience. Perhaps underscoring the increasing attention to tracking other dimensions of the pain experience, medication use and mood were monitored in 57% (28/49) and 51% (25/49) of included apps, respectively, which are higher than the 39% and 31% reported in prior research [45]. Additionally, 65% (32/49) of the apps in our study used graphs or charts to visualize pain intensity, closely matching the 61% of apps that were noted to have data visualizations in a prior study of apps that tracked general or cancer pain [41]. Given how useful graphical representations are in summarizing an overall pattern in pain presentation, it is surprising that more apps have not adopted a graphical display.

We identified several limitations of the apps reviewed in this study. Of significant concern, only 20% (10/49) to 22% (11/49) of the apps included in this study had involved either clinicians or people with persistent pain in their design process. This aligns with prior findings indicating that only 31% of general pain monitoring apps consulted health care practitioners in development, and only 5.6% involved patients [45]. Though the apparent increase in the involvement of people with persistent pain is promising, as these individuals are ultimately the end users of these apps, the relatively low involvement of these key stakeholders is concerning and suggests that a greater focus is required for user-centered design and co-design. Moreover, we discovered that 4 out of every 5 apps (39/49, 80%) assessed in this study did not provide any publicly available literature outlining their development process or reporting any evaluation of validity, user perspectives, efficacy, or usage. Although this may not be entirely unexpected, considering the considerable costs and time investments necessary to undertake such research, it hampers the ability of consumers and health care professionals to make well-informed decisions when selecting or endorsing such apps. Finally, none of the included apps allowed clinicians to tailor any management components to the individual, despite there being some evidence that tailored technology interventions [54] and individualized pain management interventions [55,56] may be more effective and are valued by patients [57].

The MARS ratings for the included apps described the overall quality and quality indicators of the apps, including engagement, functionality, aesthetics, and information quality [34]. Scores for apps were highest for the functionality subscales and lowest for the engagement subscales. This pattern in MARS ratings has been consistently observed across several prior studies that appraised apps for the self-management of pain conditions [35,36,38,44,58]. Likewise, in apps for monitoring general or cancer pain, functionality was the highest-scoring domain, and engagement was the second-lowest–scoring MARS domain after information quality [41]. Therefore, despite the apps functioning well, the lack of features for promoting engagement may result in the inconsistent or discontinued use of the app for remote pain monitoring and, in turn, an incomplete picture of the pain presentation. Only 10% (5/49) of the apps included in this study had any gamification features, and in a prior investigation, gamification was not noted in any apps for the self-management of back pain [36]. Future apps and further updates to existing apps should consider embedding features and techniques that increase engagement (ie, gamification) to promote the desired frequency of monitoring.

An emergent finding of our review was that most apps (44/49, 90%) lacked a method for direct data sharing between users and clinicians (eg, facilitating live access via a dashboard or exporting data in a standard format that can be imported to electronic medical records). Instead, the apps usually saved a data file or graph (often to be printed or attached to an email). Considering that clinicians are concerned about implementing apps in a clinical context due to workflow disruptions and time burden [59,60], this has the potential to place significant burdens on clinicians who may be unable to deal with the reports generated by their patients, especially due to the potential for varied formats from different apps. The potential burden on clinicians could be further exacerbated by apps that export data in file formats that are not designed for human readability (such as JSON and CSV), which require the clinician to have the time and skills for converting these data to a format that makes them readable to a human and clinically meaningful. Data need to be presented in a format that is useful, usable, and interpretable for clinicians and users, such as a familiar data visualization format [61,62]. Access to a live dashboard or the sharing of standardized data in usable and clinician-friendly formats is recommended for future pain monitoring apps, as well as ensuring that robust security measures are put in place to protect sensitive health data and adhere to relevant data protection laws.

Implications for Clinical Practice

Health professionals may opt to use a pain monitoring app for clients with varying outcomes in mind and for a variety of reasons. The appropriateness of using pain monitoring apps must be determined based on the client's clinical presentation. For example, clients demonstrating high levels of pain catastrophizing may not be suitable candidates for ongoing pain monitoring, especially not without supervision and reassurance from clinicians, as pain monitoring may lead to increased focus on and worry about pain symptoms [63]. On the other hand, clients who are "overactive" and experience pain flare-ups may be able to use self-monitoring to better pace themselves [57,64]. Similarly, for clients who avoid activity due to the fear of provoking pain, the regular monitoring of pain may lead to improved confidence by correcting exaggerated predictions of pain provocation [65]. In this latter case especially, a mobile app that tracks activity, as well as pain monitoring, may be necessary to demonstrate that activity does not always lead to an increase in pain [63,65] and that safe and acceptable levels of pain can occur during exercise without causing harm, which are both important understandings to facilitate participation in rehabilitation (eg, tendinopathy rehabilitation [66]). Finally, clients who are making progress may be motivated by visualizing progress in easily comprehensible data outputs, such as graphs, which can facilitate focus on past successes—a known behavior change technique [24].

When pain monitoring apps are deemed appropriate, the selection of a specific app can be facilitated by the findings of this study. Clinical decisions regarding which app would be the most suitable for individual clients can be based on various factors, including the method of pain identification used within the app (such as the selection of broad body areas vs the ability to shade on a body chart), the requirements for monitoring additional dimensions of pain symptoms (such as neuropathic pain, tingling, or numbness), the availability of additional tracking features (such as those for exercises, physical activity, or mood), and the need for clinicians to access the data (such as real-time monitoring for high-level athletes vs weekly check-ins for an outpatient clinic setting). Given the tendency for clinicians to recommend mobile apps that patients are already using and are favorable toward [67], it could also be beneficial for patients' familiarity and preferences for specific apps to be routinely considered to potentially enhance engagement.

Implications for Future Research

This study revealed that a limited number of the evaluated apps were clearly grounded in evidence or research. This highlights the need for future research into the development of pain monitoring apps that are based on established clinical guidelines [68,69] and developed in consultation with consumers and clinicians. Rigorous randomized controlled trials remain the gold standard for assessing the effectiveness of mHealth apps but can be challenging to conduct with sufficient durations and sample sizes. Although the UK National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies recommends formal trials for apps that aim to treat or diagnose health conditions [70], for health monitoring apps, the focus is on evidence of successful pilot tests within the health and care system that show relevance to current service provision or best practice. In addition to empirical evidence, app developers and researchers can reference frameworks, such as the Assessment Framework for mHealth Apps published by the Australian Digital Health Agency [71], to ensure that the app is deemed safe, trustworthy, useful, usable, and likely to be effective. Further research is also needed to investigate the real-world use and uptake of pain monitoring apps by individuals with persistent pain, as this information can inform the design and implementation of future apps to better meet the needs of users. Additionally, future research could also explore how apps can be tailored to specific populations, such as older adults, children, or individuals with specific pain conditions.

Limitations

There are several limitations to this study that should be considered when interpreting the results. First, the search was conducted over 1 year prior to the date of publication. Although the results presented nonetheless provide a valuable snapshot of the app landscape at the time, the rapidly evolving nature of this field must be considered. It is likely that this study may include apps that have since been discontinued or substantially updated, and newly published relevant apps would be missing from our review. This omission could potentially limit the future applicability of our findings to clinicians and people with persistent pain. Second, our search was limited to the Australian Apple and Google Play stores and was conducted in English, which means that apps that are only available in other countries or languages were not included in this review. Third, we excluded apps that required users to log into a patient portal or medical practice website because we were unable to access these apps. It is possible that these apps may provide real-time access to data and improved management options for clinicians, as well as potentially better privacy protections. However, the aim of this study was to evaluate publicly available apps for monitoring pain; therefore, these apps were outside the scope of our review. Finally, agreement on the quality assessment using the MARS was lower than that in prior studies. This may be due, in part, to the fact that prior studies used the same reviewers for all apps, whereas in this study, each app was rated by just 2 of the 4 reviewers, which may have introduced an additional source of variability.

Conclusions

This study reviewed mobile phone apps designed for monitoring pain intensity over time and found that while many apps with various features existed, they lacked the capacity for real-time data sharing with clinicians and were rated poorly for engagement. Many of the apps lacked any supporting research publications. This study suggests that future apps should focus on increasing engagement and providing data in a usable format for clinicians. The appropriateness of using pain monitoring apps should be determined based on the patient's clinical presentation, as well as client preferences.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Hardware and system software used by app reviewers.

[DOCX File, 19 KB - mhealth_v11i1e46881_app1.docx]

Multimedia Appendix 2

Details of data extracted from the app stores.

[XLSX File, 22 KB - mhealth_v11i1e46881_app2.xlsx]

Multimedia Appendix 3

Data extracted by reviewing each app.

[XLSX File, 27 KB - mhealth_v11i1e46881_app3.xlsx]

Multimedia Appendix 4

Additional features present in each included app.

[XLSX File, 21 KB - mhealth_v11i1e46881_app4.xlsx]

Multimedia Appendix 5

Mobile App Rating Scale scores (average of 2 reviewers) for each included app.

[XLSX File, 19 KB - mhealth_v11i1e46881_app5.xlsx]

References

1. Goldberg DS, McGee SJ. Pain as a global public health priority. BMC Public Health 2011 Oct 6;11:770. [doi: 10.1186/1471-2458-11-770] [Medline: 21978149]
2. Zimmer Z, Fraser K, Grol-Prokopczyk H, Zajacova A. A global study of pain prevalence across 52 countries: examining the role of country-level contextual factors. Pain 2022 Sep 1;163(9):1740-1750. [doi: 10.1097/j.pain.0000000000002557] [Medline: 35027516]
3. Raffaeli W, Arnaudo E. Pain as a disease: an overview. J Pain Res 2017 Aug 21;10:2003-2008. [doi: 10.2147/JPR.S138864] [Medline: 28860855]

4. Clauw DJ, Essex MN, Pitman V, Jones KD. Reframing chronic pain as a disease, not a symptom: rationale and implications for pain management. Postgrad Med 2019 Apr;131(3):185-198. [doi: 10.1080/00325481.2019.1574403] [Medline: 30700198]
5. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
6. Australian Institute of Health and Welfare. Chronic pain in Australia. 2020. URL: www.aihw.gov.au/getmedia/10434b6f-2147-46ab-b654-a90f05592d35/aihw-phe-267.pdf.aspx [accessed 2023-08-30]
7. Chapman CR, Vierck CJ. The transition of acute postoperative pain to chronic pain: an integrative overview of research on mechanisms. J Pain 2017 Apr;18(4):359.e1-359.e38. [doi: 10.1016/j.jpain.2016.11.004] [Medline: 27908839]
8. Hadi MA, McHugh GA, Closs SJ. Impact of chronic pain on patients' quality of life: a comparative mixed-methods study. J Patient Exp 2019 Jun;6(2):133-141. [doi: 10.1177/2374373518786013] [Medline: 31218259]
9. Kawai K, Kawai AT, Wollan P, Yawn BP. Adverse impacts of chronic pain on health-related quality of life, work productivity, depression and anxiety in a community-based study. Fam Pract 2017 Nov 16;34(6):656-661. [doi: 10.1093/fampra/cmx034] [Medline: 28444208]
10. Suso‐Ribera C, Yakobov E, Carriere JS, García‐Palacios A. The impact of chronic pain on patients and spouses: consequences on occupational status, distribution of household chores and care-giving burden. Eur J Pain 2020 Oct;24(9):1730-1740. [doi: 10.1002/ejp.1616] [Medline: 32533892]
11. de Sola H, Salazar A, Dueñas M, Ojeda B, Failde I. Nationwide cross-sectional study of the impact of chronic pain on an individual's employment: relationship with the family and the social support. BMJ Open 2016 Dec 23;6(12):e012246. [doi: 10.1136/bmjopen-2016-012246] [Medline: 28011806]
12. Volkow ND, Blanco C. The changing opioid crisis: development, challenges and opportunities. Mol Psychiatry 2021 Jan;26(1):218-233. [doi: 10.1038/s41380-020-0661-4] [Medline: 32020048]
13. Phillips CJ. Economic burden of chronic pain. Expert Rev Pharmacoecon Outcomes Res 2006 Oct;6(5):591-601. [doi: 10.1586/14737167.6.5.591] [Medline: 20528505]
14. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021 May 29;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
15. Hogg MN, Gibson S, Helou A, DeGabriele J, Farrell MJ. Waiting in pain: a systematic investigation into the provision of persistent pain services in Australia. Med J Aust 2012 Apr 2;196(6):386-390. [doi: 10.5694/mja12.10140] [Medline: 22471539]
16. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
17. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893. [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
18. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A systematic review of technology-assisted self-management interventions for chronic pain: looking across treatment modalities. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]
19. Mohammed KI, Zaidan AA, Zaidan BB, Albahri OS, Alsalem MA, Albahri AS, et al. Real-time remote-health monitoring systems: a review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure. J Med Syst 2019 Jun 11;43(7):223. [doi: 10.1007/s10916-019-1362-x] [Medline: 31187288]
20. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480. [doi: 10.2196/10480] [Medline: 30001997]
21. Yempally S, Singh SK, Velliangiri S. Review of an IoT-based remote patient health monitoring system. Presented at: 2021 Smart Technologies, Communication and Robotics (STCR); October 9-10, 2021; Sathyamangalam, India. [doi: 10.1109/STCR51658.2021.9589019]
22. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
23. Keogh A, Tully MA, Matthews J, Hurley DA. A review of behaviour change theories and techniques used in group based self-management programmes for chronic low back pain and arthritis. Man Ther 2015 Dec;20(6):727-735. [doi: 10.1016/j.math.2015.03.014] [Medline: 25865062]
24. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
25. Charoenpol FN, Tontisirin N, Leerapan B, Seangrung R, Finlayson RJ. Pain experiences and intrapersonal change among patients with chronic non-cancer pain after using a pain diary: a mixed-methods study. J Pain Res 2019 Jan 23;12:477-487. [doi: 10.2147/JPR.S186105] [Medline: 30774417]
26. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008 Sep 30;139(1):146-157. [doi: 10.1016/j.pain.2008.03.024] [Medline: 18455312]

27. Giske L, Sandvik L, Røe C. Comparison of daily and weekly retrospectively reported pain intensity in patients with localized and generalized musculoskeletal pain. Eur J Pain 2010 Oct;14(9):959-965. [doi: 10.1016/j.ejpain.2010.02.011] [Medline: 20363653]
28. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. Pain 2004 Jan;107(1-2):61-69. [doi: 10.1016/j.pain.2003.09.020] [Medline: 14715390]
29. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231. [doi: 10.2196/11231] [Medline: 30348633]
30. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
31. Mano R. Mobile health apps and health management behaviors: cost-benefit modeling analysis. JMIR Hum Factors 2021 Apr 22;8(2):e21251. [doi: 10.2196/21251] [Medline: 33885372]
32. Lu C, Hu Y, Xie J, Fu Q, Leigh I, Governor S, et al. The use of mobile health applications to improve patient experience: cross-sectional study in Chinese public hospitals. JMIR Mhealth Uhealth 2018 May 23;6(5):e126. [doi: 10.2196/mhealth.9145] [Medline: 29792290]
33. Thompson D, Rattu S, Tower J, Egerton T, Francis J, Merolli M. Mobile app use to support therapeutic exercise for musculoskeletal pain conditions may help improve pain intensity and self-reported physical function: a systematic review. J Physiother 2023 Jan;69(1):23-34. [doi: 10.1016/j.jphys.2022.11.012] [Medline: 36528508]
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Coe-O'Brien R, Joseph L, Kuisma R, Paungmali A, Sitilertpisan P, Pirunsan U. Outcome measures used in the smartphone applications for the management of low back pain: a systematic scoping review. Health Inf Sci Syst 2020 Jan 2;8(1):5. [doi: 10.1007/s13755-019-0097-x] [Medline: 31938540]
36. Didyk C, Lewis LK, Lange B. Availability, content and quality of commercially available smartphone applications for the self-management of low back pain: a systematic assessment. Disabil Rehabil 2022 Dec;44(24):7600-7609. [doi: 10.1080/09638288.2021.1979664] [Medline: 34612133]
37. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 9;17(24):9209. [doi: 10.3390/ijerph17249209] [Medline: 33317134]
38. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
39. Agnew JMR, Nugent C, Hanratty CE, Martin E, Kerr DP, McVeigh JG. Rating the quality of smartphone apps related to shoulder pain: systematic search and evaluation using the Mobile App Rating Scale. JMIR Form Res 2022 May 26;6(5):e34339. [doi: 10.2196/34339] [Medline: 35617008]
40. Marques J, Borges L, Andias R, Silva AG. Characterisation and assessment of the most popular mobile apps designed for neck pain self-management: a systematic search in app stores. Musculoskeletal Care 2022 Mar;20(1):192-199. [doi: 10.1002/msc.1581] [Medline: 34383379]
41. Lu DJ, Girgis M, David JM, Chung EM, Atkins KM, Kamrava M. Evaluation of mobile health applications to track patient-reported outcomes for oncology patients: a systematic review. Adv Radiat Oncol 2021;6(1):100576. [doi: 10.1016/j.adro.2020.09.016] [Medline: 33073061]
42. Bhattarai P, Newton-John TRO, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. Pain Med 2018 Mar 1;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]
43. Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e18495. [doi: 10.2196/18495] [Medline: 32706727]
44. Gamwell KL, Kollin SR, Gibler RC, Bedree H, Bieniak KH, Jagpal A, et al. Systematic evaluation of commercially available pain management apps examining behavior change techniques. Pain 2021 Mar 1;162(3):856-865. [doi: 10.1097/j.pain.0000000000002090] [Medline: 33003110]
45. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106. [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
46. El-Tallawy SN, Nalamasu R, Salem GI, LeQuang JAK, Pergolizzi JV, Christo PJ. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. Pain Ther 2021 Jun;10(1):181-209. [doi: 10.1007/s40122-021-00235-2] [Medline: 33575952]
47. Giske L, Bautz-Holter E, Sandvik L, Røe C. Relationship between pain and neuropathic symptoms in chronic musculoskeletal pain. Pain Med 2009;10(5):910-917. [doi: 10.1111/j.1526-4637.2009.00622.x] [Medline: 19453955]

48. Henderson JV, Harrison CM, Britt HC, Bayram CF, Miller GC. Prevalence, causes, severity, impact, and management of chronic pain in Australian general practice patients. Pain Med 2013 Sep;14(9):1346-1361. [doi: 10.1111/pme.12195] [Medline: 23855874]
49. Schopflocher D, Taenzer P, Jovey R. The prevalence of chronic pain in Canada. Pain Res Manag 2011;16(6):445-450. [doi: 10.1155/2011/876306] [Medline: 22184555]
50. GitHub. Google Play Store scraper. URL: github.com/digitalmethodsinitiative/google-play-scraper [accessed 2022-02-01]
51. GitHub. iTunes App Store scraper. URL: github.com/digitalmethodsinitiative/itunes-app-scraper [accessed 2022-02-01]
52. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
53. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
54. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
55. Ford J, Hahne A, Surkitt L, Chan A, Richards M. The evolving case supporting individualised physiotherapy for low back pain. J Clin Med 2019 Aug 28;8(9):1334. [doi: 10.3390/jcm8091334] [Medline: 31466408]
56. O'Keeffe M, O'Sullivan P, Purtill H, Bargary N, O'Sullivan K. Cognitive functional therapy compared with a group-based exercise and education intervention for chronic low back pain: a multicentre randomised controlled trial (RCT). Br J Sports Med 2020 Jul;54(13):782-789. [doi: 10.1136/bjsports-2019-100780] [Medline: 31630089]
57. Andrews NE, Ireland D, Deen M, Varnfield M. Clinical utility of a mHealth assisted intervention for activity modulation in chronic pain: the pilot implementation of pain ROADMAP. Eur J Pain 2023 Jul;27(6):749-765. [doi: 10.1002/ejp.2104] [Medline: 36899447]
58. MacPherson M, Bakker AM, Anderson K, Holtzman S. Do pain management apps use evidence-based psychological components? A systematic review of app content and quality. Can J Pain 2022 Jun 3;6(1):33-44. [doi: 10.1080/24740527.2022.2030212] [Medline: 35694141]
59. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an app? Oncology providers' views on mobile health apps for cancer care. JCO Clin Cancer Inform 2017 Nov;1:1-7. [doi: 10.1200/CCI.17.00107] [Medline: 30657404]
60. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 3;7(6):e13199. [doi: 10.2196/13199] [Medline: 31199343]
61. Keogh A, Johnston W, Ashton M, Sett N, Mullan R, Donnelly S, et al. "It's Not as Simple as Just Looking at One Chart": a qualitative study exploring clinician's opinions on various visualisation strategies to represent longitudinal actigraphy data. Digit Biomark 2020 Nov 26;4(Suppl 1):87-99. [doi: 10.1159/000512044] [Medline: 33442583]
62. Wanderer JP, Nelson SE, Ehrenfeld JM, Monahan S, Park S. Clinical data visualization: the current state and future needs. J Med Syst 2016 Dec;40(12):275. [doi: 10.1007/s10916-016-0643-x] [Medline: 27787779]
63. Sullivan MJL, Wideman TH. Pain catastrophizing and fear of movement: detection and intervention. In: Lynch ME, Craig KD, Peng PWH, editors. Clinical Pain Management: A Practical Guide. Hoboken, NJ: Blackwell Publishing Ltd; 2010:207-214. [doi: 10.1002/9781444329711]
64. Ireland D, Andrews N. Pain ROADMAP: a mobile platform to support activity pacing for chronic pain. Stud Health Technol Inform 2019 Aug 8;266:89-94. [doi: 10.3233/SHTI190778] [Medline: 31397307]
65. Turk DC, Wilson HD. Fear of pain as a prognostic factor in chronic pain: conceptual models, assessment, and treatment implications. Curr Pain Headache Rep 2010 Apr;14(2):88-95. [doi: 10.1007/s11916-010-0094-x] [Medline: 20425197]
66. Malliaras P, Cook J, Purdam C, Rio E. Patellar tendinopathy: clinical diagnosis, load management, and advice for challenging case presentations. J Orthop Sports Phys Ther 2015 Nov;45(11):887-898. [doi: 10.2519/jospt.2015.5987] [Medline: 26390269]
67. Karduck J, Chapman-Novakofski K. Results of the Clinician Apps Survey, how clinicians working with patients with diabetes and obesity use mobile health apps. J Nutr Educ Behav 2018 Jan;50(1):62-69.e1. [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
68. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022 Nov 4;71(3):1-95. [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]
69. National Institute for Health and Care Excellence. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. 2021. URL: www.nice.org.uk/guidance/ng193 [accessed 2023-08-30]
70. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. 2019. URL: www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-pdf-1124017457605 [accessed 2023-07-10]
71. Australian Digital Health Agency. Assessment framework for mHealth apps. 2022. URL: www.digitalhealth.gov.au/sites/default/files/documents/adha-mhealth-assessment-framework-dec.pdf [accessed 2023-08-2]

Abbreviations

ICC: intraclass correlation
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by L Buis; submitted 01.03.23; peer-reviewed by C Didyk, H Huang, M Merolli, M MacPherson; revised version received 07.08.23; accepted 10.08.23; published 13.09.23.

Please cite as:
Simmich J, Ross MH, Andrews NE, Vaezipour A, Russell TG
Content and Quality of Mobile Apps for the Monitoring of Musculoskeletal or Neuropathic Pain in Australia: Systematic Evaluation
JMIR Mhealth Uhealth 2023;11:e46881
URL: https://mhealth.jmir.org/2023/1/e46881
doi:10.2196/46881

Original Paper

Mobile Health Apps for Breast Cancer: Content Analysis and Quality Assessment

Seongwoo Yang[1,2], MPH; Cam Nhung Bui[1], MD, PhD; Kyounghoon Park[1], MD, PhD

[1]HERINGS, The Institute of Advanced Clinical & Biomedical Research, Seoul, Republic of Korea

[2]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Kyounghoon Park, MD, PhD
HERINGS, The Institute of Advanced Clinical & Biomedical Research
14F, 560, Eonju-ro
Gangnam-gu
Seoul, 06144
Republic of Korea
Phone: 82 269493523
Fax: 82 269493517
Email: khpark@heringsglobal.com

Abstract

Background: The number of mobile health apps is rapidly increasing. This means that consumers are faced with a bewildering array of choices, and finding the benefit of such apps may be challenging. The significant international burden of breast cancer (BC) and the potential of mobile health apps to improve medical and public health practices mean that such apps will likely be important because of their functionalities in daily life. As the app market has grown exponentially, several review studies have scrutinized cancer- or BC-related apps. However, those reviews concentrated on the availability of the apps and relied on user ratings to decide on app quality. To minimize subjectivity in quality assessment, quantitative methods to assess BC-related apps are required.

Objective: The purpose of this study is to analyze the content and quality of BC-related apps to provide useful information for end users and clinicians.

Methods: Based on a stepwise systematic approach, we analyzed apps related to BC, including those related to prevention, detection, treatment, and survivor support. We used the keywords "breast cancer" in English and Korean to identify commercially available apps in the Google Play and App Store. The apps were then independently evaluated by 2 investigators to determine their eligibility for inclusion. The content and quality of the apps were analyzed using objective frameworks and the Mobile App Rating Scale (MARS), respectively.

Results: The initial search identified 1148 apps, 69 (6%) of which were included. Most BC-related apps provided information, and some recorded patient-generated health data, provided psychological support, and assisted with medication management. The Kendall coefficient of concordance between the raters was 0.91 (*P*<.001). The mean MARS score (range: 1-5) of the apps was 3.31 (SD 0.67; range: 1.94-4.53). Among the 5 individual dimensions, functionality had the highest mean score (4.37, SD 0.42) followed by aesthetics (3.74, SD 1.14). Apps that only provided information on BC prevention or management of its risk factors had lower MARS scores than those that recorded medical data or patient-generated health data. Apps that were developed >2 years ago, or by individuals, had significantly lower MARS scores compared to other apps (*P*<.001).

Conclusions: The quality of BC-related apps was generally acceptable according to the MARS, but the gaps between the highest- and lowest-rated apps were large. In addition, apps using personalized data were of higher quality than those merely giving related information, especially after treatment in the cancer care continuum. We also found that apps that had been updated within 1 year and developed by private companies had higher MARS scores. This may imply that there are criteria for end users and clinicians to help choose the right apps for better clinical outcomes.

(JMIR Mhealth Uhealth 2023;11:e43522) doi:10.2196/43522

KEYWORDS

app; breast cancer; quality assessment; mobile health; mHealth; digital health; digital health intervention; cancer management; tablet; prevention; survivor; peer-support

Introduction

With the increasing use of mobile apps since 2010, the number of mobile health–related (mHealth) apps has also risen [1]. According to a 2021 IQVIA report, there is a growing number of digital health care apps, with >350,000 apps related to health and fitness or medical categories available in the App Store and Google Play [2]. In 2020, a national survey in the United States found that more than half of all mobile phone users had downloaded health-related apps, and among them, two-thirds felt that such apps helped improve their health [3]. The key features of these health-related apps are maintaining a medication log, monitoring side effects, and scheduling follow-up appointments [4]. In addition, disease-specific apps may empower patients to promote self-efficacy, and self-care behavior in daily life, via information-technology services such as educational, patient-to-patient, and electronic patient-reported outcome services [5]. Therefore, maximizing the use of these advanced smartphone functions has paved the way for the delivery of diverse health care services [6]. In this respect, clinicians may want to employ useful apps to their patients and monitor their effect on patient outcomes [7]. However, to encourage health care providers to exploit them with confidence, there is a need for unbiased and scientific assessment of mHealth apps [6]. In addition, studies on mHealth have found that consumers are faced with a “bewildering array” of such apps because of the difficulty in discerning app quality, despite their beneficial functions [8].

Furthermore, since most public reviews of apps are based on individuals’ own subjective experience, more scientific and systematic assessment of mHealth apps is important [6]. Therefore, methods for the systematic search and analysis of apps have been developed [9]. Using these methods, several studies have evaluated apps concerning general cancer care [4,10], specific cancer types (eg, prostate cancer [11,12]), skin monitoring and melanoma detection [13,14], and medication compliance [15,16].

Even though breast cancer (BC) is one of the most prevalent cancers, especially in high-income countries, and is the leading cause of death in women in most low-income and many middle-income countries [17,18], morbidity and mortality can be reduced by promoting exercise, a healthy diet, adequate access to screening services, treatment, and care management [19,20]. However, a lack of information and support has prevented many women with BC from engaging in healthier behaviors during their cancer care [17,19,20]. Additionally, increased rates of early diagnosis and treatment [21] and a better prognosis and survival than other cancer types have resulted in unmet supportive-care needs for BC survivors [22,23]. In this sense, mHealth interventions can serve as promising platforms to enhance preventive and postdiagnosis behavior change with a clear goal setting and adherence to relevant theories [24]. Therefore, in terms of the international burden of BC and the potential of mHealth apps to improve medical and public health practices, the use of BC-related apps across the cancer care continuum (CCC) is important because of their functionalities in daily life [25].

According to a systematic review, in their nascent stage, BC-specific apps focused mainly on resources for BC awareness, screening, diagnosis, and treatment, so there is a lack of evidence on their utility, effectiveness, and safety [4]. Another systematic review analyzed all BC-related apps for their content and adherence to design standards outlined by the Institute of Medicine, as well as the relationships between their content, user ratings, and price [26]. The review found that mHealth apps have not met their potential for consumer engagement with evidence-based information, and BC-specific apps represented a limited spectrum on the cancer continuum [26]. Another systematic review that targeted BC survivorship and self-management pointed out that very few relevant resources were available in the apps considered [6], despite their utility in alleviating the burden and costs for BC survivors [27]. In sum, all these studies scrutinized cancer- or BC-related apps, but their focus was on mobile app availability, and they relied on user ratings to decide app quality. Therefore, to minimize subjectivity in quality assessment, quantitative methods to assess BC-related apps are required.

mHealth apps related to BC have different contents and features; therefore, the primary goal of this study was to perform a detailed evaluation of each element. In particular, patients with BC receive long-term care; they need proper information and direction outside the hospital setting through mobile apps. Therefore, the secondary goal of the study was to analyze the quality of BC-related apps to provide useful information for users and clinicians.

Methods

Overview of Mobile Apps

This stepwise systematic approach evaluated BC-related smartphone apps available on the Android and iOS platforms that had features related to cancer prevention, detection, treatment, and the provision of survivor support. We used the keywords “breast cancer” in English and Korean to identify commercially available apps in Google Play and the App Store, using accounts in both the United States and South Korea. The search was conducted on July 1, 2022, using the app search engine AppAgg, a mobile app metadata resource that was also used in a related study [28]. We recorded each app’s title, developer, final update, description, price, and website address.

Selection Criteria

This study included English and Korean BC-related apps for women who are at risk for BC across the life stages in the relevant app categories (health and fitness, medical, social, and lifestyle) that had been updated within the previous 3 years (July 2019-July 2022) and were available free of charge. We excluded apps that did not function correctly (eg, unreadable text or a blank screen), those that merely provided lists of conditions, and those intended for medical students that used self-made flashcards. In addition, we excluded apps that were developed with specific target users in mind (eg, those for health care professionals or children), to prompt a donation, or for trial recruitment. The eligible apps for Android and iOS were installed and alternately tested by each reviewer on a Samsung Galaxy S21 (Android version 11.0; Google LLC) and an iPhone 11 (iOS version 15.5; Apple Inc), respectively.

Content Analysis

The apps were independently evaluated by 2 investigators (SY and CNB). We recorded each app's title, platform, developer, category, date of latest update, language, and description. The content and functions of the selected apps were classified using the CCC, which has been used since the mid-1970s to describe the various stages of cancer in terms of etiology, prevention, detection, diagnosis, treatment, and survivorship [29,30]. Although the CCC categories are not discrete due to their oversimplified nature, they provide useful labels based on the development of cancer biology. We adopted the coding scheme proposed by Charbonneau et al [10], which redefined the following 7 categories of cancer apps identified by Bender et al [4]: educational, fundraising, prevention, early detection, disease and treatment information, disease management, and support. By integrating these concepts with the CCC stages, we created new categories that covered the app features identified in previous studies (Multimedia Appendix 1). We assumed that the functions and content of BC-related apps would fit into these categories.

Quality Assessment

The Mobile App Rating Scale (MARS) was used to evaluate the quality of the selected apps. This scale is a commonly used and validated tool that evaluates the following 5 dimensions of mobile apps: engagement, functionality, aesthetics, information, and subjective quality [31,32]. The apps are scored on a 5-point scale (1=inadequate; 2=poor; 3=acceptable; 4=good; and 5=excellent). Two blinded reviewers evaluated the apps separately without sharing detailed information; the Kendall coefficient of concordance was calculated to evaluate the agreement between them [33]. The apps were evaluated based on the MARS scores (total and dimension), and the mean scores of the 2 raters were calculated. We further analyzed differences in MARS scores by content based on the CCC and the number of years since the last update. We also classified the apps by the type of developer such as individual, commercial (including private companies or for-profit organizations), or public institutions (including nongovernmental organizations, hospitals, government agencies, or universities). For multiple comparisons of the numbers of years since last updated and developer types, analysis of variance and the Tukey honestly significant difference test were performed. Statistical analyses were performed using R software (version 4.1.0; R Foundation for Statistical Computing).

Ethical Considerations

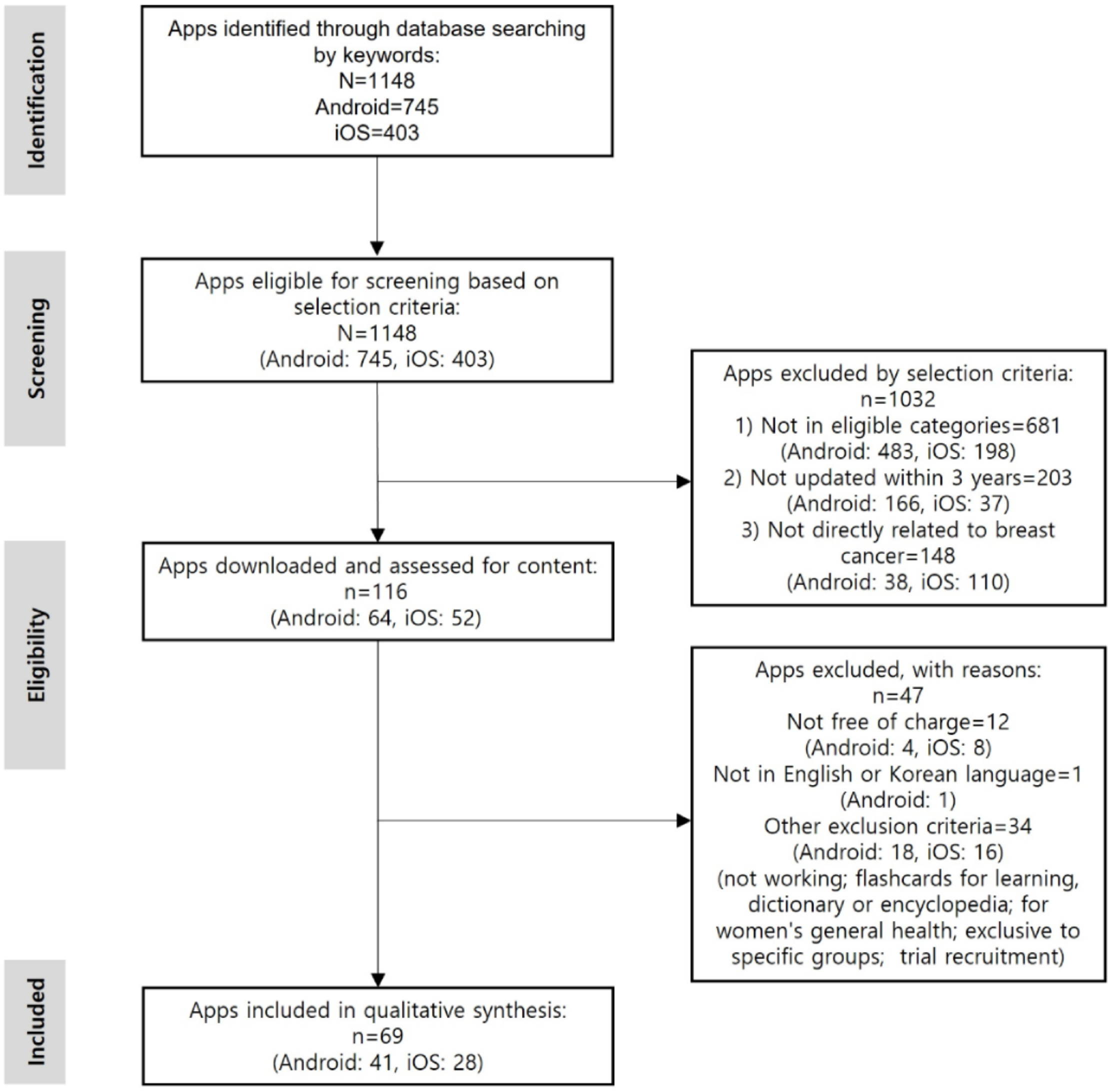
Since this study contains no primary data obtained from any experiment on human subjects, ethics approval was not required.

Results

App Identification

In total, 1148 apps identified through the database search were reviewed for eligibility for this study. After applying our exclusion screening criteria (eligible categories, updates, and relevance), 116 apps were downloaded and assessed in terms of content, price, language, and other criteria. Finally, 69 apps (n=41, 59% Android apps and n=28, 41% iOS apps) were included and subjected to content and quality assessment (see Figure 1 for the app-selection process and Multimedia Appendix 2 for the list of 69 BC-related apps included).

Figure 1. Flow diagram for the selection of breast cancer mobile health apps.



General Characteristics of the Included Apps

Of the 69 apps selected, 41 (59%) and 28 (41%) were available only on Android Google Play and the Apple App Store, respectively (Table 1), while 15 (22%) were available on both platforms. Most apps were included in the Health and Fitness category (n=48, 70%), had been updated within the previous year (n=46, 67%), were developed by a commercial organization (n=43, 62%), and were available in English (n=62, 70%).

Table 1. General characteristics of the 69 included apps.

Characteristics	Values, n (%)
Platform	
Android	41 (59.4)
iOS	28 (40.6)
Android and iOS	15 (21.7)
Category	
Health and fitness	48 (69.6)
Medicine	15 (21.7)
Lifestyle	2 (2.9)
Social networking	4 (5.8)
Updated within	
1 year	46 (66.7)
2 years	8 (11.6)
3 years	15 (21.7)
Developer	
Individual	8 (11.6)
Commercial organization[a]	43 (62.3)
Public institution[b]	18 (26.1)
Language	
English	62 (89.9)
Korean	2 (2.9)
English and Korean	5 (7.2)

[a]Such as private company or for-profit organization.

[b]Such as nongovernmental organization, hospital, government agency, or university.

Content Analysis

Table 2 shows the contents of the BC-related apps according to category and platform (Android and iOS). Their most common function was to provide information related to early detection of BC (n=34, 49%), followed by information concerning the risk factors and biological processes of BC at the prevention stage (n=28, 41%). The next most common content was information about symptoms, treatments, new advancements in BC treatment, and side effects related to BC after the treatment stage (n=27, 39%); education for lifestyle modification (n=23, 33%), and education for facts and knowledge related to BC (n=22, 32%). The apps also recorded patient-generated health data (PGHD), including tracking patients' and survivors' exercise, sleep, diet, and symptoms (n=19, 28%); provided psychological support (n=17, 25%); and assisted in medication management (n=16, 23%). However, content related to PGHD and medication management was more likely to be included in apps for iOS than in those for Android. The content of BC-related apps was also analyzed according to the number of years since the last update and the developer type (Multimedia Appendix 3). We found that 46 (67%) apps had been updated in the previous year, and 43 (62%) were developed by commercial organizations. Additionally, the apps that were outdated or had been developed by individuals mainly provided information or education regarding prevention or treatment. By contrast, those that had recently been updated and those that were developed by commercial organizations provided diverse content, including guidance for early detection, recording of PGHD, facilitation of medication management, and promotion of lifestyle modifications. Details of the apps' contents are provided in Multimedia Appendix 4.

Table 2. Content analyses of the 69 breast cancer (BC)–related mobile health apps.

Cancer control continuum and content	Android (n=41), n (%)	iOS (n=28), n (%)	Total (n=69), n (%)
Etiology and prevention			
Information on BC	19 (46.3)	9 (32.1)	28 (40.6)
Risk prediction	8 (19.5)	2 (7.1)	10 (14.5)
Education for prevention and risk factors for BC	16 (39.0)	6 (21.4)	22 (31.9)
Detection			
Guidance for early detection	21 (51.2)	13 (46.4)	34 (49.3)
Connection to professionals	5 (12.2)	5 (17.9)	10 (14.5)
Diagnosis and treatment			
Information on BC treatment	17 (41.5)	10 (35.7)	27 (39.1)
PGHD[a]	8 (19.5)	11 (39.3)	19 (27.5)
Medical records	4 (9.8)	2 (7.1)	6 (8.7)
Medication management	7 (17.1)	9 (32.1)	16 (23.2)
Consultation by a physician	3 (7.3)	1 (3.6)	4 (5.8)
Tracking appointments	4 (9.8)	5 (17.9)	9 (13.0)
Participation in decision-making	5 (12.2)	4 (14.3)	9 (13.0)
Survivorship			
Information on posttreatment care and prevention of recurrence	8 (19.5)	6 (21.4)	14 (20.3)
Education for lifestyle modification	15 (36.6)	7 (25.0)	23 (33.3)
Consultation with an expert	6 (14.6)	4 (14.3)	10 (14.5)
Psychological support	7 (17.1)	10 (35.7)	17 (24.6)
Community	7 (17.1)	6 (21.4)	13 (18.8)
Sharing information with family and caregivers	5 (12.2)	2 (7.1)	7 (10.1)
Fundraising	3 (7.3)	5 (17.9)	8 (11.6)

[a]PGHD: patient-generated health data.

Quality Assessment of the Apps

The 69 apps were evaluated by 2 raters using the MARS (Figure 2). The Kendall coefficient of concordance was 0.91 (*P*<.001), indicating a good agreement between the raters; disagreements between them were resolved by consensus. The mean total score (1-5) of the included apps was 3.31 (SD 0.67; range: 1.94-4.53). Among the MARS dimensions, functionality had the highest mean score (4.37, SD 0.42), followed by aesthetics (3.74, SD 1.14), information (3.53, SD 0.69), engagement (2.89, SD 0.92), and subjective quality (2.20, SD 0.79). The dimension scores of the apps varied widely, indicating large variations in app quality. The MARS scores of the included apps are presented in Multimedia Appendix 5. Of the 69 apps included, *OncoPower* (Android: 4.0, iOS: 4.2), *Outcomes4Me Breast Cancer Care* (Android: 4.4, iOS: 4.5), *OWise – Breast Cancer Support* (Android: 4.2, iOS: 4.1), and *War On Cancer* (Android: 4.0, iOS: 4.0) had the highest scores. Among the MARS dimensions, they scored especially highly in functionality and aesthetics (Multimedia Appendix 5).

The MARS scores were analyzed based on the number of years since the app was last updated and the developer type (Figure 3). Tukey test showed that the mean scores for apps that had been updated within the previous 1 (3.59, SD 0.53) or 2 (3.17, SD 0.68) years were not statistically different. However, outdated apps (ie, those that had been updated >2 years previously) had a significantly lower mean MARS score (2.54, SD 0.41) compared to the other apps (*P*<.001). Apps developed by individuals had a significantly lower mean MARS score (2.39, SD 0.41) than those developed by commercial organizations (3.43, SD 0.70) and public institutions (3.45, SD 0.28; *P*<.001).

Figure 2. App scores by the dimension of the Mobile App Rating Scale (MARS; n=69).

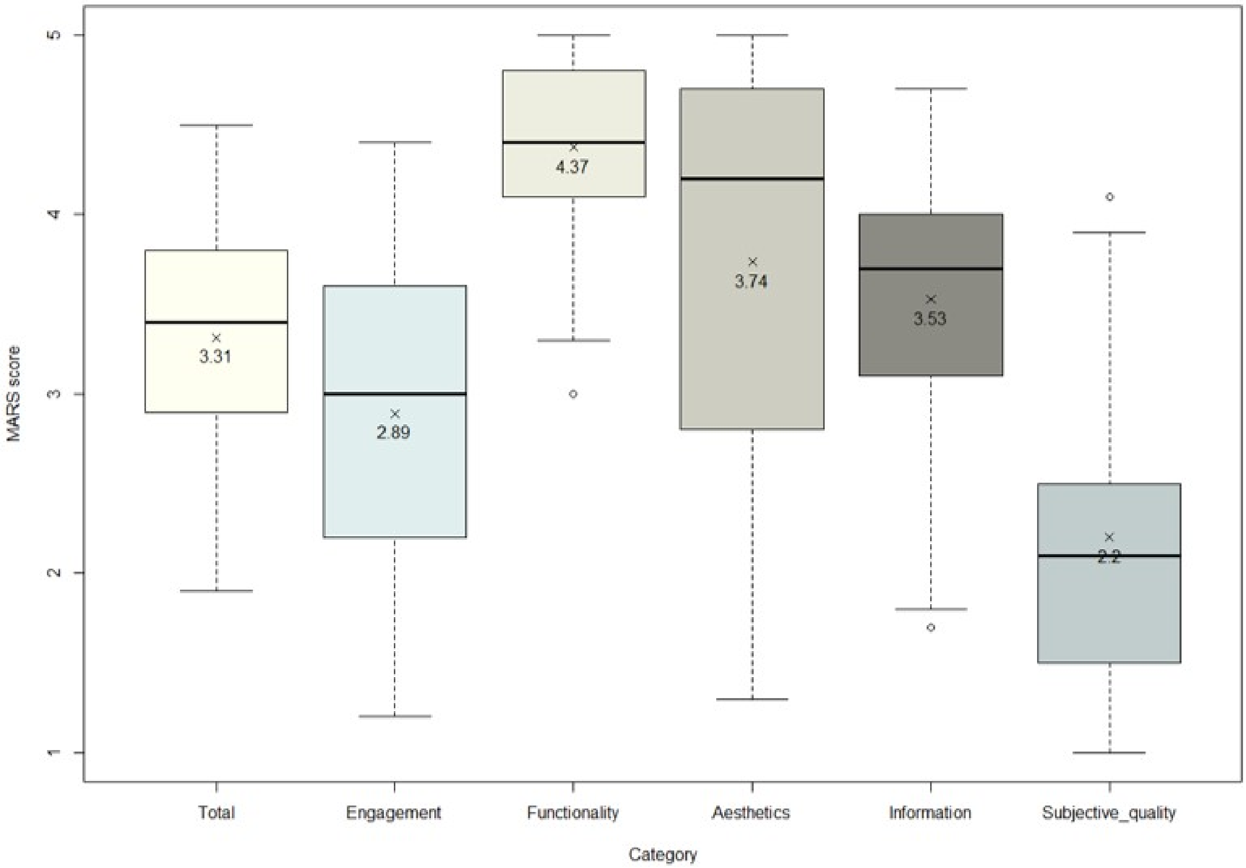
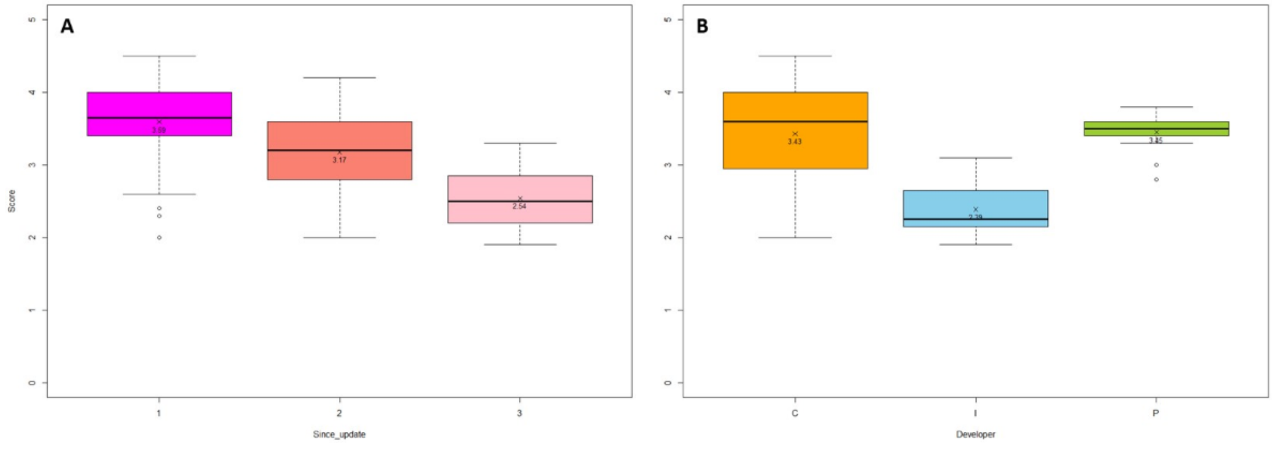


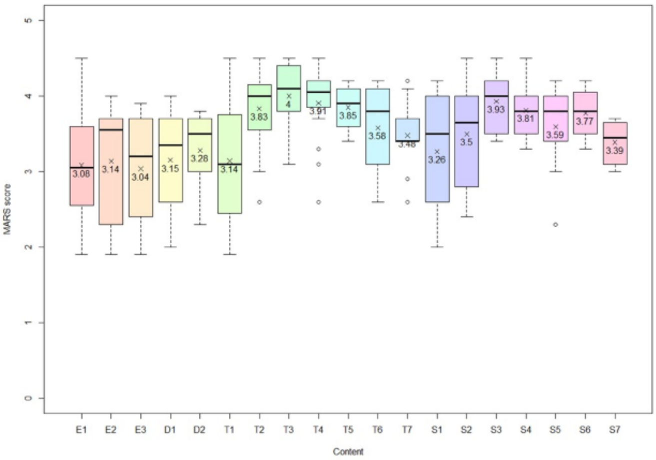
Figure 3. Mobile App Rating Scale (MARS) total scores by (A) years since apps were last updated and (B) type of developer. C: commercial; I: individual; P: public institution.



Quality Assessment of the Apps' Content

The quality of the content of the included apps was assessed using the MARS (Figure 4). Apps that tracked treatment records had the highest mean MARS score (T3; 4.0, SD 0.51), followed by those that facilitated consultations with experts (S3; 3.93, SD 0.38), medication management (T4; 3.91, SD 0.5), and consultations with physicians (T5; 3.85, SD 0.34). The mean MARS score was higher for apps that involved recording PGHD (T2; 3.83, SD 0.5), provided psychological support (S4; 3.81, SD 0.37), and shared information with family members or caregivers (S6; 3.77, SD 0.35) or the community (S5; 3.59, SD 0.53). Apps that only provided information on BC prevention or risk factors had significantly lower MARS scores than the other apps.

Figure 4. Content quality assessment (n=69). D: detection; E: etiology; MARS: Mobile App Rating Scale; S: survivorship; T: treatment.



Etiology and prevention	E1	Information on BC
	E2	Risk prediction
	E3	Education for prevention and risk factors for BC
Detection	D1	Guidance for early detection
	D2	Connection to professionals
Diagnosis and treatment	T1	Information on BC treatment
	T2	Patient-generated health data (PGHD)
	T3	Medical records
	T4	Medication management
	T5	Consultation by a physician
	T6	Tracking appointment
	T7	Participation in decision-making
Survivorship	S1	Information on posttreatment care and prevention of recurrence
	S2	Education for lifestyle modification
	S3	Consultation by an expert
	S4	Psychological support
	S5	Community
	S6	Sharing information with family and caregiver
	S7	Fundraising

Discussion

Principal Findings

We evaluated the contents and quality of BC-related mobile apps using the MARS. We organized the CCC using the results of previous studies to establish specific definitions for the content of mobile apps for women. Our results are based on 69 BC-related mHealth apps. The most frequent content of the apps was provision of information related to early detection of BC (n=34, 49%), followed by information regarding risk factors and biological processes of BC at the prevention stage (n=28, 41%). The next most frequent content was information about symptoms, treatments, new advancements in BC treatment, and side effects related to BC after the treatment stage (n=27, 39%). Education regarding lifestyle modification (n=23, 33%), prevention and risk factors of BC (n=22, 32%), and PGHD (n=19, 28%) were employed as content in BC-related apps. A total of 46 (68%) apps had been updated in the previous year, and 43 (62%) were developed by private companies.

The mean MARS score of the included apps was 3.31 out of 5, which was higher than the "acceptable" MARS score of 3 [31]. However, there was a significant gap between the highest- and lowest-rated apps (score range: 1.94-4.53). Functionality had the highest mean score (4.37, SD 0.42) among the MARS dimensions, and outdated apps (mean 2.54, SD 0.41) and apps developed by individuals (mean 2.39, SD 0.41) had significantly lower mean MARS scores. This is in agreement with the findings of a previous review study that found the functionality and usability of apps increased over a 2-year period, but content credibility did not [34]. Therefore, users might need to use their discretion if using outdated apps.

Of the 69 mHealth apps included, *OncoPower, Outcomes4Me Breast Cancer Care*, *OWise – Breast Cancer Support*, and *War On Cancer* had the highest scores. These apps are available on both Google Play and the App Store, and had high scores for aesthetics, functionality, and engagement. Although *OncoPower* and *War On Cancer* do not target patients with BC and survivors exclusively, they have multiple functions that support patients with other cancer types, such as those related to nutrition and meditation. On the other hand, *Outcomes4Me Breast Cancer Care* and *Owise – Breast Cancer Support* were developed specifically for patients with BC, so these provide evidence-based treatment options and personalized resources based on individuals' medical records to promote communication with their health care providers. The *War On Cancer* app was developed to promote social networking by patients with cancer and survivors, which may be useful given the importance of social support in cancer care [35-37]. The 4 highest-ranked apps focused mainly on providing personalized care after a diagnosis of BC and on treatment, a time at which patients may struggle due to physical and psychological impairments. Therefore, providing appropriate information on treatment options and lifestyle modification, facilitating medication management, providing psychological support, and tracking PGHD and medical records may ease the burden on patients.

Our results show approximately one-third (n=19, 28%) of the included apps are using individual data, such as PGHD. PGHD, including patient-reported outcomes, provide clinically relevant information obtained outside traditional care settings, and could be useful to improve outcomes and enhance patient-provider communication [38]. Some studies found that effective physician-patient communication improved patient health outcomes [39], as well as BC patients' depression and quality of life [40]. With the increasing use of wearable devices and advancements in technology, the use of PGHD and established medical screening and surveillance strategies may enhance long-term cancer survivorship at the individual and population levels. Furthermore, these approaches can strengthen the survivor-provider relationship [41]. Despite the benefits of PGHD, there are several barriers to their use, including a lack of technical support in patients' primary language, the reluctance of clinicians to review PGHD, a lack of access to broadband internet, and concerns related to the confidentiality of personal information [42]. Nevertheless, mobile apps that use medical records and PGHD had the highest MARS scores (4.0 and 3.83, respectively), indicating the willingness of users to use individualized health-related apps.

The large differences in MARS scores among the apps may imply a need to improve the standards used for their approval, and for quality checks at all stages of development (assessment, prototype, content, and evaluation). Some studies strived to evaluate health-related apps by establishing a practical framework based on guidelines by the US Food and Drug Administration and the UK National Health Service [43], or by organizing published studies [44]. This framework was developed by organizing app evaluation questions from 45 previous systematic reviews and verified by the patient advisory panel. It represents the pyramid shape that begins at the bottom from background information, privacy, and security, to evidence based, ease of use, and data integration. If these kinds of evaluations were implemented effectively at the time of app approval, higher-quality apps might be released. Most BC-related apps included in this study provided information or education regarding prevention and survivorship, in line with previous studies [4,26]. However, the mean MARS score of those that provided information at any CCC stage was slightly above 3, which means acceptable but not good enough. Provision of information is important but insufficient to modify multifaceted health behaviors [45]. Additionally, patients with limited health literacy are at a distinct disadvantage during BC treatment in terms of unmet information needs [46]. Moreover, information and educational content that are not based on guidelines or evidence discourage women from consistently using mobile BC apps; this may explain the lower MARS scores on those apps that merely provided information. Furthermore, the MARS scores of apps that targeted patients after a diagnosis of BC, including those that provided survivor support, were higher than those reported previously [26]. Patients who are unconcerned regarding their health while being investigated for cancer may change their attitude toward cancer after their diagnosis [47]. Therefore, such people might search for and use the helpful tools provided by mobile apps.

Strengths and Limitations

The strengths of our study include using the MARS, an objective tool used to measure app quality. Star ratings and user reviews are also valuable for developers and potential new users because they offer a crowdsourced indicator of the effectiveness and popularity of apps [48]. However, these indicators do not accurately reflect app quality [49]. The majority of public reviews of apps rely on the personal opinions of individuals' own experience and are thus highly subjective. Therefore, we did not record star ratings in this study. Additionally, a previous study found that MARS scores did not significantly correlate with users' star ratings [50]. We could not confirm this in our study because of a lack of star ratings and reviews. Therefore, we employed the MARS to assess the quality of the most credible apps. Furthermore, we used these quantitative results to identify the relationship between MARS dimensions and apps' content.

This study had several limitations. First, paid and inaccessible apps could not be downloaded because of a lack of funding and access. Such apps should be evaluated in future studies for external validity. Second, although the MARS has been widely used and validated previously, we cannot rule out subjectivity due to the nature of evaluations. However, to minimize subjectivity, we confirmed high interrater reliability between the independent raters. Lastly, qualitative measures may be needed to reflect end users' experience.

Recommendations for Future Development

The growing numbers of BC-related mHealth apps and of studies on their usefulness allow women to select those most likely to improve their quality of life. We identified certain issues that could be addressed to improve app quality. First, the quality assessment system for the digital platform needs to be improved. Our results identified limitations in certain MARS dimensions, indicating poor app quality. Therefore, a new method should be introduced to validate the MARS that reflects differences between specific diseases and users' experience [32]. Second, evidence-based content and functions are required. Although the mechanisms underlying the effects of mHealth apps are not known, efforts should be made toward elucidating them by developing a theory-based intervention that is administered via an mHealth app and tested in a clinical trial. Although some feasibility studies and trials have been conducted on the usefulness of PGHD [51,52], further studies of the clinical benefits of mHealth apps are required. The development of a set of core outcomes may be another option to measure their effectiveness and induce behavior change, as has been found by others [53]. Additionally, patient-centered considerations of the design and interface before the development of an app might be helpful to promote behavior change. One study suggested iterative development through a user-centered design approach involving the following 3 phases: analysis, design, and implementation to achieve less fragmented care [54]. This systematic approach could encourage patients, survivors, and health care providers to participate in the development and quality assessment of mobile apps. In sum, the mapping of app content against current BC guidelines and creating adequate evidence would help clinicians have more information about useful content and promote them to recommend using related mobile apps.

Conclusions

We systematically analyzed 69 BC-related mHealth apps, using literature-based content categories and the MARS for quality assessment. Generally, the quality of the apps was acceptable according to the MARS, but the gap between the highest- and lowest-rated apps was significant. Our findings indicate that the BC-related apps using personalized data were of higher quality than those that merely provided women with information on BC, especially after treatment in the CCC. We also found that apps that had been updated within 1 year and those developed by private companies had higher MARS scores. These findings provide specific criteria for women and clinicians to help them choose the right mobile BC apps for better clinical outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Content categories for mobile apps related to cancer management.
[DOCX File , 19 KB - mhealth_v11i1e43522_app1.docx]

Multimedia Appendix 2
Full list of included breast cancer apps (n=69).
[DOCX File , 23 KB - mhealth_v11i1e43522_app2.docx]

Multimedia Appendix 3
Content analyses of the 69 breast cancer mobile apps by update year and developer.
[DOCX File , 21 KB - mhealth_v11i1e43522_app3.docx]

Multimedia Appendix 4
Details on the content of the included apps (n=69).
[DOCX File , 44 KB - mhealth_v11i1e43522_app4.docx]

Multimedia Appendix 5
Details on the results of quality assessment of the included apps (n=69).
[DOCX File , 31 KB - mhealth_v11i1e43522_app5.docx]

References

1. Xu W, Liu Y. mHealthApps: A Repository and Database of Mobile Health Apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
2. Aitken M. Digital Health Trends. Danbury, CT, USA: IQVIA Institue; 2021.
3. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
4. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
5. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Empowerment of Cancer Survivors Through Information Technology: An Integrative Review. J Med Internet Res 2015 Nov 27;17(11):e270 [FREE Full text] [doi: 10.2196/jmir.4818] [Medline: 26614438]
6. Kapoor A, Nambisan P, Baker E. Mobile applications for breast cancer survivorship and self-management: A systematic review. Health Informatics J 2020 Dec 25;26(4):2892-2905 [FREE Full text] [doi: 10.1177/1460458220950853] [Medline: 32842830]
7. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
8. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a User-Centered Design, Self-Management App to Existing mHealth Apps for Persons Living With HIV. JMIR Mhealth Uhealth 2015 Sep 18;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
9. Grundy QH, Wang Z, Bero LA. Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
10. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: A content analysis of the digital health marketplace. Digit Health 2020 Feb 11;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]
11. Pereira-Azevedo NM, Venderbos LDF. eHealth and mHealth in prostate cancer detection and active surveillance. Transl Androl Urol 2018 Feb;7(1):170-181 [FREE Full text] [doi: 10.21037/tau.2017.12.22] [Medline: 29594031]
12. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile Phone Apps for Quality of Life and Well-Being Assessment in Breast and Prostate Cancer Patients: Systematic Review. JMIR Mhealth Uhealth 2017 Dec 04;5(12):e187 [FREE Full text] [doi: 10.2196/mhealth.8741] [Medline: 29203459]
13. Chao E, Meenan CK, Ferris LK. Smartphone-Based Applications for Skin Monitoring and Melanoma Detection. Dermatol Clin 2017 Oct;35(4):551-557. [doi: 10.1016/j.det.2017.06.014] [Medline: 28886812]
14. Kassianos A, Emery J, Murchie P, Walter F. Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review. Br J Dermatol 2015 Jun 06;172(6):1507-1518. [doi: 10.1111/bjd.13665] [Medline: 25600815]
15. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]

16. Jupp JCY, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer 2018 Nov 26;65(11):e27278. [doi: 10.1002/pbc.27278] [Medline: 29943893]
17. Ginsburg O, Bray F, Coleman MP, Vanderpuye V, Eniu A, Kotha SR, et al. The global burden of women's cancers: a grand challenge in global health. The Lancet 2017 Feb;389(10071):847-860. [doi: 10.1016/s0140-6736(16)31392-7]
18. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
19. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin 2017 Nov 03;67(6):439-448 [FREE Full text] [doi: 10.3322/caac.21412] [Medline: 28972651]
20. Torre L, Siegel R, Ward E, Jemal A. Global Cancer Incidence and Mortality Rates and Trends--An Update. Cancer Epidemiol Biomarkers Prev 2016 Jan;25(1):16-27. [doi: 10.1158/1055-9965.EPI-15-0578] [Medline: 26667886]
21. Ginsburg O, Yip C, Brooks A, Cabanes A, Caleffi M, Dunstan Yataco JA, et al. Breast cancer early detection: A phased approach to implementation. Cancer 2020 May 15;126 Suppl 10(Suppl 10):2379-2393 [FREE Full text] [doi: 10.1002/cncr.32887] [Medline: 32348566]
22. Cheng K, Darshini Devi R, Wong W, Koh C. Perceived symptoms and the supportive care needs of breast cancer survivors six months to five years post-treatment period. Eur J Oncol Nurs 2014 Feb;18(1):3-9. [doi: 10.1016/j.ejon.2013.10.005] [Medline: 24315730]
23. Harrison JD, Young JM, Price MA, Butow PN, Solomon MJ. What are the unmet supportive care needs of people with cancer? A systematic review. Support Care Cancer 2009 Aug 25;17(8):1117-1128. [doi: 10.1007/s00520-009-0615-5] [Medline: 19319577]
24. Kalke K, Ginossar T, Bentley JM, Carver H, Shah SFA, Kinney AY. Use of Evidence-Based Best Practices and Behavior Change Techniques in Breast Cancer Apps: Systematic Analysis. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e14082 [FREE Full text] [doi: 10.2196/14082] [Medline: 32012084]
25. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2023-02-16]
26. Ginossar T, Shah SFA, West AJ, Bentley JM, Caburnay CA, Kreuter MW, et al. Content, Usability, and Utilization of Plain Language in Breast Cancer Mobile Phone Apps: A Systematic Analysis. JMIR Mhealth Uhealth 2017 Mar 13;5(3):e20 [FREE Full text] [doi: 10.2196/mhealth.7073] [Medline: 28288954]
27. Earle CC. Failing to Plan Is Planning to Fail: Improving the Quality of Care With Survivorship Care Plans. JCO 2006 Nov 10;24(32):5112-5116. [doi: 10.1200/jco.2006.06.5284]
28. Odigie E, Andreadis K, Chandra I, Mocchetti V, Rives H, Cox S, et al. Are Mobile Applications in Laryngology Designed for All Patients? Laryngoscope 2022 Nov 01:Online ahead of print. [doi: 10.1002/lary.30465] [Medline: 36317789]
29. Rowland JH, Bellizzi KM. Cancer survivors and survivorship research: a reflection on today's successes and tomorrow's challenges. Hematol Oncol Clin North Am 2008 Apr;22(2):181-200, v. [doi: 10.1016/j.hoc.2008.01.008] [Medline: 18395144]
30. Cancer Control Continuum. National Cancer Institute. 2021. URL: https://cancercontrol.cancer.gov/about-dccps/about-cc/cancer-control-continuum [accessed 2022-11-06]
31. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
32. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
33. Salehinejad S, Niakan Kalhori SR, Hajesmaeel Gohari S, Bahaadinbeigy K, Fatehi F. A review and content analysis of national apps for COVID-19 management using Mobile Application Rating Scale (MARS). Inform Health Soc Care 2021 Mar 02;46(1):42-55. [doi: 10.1080/17538157.2020.1837838] [Medline: 33164594]
34. Musgrave LM, Kizirian NV, Homer CSE, Gordon A. Mobile Phone Apps in Australia for Improving Pregnancy Outcomes: Systematic Search on App Stores. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e22340 [FREE Full text] [doi: 10.2196/22340] [Medline: 33196454]
35. Krishnasamy M. Social support and the patient with cancer: a consideration of the literature. J Adv Nurs 1996 Apr;23(4):757-762. [doi: 10.1111/j.1365-2648.1996.tb00048.x] [Medline: 8675894]
36. Taylor SE, Falke RL, Shoptaw SJ, Lichtman RR. Social support, support groups, and the cancer patient. Journal of Consulting and Clinical Psychology 1986;54(5):608-615. [doi: 10.1037/0022-006x.54.5.608]
37. Usta YY. Importance of social support in cancer patients. Asian Pac J Cancer Prev 2012 Aug 31;13(8):3569-3572 [FREE Full text] [doi: 10.7314/apjcp.2012.13.8.3569] [Medline: 23098436]
38. Shapiro M, Johnston D, Wald J, Mon D. Patient-generated health data. RTI International. 2012. URL: https://www.rti.org/publication/patient-generated-health-data-white-paper/fulltext.pdf [accessed 2023-02-16]

39. Stewart M. Effective physician-patient communication and health outcomes: a review. CMAJ 1995 May 01;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
40. Vogel BA, Leonhart R, Helmes AW. Communication matters: the impact of communication and participation in decision making on breast cancer patients' depression and quality of life. Patient Educ Couns 2009 Dec;77(3):391-397. [doi: 10.1016/j.pec.2009.09.005] [Medline: 19796910]
41. Petersen C. Patient-generated health data: a pathway to enhanced long-term cancer survivorship. J Am Med Inform Assoc 2016 May;23(3):456-461. [doi: 10.1093/jamia/ocv184] [Medline: 26714765]
42. Omoloja A, Vundavalli S. Patient generated health data: Benefits and challenges. Curr Probl Pediatr Adolesc Health Care 2021 Nov;51(11):101103. [doi: 10.1016/j.cppeds.2021.101103] [Medline: 34799255]
43. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
44. Llorens-Vernet P, Miró J. Standards for Mobile Health-Related Apps: Systematic Review and Development of a Guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
45. Ng TWH, Feldman DC. Age and innovation-related behavior: The joint moderating effects of supervisor undermining and proactive personality. J. Organiz. Behav 2012 May 23;34(5):583-606. [doi: 10.1002/job.1802]
46. Halbach SM, Ernstmann N, Kowalski C, Pfaff H, Pförtner TK, Wesselmann S, et al. Unmet information needs and limited health literacy in newly diagnosed breast cancer patients over the course of cancer treatment. Patient Educ Couns 2016 Sep;99(9):1511-1518. [doi: 10.1016/j.pec.2016.06.028] [Medline: 27378079]
47. Leydon G, Bynoe-Sutherland J, Coleman M. The journey towards a cancer diagnosis: the experiences of people with cancer, their family and carers. Eur J Cancer Care (Engl) 2003 Dec;12(4):317-326. [doi: 10.1046/j.1365-2354.2003.00418.x] [Medline: 14982310]
48. Schumer H, Amadi C, Joshi A. Evaluating the Dietary and Nutritional Apps in the Google Play Store. Healthc Inform Res 2018 Jan;24(1):38-45 [FREE Full text] [doi: 10.4258/hir.2018.24.1.38] [Medline: 29503751]
49. Choi J, Chung C, Woo H. Diet-Related Mobile Apps to Promote Healthy Eating and Proper Nutrition: A Content Analysis and Quality Assessment. Int J Environ Res Public Health 2021 Mar 28;18(7):3496 [FREE Full text] [doi: 10.3390/ijerph18073496] [Medline: 33800531]
50. Rodrigues AT, Sousa CT, Pereira J, Figueiredo IV, Lima TDM. Mobile Applications (Apps) to Support the Hepatitis C Treatment: A Systematic Search in App Stores. Ther Innov Regul Sci 2021 Jan 22;55(1):152-162. [doi: 10.1007/s43441-020-00201-8] [Medline: 32700147]
51. Mangold KN, Gallicchio L, Gaysynsky A, Kent EE, Constantine K, Hesse BW. Feasibility of collecting patient-generated health data to enhance cancer registry surveillance. J Cancer Surviv 2021 Oct 05;15(5):785-791. [doi: 10.1007/s11764-020-00970-z] [Medline: 33400151]
52. Miyaji T, Kawaguchi T, Azuma K, Suzuki S, Sano Y, Akatsu M, et al. Patient-generated health data collection using a wearable activity tracker in cancer patients-a feasibility study. Support Care Cancer 2020 Dec 12;28(12):5953-5961. [doi: 10.1007/s00520-020-05395-z] [Medline: 32281031]
53. Musgrave LM, Baum A, Perera N, Homer CS, Gordon A. Baby Buddy App for Breastfeeding and Behavior Change: Retrospective Study of the App Using the Behavior Change Wheel. JMIR Mhealth Uhealth 2021 Apr 15;9(4):e25668 [FREE Full text] [doi: 10.2196/25668] [Medline: 33856351]
54. Kim M, Kim Y, Choi M. Mobile health platform based on user-centered design to promote exercise for patients with peripheral artery disease. BMC Med Inform Decis Mak 2022 Aug 02;22(1):206 [FREE Full text] [doi: 10.1186/s12911-022-01945-z] [Medline: 35918702]

Abbreviations

BC: breast cancer
CCC: cancer care continuum
MARS: Mobile App Rating Scale
mHealth: mobile health
PGHD: patient-generated health data

Edited by L Buis, G Eysenbach; submitted 14.10.22; peer-reviewed by L Musgrave, Y Lee; comments to author 14.11.22; revised version received 11.12.22; accepted 10.02.23; published 23.02.23.

Please cite as:
Yang S, Bui CN, Park K
Mobile Health Apps for Breast Cancer: Content Analysis and Quality Assessment
JMIR Mhealth Uhealth 2023;11:e43522
URL: https://mhealth.jmir.org/2023/1/e43522
doi:10.2196/43522
PMID:36821352

Original Paper

Menstrual Tracking Mobile App Review by Consumers and Health Care Providers: Quality Evaluations Study

Siyeon Ko[1], BSc; Jisan Lee[2,3], PhD; Doyeon An[1], MPH; Hyekyung Woo[1,4], PhD

[1]Department of Health Administration, College of Nursing & Health, Kongju National University, Gongju-Si, Chungcheongnam-do, Republic of Korea
[2]Department of Nursing Science, College of Life & Health Sciences, Hoseo University, Asan-si, Chungcheongnam-do, Republic of Korea
[3]Department of Nursing, Gangneung-Wonju National University, Wonju-si, Republic of Korea
[4]Institute of Health and Environment, Kongju National University, Gongju-si, Chungcheongnam-do, Republic of Korea

Corresponding Author:
Hyekyung Woo, PhD
Department of Health Administration
College of Nursing & Health
Kongju National University
56 Gongjudaehak-ro
Gongju-Si, Chungcheongnam-do
Republic of Korea
Phone: 82 10 3350 3486
Email: hkwoo@kongju.ac.kr

Abstract

Background: Women's menstrual cycle is an important component of their overall health. Physiological cycles and associated symptoms can be monitored continuously and used as indicators in various fields. Menstrual apps are accessible and can be used to promote overall female health. However, no study has evaluated these apps' functionality from both consumers' and health care providers' perspectives. As such, the evidence indicating whether the menstrual apps available on the market provide user satisfaction is insufficient.

Objective: This study was performed to investigate the key content and quality of menstrual apps from the perspectives of health care providers and consumers. We also analyzed the correlations between health care provider and consumer evaluation scores. On the basis of this analysis, we offer technical and policy recommendations that could increase the usability and convenience of future app.

Methods: We searched the Google Play Store and iOS App Store using the keywords "period" and "menstrual cycle" in English and Korean and identified relevant apps. An app that met the following inclusion criteria was selected as a research app: nonduplicate; with >10,000 reviews; last updated ≤180 days ago; relevant to this topic; written in Korean or English; available free of charge; and currently operational. App quality was evaluated by 6 consumers and 4 health care providers using Mobile Application Rating Scale (MARS) and user version of the Mobile Application Rating Scale (uMARS). We then analyzed the correlations among MARS scores, uMARS scores, star ratings, and the number of reviews.

Results: Of the 34 apps, 31 (91%) apps could be used to predict the menstrual cycle, and 2 (6%) apps provided information pertinent to health screening. All apps that scored highly in the MARS evaluation offer a symptom logging function and provide the user with personalized notifications. The "Bom Calendar" app had the highest MARS (4.51) and uMARS (4.23) scores. The MARS (2.22) and uMARS (4.15) scores for the "Menstrual calendar—ovulation & pregnancy calendar" app were different. In addition, there was no relationship between MARS and uMARS scores (r=0.32; P=.06).

Conclusions: We compared consumer and health care provider ratings for menstrual apps. Continuous monitoring of app quality from consumer and health care provider perspectives is necessary to guide their development and update content.

(JMIR Mhealth Uhealth 2023;11:e40921) doi:10.2196/40921

KEYWORDS

mobile app; period; menstrual cycle; mHealth; mobile health; evaluation; women's health; health care provider; consumer; menstrual app; digital health app; health screening; consumer satisfaction

Introduction

Background

Women's menstrual cycles are important to their overall health [1,2] and are characterized by predictable and recurring symptoms. Continuous tracking of the menstrual cycle can aid in health management. Monitoring systems are needed to optimize menstrual health and provide easily accessible health information for women [3].

Mobile health (mHealth) apps facilitate personalized health monitoring and management [4], and menstrual apps are the most important mHealth apps for women. Such apps are typically highly accessible for most women and provide indicators relevant to various health domains [5].

Typically, menstrual apps also allow the user to log their symptoms, mood changes, and body temperature, and they visually represent statistical data via graphs and tables. Some apps offer professional-level information through communities and links, that is, they promote women's health care by facilitating smooth communication with medical staff through information-sharing services [6]. The apps' menstrual cycle tracking functions facilitate health care planning and management in various domains, including contraception, fertility, preparation with respect to pregnancy and ensuring adequate "menstrual supplies," leisure activities, and travel [7].

Women use these functions for various purposes, such as tracking pregnancy, preventing pregnancy, and managing menstruation periods [8]. The functions desired by consumers depend on the intended purpose of the app. Currently, indicators to help consumers identify and select menstrual apps according to the desired functions are lacking [9]. Recently developed systems recommend apps based on consumer requirements [10,11]. Menstrual apps with various functions have been developed, and related research is being actively conducted. However, most of the existing studies are content reviews or expert evaluations [12-15]. Consumer-centered studies have also started to appear [5,16-18], but quality evaluations of menstrual apps remain scarce.

Menstrual apps are directly relevant to women's health, so it is necessary for experts and health care providers to evaluate these apps [8]. Health care provider quality evaluations can contribute to the app development, which is important to ensure that consumers have access to high-quality apps [19]. Consumer quality evaluations provide feedback on apps, such as consumer preferences (eg, for easy-to-use content) [20,21]. Consistent use of an app is critical given the recurring nature of the menstrual cycle, but research on this topic is lacking from the developer's standpoint. To promote sustained app use, evaluations from consumers and health care providers' perspective are required. However, it is unclear whether current commercially available apps satisfy the quality standards of consumers and health care providers. If we find the differences between consumers and health care providers, they need to be discussed importantly to develop apps that can be used to promote women's health that satisfy both consumers' wants and health care providers' needs.

Purpose

This study examines currently available menstrual apps in term of their main contents and quality based on evaluations by health care providers and consumers. We also correlated the two sets of evaluation scores. The study's findings could improve the utility and convenience of future mHealth apps.

Methods

App Selection

We searched for keywords related to the development and evaluation of menstrual apps commonly included in previous studies, such as "period" and "menstrual cycle" in both English and Korean. The Google Play Store and the iOS App Store were searched from April 8 to April 15, 2021. Up to 150 apps were screened for each keyword. Since then, to secure the appropriated number of apps that can be statistically analyzed, the following app inclusion criteria have been set, based on previous studies [14,22,23]:

- The app is nonduplicate.
- It has more than 10,000 reviews.
- It has been last updated ≤180 days ago.
- It is relevant to this topic.
- The app language is Korean or English.
- The app is free to use.
- It is currently in operation.

Analysis of App Contents

To examine the apps' main contents, we performed a pilot study of 14 representative apps. The apps' main contents were classified as follows: menstrual cycle management, education and knowledge, sharing information, and notifications. Frequency analysis was performed to determine the number of apps providing these functions.

Evaluation of App Quality

The quality of 34 menstrual apps were evaluated using Mobile Application Rating Scale (MARS) and user version of the Mobile Application Rating Scale (uMARS). The MARS was developed to evaluate mobile apps; it is a reliable tool with a high internal consistency and interrater reliability [24]. The uMARS was subsequently developed, based on the MARS, to allow consumers to evaluate the quality of mHealth apps. uMARS has excellent internal consistency [25]. Both scales comprise the following five categories: engagement, aesthetics, functionality, information, and subjective quality (Table 1). The MARS has been used to evaluate various health care–related apps, such as apps related to chronic disease, COVID-19, and physical activity, as well as apps related to allergy, hepatitis treatment support, and breast cancer [14,15,19,22,23,26,27]. The uMARS has recently been used to evaluate various types of mHealth apps, including apps pertaining to weight loss and nutrition, rheumatic diseases, and the management of ankylosing spondylitis [28-31].

Table 1. Mobile Application Rating Scale (MARS) and user version of the Mobile Application Rating Scale (uMARS) evaluation items.

Measures	Characteristics
Objective measures	
Engagement	• Entertainment[a,b] • Interest[a,b] • Customization[a,b] • Interactivity[a,b] • Target group[a,b]
Functionality	• Performance[a,b] • Ease of use[a,b] • Navigation[a,b] • Gestural design[a,b]
Information	• Accuracy of app design[a] • Goal[a] • Credibility[a] • Evidence bases[a] • Quality of information[a,b] • Quantity of information[a,b] • Visual information[a,b] • Credibility of source[b]
Aesthetics	• Layout[a,b] • Graphics[a,b] • Visual appeal[a,b]
Subjective measures	
Subjective quality	• Recommendation[a,b] • Frequency of use[a,b] • Payment for expenses[a,b] • Star rating[a,b]

[a]MARS evaluation items.

[b]uMARS evaluation items.

A total of 6 consumers completed the uMARS between July 9 and July 22, 2021, and 4 nurses (ie, health care providers) majoring in health care and working in medical centers completed the MARS between July 21 and July 30, 2021. Each evaluator was asked to use the app for more than 10 minutes every day and the evaluation was conducted in a blind test. The apps were randomly assigned to evaluators to prevent bias related to subjectivity. Each app was crossevaluated by at least two evaluators.

Comparative Analysis of Consumer and Health Care Provider Data

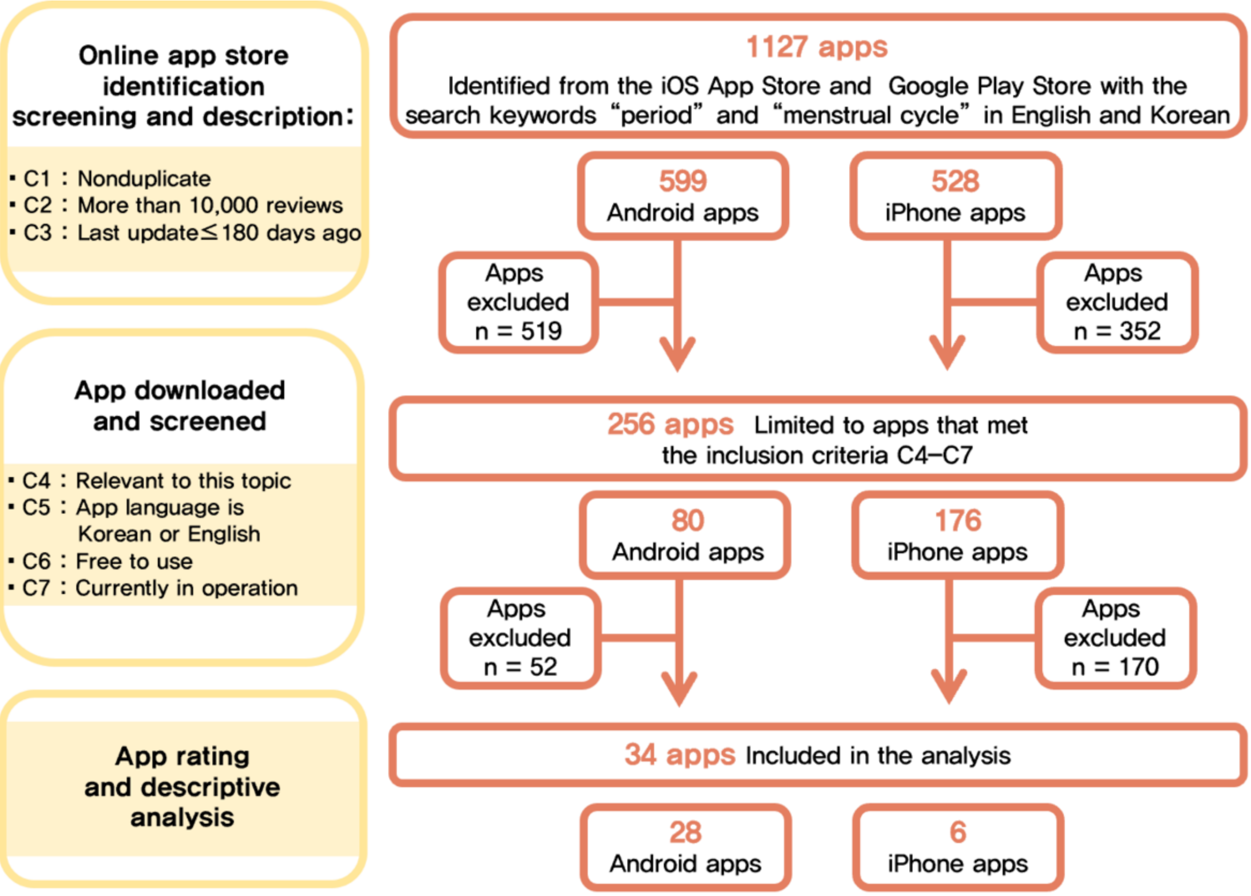
The MARS results represent health care providers' perspectives, as stated above. The uMARS results, star ratings, and number of reviews were analyzed from the consumers' perspective. After normalization, Pearson correlation was used to correlate app content, MARS and uMARS scores, star ratings, and reviews and to correlate perspectives of health care providers and consumers. We identified the top and bottom five apps based on the MARS and uMARS scores and compared app preferences between consumers and health care providers. *P* values <.05 were considered significant. R software (version 4.1.2; R Core Team,) was used for the analysis.

Results

App Selection

A total of 1127 menstrual apps were initially identified via the keyword search, and 34 apps (Android: n=28; iPhone: n=6) met all of the study criteria and were included in the final analysis (Figure 1).

Figure 1. Flow diagram of the app review and selection process. C1-7 denotes Criterion 1-7.



The operating system (OS) of apps is linked with the app store, so that the app is updated simultaneously with updates of the OS [32]. Therefore, the app version and function may vary depending on the timing of the OS update. The number of reviews and star ratings differed among the apps in this study. For example, Bom Calendar had star ratings of 4.8 and 4.4 for the Android and iOS versions, respectively (23,437 and 76,258 reviews, respectively). In instances where the same apps were available for different OSs, each version was considered to be a unique app in the analysis.

Analysis of App Contents

Most apps (n=31, 91%) offered a menstrual cycle prediction function. Some apps (n=14, 41%) offered menstruation and fertility period notifications, while others had no specific functions (n=3, 9%). Most apps were confidential (n=29, 85%), allowed data export (n=28, 82%), and had a log-in function (n=25, 74%). However, few apps provided education or knowledge (n=10, 29%), screening-related information (n=2, 6%), or advice (n=4, 12%; Table 2).

Table 2. App contents analysis result.

Contents	App, n (%)		
	Android (n=28)	iPhone (n=6)	Total (N=34)
Menstrual cycle management			
Symptoms (pain)	4 (14)	0 (0)	4 (12)
Additional symptom	18 (64)	6 (100)	24 (71)
Ovulation management			
Calculate pregnancy probability	4 (14)	0 (0)	4 (12)
Contraception methods	3 (11)	0 (0)	3 (9)
Both	12 (43)	6 (100)	18 (53)
Last update (months)			
<3	21 (75)	5 (83)	26 (76)
3~6	7 (25)	1 (17)	8 (24)
Function s			
Graphical chart	17 (61)	4 (67)	21 (62)
Lock	23 (82)	6 (100)	29 (85)
Advice provision	2 (7)	2 (33)	4 (12)
Data export	23 (82)	5 (83)	28 (82)
Predictions	25 (89)	6 (100)	31 (91)
Log-in	20 (71)	5 (83)	25 (74)
Education or knowledge			
General health information	5 (18)	1 (17)	6 (18)
Personalized information	1 (4)	2 (33)	3 (9)
Both	1 (4)	0 (0)	1 (3)
Health screening	1 (4)	1 (17)	2 (6)
Sharing information (with health care professionals)			
All information	4 (14)	3 (50)	7 (21)
Only information specified by the consumer	1 (4)	0 (0)	1 (3)
Visualization			
Menstruation or ovulation	2 (7)	0 (0)	2 (6)
Menstrual cycle	9 (32)	0 (0)	9 (26)
All data	16 (57)	6 (100)	22 (65)
Notifications			
Menstruation or fertility	2 (7)	0 (0)	2 (56)
Both	14 (50)	0 (0)	14 (41)
Personalized alarms	9 (32)	6 (100)	15 (44)
Other features			
Community	4 (14)	2 (33)	6 (18)
Shopping	2 (7)	1 (17)	3 (9)

Evaluation of App Quality

The MARS and uMARS scores of all apps were obtained through the average of the evaluator's evaluation scores. The average MARS and uMARS scores (ie, health care provider and consumer scores, respectively) were 3.06 (SD 0.62) and 3.33 (SD 0.57), respectively. The iPhone "Bom Calendar" app had the highest score for both the MARS (4.51, SD 0.22) and uMARS (4.23, SD 0.27). The Android "Period calendar—Women's menstrual calendar❤" app had the second lowest score for both the MARS (2.05, SD 0.45) and uMARS (2.09, SD 0.05), despite its high star rating of 4.8. The Android

"Menstrual calendar—ovulation & pregnancy calendar" app showed contrasting scores between the MARS (2.22, SD 0.07) and uMARS (4.15, SD 0.46); it had the third highest overall uMARS score, with an engagement score of 3.86 (SD 0.49), functionality score of 4.04 (SD 0.49), aesthetics score of 4.25 (SD 0.47), information score of 4.25 (SD 0.47), and subjective quality score of 4.37 (SD 0.41). However, for the MARS, its overall score was the third lowest, with an engagement score of 2.40 (SD 0.20), functionality score of 2.50 (SD 0.50), aesthetics score of 2.00 (SD 0.00), information score of 2.58 (SD 0.42), and subjective quality score of 1.63 (SD 0.38; Table 3 and Table 4). MARS and uMARS scores of all 34 apps are presented in Multimedia Appendix 1.

Table 3. The five highest- and lowest-scoring apps (N=34) based on the user version of the Mobile Application Rating Scale (uMARS). All the values are mean (SD).

App	Engagement	Functionality	Aesthetics	Information	Subjective quality	uMARS[a]
Top five[b]						
	3.87 (0.33)	4.17 (0.26)	4.33 (0.32)	4.27 (0.29)	4.51 (0.22)	4.23 (0.27)
	4.03 (0.38)	4.16 (0.44)	4.23 (0.35)	4.31 (0.27)	4.40 (0.25)	4.23 (0.33)
	3.86 (0.49)	4.04 (0.49)	4.25 (0.47)	4.25 (0.47)	4.37 (0.41)	4.15 (0.46)
	3.80 (1.01)	3.97 (0.78)	4.30 (0.51)	4.30 (0.51)	4.29 (0.49)	4.13 (0.66)
	3.83 (0.60)	4.06 (0.43)	4.16 (0.30)	4.11 (0.35)	4.39 (0.36)	4.11 (0.38)
Bottom five[c]						
	1.91 (0.34)	1.96 (0.39)	1.91 (0.54)	1.94 (0.75)	1.90 (0.72)	1.93 (0.54)
	1.76 (0.11)	2.04 (0.13)	2.17 (0.07)	2.13 (0.10)	2.33 (0.10)	2.09 (0.05)
	1.93 (0.38)	2.11 (0.53)	2.46 (0.47)	2.53 (0.37)	2.46 (0.51)	2.30 (0.44)
	2.30 (0.30)	2.50 (0.50)	2.70 (0.30)	2.74 (0.16)	2.90 (0.10)	2.63 (0.23)
	2.24 (1.05)	2.43 (1.24)	3.06 (1.07)	3.19 (1.19)	3.33 (1.33)	2.85 (1.17)

[a]Average of uMARS scores evaluated by consumers.

[b]Top five apps: (1) Bom Calendar (iOS), (2) Pregnancy planning & management (Android), (3) Menstrual calendar—ovulation & pregnancy calendar (Android), (4) Flo (Android), and (5) Flo (iOS).

[c]Bottom five apps: (1) Maya—My Period Tracker (Android); (2) Period calendar—Women's menstrual calendar❤ (Android); (3) Women's menstrual calendar, menstrual & ovulation day calculator, childbearing age, pregnancy planning (Android); (4) My Days—Ovulation Calendar & period Tracker (Android); and (5) Clover: Period & Cycle Tracker (Android).

Table 4. The five highest- and lowest-scoring apps (N=34) based on the Mobile Application Rating Scale (MARS).

App and operating system	Engagement	Functionality	Aesthetics	Information	Subjective quality	MARS[a]
Top five[b]						
	4.18 (0.63)	4.75 (0.25)	5.00 (0.00)	4.29 (0.11)	4.36 (0.25)	4.51 (0.14)
	4.10 (0.10)	4.06 (0.21)	4.67 (0.33)	4.92 (0.08)	3.93 (0.24)	4.34 (0.15)
	3.85 (0.46)	4.13 (0.63)	4.42 (0.43)	4.00 (0.33)	3.58 (0.58)	3.99 (0.48)
	3.65 (0.75)	4.19 (0.57)	4.00 (1.00)	4.42 (0.58)	3.68 (0.07)	3.99 (0.59)
	4.00 (0.00)	3.56 (0.37)	4.83 (0.17)	4.33 (0.35)	3.02 (0.61)	3.95 (0.29)
Bottom five[c]						
	2.05 (0.65)	2.13 (0.63)	1.50 (0.50)	2.38 (0.71)	1.53 (0.53)	1.92 (0.60)
	2.10 (0.71)	2.81 (0.48)	2.08 (0.60)	2.25 (0.25)	1.01 (0.26)	2.05 (0.45)
	2.40 (0.20)	2.50 (0.50)	2.00 (0.00)	2.58 (0.42)	1.63 (0.38)	2.22 (0.07)
	2.00 (0.20)	2.94 (0.57)	2.42 (0.43)	2.63 (0.14)	1.34 (0.09)	2.26 (0.15)
	2.00 (0.00)	2.75 (0.25)	2.33 (0.33)	2.75 (0.08)	1.81 (0.56)	2.33 (0.01)

[a]Average of MARS scores evaluated by health care providers.

[b]Top five apps: (1) Bom Calendar (iOS), (2) Pink Diary (Android), (3) Bom Calendar (Android), (4) Clue Period, Ovulation Tracker (iOS), and (5) Femometer—Fertility Tracker (Android).

[c]Bottom five apps: (1) My Days—Ovulation Calendar & period Tracker (Android), (2) Period calendar—Women's menstrual calendar❤ (Android), (3) Menstrual calendar—ovulation & pregnancy calendar (Android), (4) My Menstrual Diary (Android), and (5) Period Tracker (Android).

Further Comparative Analysis of Consumer and Health Care Provider Data

Table 5 shows the results of correlation analysis of the MARS and uMARS scores, star ratings, number of reviews, and app content. The number of reviews was not correlated with app content and menstrual cycle management (r=0.53; P=.001), and visualization (r=0.51; P=.002) had the highest correlation with star ratings. Among the evaluation scores, the highest correlation was found between uMARS and notification (r=0.39; P=.02) as well as between MARS and ovulation date management (r=0.49; P=.003).

Multimedia Appendix 2 shows content comparison between the top and bottom five apps in consumer and health care provider evaluation scores. Personalized alarms could be set in the top five apps. In addition, they provided a function to visualize all information through a calendar or to specify and manage ovulation days. On the contrary, in the bottom five apps did not have functions for managing or predicting the menstrual cycle.

Figure 2 shows the correlations among the MARS and uMARS scores, star ratings, and number of reviews to compare the perspective of health care providers and consumers. Figure 2 shows no correlation between MARS and uMARS scores of the health care providers and consumers (r=0.32; P=.06). uMARS scores and star rating (r=0.11; P=.54) as well as uMRAS scores and number of reviews (r=0.07; P=.67) also showed no significant correlations. The number of reviews and star rating (r=0.39; P=.02) showed a very low correlation.

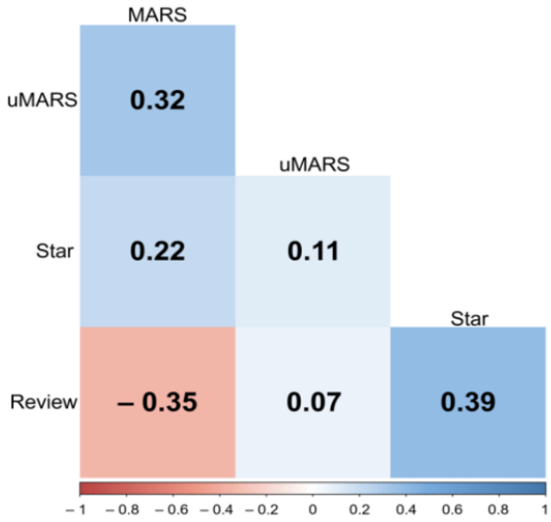
Table 5. Results of correlation analysis of the Mobile Application Rating Scale (MARS) and the user version of the Mobile Application Rating Scale (uMARS) scores, star rating, number of reviews, and app content types.

App contents	Health care provider and consumer perspective			
	MARS	uMARS	Star rating	Number of reviews
Menstrual cycle management	0.39[a]	0.18	0.53[b]	0.21
Ovulation date management	0.49[b]	0.33	0.41[a]	0.16
Functions				
Graphic chart	–0.06	–0.10	–0.54[b]	–0.31
Lock	–0.37[a]	–0.40[a]	–0.32	–0.10
Advice provision	–0.05	–0.31	–0.25	–0.01
Data export	–0.22	–0.20	–0.53[b]	–0.10
Prediction	–0.43[a]	–0.17	–0.79[b]	–0.22
Log-in	–0.12	–0.01	–0.47[b]	–0.13
Educational or knowledge	0.23	0.04	–0.02	–0.14
Health screening	0.25	0.13	0.12	0.01
Sharing information	0.16	0.14	0.04	–0.13
Visualization	0.32	0.22	0.51[b]	0.19
Notification	0.32	0.39[a]	0.24	0.06
Other features				
Community	–0.14	–0.19	–0.16	–0.10
Shopping	–0.35[a]	–0.05	–0.14	0.12

[a] $P<.05$.

[b] $P<.01$.

Figure 2. Result of the correlation analysis of the Mobile Application Rating Scale (MARS) and user version of the Mobile Application Rating Scale (uMARS) scores, star ratings, and number of reviews.



Discussion

Principal Findings and Comparison With Prior Work

Interest in mHealth apps has recently increased, and new apps are continuously being developed, including menstrual apps [33]. However, the needs of consumers and health care providers are different, and studies evaluating whether the available menstrual apps satisfy both of these groups are difficult to find. This study obtained quality evaluations of relevant apps from consumers and health care providers using the uMARS and MARS, respectively; the consistency of the evaluations of these two groups in terms of key app contents was analyzed. The health care providers valued engagement, functionality, and aesthetics when evaluating apps, while consumers valued aesthetics and information provision the most.

The MARS and uMARS scores were not correlated in this study. For example, the app with the third lowest MARS score had the third highest uMARS score. A significant difference was observed in app aesthetics scores between the consumers and health care providers; the scores for this attribute showed the largest group difference. Previous studies on health services have reported disparities between consumers and health care providers, and these results affect the implementation of consumer-centered services. Data from health care providers can provide a basis for high-quality apps [20], while consumer data serves as feedback on app quality [19]. To ensure high app quality and consumer satisfaction, app quality should be continuously monitored from the perspective of both health care providers and consumers. Monitoring can identify the needs of consumers and health care providers, which can in turn help in app development and update [10]. This study focused on evaluating app quality using MARS and uMARS for consistency, but could be extended to qualitative studies, including interviews, to collect in-depth answers in the future [34].

According to our findings, the uMARS (ie, consumer) scores, star ratings, and number of reviews were unrelated variables. The uMARS allows for direct assessment of mHealth apps and is a reliable measure of app quality [25]. However, reviews and star ratings are subjective indicators [24]. The currently available mHealth apps have been evaluated in a simplistic manner, such as through star ratings and reviews, even though they differ significantly with respect to content; thus, appropriate guidelines to aid app selection are lacking. By using the uMARS to guide app selection, the limitations of reviews and star ratings can be overcome such that consumers will likely select more useful apps to meet their particular needs. However, it is difficult for consumers to evaluate the quality of the apps using uMARS every time they download one. New indicators to guide app selection are needed so that consumers can make decisions based on objective evaluation results.

Most of the five apps that scored highly in the quality evaluation in this study included personalized monitoring functions. Menstrual apps are becoming increasingly popular, and apps that include self-monitoring functions and provide related information are continuously being developed [28]. Personalized monitoring can improve user well-being by encouraging them to check for signs and symptoms of health issues. Health-related information can also promote consumer health. mHealth apps that include personalized content of this nature are particularly useful for consumers [23,35]. However, only a few of the apps evaluated in this study facilitated consultations with specialists or provided information relating to women's health. To increase the utility of apps, notifications, symptom recording functions, and the provision of knowledge should be prioritized.

mHealth apps should provide customized content for individual consumers [36]. However, the five bottom-scoring apps in this study did not meet the needs of the women who were using them. Consumers use apps to predict menstrual cycles and ovulation dates, and to monitor their general health [37]. However, most of the five bottom-scoring apps did not provide content enhancing consumer convenience, such as functions for menstrual cycle and ovulation day management, and some apps also lacked predictive functions. Such apps must be updated to include content allowing for the prediction and management of menstrual cycles based on accumulated menstrual cycle- and health-related information.

Menstrual apps collects personal information from consumers, such as name, date of birth, menstrual cycle, and medical history [38]. Personal data must be protected because it is sensitive information [12,39], but some apps do not provide locking functions, and few apps provide icon change functions for protecting personal information. Most fitness apps that record the number of steps do not consider privacy issues, and the data protection of mHealth apps related to women's health is typically poor [39]. Therefore, regulations pertaining to app management of private data are necessary [40,41]. In fact, there are existing regulations protecting personal information, such as the European Unions's General Data Protection Regulation, but no standard regulations are enforced worldwide. mHealth apps that protect personal information tend to be favored by consumers [21]. App developers should improve data protection–related functions to protect the personal information of consumers.

The MARS and uMARS were developed specifically for evaluating mHealth apps that aim to improve consumers' health. Meanwhile, menstrual apps were designed to help consumers keep track of their current health rather than improve it. Health apps must provide solutions customized to individual consumers [36]; reliability may be key in this respect [42]. This study's results indicate that current health apps do not fully meet consumers' requirements or desires with respect to content. To better identify consumer objectives and take account of them during the development of menstrual apps, a new evaluation scale is needed to evaluate menstrual apps.

Limitations

This study had several limitations. First, the results cannot be generalized to all menstrual apps because a small number of evaluators evaluated only the most popular apps. Second, the apps were selected based on App Store searches with a limited timeframe. Updates to apps may result in differences between the analyzed content and that in the future. Third, the database is not an electronic database but the App Store. The App Store's app recommendation function may have compromised an inconsistent search accuracy.

Conclusions

In this study, consumer and health care provider ratings of menstrual apps were obtained using validated scales. Consumer preferred app had high scores of aesthetics and information, and evaluation scores differed between consumers and health care providers. The findings highlight the importance of consumer participation in menstrual app development and evaluation. This study is significant in that it is the first to compare health care providers' and consumers' menstrual app quality ratings. We expect our results to guide future mHealth app development and provide consumers with information on menstrual app content and quality. To provide high-quality apps for consumers, continuous quality evaluation research needs to be conducted, and the perspectives of both consumers and health care providers should be taken into account.

Acknowledgments

This work was supported by a 2022 research grant from Kongju National University (KNU) and the National Research Foundation of Korea (NRF) funded by the Korean government (NRF-2020R1C1C1009679). This study was presented at the 2021 Postacademic Conference of the Korean Academy of Health Policy and Management.

JL was affiliated with the Department of Nursing Science, College of Life & Health Sciences at Hoseo University at the time of the study and is currently affiliated with the Department of Nursing at Gangneung-Wonju National University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile Application Rating Scale (MARS) and user version of the Mobile Application Rating Scale (uMARS) score of all 34 apps.
[XLSX File (Microsoft Excel File), 14 KB - mhealth_v11i1e40921_app1.xlsx]

Multimedia Appendix 2
Content comparison between the top and bottom five apps (ranked according to the consumer and health care provider evaluation scores).
[PDF File (Adobe PDF File), 74 KB - mhealth_v11i1e40921_app2.pdf]

References

1. Bae J, Park S, Kwon J. Factors associated with menstrual cycle irregularity and menopause. BMC Womens Health 2018 Feb 06;18(1):36 [FREE Full text] [doi: 10.1186/s12905-018-0528-x] [Medline: 29409520]
2. Critchley HO, Babayev E, Bulun SE, Clark S, Garcia-Grau I, Gregersen PK, et al. Menstruation: science and society. AJOG 2020 Nov;223(5):624-664. [doi: 10.1016/j.ajog.2020.06.004]
3. Park H. Health behavior and health promotion in women. In: Health and Welfare Policy Forum. South Korea: Korea Institute for Health and Social Affairs; Sep 01, 2021:2-4.
4. Derbyshire E, Dancey D. Smartphone medical applications for women's health: what is the evidence-base and feedback? Int J Telemed Appl 2013;2013:782074 [FREE Full text] [doi: 10.1155/2013/782074] [Medline: 24454354]
5. Moglia ML, Nguyen HV, Chyjek K, Chen KT, Castaño PM. Evaluation of smartphone menstrual cycle tracking applications using an adapted APPLICATIONS scoring system. Obstet Gynecol 2016 Jun;127(6):1153-1160. [doi: 10.1097/AOG.0000000000001444] [Medline: 27159760]
6. Levy J, Romo-Avilés N. "A good little tool to get to know yourself a bit better": a qualitative study on users' experiences of app-supported menstrual tracking in Europe. BMC Public Health 2019 Sep 03;19(1):1213 [FREE Full text] [doi: 10.1186/s12889-019-7549-8] [Medline: 31481043]
7. Epstein DA, Lee NB, Kang JH, Agapie E, Schroeder J, Pina LR, et al. Examining menstrual tracking to inform the design of personal informatics tools. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May 02;2017:6876-6888 [FREE Full text] [doi: 10.1145/3025453.3025635] [Medline: 28516176]
8. Earle S, Marston HR, Hadley R, Banks D. Use of menstruation and fertility app trackers: a scoping review of the evidence. BMJ Sex Reprod Health 2021 Apr 06;47(2):90-101. [doi: 10.1136/bmjsrh-2019-200488] [Medline: 32253280]
9. DeNicola N, Marko K. Connected health and mobile apps in obstetrics and gynecology. Obstet Gynecol Clin North Am 2020 Jun;47(2):317-331. [doi: 10.1016/j.ogc.2020.02.008] [Medline: 32451020]
10. Lee J, Kim J. Can menstrual health apps selected based on users' needs change health-related factors? A double-blind randomized controlled trial. J Am Med Inform Assoc 2019 Jul 01;26(7):655-666 [FREE Full text] [doi: 10.1093/jamia/ocz019] [Medline: 30946478]
11. Lee J, Kim J. Method of app selection for healthcare providers based on consumer needs. Comput Inform Nurs 2018 Jan;36(1):45-54. [doi: 10.1097/CIN.0000000000000399] [Medline: 29084030]
12. Dantas LO, Carvalho C, Pena C, Breda CC, Driusso P, Ferreira CHJ, et al. Mobile health technologies for the monitoring of menstrual cycle: a systematic review of online stores in Brazil. J Obstet Gynaecol Res 2022 Jan 27;48(1):5-14. [doi: 10.1111/jog.15082] [Medline: 34706123]
13. Starič KD, Trajkovik V, Belani H, Vitagliano A, Bukovec P. Smart phone applications for self-monitoring of the menstrual cycle: a review and content analysis. Clin Exp Obstet Gynecol 2019;46(5):731-735. [doi: 10.12891/ceog4830.2019]

14. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
15. Rodrigues AT, Sousa CT, Pereira J, Figueiredo IV, Lima TDM. Mobile applications (Apps) to support the hepatitis c treatment: a systematic search in app stores. Ther Innov Regul Sci 2021 Jan;55(1):152-162. [doi: 10.1007/s43441-020-00201-8] [Medline: 32700147]
16. Gambier-Ross K, McLernon DJ, Morgan HM. A mixed methods exploratory study of women's relationships with and uses of fertility tracking apps. Digit Health 2018;4:2055207618785077 [FREE Full text] [doi: 10.1177/2055207618785077] [Medline: 30225095]
17. Starling MS, Kandel Z, Haile L, Simmons RG. User profile and preferences in fertility apps for preventing pregnancy: an exploratory pilot study. Mhealth 2018;4:21 [FREE Full text] [doi: 10.21037/mhealth.2018.06.02] [Medline: 30050917]
18. Saparamadu AADNS, Fernando P, Zeng P, Teo H, Goh A, Lee JMY, et al. User-centered design process of an mhealth app for health professionals: case study. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e18079 [FREE Full text] [doi: 10.2196/18079] [Medline: 33769297]
19. Narrillos-Moraza, Gómez-Martínez-Sagrera P, Amor-García, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2022 Feb 16;10(2):e32826 [FREE Full text] [doi: 10.2196/32826] [Medline: 35171109]
20. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
21. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
22. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
23. Karasneh RA, Al-Azzam SI, Alzoubi KH, Muflih SM, Hawamdeh SS. Smartphone applications for period tracking: rating and behavioral change among women users. Obstet Gynecol Int 2020 Sep 01;2020:2192387-2192389 [FREE Full text] [doi: 10.1155/2020/2192387] [Medline: 32952563]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
26. Lee JB, Woo H. Quality evaluation of smartphone applications for physical activity promotion. KOSHEP_EN 2019 Oct 01;36(4):67-76. [doi: 10.14367/kjhep.2019.36.4.67]
27. Wright A. Evaluation of two mobile health apps for patients with breast cancer using the Mobile Application Rating Scale. Mhealth 2021;7:60 [FREE Full text] [doi: 10.21037/mhealth-20-161] [Medline: 34805391]
28. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836 [FREE Full text] [doi: 10.2196/mhealth.9836] [Medline: 30672742]
29. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the china app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261 [FREE Full text] [doi: 10.2196/13261] [Medline: 31364606]
30. Lambrecht A, Vuillerme N, Raab C, Simon D, Messner E, Hagen M, et al. Quality of a supporting mobile app for rheumatic patients: patient-based assessment using the user version of the Mobile Application Scale (uMARS). Front Med (Lausanne) 2021;8:715345 [FREE Full text] [doi: 10.3389/fmed.2021.715345] [Medline: 34368202]
31. Song Y, Chen H. Evaluating chinese mobile health apps for ankylosing spondylitis management: systematic app search. JMIR Mhealth Uhealth 2021 Jul 14;9(7):e27234 [FREE Full text] [doi: 10.2196/27234] [Medline: 34259644]
32. Novac O, Novac M, Gordan C, Berczes T, Bujdosó G. Comparative study of Google Android, Apple iOS and Microsoft Windows Phone mobile operating systems. 2017 Presented at: 2017 14th International Conference on Engineering of Modern Electric Systems (EMES); June 1-2; Oradea, Romania p. 154-159. [doi: 10.1109/emes.2017.7980403]
33. Symul L, Wac K, Hillard P, Salathé M. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. NPJ Digit Med 2019;2:64 [FREE Full text] [doi: 10.1038/s41746-019-0139-4] [Medline: 31341953]
34. Namey E, Guest G, McKenna K, Chen M. Evaluating Bang for the Buck. AJE 2016 Jul 09;37(3):425-440. [doi: 10.1177/1098214016630406]
35. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

36. Anderson K, Emmerton LM. Contribution of mobile health applications to self-management by consumers: review of published evidence. Aust Health Rev 2016 Nov;40(5):591-597. [doi: 10.1071/AH15162] [Medline: 26681206]
37. Adnan T, Coull BA, Jukic AM, Mahalingaiah S. The real-world applications of the symptom tracking functionality available to menstrual health tracking apps. Curr Opin Endocrinol Diabetes Obes 2021 Dec 01;28(6):574-586 [FREE Full text] [doi: 10.1097/MED.0000000000000682] [Medline: 34560714]
38. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
39. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mhealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735]
40. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
41. Costa Figueiredo M, Huynh T, Takei A, Epstein DA, Chen Y. Goals, life events, and transitions: examining fertility apps for holistic health tracking. JAMIA Open 2021 Jan;4(1):ooab013 [FREE Full text] [doi: 10.1093/jamiaopen/ooab013] [Medline: 33718804]
42. Pichon A, Jackman K, Winkler I, Bobel C, Elhadad N. The messiness of the menstruator: assessing personas and functionalities of menstrual tracking apps. J Am Med Inform Assoc 2022 Jan 12;29(2):385-399 [FREE Full text] [doi: 10.1093/jamia/ocab212] [Medline: 34613388]

Abbreviations

MARS: Mobile Application Rating Scale
mHealth: mobile health
OS: operating system
uMARS: user version of the Mobile Application Rating Scale

Edited by G Eysenbach, L Buis; submitted 09.07.22; peer-reviewed by M Bardus, V Buss; comments to author 29.07.22; revised version received 31.10.22; accepted 20.12.22; published 01.03.23.

Please cite as:
Ko S, Lee J, An D, Woo H
Menstrual Tracking Mobile App Review by Consumers and Health Care Providers: Quality Evaluations Study
JMIR Mhealth Uhealth 2023;11:e40921
URL: https://mhealth.jmir.org/2023/1/e40921
doi:10.2196/40921
PMID:36857125

Review

Planting Seeds for the Future: Scoping Review of Child Health Promotion Apps for Parents

Sarah B Blakeslee[1], Dr rer med; Kristin Vieler[1], MSc; Ingo Horak[2], Dipl; Wiebke Stritter[1], Dr rer med; Georg Seifert[1], Dr med, PhD

[1]Department of Pediatrics, Division of Oncology and Hematology, Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Digital Health Entrepreneur, Berlin, Germany

Corresponding Author:
Sarah B Blakeslee, Dr rer med
Department of Pediatrics
Division of Oncology and Hematology
Charité - Universitätsmedizin Berlin
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30 450 566058
Email: sarah.blakeslee@charite.de

Abstract

Background: Increasingly, parents use child health promotion apps to find health information. An overview of child health promotion apps for parents currently does not exist. The scope of child health topics addressed by parent apps is thus needed, including how they are evaluated.

Objective: This scoping review aims to describe existing reported mobile health (mHealth) parent apps of middle- to high-income countries that promote child health. The focus centers on apps developed in the last 5 years, showing how the reported apps are evaluated, and listing reported outcomes found.

Methods: A scoping review was conducted according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) guidelines to identify parent apps or web-based programs on child health promotion published between January 2016 and June 2021 in 5 databases: PubMed, ERIC, IEEE Xplore, Web of Science, and Google Scholar. Separate sources were sought through an expert network. Included studies were summarized and analyzed through a systematic and descriptive content analysis, including keywords, year of publication, country of origin, aims/purpose, study population/sample size, intervention type, methodology/method(s), broad topic(s), evaluation, and study outcomes.

Results: In total, 39 studies met the inclusion criteria from 1040 database and 60 expert-identified studies. Keywords reflected the health topics and app foci. About 64% (25/39) of included studies were published after 2019 and most stemmed from the United States, Australian, and European-based research. Studies aimed to review or evaluate apps or conducted app-based study interventions. The number of participants ranged from 7 to 1200. Quantitative and qualitative methods were used. Interventions included 28 primary studies, 6 app feasibility studies, and 5 app or literature reviews. Eight separate topics were found: parental feeding and nutrition, physical activity, maternal-child health, parent-child health, healthy environment, dental health, mental health, and sleep. Study intervention evaluations cited behavior change theories in 26 studies and evaluations were carried out with a variety of topic-specific, adapted, self-developed, or validated questionnaires and evaluation tools. To evaluate apps, user input and qualitative evaluations were often combined with surveys and frequently rated with the Mobile App Rating Scale. Outcomes reported some positive effects, while several intervention studies saw no effect at all. Effectively evaluating changes in behavior through apps, recruiting target groups, and retaining app engagement were challenges cited.

Conclusions: New parents are a key target group for child health apps, but evaluating child health promotion apps remains a challenge. Whether tailored to parent needs or adapted to the specific topic, apps should be rooted in a transparent theoretical groundwork. Applicable lessons for parent apps from existing research are to tailor app content, include intuitive and adaptive features, and embed well-founded parameters for long-term effect evaluation on child health promotion.

(JMIR Mhealth Uhealth 2023;11:e39929) doi:10.2196/39929

KEYWORDS
scoping review; child health promotion; parents; mobile apps; health apps; digital prevention; behavior change; mHealth

Introduction

Digital health is a growing field and apps are used regularly to target health prevention. eHealth measures have steadily gained popularity and are increasingly available in the app form. For the promotion and maintenance of health, digital interventions have been examined for their ability to work as a preventive measure [1]. An increasing number of apps target parents and children for child health promotion and well-being, yet little is known about their impact. Research is conclusive that health promotion activities for child health have a long-term impact on health, whether it be mental health, physical activity, nutrition, or risk behavior prevention [2-6]. Smartphones are estimated to be owned by over 50% of the world's population (~4.3 billion people by 2023) [7], with smartphone ownership averaging over 75% in countries with high-level economies such as the United States and the European Union [8]. Nearly all adults (96%) aged 18-29 own a smartphone in the United States [9] and in Europe on average 75% of people in this age bracket use the internet every day [10]. Current parents and the next generation of parents are seeking health information from digital sources and increasingly from apps, demonstrating the opportunity for health promotion through app use [11].

Stemming from different theoretical approaches from health psychology and fields studying social behavior [12-16], a need to evaluate the ability of illness prevention and health promotion interventions to change behaviors led to the development of behavior change techniques (BCTs) [17]. These are categories of evaluable information, termed taxonomies, that track and measure how effective health promotion interventions can be [18]. The application of such evaluative measures in digital interventions has become a well-established method to evaluate changes in behavior over the last decade [19,20]. For instance, there has been some evidence demonstrating moderate effects of health apps on physical activity and diet in pregnant women [17,21], adults [22], or children [23]. A recent meta-analysis of apps directed at health promotion and illness management described the need for stronger evidence to underscore their effects [24]. At the same time, when it comes to the promotion of health, not enough is known about how or if the use of apps has an effect on behavior change, nor to what extent the evaluation of such apps is undertaken [25], nor how this relates to the actual use of such health apps [26]. Despite the potential and opportunity for combining prevention activities into digital health apps, evaluation of behaviors to measure the effectiveness of mobile interventions is imperative to demonstrate any impact on well-being.

New parents bestow both the genetic makeup and the preliminary foundation for health to their children—from pregnancy to independent adulthood. Despite being an essential cornerstone and stakeholder of child health promotion and well-being, parents often feel unprepared for parenthood [27] and ill-informed about their child's development [28]. There has been no review to our knowledge that assesses if and how child health promotion broadly targeted in parent-based interventions is being evaluated. In an ever-changing digital landscape with continually developed new apps, establishing what apps exist to target parenting and childhood health promotion as well as how they are evaluated is an area of interest.

A preliminary search of literature confirmed that reviews have systematically looked at the impact of apps on behavior [29], and also specific areas of health promotion have been systematically addressed for adults and children, such as nutrition or physical activity [17,30-32], literacy [33], pregnancy [32], and even general well-being [29]. However, a comprehensive compendium of apps that apply to parents for the health promotion activities in children does not exist nor are the evaluative effects of such apps clear. The need to better understand the scope of what apps exist and how they are currently evaluated provides the rationale for this review. The aim of this scoping review is therefore to address this gap by reviewing the existing studies on mobile health (mHealth) prevention apps that target parents for promoting the health of their children. The primary objective of this review is to describe existing reported mHealth parent apps of middle- to high-income countries that promote child health, with a focus on the parent apps developed in the last 5 years. To achieve the objective, this paper intends to give an overview and details on the topic areas of health promotion that parent apps cover and presents the scope of apps that are reported on (keywords, year of publication, country of origin, aims/purpose, study population and sample size, intervention type, and methods). The secondary objective of this review is to compile a list of how the reported apps are evaluated by listing and describing health measures found. The research questions that guided this review were as follows: What current parent mHealth apps exist in middle- to upper-income countries for promoting child health and how, when, and where are they reported on? What topics do they cover? How are child promotion apps for parents evaluated and what outcomes are described in terms of their effectiveness and efficacy? This scoping review aims to shed light on and give a comprehensively reported overview of existing parent apps to promote children's health.

Methods

Design and Overview

A scoping review method was chosen as the appropriate review type to give a broad overview of the existing apps on child health available for parents because this field has not yet been comprehensively mapped and ever-emerging evidence rapidly changes. A planned 3-step search strategy study protocol was registered with the Open Science Forum [34] and used with an established scoping framework [35-37] to search for apps geared toward parents for health promotion in children. The scoping review reporting was supported throughout by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) checklist [38].

Parental Mobile App Study Search Strategy

In a first step, from May 26, 2021 to May 28, 2021, 4 available databases were searched in 2 rounds to include the fields of health, education, and technology: PubMed, ERIC, IEEE Xplore, and Web of Science. After the first-round search with Google Scholar (Google Inc.), too many undifferentiated resources outside the inclusion were found for the search terms, and thus we decided to strategically limit the search to 2021 to find the most recent publications that may be found in the first months after publication, but before these are added to other databases. Search terms combined the keywords "health promotion," "parent*," "child*," and "app," "eHealth" and "mHealth," "mobile health prevention," and "digital health" (Multimedia Appendix 1). Inclusion and exclusion were described and then tailored after the initial search with the study team (SBB, WS, IH, and GS).

In a parallel organizational step to include health expert input from May to August.20, 2021, the third author (IH), gathered stakeholder inputs with authors and health experts located in Germany and Europe to identify parenting studies or apps that may not have been included. This was conducted first through a LinkedIn (Microsoft Corporation) post from a well-established networking account asking for expert input(s) on apps or research projects aimed at young parents to promote the health of their children from birth and how these have been assessed or evaluated. From the expert responses, this information was followed up on to elicit more detailed information on known apps.

Eligibility and Exclusion Criteria

Apps or projects that met the inclusion criteria (Table 1) were assessed further. Study inclusion and exclusion were documented at each step (Figure 1). We aimed to include studies, evaluations, and assessments of digital apps developed toward parents for child health promotion. Studies of all types, reports, and assessments were included if they were (1) digital apps (2) used primarily by parents or expectant parents for (3) health promotion of children without a diagnosis or risk.

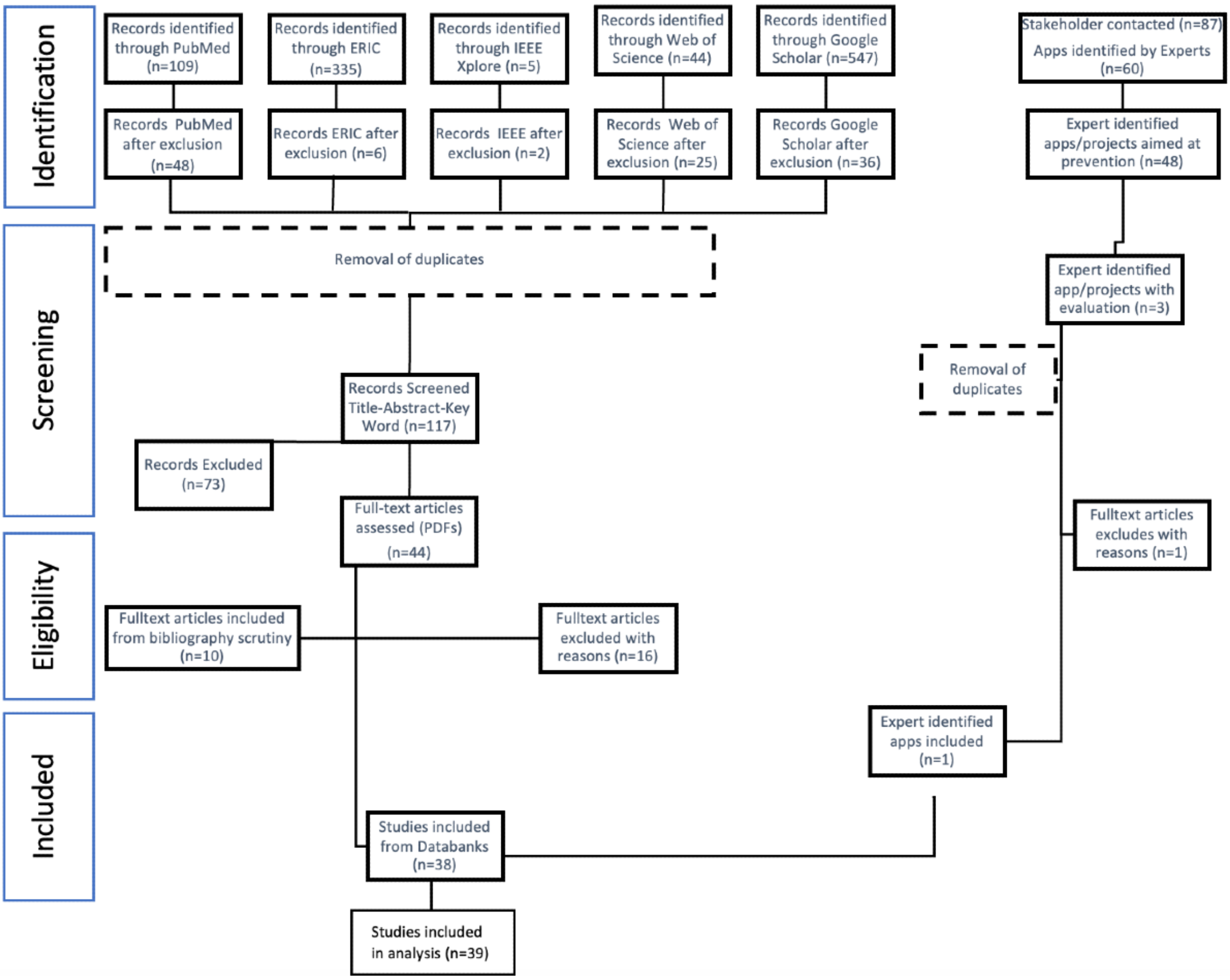
We included both primary studies and reviews of studies and apps. Gray literature was included as long as there was an evaluative component to the work. The apps could be web or mobile-based programs. Based on content, we allowed for a broad range of study interest as it applied to both programs and the people these programs were applied to, including app feasibility or design, evaluation of the apps themselves, evaluation of the potential or actual effect on behaviors, or discussed evaluation strategies. For the expert input, we included studies collected from German or European digital health experts, child health experts, educational experts, or study authors. Only studies based in a middle- or high-income country and published in or after 2016 were included because we were particularly interested in the most recent apps and contexts most resembling the German context of our own research.

All studies that aimed to manage illness or high risk of illness were excluded. Exclusion was applied to any apps or programs aimed solely toward professions or children or where parents were simply gatekeepers. Additionally, studies on apps that were only used as health monitoring, tracking, product-based devices, or as information communication tools such as for text messaging/SMS transmission, videoconferencing, or telehealth were removed from review.

Table 1. Inclusion and exclusion criteria overview.

Selection category	Inclusion criteria	Exclusion criteria
Study population	Expectant parents, parents, parents and children together	Professionals use in work setting, primary use by children with parents only as an app gatekeeper
Health area	All areas of illness prevention/health promotion	Apps for active management of diagnosis, illness, secondary disease prevention, sexual health, those that are institution based, or those recruiting high-risk patients
App type	Smartphone/tablet/desktop	Telehealth, text messaging/SMS-based health support, videoconferencing, health product–based, app only for tracking device facilitation, virtual reality
Publication type	Empirical studies, reports, reviews, study synthesis, meta-analysis, theses, study protocols	Guidelines, handbooks, instructional manuals, user-based information, technical or specialist publications, commentary, product description
Content of interest	App design, reports on app functionality, evaluations of apps and study reviews, behavior change techniques reporting or evaluation, evaluation strategies, structured digital application	Review of app functionality, usability survey results
Countries of interest	All upper-middle or high-income country context [39]	≤Lower- to middle-income country contexts
Stakeholder input	Digital health experts, child health experts, educational experts, study authors (focus on Germany and Europe)	No restrictions applicable
Timeframe	≥2016 (Google Scholar >2021)	<2015

Figure 1. PRISMA-ScR flowchart. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews.



Study Selection

The search took place following an initial identification of studies through the databases. Then, we performed a screening of the title, abstract, and keywords for applicability according to the inclusion and exclusion criteria and studies were imported into EndNote X9 (Clarivate) [40].

In the next screening step, the first author (SBB) applied the inclusion and exclusion criteria according to study abstracts, eliminated duplicates, and added full-text PDFs of all studies fitting the inclusion criteria. All expert contributions were controlled for documentary evaluation or assessment of the apps or projects, ensuring they fit within the inclusion/exclusion criteria and removing duplication. The resulting full-text studies and corresponding research information system (RIS) files that compiled bibliographic data information were imported into the analysis management software MAXQDA (version 20; VERBI GmbH) [41].

All studies that passed the original screening were reviewed in full text, coded deductively with the bibliographic RIS content, and systematically evaluated according to the paper sections. After full-text scrutiny, studies not meeting the inclusion criteria were excluded and adjustments were discussed, justified, and made within the whole team when necessary, based on the refinement of the inclusion criteria. Additionally, scrutiny of the included bibliographies, especially topically relevant reviews, was culled for additional studies.

Summarizing the Data

The included studies summarized the key information as suggested by Peters et al [35] and this key information was analyzed through a systematic and descriptive content analysis based on Mayring and Fenzl [42] using a combined deductive and inductive approach. Deductive coding and descriptive analysis were conducted on all the included studies to compile and describe the following information: (1) keywords, (2) year of publication, (3) country of origin, (4) aims/purpose, (5) study population and sample size, (6) intervention type, (7) methodology/method(s), (8) broad topic(s), (9) evaluation, and (10) outcomes and details. Following this, key findings that related to the scoping review questions were coded inductively within the deductive descriptive categories: broad paper topics and evaluation. To verify the reliability of the coding of the study types and topics, the second author (KV) reviewed all studies based on inductively developed definitions. Discrepancies were discussed within the team and code definitions were adjusted accordingly. A descriptive summary of how apps and behaviors were evaluated are summarized in Table 2.

Table 2. List of evaluation tools found in included studies.

Broad paper topics and evaluation tool category	Evaluation tool type or name [reference]
Physical activity	• Assessment of subcategories: changes in physical activity, adult physical activity, family and social group physical activity, children's physical activity evaluation, and tracking physical activity and real-time measurements
Moderators of physical activity	• Ecological Momentary Assessment (EMA); Behavioral Regulation in Exercise Questionnaire; Self-Efficacy Scale; intention to participate in physical activity and to eat healthy foods [43] • Barriers to Being Active Quiz; Self-Efficacy for Physical Activity Scale; Physical Activity Stages of Change [44]
Adult physical activity	• Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH) [45] • International Physical Activity Questionnaire [43] • Stanford Brief Physical Activity Survey [44] • WHO[a] physical activity criteria [46]
Family and social group physical activity	• Modified National Board of Health and Welfare's survey [47] • Parental Support for Physical Activity Scale [48] • The Social Support and Exercise Survey [44] • Family Health Climate Scale [43] • Family physical activity goal setting [47]
Children's physical activity evaluation	• Test of Gross Motor Development 2nd Edition (TGMD-2); The Burdette Outdoor Playtime Checklist [49]
Tracking physical activity	• Average minutes per day physical activity [47] • Accelerometers or pedometers [43,44,50] • Physical Activity Diary [43]
Body measurements for physical activity	• BMI or height and weight [43-45,47-50]
Parent feeding and nutrition	• Assessment of subcategories: food types and quality, parent feeding and food acceptance, food environment, food and body measurements, and breastfeeding
Food types and quality	• Youth Risk Behavior Survey questions; The Behavioral Risk Factor Surveillance System Questions [51] • The Willett Questionnaire Harvard Food Frequency; Healthy Eating Index (HEI) [45] • Healthy Kids Survey [52] • Food Frequency Questionnaire [43,45,48,53-55] • Consumption of fruit, vegetables, water, soft drinks, and snacks [48]
Parent feeding and food acceptance	• Infant Feeding Questionnaire [56,57] • Parent Feeding Practices Scale; Child Feeding Questionnaire (CFQ) [57] • Norwegian Mother and Child Cohort Study (MoBa) Questions; Children's Eating Behaviour Questionnaire (CEBQ); Child Food Neophobia Scale (CFNS) [54] • Infant Food Exposure and Parental Intentions to Offer Foods [56]
Food environment	• The Family Eating and Activity Habits Questionnaire [50,57] • Self-efficacy scales, food insecurity [52] • Regulation of Eating Behavior Scale [43] • Postpartum Partner Support Scale (PPSS) [58] • Parenting Strategies for Eating and Activity Scale, Parenting Feeding Style Questionnaire [48] • Parenting Practices Questionnaire, Parent Modelling Questionnaire, Family Support [57] • Menu planning and shopping practices, healthy restaurant selection practices, family food preparation practices [51] • Australian NOURISH study questionnaire [54] • Environment and Policy Assessment Observation [59] • School Food Checklist [60]
Food measurements	• Fruit and Vegetable Intake Diary [43,53] • 24-hour dietary recall of foods and beverages [57] • Food photography and weighed food records [59,60] • Caloric counting in kilojoules [59,60]

Broad paper topics and evaluation tool category	Evaluation tool type or name [reference]
Body measurements for nutrition	• Weight reporting [51] • BMI or height and weight [45,47,48,50,51,54-56,61] • Waist circumference [50,57]
Breastfeeding	• WHO duration of exclusive breastfeeding, Breastfeeding Self-Efficacy Scale (BSES-SF) [58] • Baby Eating Behaviour Questionnaire (BEBQ) [56]
Dental health	• Dental Knowledge Attitudes and Practices Questionnaire [62] • Oral health behaviors in children and determinants of the Theory of Planned Behavior [63] • Purposively sampled qualitative interviews [63]
Sleep	• Customized Sleep Profile (CSP); Brief Infant Sleep Questionnaire–Revised (BISQ-R) [64] • Familial risk moderates the association between sleep and zBMI[b]; activity-based sleep-wake identification; Sleep Habits Questionnaire [57]
Mental health	• Center for Epidemiological Studies Depression Scale [44] • Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) [65] • Edinburgh Postnatal Depression Scale (EPDS), State Trait Anxiety Inventory (STAI) [66]
Parent child health	• Patient activation measure (PAM); Functional, Communicative, Critical Health Literacy Scale [67] • Patient Education Materials Assessment Tool (PEMAT) [68] • The 21-item Asian Self-Identity Acculturation Scale [50]
Healthy environment	• Safety behaviors and behavioral intentions [69] • Safety knowledge [69,70] • Hot beverage scald risk and burn first-aid knowledge [71]
Maternal health and parenting	• Prenatal Interpersonal Processes of Care (PIPC) Scale [67] • Maternal Self-Efficacy Scale [50] • Pregnancy Discomfort Checklist [44] • Parenting Self-Efficacy (Tool to Measure Parenting Self-Efficacy [TOPSE]) [65] • Parenting Efficacy Scale, What Being a Parent of a Baby Is Like (WPBL), Perceived Social Support for Parenting (PSSP), Parent-to-Infant Bonding Questionnaire (PIBQ) [66]
App evaluation	• Assessment of subcategories: app quality, app usability, and app coverage
App quality	• Suitability Assessment of Materials (SAM) [72] • Persuasive System Design Model [73] • Mobile App Rating Scale (MARS) [53,72,74] • Semistructured and structured interviews [74,75] • Participant app testing [74,75]
App usability	• Just-in-time adaptive interventions [43] • Push notifications [43,44,58] • Gamification [43,46,58,71,74] • Technology Acceptance Model (TAM) [76] • Engagement Index Tool [75] • The System Usability Scale (SUS) [74]
App coverage	• Health-Related Website Evaluation Form: Developed Quantitative Tool for App Coverage [72]

[a]WHO: World Health Organization.

[b]zBMI: sex- and age-standardized BMI.

Collating, Summarizing, and Reporting Results

The analysis of keywords (1) was conducted from the bibliographic RIS data according to their frequency of appearance. Presentation of the overall findings from the deductive analysis of the study information 2-7 was summarized and detailed in Multimedia Appendix 2. Within the broad topic(s), ways apps and behaviors were evaluated and study-described outcomes 8-10 and details were analyzed, and then described and summarized in an iterative, inductive process used for the included studies, including a cross-reference between topics and evaluation tools listed within the studies (Table 2).

Reviews were included in this scoping review. For pragmatic organizational reasons, and because some of the primary source data did not fit the scope of our review objectives or fit our inclusion criteria, only the findings of the reviews themselves were included, not the primary literature that they were based on.

Results

Overview

Of the 39 studies included in this review of child health apps for parent use, most stemmed from US-, Australian-, and European-based research. A total of 8 overlapping health promotion topics that were addressed in 28 primary intervention studies, assessed in 6 app feasibility studies, and reviewed in 5 app or literature reviews were identified. The topics found in the inductive analysis were parental feeding and nutrition, physical activity, maternal-child health, parent-child health, healthy environment, dental health, mental health, and sleep. In primary intervention studies, behavior change theories were embedded in 26 studies and evaluations were carried out with a variety of topic-specific, adapted, self-developed, or validated questionnaires and evaluation tools. Methodologically, included studies were summarized and the effects, if any, of interventions were described. Reported study effects varied and used diverse tools to evaluate intervention effects. Alternatively, the feasibility of apps or health behaviors was assessed with a described combination of quantitative evaluation and survey tools along with user input. Included studies cited challenges in assessing healthy behaviors of children though parent apps, specifically in finding the appropriate way to evaluate changes in behavior through apps, recruiting target groups, and retaining app engagement.

Overall, 1040 studies from the 5 selected databases were analyzed and 60 apps and programs were gathered through the expert network. After screening for eligibility and duplication, and adding resources from reviews, 39 studies were included in total; 28 of these were found from databases, 10 were discovered by scrutinizing the bibliographies of included sources, and 1 resource was included from the expert input. An overview of study inclusion can be seen in the PRISMA-ScR flowchart (Figure 1).

Keywords

Keywords of all included studies demonstrated the following terms according to the bibliographic RIS information from the studies. The 11 most frequently used keywords listed in 9 or more included publications (with listed frequency of appearance) were humans (n=19), female (n=14), child (n=13), health promotion (n=12), male (n=12), parents (n=12), mHealth (n=11), smartphone/s (n=10), mobile apps (n=10), adult (n=9) and infant/s (n=9; Multimedia Appendix 3).

Year of Publication and Country of Origin

The included studies were published between 2016 and 2021, with two-thirds published between 2019 and 2021 and an uptick observed in 2019 (Multimedia Appendix 4). Among the upper-middle and high-income countries included, the majority came from the United States (n=15) [44,51,52,77-79], followed by Australia (n=13) [45,49,53,56,58-60,72-74,80-82] and then the European region (n=9) [43,46-48,54,55,57,83]. Included European countries with 1 study each were Belgium [48], the Netherlands [83], Portugal [57], Sweden [47], and the United Kingdom [83], with 2 studies each in Norway [54,55] and Germany [43,46]. Only 2 studies came from countries outside the global North (Singapore [66] and Iran [62]).

Aim, Sample Size, and Intervention Type

Specific aims of the studies were diverse and ranged from creating a topic overview of existing studies or apps, assessing the feasibility of developed apps, to evaluating the effectiveness of a child health promoting intervention involving app or web-based content. There were 3 types of interventions that were included in our review: 28 primary studies [43-52,54-56,58-67,69,71,78,79,82], 6 app feasibility studies [70,73-75,83,84], and 5 reviews, of which 2 were literature reviews [53,57] and the remaining 3 were app reviews [68,72,85]. In the studies, the number of participants ranged from 7 to 1200. The review of apps included between 29 and 47 apps and the review of studies included 11 studies each. Methodologically, the studies were heterogenous in design and evaluation method. The clinical trial was the most frequent study design type for 21 studies [43-45,47-51,54-56,58-62,66,67,69,79,80] with most using the randomized controlled trial (n=15) and others with pilot, nonrandomization or experimental designs (n=6). Four of the included studies [43,47,59,69] published protocols of studies yet to be undertaken. The second most frequently undertaken type of evaluation was feasibility studies connected to the evaluation of app design features, testing, and functioning [53,70,73-75,83,84]. Quantitative and qualitative results were combined in the mixed method designs of 7 of the included primary (n=3) [63,65,70] and feasibility (n=4) [73,75,76,83] studies. A predominantly qualitative design was undertaken by 2 studies [63,84]. Of the 32 single studies, 25 individual project names were listed, of which 3 projects had 2 publications (Make Safe Happen [69,70], Swap It [59,60], and the Growing Healthy Program [56,76]) and 4 did not list a specific name [62,63,67,78]. An overview and summary of the included studies can be found in Multimedia Appendix 2.

Broad Topics

The studies included could be sorted into 8 main prevention and child health promoting topics: parental feeding and nutrition (n=19) [43,45,47,48,51-53,55-61,72-74,82,86], physical activity (n=8) [43-49,84], maternal-child health (n=6) [44,45,65,67,75,85], parent-child health (n=5) [66,68,78,79,83], healthy environment (n=3) [69-71], dental health (n=2) [62,63], mental health (n=1) [66], and sleep (n=1) [64]. A crossover of these inductively derived topics occurred in some studies and these were not mutually exclusive; if a study descriptively included more than 1 topic, then the study was included in both topics. This occurred most frequently with studies that addressed parental feeding or nutrition and physical activity: this combination of topics was found for 7 of the studies. In 2 studies physical activity was addressed in combination with maternal health. Parental feeding and nutrition addressed nutritional intake for a range of ages: starting with nutrition in pregnancy

[45]; feeding practices and nutrition for infants and young children, whether through breastfeeding or solid food [47,48,54-57,61,72,73,76]; or promotion of healthier school meals or family nutrition [43,51-53,59,60,74,87]. Included studies that broached physical activity were interested in either tracking the movement as part of the app-based intervention [44,45,47] or physical activity as part of obesity prevention, comprehensive child fitness, or overall family health [43,46,48-50,84]. All studies with a topical focus on maternal-child health targeted women in pregnancy. The parent-child health app studies included had an educative or informational focus on parenting and child health. Included apps promoting a healthy environment targeted home safety and accident prevention, while studies addressing dental health were concerned with caries prevention and dental hygiene. Mental health was addressed from the standpoint of overall child well-being and the sleep app studies included assessed the parent tracking of infant sleep schedules.

Parent Mobile App Evaluation

Evaluation of Behavior Change in Apps

Many of the study evaluations assessed changes in intentions, knowledge, or behavior over time. In total, 26 studies listed at least one specific behavior change theory that the study evaluation was based on: Social Cognitive Theory was mentioned in 9 studies [44,47,48,51,54,55,58,66,79] and in 1 meta-analysis [57]; Self-efficacy Theory was mentioned in 3 studies [53,65,66]; Social Determination Theory also in 3 studies [43,48,67]; and the Behavior Change Wheel in 4 studies [46,59,61,74]. Some studies also used BCTs in their interventions (n=6) [43,47,57,59,78,85]. While most studies do not explicitly name the individual BCTs (n=20), 10 of these studies used BCTs. Among studies that mentioned techniques of behavior change, the most frequently cited were the BCT taxonomy by Michie et al [88], which was cited in 2 studies [47,57], and the mHealth theory–based taxonomy for mobile apps, which was also cited in 2 studies [78,85]. Individual BCTs mentioned in the included studies were shaping knowledge, identification of self as a role model, demonstration of the behavior, self-monitoring of behavior, self-belief, prompts/cues, goal setting (behavior and outcome), identity, and social support.

To measure the potential for change in behavior, multiple questionnaires were used that cut across topics. Some questionnaires that assessed changes in behavior were self-developed [51,54,55,85] or developed out of other validated questionnaires [48,59,62]. As an essential part of most behavior change models, the most frequently used validated questionnaires in the studies assessed self-efficacy as a predictor for changes in behavior for different topics such as motherhood, nutrition, breastfeeding, and physical activity. Measures for changes in self-efficacy or knowledge before and after the intervention were described to give an outlook for the continuation of the new behavior. Listed validated questionnaires used to evaluate behavior changes were the 10-item COM-B Self-Evaluation Survey (healthy family meals) [74], Maternal Self-Efficacy Scale (a 12-item scale measuring the mother's self-efficacy for promoting healthy eating, physical activity, and in limiting noncore foods) [50], the 14-item short form Breastfeeding Self-Efficacy Scale [58] assessing breastfeeding confidence, Self-Efficacy for Physical Activity [44], the 10-item Parenting Efficacy Scale [66], and 36-item Parenting Self-Efficacy (Tool to Measure Parenting Self-Efficacy [TOPSE]) [65]. Increasing knowledge cut across topics, ranging from a healthy environment [69,70,80], physical activity or nutrition [47,52,54,78], dental health [62,63] parenting for health [50,65], or sleep [64,78]. Despite the objective to increase health knowledge of parents, not all studies undertook explicit evaluations to measure knowledge change.

Assessment tools were mentioned and used for specific topics. An entire overview of assessment tools for evaluating data and parameters can be found in Table 2.

Physical Activity

Physical activity was assessed through different means: 10 studies used physical activity measures [43-50,78,84]. We identified 21 separate measures that evaluated physical activity in 3 ways: specific behaviors as they related to quantified movement (ie, accelerometer), those that predicted or moderated the physical activity undertaken (ie, self-efficacy), and measures of the outcomes of physical activity (ie, BMI or weight over time). Of these tools, 17 used validated measures to assess physical activity. Wunsch et al [43] and Choi et al [44] measured the self-efficacy of physical activity specifically. Accelerometer to track steps and physical movement were used or planned in several studies [43,44,50]. BMI calculations were investigated in 6 studies [44,45,47-50] evaluating physical activity, especially when combined with the topic of nutrition and as a secondary parameter. In studies with small children, the evaluation measurements and intervention for physical activity were frequently given by the parents or primary caregivers. For instance, in the studies by Trost and Brookes [49] and De Lepeleere et al [48], the parental support for Physical Activity Scale was used. A strong connection of studies researching the topics of nutrition and physical activity demonstrated a crossover in evaluation tools used for body measurement, such as BMI calculated from height and weight [43,45,47,48,50]. Combined nutrition and physical activity likewise evaluated parent preferences within theory-guided domains for healthy goal setting [78].

Parent Feeding and Nutrition

In total, 20 studies [43,45,47,48,51-61,72-74,76,78] fell into the topic of parent feeding or nutrition and had the largest number of individual assessments. Overall, we were able to identify 41 assessment tools used in the studies that fit into 1 of 6 separate evaluative purposes (see as referenced in Table 2): measuring food amounts, taking body measurements for nutrition (often also for evaluating physical activity), assessing the ways and environment in which food is consumed, evaluating the quality of food consumed, examining parent feeding and young child food acceptance, or assessing breastfeeding-specific practice. Of the 41 assessment tools and questionnaires used, the majority (n=32) were validated tools. Six tools were self-developed specifically for the study and 3 further assessments were listed in the reviews and their origin was unclear. The Child Feeding Questionnaire was found to be the most frequently used questionnaire to assess parental feeding practices [50,54,57]. An instrument most frequently used for evaluating nutrition was the Food Frequency Questionnaire [43,48,54,55,57,59].

Dental Health

Four studies evaluated parameters of dental health. In the dental study by Zolfaghari et al [62], for instance, the authors used a self-developed questionnaire to assess parent knowledge and practices that combined the self-developed questions with other validated questionnaires [89-91]. A 24-item validated questionnaire designed by Van den Branden et al [92] to measure oral health behaviors in children and the Theory of Planned Behavior determinants was used, with permission, prior to and following use of the app [63].

Sleep

Only 1 study [64] specifically evaluated sleep as an mHealth intervention. This specifically assessed the sleep of infants and babies with a Brief Infant Sleep Questionnaire-Revised. However, an evaluation of the sex- and age-standardized BMI (zBMI) was found in Gomes et al’s [57] review of parental feeding practices and as part of a parent information needs assessment [78].

Mental Health

Mental health was assessed in 3 of the included studies [44,65,66]. The Warwick-Edinburgh Mental Well-Being Scale, a validated measure, was used by Deave et al [65], using a 14-item scale of subjective mental well-being and psychological functioning. Choi et al [44] used the Center for Epidemiological Studies Depression Scale to assess the mental health.

Parent Child Health

A total of 8 studies [50,66-68,71,78,79,83] were found to address parent-child health interactions, including the health of families, identity, and family-based evaluations. None of the evaluation tools broadly assessed the parent-child health interactions, but rather concentrated on the specific topic of interest for the parent-child interaction. For instance, Knowlden and Sharma [79] used the most general assessment. The authors developed separate evaluations of maternal-facilitated and child-behavior constructs based on Social Cognitive Theory to evaluate the parent-child health interaction [79] with an aim to address healthy child nutrition and physical activity. Other topic-oriented parent-child health parameters were also found that focused on evaluating educative [66-68,71,83] or identity parameters [50].

Healthy Environment

Three studies [69,70,80] specifically evaluated healthy environment through evaluations of safety behavior and first-aid knowledge.

Maternal Health and Parenting

Six studies [44,45,65,67,75,85] addressed evaluations of maternal health and 7 studies [48,65,66,68,75,78,79] looked at specific parenting parameters. In 1 study [65], the parenting self-efficacy was measured with the TOPSE. The TOPSE was used to compare mothers at 3 months after birth who had downloaded the Baby Buddy app with those who had not downloaded the app, controlling for confounding factors. The postnatal mental state was measured in Shorey et al [66] with a crossover of mental health and parenting and infant bonding tools.

App Feasibility (Quality and Usability)

The most frequent way by which child health apps for parents were assessed was through the Mobile App Rating Scale [53,72,74], developed by Stoyanov and colleagues [93]. To further assess the feasibility and quality of parent apps, a mixed methods approach was used for further development and contextual adaptation of feedback through interviews, where mostly semistructured interviews were conducted [73-75,83,84]. Qualitative assessments of the apps used in in-person, online, and telephone [73] semistructured interviews or focus groups were analyzed by a stated inductive or thematic analysis. Whereas app development approaches guided the qualitative interview data collection [73,75], explicit stating of the qualitative theoretical approaches for the interviews themselves was notably lacking in some studies [83,84]. Braun and Clark was the most frequently cited theoretical approach [70,74,75]. Furthermore, data analytic tools for coverage, usability, and engagement were used by several studies of apps [72,74-76]. Additionally, features of apps such as push notifications, gamification, and just-in-time adaptive interventions were used or listed for apps to retain engagement [43,44,46,58,71,74].

Parent Mobile App Outcomes

Reported Evaluation Outcomes Based on Topics

The manner in which parent-based apps and interventions reported on outcomes in the primary studies was mixed. The study-reported effectiveness of an intervention was cited by many to depend on the length of the intervention, the intended intervention that was targeted, and whether an app included in-person support. Apps increasing knowledge seemed to be a particularly effective means to create a healthy environment with children [70,71] or to increase knowledge on child oral health [62]. An increase in physical activity of pregnant women was cited by 2 studies [44,50] and an 8-week app intervention was able to increase the physical activity performed by children, but this was not a significant outcome [49]. Increasing knowledge on nutrition was demonstrated in 1 study [52]; however, this intervention was coupled with in-person support classes. For nutrition outcomes, a reported increase in motivation or the consumption of fruit and vegetables in a child’s diet was reported by several studies [48,51,55] and healthier lunches saw less discretionary foods packed by parents who used an app [60]. Most improved outcomes with the interventions were not simply attributed to the use of the app alone, however. For example, a trial on dental hygiene demonstrated improvement for app users with a high level of perceived behavioral control, especially when coupled with regular dental checkups [63]. App-only outcomes demonstrated some positive effects for new parents of infants with sleep problems [64] and for improving parent bonding and self-efficacy after birth [66]. Outcomes in nutrition studies that relied on longer term growth outcomes saw little sustained or no positive effect over time with app use [54,56,61,79]. Indeed, studies on app-based interventions for baby food introduction and sustained healthy eating in early childhood highlighted the difficulty of achieving any sustained positive effect over time [54-56]. Across other topics, app support for partners of breastfeeding women or lifestyle advice for pregnant women resulted in no changed outcome with the apps and eHealth interventions [45,58], or even saw negative outcomes in the group receiving an app-supported intervention (ie, intervention group) to aid pregnant women decision-making [67]. This outcome supports a recommendation given in multiple interventions to use real-world interaction and support interventions in conjunction with the app [50,55,61,65,66]. Recruitment posed its own challenges. Particularly, in studies that aimed at healthier behaviors for children that were facilitated and necessitated parental support, authors employed several strategies: some recruited children but evaluated data from parents [59], some spoke of parent-child dyads [50,55,61], while others focused on the recruitment of families [51]. Some studies reported parents having higher education levels and potentially greater willingness to engage with the technology than a targeted population that would most benefit from the intervention [45,48,54,58,61,63,79].

App Evaluations of Behavior Changes and Parent Experience

A few studies highlighted the difficulty of customizing BCTs to their app content that combined the aims of the intervention with potential needs of parents and the ability to effectively evaluate these measures [56,65,78], a point that was discussed in additional detail in the reviews by both Gomes et al [57] and Biviji et al [85]. Particularly, the app reviews and a few studies underscored the gap of evidence-based apps with best practices among available apps for parents across health promotion topics [72,78,83]. Tracking of growth, pregnancy development, breastfeeding, dental hygiene, and diet were features that parents enjoyed, especially if these contents were tailored to the health parameters [53,63,77,83]. At the same time, features such as chat functions [53,73] or diaries [44] had mixed reviews or negative desirability by parents in the studies.

App Content Delivery and Technical Features

Keeping parents motivated to use the app was a challenge reported in multiple studies [45,56]. Other content delivery mechanisms, such as audio recordings (podcasts) [75] or videos [48], saw a high level of adherence in terms of the content consumption. Technical problems, interface challenges, or the inability to appropriately tailor app features were feedback highlighted by several studies [56,58,61]. The engagement with the apps by parents was described in a few studies to have the highest relevance for first-time parents [66,76] and retaining app or program engagement, particularly for the group targeted, was a challenge cited in multiple studies [46,56,61]. Features such as push notifications were seen as helpful delivery tools to maintain engagement with the app [44,58,60,61,76] and gamification was seen to have some success in achieving this goal [46,62,71]. Future designs for engaging parents reference increasingly developed "just-in-time" features to enhance practicability and interaction [43,74,76].

Discussion

Principal Findings

The 39 studies that met the inclusion criteria for this review reflected a wide range of child health topics: parental feeding and nutrition, physical activity, maternal-child health, parent-child health, healthy environment, dental health, mental health, and sleep. The 8 individual topics were concluded by an inductive analysis. Behavior change theories guided the research of 26 studies and topic-specific, adapted, self-developed, or validated questionnaires and evaluation tools were used to assess and report study outcomes. At the same time, challenges were reported in effectively evaluating changes in behavior through apps, recruiting target groups, and retaining app engagement.

An overall increase of publications on the topic may reflect the growing number of apps developed in general. The lower number of the published studies during 2020 may be an influence of the COVID-19 pandemic, a trend that we saw increase in a swift subsequent spot search in each of the included databases (see Multimedia Appendix 1). Since this review was conducted, 3 additional study results from included study protocols were published [94-96]. The demand and need for addressing child health promotion have only grown since the start of the COVID-19 pandemic [97] and digital mHealth solutions are forecasted to continue to grow [98]. The greater opportunity to digitally support child health through parents solidifies the need to make sure that parents have access to health promotion apps that are embedded in scientific evidence and best practices. Generally, the regulation of recruitment strategies was very bound to the study context and was a challenge highlighted by the studies in our findings. Varied descriptions of how potential participants were recruited and who was recruited detailed a level of complexity requiring consideration for study designs with multiple sites (homes and schools, for instance) and studied parties (children and parents).

Our findings highlighted the complexity of compiling evidence of behavior changes that are supported by apps and web-based programs for child health. When app interventions evaluated parents' knowledge after use as a primary outcome, evaluation of the knowledge increase was easily assessed [52,62,69-71]. Evaluating the effectiveness of more complex interventions of health promotion as described in the included studies requires multiple evaluation tools and behavior-specific tailoring in order to see potential effects that may or may not continue in the long term. Prevention interventions in primary care with young children have been found to be exceptionally challenging to sustain over time, requiring complex interventions and involvement of multiple actors [99]. One additional impediment for long-term measurable changes could derive from the need for a clear theoretical underpinning and health mode within health promotion apps. With the absence of illness in the prevention setting, apps for health promotion could benefit from a health psychology theory–based development with a systematic evaluation in order to lead to substantial positive changes in behaviors [100]. The studies included in this review had varying degrees of theory embedded into the app design, which can provide a framework for evaluation. The most frequently used framework in the included studies was the Behavior Change Taxonomy [88] and its adapted version for mobile apps [20,101], which was itself developed from an expert collaboration. Many of the included studies were not transparent in reporting the link between the theory of behaviors and the evaluation parameters assessed or app features developed. On the whole, the multipronged strategies required for developing and evaluating apps for parents exhibit methodological agility and interdisciplinary collaboration. Interventions with demonstratable effectiveness were able to do this, as was markedly evident in the included studies compiled and reviewed on the topics of maternal child health, parent feeding, and lunch box nutrition [53,57,77].

Involvement of stakeholders is an imperative first step in the development of apps. Health experts bring expertise and scientific basis to the interventions for child health promotion and such expertise can be built on to further develop and adapt apps to changing evidence and circumstances. An example of this adaptation is the Growing Healthy program, where an initial study on childhood obesity prevention starting in infancy was published [61,102] and then compared in an upscaled study with another intervention [56] and followed by parent insights and feedback that were able to be integrated back into the app development in order to make them more intuitive and adaptive to specific engagement levels and identified target groups [76]. Parent feedback demonstrates that the apps are used most when the intuitive apps and features can address their needs and questions they have about their child's health at the point when they need answers. While parents in the included studies were not always able to imagine what theoretical features would be useful [46,78], they provided strong feedback when asked for (for instance, [53,70,74,83,84]).

Strengths and Limitations

This scoping review provides the first comprehensive overview of available mobile apps and web-based programs for use by parents aimed at the health promotion of their children. The 39 included studies were systematically categorized, provide a thorough summary of current evidence, describe some of the best practices for app development on this topic, and give a strong foundation for further research.

Despite this, this review is not without limitations. Inclusion criteria for this review were purposefully phrased broadly to be as inclusive as possible for apps aimed at parents. However, the multiplicity of study types was not foreseen and may have been more succinctly described. For instance, only including primary studies may have facilitated greater clarity in study summary. This methodological choice also hindered greater comparison between the studies. This study did not include an evaluation of outcomes, a step that would be helpful in future research to evaluate measured changes in behavior or effectiveness that the parent apps had. We also purposefully only included apps and programs from middle- and upper-income countries, apps that targeted healthy children without a diagnosis, and only studies published after 2015. This limitation may have therefore excluded apps or programs in other contexts that may have had broader and more global application. A future review would benefit from a systematic evaluation of app outcomes that includes only primary studies with inclusion of middle- and lower-income countries to be more generalizable and relevant to a larger population. Despite our attempts to include potential gray literature and expert input, no unpublished app evaluations were found. Despite our best efforts to include studies from other disciplines, most apps for parents, which were aimed at the health of their children, were found and evaluated within the health field. Access to published analysis of apps with detailed information evaluation is likely a further limitation of this study, because of the assumption that most apps developed in a scientific context are motivated to publish on the development and evaluation findings. It must be recognized that apps are developed out of many contexts and future reviews would benefit from the inclusion of parent apps developed from other fields (eg, marketing, industry, governmental or nongovernmental organizations, or other interest groups). Our own attempt to bridge this gap with the addition of extending and tapping into an expert network only saw limited methodological success.

Conclusions

Existing apps and web-based programs aimed at parents to promote the health of their children cover a broad range of topics. Most aim to modify the nutrition and physical activity behavior—important for lifelong prevention of illness. New parents are a key target group for apps, whether to increase their knowledge or parental self-efficacy. Evaluating apps for child health promotion provides a special challenge and must be tailored to the needs of parents, context of the topic, and are ideally rooted in a transparent theoretical framework. Given the increasing digitalization of health and expanding focus of health policy on prevention measures, parent apps are guaranteed a role in our lives. Lessons learned can be garnered from existing research studies that tailor developed content to target group needs, include intuitive and adaptive features, and embed well-founded parameters for evaluations able to investigate long-term effects of parent apps on child health.

Acknowledgments

We thank Janina Everding for her tenacious help with the manuscript formatting. Funding for this scoping review was supported by funding from the Federal Ministry for Economic Affairs and Climate Action "Zentrales Innovationsprogramm Mittelstand" Program for small- and medium-sized enterprises under the grant number 16KN0732.

Authors' Contributions

SBB was responsible for methodology, validation, formal analysis, investigation, data curation, writing, review and editing, visualization, and project administration. KV performed validation, formal analysis, writing, review and editing. IH performed investigation and data curation. WS was responsible for methodology, conceptualization, validation, formal analysis, review and editing, supervision, and project administration. GS was responsible for conceptualization, validation, resources, review and editing, supervision, funding acquisition, and project administration.

Conflicts of Interest

None to declare.

Multimedia Appendix 1
Search strategy details.
[DOCX File , 24 KB - mhealth_v11i1e39929_app1.docx]

Multimedia Appendix 2
Detailed summary of results.
[DOCX File , 155 KB - mhealth_v11i1e39929_app2.docx]

Multimedia Appendix 3
Research information system (RIS) keywords from all included studies (occurrence in publications ≥2).
[PNG File , 220 KB - mhealth_v11i1e39929_app3.png]

Multimedia Appendix 4
Included scoping review publications (2016-2021).
[PNG File , 45 KB - mhealth_v11i1e39929_app4.png]

References

1. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
2. Mead E, Brown T, Rees K, Azevedo LB, Whittaker V, Jones D, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese children from the age of 6 to 11 years. Cochrane Database Syst Rev 2017 Jun 22;6(6):CD012651 [FREE Full text] [doi: 10.1002/14651858.CD012651] [Medline: 28639319]
3. Martin A, Booth JN, Laird Y, Sproule J, Reilly JJ, Saunders DH. Physical activity, diet and other behavioural interventions for improving cognition and school achievement in children and adolescents with obesity or overweight. Cochrane Database Syst Rev 2018 Jan 29;1(1):CD009728 [FREE Full text] [doi: 10.1002/14651858.CD009728.pub3] [Medline: 29376563]
4. Martin A, Saunders DH, Shenkin SD, Sproule J. Lifestyle intervention for improving school achievement in overweight or obese children and adolescents. Cochrane Database Syst Rev 2014 Mar 14(3):CD009728 [FREE Full text] [doi: 10.1002/14651858.CD009728.pub2] [Medline: 24627300]
5. Al-Khudairy L, Loveman E, Colquitt JL, Mead E, Johnson RE, Fraser H, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. Cochrane Database Syst Rev 2017 Jun 22;6(6):CD012691 [FREE Full text] [doi: 10.1002/14651858.CD012691] [Medline: 28639320]
6. Weissberg RP, Kumpfer KL, Seligman MEP. Prevention that works for children and youth: An introduction. American Psychologist 2003 Jun;58(6-7):425-432. [doi: 10.1037/0003-066x.58.6-7.425]
7. O'Dea S. Telecommunications Smartphone users worldwide 2016-2023. Statista. 2020. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2023-06-22]
8. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019 Feb 5. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-05-08]
9. Mobile Fact Sheet: Mobile phone ownership over time. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-06-22]
10. Being young in Europe today - digital world. Statistics Explained. 2020. URL: https://tinyurl.com/5n8dfzrf [accessed 2023-06-22]
11. Kubb C, Foran HM. Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]
12. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]

13. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]
14. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
15. Pelikan J. The Application of Salutogenesis in Healthcare Settings. In: Mittelmark MB, Sagy S, Eriksson M, Bauer GF, Pelikan JM, Lindström B, editors. The Handbook of Salutogenesis. Cham, Switzerland: Springer; 2017:261-266.
16. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
17. Hayman M, Alfrey K, Cannon S, Alley S, Rebar AL, Williams S, et al. Quality, Features, and Presence of Behavior Change Techniques in Mobile Apps Designed to Improve Physical Activity in Pregnant Women: Systematic Search and Content Analysis. JMIR Mhealth Uhealth 2021 Apr 07;9(4):e23649 [FREE Full text] [doi: 10.2196/23649] [Medline: 33825693]
18. Michie S, Wood C, Johnston M, Abraham C, Francis J, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
19. Wang Y, Fadhil A, Lange J, Reiterer H. Integrating Taxonomies Into Theory-Based Digital Health Interventions for Behavior Change: A Holistic Framework. JMIR Res Protoc 2019 Jan 15;8(1):e8055 [FREE Full text] [doi: 10.2196/resprot.8055] [Medline: 30664477]
20. Yardley L, Spring B, Riper H, Morrison L, Crane D, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
21. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul 06;15(3):e12768 [FREE Full text] [doi: 10.1111/mcn.12768] [Medline: 30569549]
22. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Ann Behav Med 2017 Apr 18;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
23. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
24. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of Mobile Apps to Promote Health and Manage Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
25. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
26. Mano R. Mobile Health Apps and Health Management Behaviors: Cost-Benefit Modeling Analysis. JMIR Hum Factors 2021 Apr 22;8(2):e21251 [FREE Full text] [doi: 10.2196/21251] [Medline: 33885372]
27. Entsieh AA, Hallström IK. First-time parents' prenatal needs for early parenthood preparation-A systematic review and meta-synthesis of qualitative literature. Midwifery 2016 Aug;39:1-11 [FREE Full text] [doi: 10.1016/j.midw.2016.04.006] [Medline: 27321714]
28. Cashin M, Wroe J, Campbell LE. What parents want to know in the first postnatal year: A Delphi consensus study. Child Care Health Dev 2021 Jan 21;47(1):47-56. [doi: 10.1111/cch.12806] [Medline: 32881020]
29. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
30. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: A systematic literature review. J Telemed Telecare 2015 Feb 12;21(5):243-253. [doi: 10.1177/1357633x15572203]
31. Quelly SB, Norris AE, DiPietro JL. Impact of mobile apps to combat obesity in children and adolescents: A systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17. [doi: 10.1111/jspn.12134] [Medline: 26494019]
32. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
33. Lin Y, Lou M. Effects of mHealth-based interventions on health literacy and related factors: A systematic review. J Nurs Manag 2021 Apr;29(3):385-394. [doi: 10.1111/jonm.13175] [Medline: 33040429]
34. Blakeslee S, Stritter W, Horak I, Seifert G. Protocol Registration: Planting Seeds for the Future -- A Scoping Review of Digital Prevention Apps for Parents. osf. Open Science Forum. 2021 May 19. URL: https://osf.io/pn36g/ [accessed 2023-06-22]

35. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
36. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
37. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
38. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
39. The World Bank. World Bank Country and Lending Groups. The World Bank. Washington, D.C: The World Bank Group; 2021. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2021-05-20]
40. EndNote X9 [software]. Philadelphia, PA: Clarivate; 2013. URL: https://endnote.com/downloads [accessed 2023-06-22]
41. MAXQDA 2020 [software]. In: MAXQDA/VERBI. Berlin, Germany: VERBI Software; 2019.
42. Mayring P, Fenzl T. Qualitative inhaltsanalyse. In: Handbuch Methoden der empirischen Sozialforschung. Berlin, Germany: Springer; 2014:543-556.
43. Wunsch K, Eckert T, Fiedler J, Cleven L, Niermann C, Reiterer H, et al. Effects of a Collective Family-Based Mobile Health Intervention Called "SMARTFAMILY" on Promoting Physical Activity and Healthy Eating: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 Nov 11;9(11):e20534 [FREE Full text] [doi: 10.2196/20534] [Medline: 33174849]
44. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth Physical Activity Intervention: A Randomized Pilot Study in Physically Inactive Pregnant Women. Matern Child Health J 2016 May;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
45. Dodd JM, Louise J, Cramp C, Grivell RM, Moran LJ, Deussen AR. Evaluation of a smartphone nutrition and physical activity application to provide lifestyle advice to pregnant women: The SNAPP randomised trial. Matern Child Nutr 2018 Jan;14(1):e12502 [FREE Full text] [doi: 10.1111/mcn.12502] [Medline: 28836373]
46. Meixner C, Baumann H, Fenger A, Wollesen B. Gamification in health apps to increase physical activity within families. New York, NY: IEEE; 2019 Presented at: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); October 21-23, 2019; Barcelona, Spain p. 21-23. [doi: 10.1109/wimob.2019.8923332]
47. Henriksson H, Alexandrou C, Henriksson P, Henström M, Bendtsen M, Thomas K, et al. MINISTOP 2.0: a smartphone app integrated in primary child health care to promote healthy diet and physical activity behaviours and prevent obesity in preschool-aged children: protocol for a hybrid design effectiveness-implementation study. BMC Public Health 2020 Nov 23;20(1):1756 [FREE Full text] [doi: 10.1186/s12889-020-09808-w] [Medline: 33228572]
48. De Lepeleere S, De Bourdeaudhuij I, Cardon G, Verloigne M. The effect of an online video intervention 'Movie Models' on specific parenting practices and parental self-efficacy related to children's physical activity, screen-time and healthy diet: a quasi experimental study. BMC Public Health 2017 Apr 27;17(1):366 [FREE Full text] [doi: 10.1186/s12889-017-4264-1] [Medline: 28449658]
49. Trost SG, Brookes DSK. Effectiveness of a novel digital application to promote fundamental movement skills in 3- to 6-year-old children: A randomized controlled trial. J Sports Sci 2021 Feb 27;39(4):453-459. [doi: 10.1080/02640414.2020.1826657] [Medline: 32985373]
50. Sun A, Cheng J, Bui Q, Liang Y, Ng T, Chen J. Home-Based and Technology-Centered Childhood Obesity Prevention for Chinese Mothers With Preschool-Aged Children. J Transcult Nurs 2017 Nov;28(6):616-624. [doi: 10.1177/1043659617719139] [Medline: 28826348]
51. Cullen KW, Thompson D, Chen T. Outcome Evaluation of Family Eats. Health Educ Behav 2017 Feb;44(1):32-40. [doi: 10.1177/1090198116643917] [Medline: 27198535]
52. Au LE, Whaley S, Rosen NJ, Meza M, Ritchie LD. Online and In-Person Nutrition Education Improves Breakfast Knowledge, Attitudes, and Behaviors: A Randomized Trial of Participants in the Special Supplemental Nutrition Program for Women, Infants, and Children. J Acad Nutr Diet 2016 Mar;116(3):490-500. [doi: 10.1016/j.jand.2015.10.012] [Medline: 26669795]
53. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
54. Helle C, Hillesund ER, Wills AK, Øverby NC. Examining the effects of an eHealth intervention from infant age 6 to 12 months on child eating behaviors and maternal feeding practices one year after cessation: The Norwegian randomized controlled trial Early Food for Future Health. PLoS One 2019;14(8):e0220437 [FREE Full text] [doi: 10.1371/journal.pone.0220437] [Medline: 31442241]
55. Røed M, Medin AC, Vik FN, Hillesund ER, Van Lippevelde W, Campbell K, et al. Effect of a Parent-Focused eHealth Intervention on Children's Fruit, Vegetable, and Discretionary Food Intake (Food4toddlers): Randomized Controlled Trial. J Med Internet Res 2021 Feb 16;23(2):e18311 [FREE Full text] [doi: 10.2196/18311] [Medline: 33591279]

56. Russell CG, Denney-Wilson E, Laws RA, Abbott G, Zheng M, Lymer SJ, et al. Impact of the Growing Healthy mHealth Program on Maternal Feeding Practices, Infant Food Preferences, and Satiety Responsiveness: Quasi-Experimental Study. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e77 [FREE Full text] [doi: 10.2196/mhealth.9303] [Medline: 29695373]
57. Gomes AI, Pereira AI, Roberto MS, Boraska K, Barros L. Changing parental feeding practices through web-based interventions: A systematic review and meta-analysis. PLoS One 2021;16(4):e0250231 [FREE Full text] [doi: 10.1371/journal.pone.0250231] [Medline: 33909666]
58. Scott JA, Burns SK, Hauck YL, Giglia RC, Jorgensen AM, White BK, et al. Impact of a Face-To-Face Versus Smartphone App Versus Combined Breastfeeding Intervention Targeting Fathers: Randomized Controlled Trial. JMIR Pediatr Parent 2021 Apr 12;4(2):e24579 [FREE Full text] [doi: 10.2196/24579] [Medline: 33843604]
59. Pond N, Finch M, Sutherland R, Wolfenden L, Nathan N, Kingsland M, et al. Cluster randomised controlled trial of an m-health intervention in centre-based childcare services to reduce the packing of discretionary foods in children's lunchboxes: study protocol for the 'SWAP IT Childcare' trial. BMJ Open 2019 Jun 01;9(5):e026829 [FREE Full text] [doi: 10.1136/bmjopen-2018-026829] [Medline: 31154306]
60. Sutherland R, Nathan N, Brown A, Yoong S, Finch M, Lecathelinais C, et al. A randomized controlled trial to assess the potential efficacy, feasibility and acceptability of an m-health intervention targeting parents of school aged children to improve the nutritional quality of foods packed in the lunchbox 'SWAP IT'. Int J Behav Nutr Phys Act 2019 Jul 02;16(1):54 [FREE Full text] [doi: 10.1186/s12966-019-0812-7] [Medline: 31266506]
61. Laws RA, Denney-Wilson EA, Taki S, Russell CG, Zheng M, Litterbach E, et al. Key Lessons and Impact of the Growing Healthy mHealth Program on Milk Feeding, Timing of Introduction of Solids, and Infant Growth: Quasi-Experimental Study. JMIR Mhealth Uhealth 2018 Apr 19;6(4):e78 [FREE Full text] [doi: 10.2196/mhealth.9040] [Medline: 29674313]
62. Zolfaghari M, Shirmohammadi M, Shahhosseini H, Mokhtaran M, Mohebbi SZ. Development and evaluation of a gamified smart phone mobile health application for oral health promotion in early childhood: a randomized controlled trial. BMC Oral Health 2021 Jan 07;21(1):18 [FREE Full text] [doi: 10.1186/s12903-020-01374-2] [Medline: 33413304]
63. Lozoya CJS, Giblin-Scanlon L, Boyd LD, Nolen S, Vineyard J. Influence of a Smartphone Application on the Oral Health Practices and Behaviors of Parents of Preschool Children. J Dent Hyg 2019 Oct;93(5):6-14. [Medline: 31628171]
64. Leichman ES, Gould RA, Williamson AA, Walters RM, Mindell JA. Effectiveness of an mHealth Intervention for Infant Sleep Disturbances. Behav Ther 2020 Jul;51(4):548-558 [FREE Full text] [doi: 10.1016/j.beth.2019.12.011] [Medline: 32586429]
65. Deave T, Ginja S, Goodenough T, Bailey E, Piwek L, Coad J, et al. The Bumps and BaBies Longitudinal Study (BaBBLeS): a multi-site cohort study of first-time mothers to evaluate the effectiveness of the Baby Buddy app. Mhealth 2019 Sep;5:42-42 [FREE Full text] [doi: 10.21037/mhealth.2019.08.05] [Medline: 31620469]
66. Shorey S, Ng YPM, Ng ED, Siew AL, Mörelius E, Yoong J, et al. Effectiveness of a Technology-Based Supportive Educational Parenting Program on Parental Outcomes (Part 1): Randomized Controlled Trial. J Med Internet Res 2019 Feb 13;21(2):e10816 [FREE Full text] [doi: 10.2196/10816] [Medline: 30758289]
67. Ledford CJW, Womack JJ, Rider HA, Seehusen AB, Conner SJ, Lauters RA, et al. Unexpected Effects of a System-Distributed Mobile Application in Maternity Care: A Randomized Controlled Trial. Health Educ Behav 2018 Jun 16;45(3):323-330. [doi: 10.1177/1090198117732110] [Medline: 28918669]
68. Davis DW, Logsdon MC, Vogt K, Rushton J, Myers J, Lauf A, et al. Parent Education is Changing: A Review of Smartphone Apps. MCN Am J Matern Child Nurs 2017;42(5):248-256. [doi: 10.1097/NMC.0000000000000353] [Medline: 28639997]
69. McKenzie LB, Roberts KJ, Clark R, McAdams R, Abdel-Rasoul M, Klein EG, et al. A randomized controlled trial to evaluate the Make Safe Happen® app-a mobile technology-based safety behavior change intervention for increasing parents' safety knowledge and actions. Inj Epidemiol 2018 Mar 12;5(1):5 [FREE Full text] [doi: 10.1186/s40621-018-0133-3] [Medline: 29527644]
70. Roberts KJ, McAdams RJ, Kristel OV, Szymanski AM, McKenzie LB. Qualitative and Quantitative Evaluation of the Make Safe Happen App: Mobile Technology-Based Safety Behavior Change Intervention for Parents. JMIR Pediatr Parent 2019 Mar 14;2(1):e12022 [FREE Full text] [doi: 10.2196/12022] [Medline: 31518322]
71. Burgess J, Watt K, Kimble RM, Cameron CM. Combining Technology and Research to Prevent Scald Injuries (the Cool Runnings Intervention): Randomized Controlled Trial. J Med Internet Res 2018 Oct 10;20(10):e10361 [FREE Full text] [doi: 10.2196/10361] [Medline: 30305263]
72. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and Quality of Infant Feeding Smartphone Apps: Five-Year Update on a Systematic Search and Evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
73. Meedya S, Win K, Yeatman H, Fahy K, Walton K, Burgess L, et al. Developing and testing a mobile application for breastfeeding support: The Milky Way application. Women Birth 2021 Mar;34(2):e196-e203 [FREE Full text] [doi: 10.1016/j.wombi.2020.02.006] [Medline: 32081557]
74. Mauch CE, Laws RA, Prichard I, Maeder AJ, Wycherley TP, Golley RK. Commercially Available Apps to Support Healthy Family Meals: User Testing of App Utility, Acceptability, and Engagement. JMIR Mhealth Uhealth 2021 May 07;9(5):e22990 [FREE Full text] [doi: 10.2196/22990] [Medline: 33960951]

75. Militello L, Sezgin E, Huang Y, Lin S. Delivering Perinatal Health Information via a Voice Interactive App (SMILE): Mixed Methods Feasibility Study. JMIR Form Res 2021 Mar 01;5(3):e18240 [FREE Full text] [doi: 10.2196/18240] [Medline: 33646136]
76. Taki S, Russell CG, Lymer S, Laws R, Campbell K, Appleton J, et al. A Mixed Methods Study to Explore the Effects of Program Design Elements and Participant Characteristics on Parents' Engagement With an mHealth Program to Promote Healthy Infant Feeding: The Growing Healthy Program. Front Endocrinol (Lausanne) 2019;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]
77. Biviji R. Evaluating the Quality of Mobile Health Apps for Maternal and Child Health (MCH). Indianapolis, IA: Indiana University-Purdue University Indianapolis; 2018.
78. Cushing CC, Fedele DA, Brannon EE, Kichline T. Parents' Perspectives on the Theoretical Domains Framework Elements Needed in a Pediatric Health Behavior App: A Crowdsourced Social Validity Study. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e192 [FREE Full text] [doi: 10.2196/mhealth.9808] [Medline: 30578173]
79. Knowlden A, Sharma M. One-Year Efficacy Testing of Enabling Mothers to Prevent Pediatric Obesity Through Web-Based Education and Reciprocal Determinism (EMPOWER) Randomized Control Trial. Health Educ Behav 2016 Feb;43(1):94-106. [doi: 10.1177/1090198115596737] [Medline: 26272782]
80. Burgess JD, Watt KA, Kimble RM, Cameron CM. Knowledge of childhood burn risks and burn first aid: Cool Runnings. Inj Prev 2019 Aug;25(4):301-306. [doi: 10.1136/injuryprev-2017-042650] [Medline: 29386371]
81. Laws RA, Litterbach EV, Denney-Wilson EA, Russell CG, Taki S, Ong K, et al. A Comparison of Recruitment Methods for an mHealth Intervention Targeting Mothers: Lessons from the Growing Healthy Program. J Med Internet Res 2016 Sep 15;18(9):e248 [FREE Full text] [doi: 10.2196/jmir.5691] [Medline: 27634633]
82. Taki S, Russell CG, Wen LM, Laws RA, Campbell K, Xu H, et al. Consumer Engagement in Mobile Application (App) Interventions Focused on Supporting Infant Feeding Practices for Early Prevention of Childhood Obesity. Front Public Health 2019;7:60 [FREE Full text] [doi: 10.3389/fpubh.2019.00060] [Medline: 30984732]
83. Otte RA, van Beukering AJE, Boelens-Brockhuis L. Tracker-Based Personal Advice to Support the Baby's Healthy Development in a Novel Parenting App: Data-Driven Innovation. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e12666 [FREE Full text] [doi: 10.2196/12666] [Medline: 31342901]
84. Venezia AP, Wirth CK, Vinci DM. Active Families at Home: The Development of a Let’s Wiggle With 5-2-1-0 App. TPE 2019;76(1):224-237. [doi: 10.18666/tpe-2019-v76-i1-8423]
85. Biviji R, Vest J, Dixon B, Cullen T, Harle C. Content analysis of behavior change techniques in maternal and infant health apps. Transl Behav Med 2021 Mar 16;11(2):504-515. [doi: 10.1093/tbm/ibaa039] [Medline: 32491165]
86. Helle C, Hillesund ER, Omholt ML, Øverby NC. Early food for future health: a randomized controlled trial evaluating the effect of an eHealth intervention aiming to promote healthy food habits from early childhood. BMC Public Health 2017 Sep 20;17(1):729 [FREE Full text] [doi: 10.1186/s12889-017-4731-8] [Medline: 28931384]
87. Meinert E, Rahman E, Potter A, Lawrence W, Van Velthoven M. Acceptability and Usability of the Mobile Digital Health App NoObesity for Families and Health Care Professionals: Protocol for a Feasibility Study. JMIR Res Protoc 2020 Jul 22;9(7):e18068 [FREE Full text] [doi: 10.2196/18068] [Medline: 32706703]
88. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
89. Mohebbi SZ, Rabiei S, Yazdani R, Nieminen P, Virtanen JI. Evaluation of an educational intervention in oral health for primary care physicians: a cluster randomized controlled study. BMC Oral Health 2018 Dec 14;18(1):218 [FREE Full text] [doi: 10.1186/s12903-018-0676-2] [Medline: 30547799]
90. S B, Peedikayil FC, Kottayi S, Narasimhan D. The knowledge, attitude and practice regarding the pediatric preventive dental care amongst medical practitioners: a cross sectional study. J. res. dent 2017 Nov 15;5(1):11. [doi: 10.19177/jrd.v5e1201611-15]
91. Jain R, Oswal K, Chitguppi R. Knowledge, attitude and practices of mothers toward their children’s oral health: A questionnaire survey among subpopulation in Mumbai (India). J Dent Res Sci Develop 2014;1(2):40-45. [doi: 10.4103/2348-3407.135073]
92. Van den Branden S, Van den Broucke S, Leroy R, Declerck D, Hoppenbrouwers K. Predicting oral health-related behaviour in the parents of preschool children: An application of the Theory of Planned Behaviour. Health Education Journal 2014 May 05;74(2):221-230. [doi: 10.1177/0017896914530585]
93. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
94. Pearson N, Finch M, Sutherland R, Kingsland M, Wolfenden L, Wedesweiler T, et al. An mHealth Intervention to Reduce the Packing of Discretionary Foods in Children's Lunch Boxes in Early Childhood Education and Care Services: Cluster Randomized Controlled Trial. J Med Internet Res 2022 Mar 17;24(3):e27760 [FREE Full text] [doi: 10.2196/27760] [Medline: 35297768]

95. Fiedler J, Seiferth C, Eckert T, Woll A, Wunsch K. Sleep quality, valence, energetic arousal, and calmness as predictors of device-based measured physical activity during a three-week mHealth intervention. Ger J Exerc Sport Res 2022 Apr 14;52(2):237-247. [doi: 10.1007/s12662-022-00809-y]
96. Alexandrou C, Henriksson H, Henström M, Henriksson P, Delisle Nyström C, Bendtsen M, et al. Effectiveness of a Smartphone App (MINISTOP 2.0) integrated in primary child health care to promote healthy diet and physical activity behaviors and prevent obesity in preschool-aged children: randomized controlled trial. Int J Behav Nutr Phys Act 2023 Feb 21;20(1):22 [FREE Full text] [doi: 10.1186/s12966-023-01405-5] [Medline: 36810069]
97. Lachman P. Where to make a difference: research and the social determinants in pediatrics and child health in the COVID-19 era. Pediatr Res 2021 Jan 10;89(2):259-262 [FREE Full text] [doi: 10.1038/s41390-020-01253-0] [Medline: 33173180]
98. Grand View Research. Digital Health Market Size, Share & Trends Analysis Report By Technology (Healthcare Analytics, mHealth, Tele-healthcare, Digital Health Systems), By Component (Software, Hardware, Services), By Region, And Segment Forecasts, 2023-2030. San Francisco, CA: Grand View Research; Nov 2022:1-100.
99. Alexander KE, Brijnath B, Biezen R, Hampton K, Mazza D. Preventive healthcare for young children: A systematic review of interventions in primary care. Prev Med 2017 Jun;99:236-250. [doi: 10.1016/j.ypmed.2017.02.024] [Medline: 28279679]
100. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res 1988 Jun;23(2):203-235 [FREE Full text] [Medline: 3384669]
101. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
102. Denney-Wilson E, Laws R, Russell CG, Ong K, Taki S, Elliot R, et al. Preventing obesity in infants: the Growing healthy feasibility trial protocol. BMJ Open 2015 Nov 30;5(11):e009258 [FREE Full text] [doi: 10.1136/bmjopen-2015-009258] [Medline: 26621519]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
PES: Tool to Measure Parenting Self-Efficacy
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews
RIS: research information system
zBMI: sex- and age-standardized BMI

Edited by L Buis; submitted 27.05.22; peer-reviewed by A Gomes, MS Liew; comments to author 25.07.22; revised version received 12.11.22; accepted 02.05.23; published 20.07.23.

Please cite as:
Blakeslee SB, Vieler K, Horak I, Stritter W, Seifert G
Planting Seeds for the Future: Scoping Review of Child Health Promotion Apps for Parents
JMIR Mhealth Uhealth 2023;11:e39929
URL: https://mhealth.jmir.org/2023/1/e39929
doi:10.2196/39929
PMID:37471125

Review

Stress Management Apps: Systematic Search and Multidimensional Assessment of Quality and Characteristics

Sarah Paganini[1], PhD; Evelyn Meier[2], MSc; Yannik Terhorst[3], MSc; Ramona Wurst[1], PhD; Vivien Hohberg[4], MA; Dana Schultchen[5], PhD; Jana Strahler[1], Prof Dr; Max Wursthorn[6], MSc; Harald Baumeister[3], Prof Dr; Eva-Maria Messner[3], PhD

[1]Department of Sport Psychology, Institute of Sports and Sport Science, University of Freiburg, Freiburg, Germany

[2]University of Education Freiburg, Freiburg, Germany

[3]Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany

[4]Department of Sport, Exercise and Health, Faculty of Medicine, University of Basel, Basel, Switzerland

[5]Clinical and Health Psychology, Ulm University, Ulm, Germany

[6]Department of Public and Nonprofit Management, University of Freiburg, Freiburg, Germany

Corresponding Author:
Sarah Paganini, PhD
Department of Sport Psychology
Institute of Sports and Sport Science
University of Freiburg
Sandfangweg 4
Freiburg, 79102
Germany
Phone: 49 76120345
Email: sarah.paganini@sport.uni-freiburg.de

Abstract

Background: Chronic stress poses risks for physical and mental well-being. Stress management interventions have been shown to be effective, and stress management apps (SMAs) might help to transfer strategies into everyday life.

Objective: This review aims to provide a comprehensive overview of the quality and characteristics of SMAs to give potential users or health professionals a guideline when searching for SMAs in common app stores.

Methods: SMAs were identified with a systematic search in the European Google Play Store and Apple App Store. SMAs were screened and checked according to the inclusion criteria. General characteristics and quality were assessed by 2 independent raters using the German Mobile Application Rating Scale (MARS-G). The MARS-G assesses quality (range 1 to 5) on the following four dimensions: (1) engagement, (2) functionality, (3) esthetics, and (4) information. In addition, the theory-based stress management strategies, evidence base, long-term availability, and common characteristics of the 5 top-rated SMAs were assessed and derived.

Results: Of 2044 identified apps, 121 SMAs were included. Frequently implemented strategies (also in the 5 top-rated SMAs) were psychoeducation, breathing, and mindfulness, as well as the use of monitoring and reminder functions. Of the 121 SMAs, 111 (91.7%) provided a privacy policy, but only 44 (36.4%) required an active confirmation of informed consent. Data sharing with third parties was disclosed in only 14.0% (17/121) of the SMAs. The average quality of the included apps was above the cutoff score of 3.5 (mean 3.59, SD 0.50). The MARS-G dimensions yielded values above this cutoff score (functionality: mean 4.14, SD 0.47; esthetics: mean 3.76, SD 0.73) and below this score (information: mean 3.42, SD 0.46; engagement: mean 3.05, SD 0.78). Most theory-based stress management strategies were regenerative stress management strategies. The evidence base for 9.1% (11/121) of the SMAs could be identified, indicating significant group differences in several variables (eg, stress or depressive symptoms) in favor of SMAs. Moreover, 38.0% (46/121) of the SMAs were no longer available after a 2-year period.

Conclusions: The moderate information quality, scarce evidence base, constraints in data privacy and security features, and high volatility of SMAs pose challenges for users, health professionals, and researchers. However, owing to the scalability of SMAs and the few but promising results regarding their effectiveness, they have a high potential to reach and help a broad audience. For a holistic stress management approach, SMAs could benefit from a broader repertoire of strategies, such as more instrumental and mental stress management strategies. The common characteristics of SMAs with top-rated quality can be used as guidance for potential users and health professionals, but owing to the high volatility of SMAs, enhanced evaluation frameworks are needed.

(JMIR Mhealth Uhealth 2023;11:e42415) doi:10.2196/42415

KEYWORDS

stress management; mobile app; mHealth; mobile health; quality assessment; review; evidence base; availability

Introduction

Stress is a public health problem that poses high risks for physical and mental well-being and is increasing in industrial societies where individuals are exposed to complex demands at work and in daily life [1-5]. The results of an American survey revealed that 75% of the participants felt significantly stressed [6], and a representative sample of the German population showed a point prevalence of perceived high chronic stress of 11% [1]. There are multiple reasons for the broad impact of stress as it affects many dimensions, such as cognition (eg, negative attributional style), affect (eg, affective dysregulation, such as increase in anxiety), physiology (eg, dysregulation of the endocrine response system), and behavior (eg, harmful behavioral changes, such as smoking or physical inactivity) [3]. As a result, chronic stress causes a higher risk for various somatic diseases and mental disorders, such as gastric ulcers, migraine, hypertension, type 2 diabetes mellitus, and depression [3,5,7-11]. In addition to the substantial impact on health, work-related stress results in high costs for society, especially through productivity-related losses [12].

Due to these negative consequences, several stress management strategies have been developed and evaluated over the past decades [13,14]. Most of them refer to transactional stress models in which stress reactions are mainly determined by a subjective interpretation and the types of coping strategies employed [13-16]. Even though “stress management” is a widely and variably used term [17], Kaluza [13] proposed three categories of strategies: (1) instrumental stress management strategies with a focus on preventing and reducing stress in everyday life (eg, self-management or seeking support); (2) mental stress management strategies aiming at changing personal stress amplifiers (eg, acceptance or gratitude); and (3) regenerative stress management strategies aiming at recovery after stress exposure (eg, relaxation techniques or health behavior) [17-19]. Thereby, effective stress management seems to be characterized by a broad repertoire and a balance between instrumental, mental, and regenerative strategies [15]. Interventions use a variety of these different strategies and are often delivered in a group setting [17,18]. In particular, cognitive or behavioral-based interventions for stress management (which typically include all 3 categories of strategies) have been shown to be effective for reducing stress in different settings (eg, the occupational setting; Cohen d=1.16, 95% CI 0.46-1.87 [20]) and for different target groups (eg, university students; standardized mean difference=−0.77, 95% CI −0.97 to −0.57 [21]). The same is true for mindfulness-based stress reduction in healthy individuals [22-24] (eg, Hedges g=0.53, 95% CI 0.41-0.64 [22]). Relaxation training (with a focus on regenerative strategies) has been shown to be effective in healthy individuals [24] and in occupational settings (Cohen d=0.50, 95% CI 0.31-0.69 [20]) but appears to be inferior to cognitive-behavioral interventions [20]. Implementing previously learned health-related strategies in daily life is essential for their short- and long-term health benefits [25]. Internet- and mobile-based health interventions can help to integrate stress management strategies into daily routines and to overcome the barriers of face-to-face interventions, such as limited accessibility, location, time, and high costs [26,27]. As a result, the relevance of mobile phones for monitoring and delivering health interventions has increased over the last decade [27]. From 2013 to 2018, the number of downloaded health apps per year increased from 1.7 to 4.1 billion worldwide [28]. Regarding stress management apps (SMAs), about 6% of an American sample reported that they already use SMAs regularly and about 50% could imagine using them in the future [29]. An observational study showed that compared to a website, delivering a stress-management intervention via an app could offer the added benefit of more frequent use and access to more intervention content [30].

Accordingly, there is a broad and growing body of SMA research. Reviews already exist; however, they all focus on specific aspects and some might be outdated. Pre-existing SMA reviews focus on content alone [31,32], content in combination with transparency and functionality [33], efficacy [34], gamification elements [35], persuasive and behavior change strategies [36], or quality of apps, with a focus exclusively on mindfulness apps [37]. Regarding content, it was shown that mindfulness and meditation were the most commonly used strategies in the reviewed SMAs (34% to 78% of all apps included these strategies) [31,33,34], followed by breathing [31,33] or goal setting [34]. Further common strategies were personalization and self-monitoring, while social support strategies were rarely used [36]. The implementation of gamification elements is relatively scarce (on average 0.5 elements per app) [35]. Concerning the evidence base, Lau et al [34] revealed that among more than 1000 screened apps for well-being and stress management, only 2% were scientifically evaluated. The 2 studies that looked at data privacy and security features, such as privacy policy, contact information, and disclosures, revealed that only half of these criteria could be met on average [33]. In addition, most of the evaluated apps showed a lack of data privacy and security [37]. This confirms the results of other health, wellness, and medicine-related apps (eg, smoking cessation and diabetes) [38-40], showing major privacy and security risks, missing transparency, or data sharing with third parties, even when they were accredited [41].

Considering the quality of SMAs, only one of the existing reviews (which was exclusively performed for mindfulness apps and not for SMAs in general) [37] employed a valid scientific measure, that is, the Mobile Application Rating Scale (MARS), which is an instrument for assessing app quality on multiple dimensions [42,43]. This is of relevance as user star ratings and app store descriptions can be manipulated in favor of commercial interests [33].

Another challenge for users and health professionals who are seeking apps for long-term use, as well as for researchers aiming to present the most recent state of research, is the high volatility of apps [44,45]. In a study considering mental health apps, only 50% of the search results were available at the end of a 9-month period [44]. Considering the excessive supply of health apps in app stores, their high update rate, and their uncertain long-term availability, as well as the current lack of transparency of app quality [46], the question arises as to which SMAs should be used and recommended.

In light of all these gaps and issues, the aim of this study was to systematically search for SMAs, to assess their quality on multiple dimensions in a scientific manner, and to give a comprehensive overview of SMAs concerning their general characteristics, theory-based stress management strategies, evidence base, and long-term availability. A further aim was to inform potential users or health professionals about common characteristics that might indicate high quality of SMAs. The following research questions were addressed:

1. What are the general characteristics of SMAs, such as descriptive information, technical aspects, strategies, and functions?
2. What is the quality of SMAs regarding multiple dimensions (ie, engagement, functionality, esthetics, and information)?
3. Which theory-based stress management strategies are used in SMAs?
4. What is the evidence base of SMAs?
5. How reliable are SMAs in terms of their long-term availability?
6. What are the common characteristics of SMAs with top-rated quality?

Methods

Overview

This study involved a systematic search and assessment of the quality and characteristics of SMAs. It was registered in the Open Science Framework (OSF) of the Center for Open Science [47].

Search Strategy and Procedure

The search terms were generated through a 3-step process. First, a narrative literature search was conducted to collect terms and keywords that were used in studies focusing on SMA interventions for the general population. Second, relevant search terms were identified based on interest group interviews with 3 psychotherapists and 3 potential SMA users. Third, the identified search terms from the literature search and results of the interest group interviews were merged, leading to the following search terms: “stress,” “stress management,” “stress reduction,” “stress prevention,” “stress coach,” “stress recovery,” “relaxation,” and “relaxation training.” An automated search using these search terms (in English and German language) was conducted in the European Apple App Store and Google Play Store with a search engine (web crawler). It was developed as part of the Mobile Health App Database Project [48], and it automatically extracts information, such as app name, description, and user rating, from the stores (for further details, please see [49]).

Apps from both app stores were identified and listed in a central database. Duplicates were automatically removed. In the first step and based on the description in the app stores, apps were screened for the following inclusion criteria: (1) the word “stress” was included in the title or in the app store description; (2) the app was developed for adults in the general population without mental or somatic disorders; (3) the focus of the app was primarily on stress management; (4) at least two different stress management strategies were applied with the aim of including apps that potentially take a holistic approach to stress management; (5) the app could be used without further equipment, devices, or programs; (6) the app was free of cost in the basic version; and (7) the app was provided in German or English language. In the second step, the app was downloaded and rechecked for criteria 1 to 7. Apps that did not work after the download were excluded.

Data Collection of General Characteristics and Quality Assessment

The general characteristics and quality of each SMA were collected and rated between March and May 2020 by 2 independent raters (EM, SP, Hannah Besel, RW, or VH) with the German Mobile Application Rating Scale (MARS-G) [42]. All raters had a psychology or sports science degree and completed an online training, which included the following components: (1) background information on the development of the MARS-G; (2) description of the dimensions and items; (3) application instructions; and (4) an exercise example [50]. Subsequently, 3 SMAs were assessed per rater in order to compare and discuss the results and to ensure a common standard as well as high data quality. According to the standardized procedure, each rater had to test an app for at least 15 minutes. Data collection and quality assessment, including the actual average time spent per rating, have been fully documented.

General Characteristics

General characteristics were mainly based on the classification section of the original MARS and MARS-G [42,43] and included (1) descriptive information on the app (ie, app name, URL, platform, user star rating, full version price, content-related app category, declared aims of the app, theoretical background, and certification); (2) technical aspects (eg, links to social media or type of support); (3) strategies (eg, relaxation or goal setting); and (4) functions (eg, feedback or reminder).

Owing to the high relevance of data privacy and security, the list of the MARS-G [42] has been supplemented with 7 additional features (ie, passive informed consent; complex passwords; anonymization or pseudonymization; creation of an access token; automatic display of the privacy policy; permanent availability of the privacy policy; and transparency regarding the right of withdrawal [51,52]). The privacy policy of each included SMA was reviewed against the listed features. It was assessed whether information was provided for each feature ("yes" or "no"), but not whether it was technically and legally realized and complied with by the respective providers.

Quality Assessment

The multidimensional quality evaluation based on the MARS-G [42] comprised 4 subscales: user engagement (5 items: entertainment, interest, customization, interactivity, and target group), functionality (4 items: performance, ease of use, navigation, and gestural design), esthetics (3 items: layout, graphics, and visual appeal), and information quality (7 items: accuracy of app description, goals, quality of information, quantity of information, quality of visual information, credibility, and evidence base). For the interpretation, the cutoff score of 3.5 (indicating above-average quality) defined by Terhorst et al [53] was used.

Additionally, the subjective quality (4 items: expected use frequency within 1 year, willingness to pay for the app, willingness to recommend the app, and subjective star rating) and the perceived impact on the user (6 items: awareness, knowledge, attitudes, intention to change, help-seeking, and behavioral change) were assessed. All items were rated on a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). The MARS is a well-validated instrument [43,54]. The validation of the German version also yielded excellent internal consistency (ω=0.84, 95% CI 0.77-0.88) and high levels of interrater reliability (interclass correlation [ICC]=0.83, 95% CI 0.82-0.85 [42]).

Theory-Based Stress Management Strategies

To depict the variety of existing stress management strategies in more detail, we developed a list of theory-based stress management strategies [13,31]. This list contains 23 instrumental, mental, and regenerative stress management strategies (see Textbox 1). Some of the included theory-based strategies (eg, breathing, hypnosis, and mindfulness) overlapped with the strategies covered in the MARS-G. The assessment of theory-based stress management strategies was performed for each SMA by 2 independent raters.

Textbox 1. List of theory-based stress management strategies (adapted from Kaluza [13] and Christmann et al [31]).

Instrumental stress management strategies

- Enhancing professional competencies (eg, learning)
- Seeking support (eg, network)
- Developing social-communicative skills (eg, self-assertion)
- Self-management

Mental stress management strategies

- Accepting reality (also included in the German Mobile Application Rating Scale [MARS-G])
- Seeing difficulties as challenges (not as threats)
- Changing personal stress amplifiers
- Self-efficacy

Regenerative stress management strategies

- Acupressure
- Autogenic training (the MARS-G includes the category "relaxation," which is differentiated in more detail here)
- Biofeedback
- Breathing (also included in the MARS-G)
- Euthymic methods
- Food or nutrition
- Guided imagination or visualization
- Hypnosis or self-hypnosis (also included in the MARS-G)
- Meditation or mindfulness (also included in the MARS-G)
- Music
- Muscle relaxation (the MARS-G includes the category "relaxation," which is differentiated in more detail here)
- Physical stress relief techniques
- Self-massage
- Sounds
- Sport (also included in the MARS-G)

Evidence Base of Included SMAs and Long-Term Availability

All included SMAs were unsystematically searched in a common web search engine for scholarly literature by applying the app name and screening the first pages of the results. Information on study design, app usage in weeks, sample and target groups, age, gender, measurement time points, measured variables, and main results were assessed from the studies found (with the exception of pilot studies).

In terms of long-term availability, all included SMAs were searched again in the app stores in August 2022. It was checked whether the app was still available (on the original platform), when the last update was made, and whether the basic version was still free of cost.

Characteristics of the 5 Top-Rated SMAs

Owing to the multitude of information, a concise overview of the common characteristics of the 5 top-rated SMAs (based on the MARS-G overall quality score) has been provided. This overview contains information on quality ratings, technical aspects, strategies and functions (all derived from the MARS-G), theory-based stress management strategies, evidence base, and long-term availability.

Data Analyses

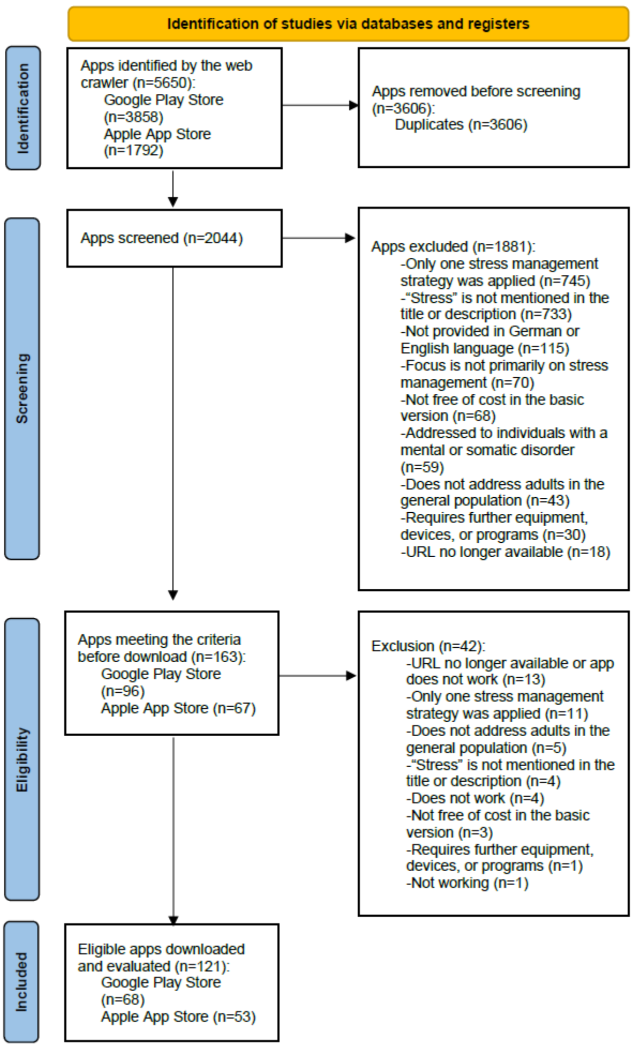
To ensure consistency between raters, the ICC (2-way mixed) was calculated according to the report by Koo and Li [55]. An ICC below 0.50 is considered poor, 0.51 to 0.75 is moderate, 0.76 to 0.89 is good, and above 0.90 is excellent [56]. For all descriptive data (such as aims, background, and data security features), frequency and percentage were calculated. The mean score and standard deviation have been presented for each dimension of the MARS-G. All analyses were performed using IBM SPSS Statistics (Version 21; IBM Corp).

Results

Search Results

The web crawler identified 5650 potential SMAs (Google Play Store, n=3580; Apple App Store, n=1792). After removing duplicates, 2044 apps were screened. This screening resulted in 163 apps, of which 121 were eligible for inclusion after the download (Figure 1 [57]). On average, each SMA was used and evaluated for 30 minutes by each rater (mean 30.2, SD 4.0 minutes).

Figure 1. Flowchart according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement 2020.



General Characteristics and Quality Rating

Descriptive Information

Of the 121 included SMAs, 68 (56.2%) were derived from the Google Play Store and 53 (43.8%) from the Apple App Store. The user star rating in the app stores could be identified for 96 (79.3%) apps. The mean user star rating was 4.27 (SD 0.56), and the number of ratings per app ranged from 1 to 126,183. Of all rated SMAs, 83 (68.6%) could be upgraded to a premium version (from 1 month [with costs between 0.99 EUR and 15.99 EUR] up to a permanent upgrade [with costs between 1.09 EUR and 449.99 EUR]; 1 EUR=1.1197 USD). Regarding content-related categories, most apps were listed under "health and fitness" (107/121, 88.4%). Further assigned categories were "lifestyle" (6/121, 5.0%), "medicine" (6/121, 5.0%), and others (3/121, 2.5%; including "learning," "entertainment," and "audio/music"). According to the description, all apps aimed at reducing stress (121/121, 100%). Further aims were improvement of well-being (113/121, 93.4%), reduction of anxiety (90/121, 74.4%), improvement of physical health (50/121, 41.3%), and emotion regulation (58/121, 47.9%). Some SMAs focused on the reduction of depressive symptoms (33/121, 27.3%), behavioral change (21/121, 17.4%), and entertainment (3/121, 2.5%). Moreover, 81 (66.9%) apps reported additional goals, such as relaxation, increasing motivation and focus, improvement of sleep, and concentration or self-awareness. The most often assigned theoretical background was third-wave behavioral therapy (106/121, 87.6%), followed by behavioral therapy (30/121, 24.8%) and cognitive behavioral therapy (24/121, 19.8%). Overall, more than 100 SMAs (111/121, 91.7%) were developed with a commercial background, 5 (4.1%) were developed by a nongovernmental organization, and only few SMAs were developed by a university (2/121, 1.7%) or a governmental institution (1/121, 0.8%). No SMA was certified according to the European Union Medical Device Regulation [58].

Technical Aspects

Data exchange with other users (eg, via social media) was possible in 40 (33.1%) SMAs, and an app community existed in 20 (16.5%) SMAs. SMAs were unguided (74/121, 61.2%), technically guided (70/121, 57.9%), asynchronously guided by humans (4/121, 3.3%), or synchronously guided by humans (1/121, 0.8%). All data privacy and security features are presented in Table 1. The 3 most common features were provision of privacy policy (111/121, 91.7%), provision of contact details or imprint (111/121, 91.7%), and passive informed consent (80/121, 66.1%).

Table 1. Frequency of declared data privacy and security features based on the German Mobile Application Rating Scale [42].[a]

Data privacy and security feature	Number of apps that specify this feature (N=121), n (%)
Provision of a privacy policy	111 (91.7)
Contact or imprint	111 (91.7)
Passive informed consent[b]	80 (66.1)
Automatic display of the privacy policy[b]	76 (62.8)
Allows password protection	74 (61.1)
Requires login	66 (54.5)
Security of data transfer	64 (52.9)
Permanent availability of the privacy policy[b]	55 (45.5)
Active confirmation of informed consent	44 (36.4)
Financial background/conflict of interest	42 (34.7)
Transparency regarding the right of withdrawal[b]	34 (28.1)
Creation of an access token[b]	26 (21.5)
Complex passwords[b]	25 (20.7)
Data sharing with third parties	17 (14.0)
Security strategies in case of device loss	17 (14.0)
Emergency function	8 (6.6)
Anonymization or pseudonymization[b]	3 (2.5)
Place of storage	2 (1.7)

[a]A more descriptive presentation of the data can be found in Multimedia Appendix 1.

[b]Additional feature that has been added to the original list.

Strategies

As shown in Table 2, more than half of all 121 SMAs included the strategies breathing (95/121, 78.5%), relaxation (94/121, 77.7%), mindfulness or gratitude (91/121, 75.2%), information or education (83/121, 68.6%), and tips or advice (70/121, 57.9%).

Table 2. Frequency of the implemented strategies according to the German Mobile Application Rating Scale.[a]

Strategy	Number of apps that include the strategy (N=121), n (%)
Breathing	95 (78.5)
Relaxation	94 (77.7)
Mindfulness/gratitude	91 (75.2)
Information, education	83 (68.6)
Tips, advice	70 (57.9)
Acceptance	49 (40.5)
Physical exercises	38 (31.4)
Gamification	33 (27.3)
Skills, training	30 (24.8)
Resource orientation	24 (19.8)
Goal setting	22 (18.2)
Serious games	5 (4.1)
Hypnosis	4 (3.3)
Exposure	0 (0.0)

[a]A more descriptive presentation of the data can be found in Multimedia Appendix 2.

Functions

The most included function was monitoring or tracking (78/121, 64.5%), followed by reminder (76/121, 62.8%), data collection (49/121, 40.5%), feedback (48/121, 39.7%), and tailored intervention or real-time feedback (22/121, 18.2%).

Quality Rating

The agreement between raters was good (ICC=0.82, 95% CI 0.81-0.82). The mean overall quality score for the SMAs was 3.59 (SD 0.50; range 2.13-4.37), indicating acceptable to good quality exceeding the cutoff score of 3.5. The mean scores of the different dimensions were as follows: engagement, 3.05 (SD 0.78; range 1.40-4.60); functionality, 4.14 (SD 0.47; range 2.63-5.00); esthetics, 3.76 (SD 0.73; range 1.50-5.00); and information, 3.42 (SD 0.46; range 1.00-4.00). The mean score for subjective quality was 2.53 (SD 0.78; range 1.00-4.00) and that for perceived impact on the user was 2.64 (SD 0.77; range 1.25-4.25). The quality ratings of all SMAs can be found in Multimedia Appendix 3.

Theory-Based Stress Management Strategies

The most common theory-based stress management strategies were meditation or mindfulness (80/121, 66.1%), breathing (61/121, 50.4%), music (31/121, 25.6%), guided imagination or visualization (26/121, 21.5%), accepting reality (25/121, 20.7%), and enhancing professional competencies (21/121, 17.4%). In the 121 SMAs, theory-based stress management strategies were included 355 times. The most implemented strategies were regenerative stress management strategies (on average, each strategy [n=15] was mentioned 17 times and implemented 248 times, ie, 70%), followed by mental stress management strategies (on average, each strategy [n=4] was mentioned 14 times and implemented 57 times, ie, 16%) and instrumental stress management strategies (on average, each strategy [n=4] was mentioned 13 times and implemented 50 times, ie, 14%). Table 3 shows the frequencies of the investigated theory-based stress management strategies.

Table 3. Theory-based stress management strategies according to Kaluza [13] and Christmann et al [31].[a]

Theory-based stress management strategy	Number of apps that include the strategy (N=121), n (%)
Regenerative stress management strategy	
Meditation or mindfulness[b]	80 (66.1)
Breathing[b]	61 (50.4)
Music	31 (25.6)
Guided imagination or visualization	26 (21.5)
Sounds	20 (16.5)
Food and nutrition	8 (6.6)
Muscle relaxation	7 (5.8)
Sport[b]	3 (2.5)
Techniques for physical stress relief	3 (2.5)
Autogenic training	3 (2.5)
Self-massage	2 (1.7)
Hypnosis or self-hypnosis[b]	2 (1.7)
Euthymic methods	1 (0.8)
Acupressure	1 (0.8)
Biofeedback	0 (0.0)
Mental stress management strategy	
Accepting reality[b]	25 (20.7)
Seeing difficulties as challenges (not as threats)	17 (14.0)
Self-efficacy	8 (6.6)
Changing personal stress amplifiers	7 (5.8)
Instrumental stress management strategy	
Enhancing professional competencies	21 (17.4)
Self-management	18 (14.9)
Developing social-communicative skills	6 (5.0)
Seeking support	5 (4.1)

[a]A more descriptive presentation of the data can be found in Multimedia Appendix 4.

[b]Theory-based stress management strategy that is also included in the German Mobile Application Rating Scale.

Evidence Base and Long-Term Availability

Scientific evaluations could be found in 11 (9.1%) of the 121 apps. Study designs varied and included randomized controlled trials (n=5), a partially randomized trial (n=1), a panel study (n=1), and pilot studies (n=3). One app was tested for its quality, and the results were summarized in a published conference paper (n=1). Target groups were university students (n=5), the general population (n=2), employed individuals (n=1), caregivers (n=1), adults with mild to moderate anxiety or depression (n=1), and nurses (n=1). Different health outcome variables were studied. In the 5 randomized controlled trials, significant group differences at postintervention in favor of the app could be found for the variables stress (n=2), self-efficacy (n=2), mindfulness (n=2), anxiety symptoms (n=4), and depression symptoms (n=4). Details of the evaluations (excluding the pilot studies and the conference paper) can be found in Multimedia Appendix 5 [59-77].

Two years after screening, 46 (38.0%) of the 121 SMAs were no longer available in the 2 app stores. Three apps did not exist anymore in English and were only available in another language (German). Nine apps were only available through other platforms that had less stringent review procedures compared with the official app stores. Among the 75 SMAs that were still accessible, 10 apps now had costs even in the basic version. Of the 121 SMAs, 46 (38.0%) had their last update in 2022 and 11 (9.1%) had not been updated since 2020.

Characteristics of the 5 Top-Rated SMAs

The 5 apps with the highest overall MARS-G ratings are presented in Table 4. None of these apps was developed by a public institution (such as a government or university). However, in all apps, it was emphasized that they were developed by different experts (such as psychologists, psychotherapists, and neuroscientist), or researchers with experience in mindfulness, meditation, or coaching. Furthermore, 3 of these 5 SMAs were part of scientific studies (study design: randomized controlled trial). All apps integrated different forms of psychoeducation and information via text or audio, provided advice, and implemented breathing and mindfulness. Monitoring or tracking and reminders were also included in all apps. Additional theory-based stress management strategies were guided imagination or visualization and music. All apps were technically guided. Moreover, 3 of the 5 apps were tailored to the users' needs based on a screening at the beginning or provided real-time feedback. Additionally, 1 SMA included messenger coaching and a contact list with therapists in different US states. Furthermore, 3 of the 5 apps offered an app community. All apps provided a specific login area (including password), privacy policy, and contact information or imprint. Moreover, 2 of the 5 apps offered an emergency function.

Table 4. Overview of the 5 top-rated apps.

Variable	App[a]				
	App 1	App 2[b]	App 3	App 4	App 5
Overall quality (MARS-G[c])	4.36	4.22	4.21	4.19	4.16
Quality dimensions (MARS-G)					
Engagement	4.30	4.10	3.80	4.00	4.30
Functionality	4.63	4.38	3.50	4.75	4.25
Esthetics	4.50	4.83	4.83	4.17	4.50
Information	4.00	3.57	3.71	3.86	3.57
Technical aspects (MARS-G)[d]					
Privacy and security features[e]	10	10	8	10	8
Technical guidance	✓	✓	✓	✓	✓
Tailored interventions, real time feedback	✓	✓		✓	
App community	✓	✓			✓
Strategies (MARS-G)[d]					
Information, education	✓	✓	✓	✓	✓
Tips, advice	✓	✓	✓	✓	✓
Breathing[f]	✓	✓	✓	✓	✓
Mindfulness, gratitude[f]	✓	✓	✓	✓	✓
Relaxation[f]		✓	✓	✓	
Acceptance[f]	✓		✓		✓
Functions (MARS-G)[d]					
Monitoring, tracking	✓	✓	✓	✓	✓
Reminder	✓	✓	✓	✓	✓
Data collection	✓	✓		✓	
Automated feedback	✓	✓		✓	✓
Theory-based stress management strategies[g]					
Guided imagination, visualization (RSMS[h])	✓				✓
Music (RSMS)		✓			✓
Evidence base[i]	✓	✓	✓		
Long-term availability					
App still available after 2 years[j]	✓	✓	✓	✓	✓
Year of the last update	2022	2022	2022	2020	2022

[a]App 1, Happify: bei Ärger und Stress (English translation: Happify: Anger and Stress); App 2, Sanvello: Stress & Anxiety Help; App 3, Headspace: Meditation & Schlaf (English translation: Headspace: Meditation & Sleep); App 4, go4health – gesund leben (English translation: go4health – living healthy); App 5, BamBu: Meditation & Achtsamkeit (English translation: BamBu: Meditation & Mindfulness).

[b]Name of this app after the second search in August 2022: "Sanvello: Anxiety and Depression." The app may no longer meet inclusion criterion 2 ("the app was developed for adults in the general population without mental or somatic disorders"). At the time of the screening process in 2020, we listed this app as "Sanvello: Stress & Anxiety Help," and it met the inclusion criteria.

[c]MARS-G: German Mobile Application Rating Scale.

[d]With the exception of "privacy and security features," all general characteristics of the categories "technical aspects," "strategies," and "functions" of the MARS-G are listed, which were included in at least three of the top 5 apps.

[e]The number of statements made regarding 19 possible security features is provided.

[f]Strategies that were included in the MARS-G and also in the list of theory-based stress management strategies.

[g]All theory-based stress management strategies (according to Kaluza [13] and Christmann et al [31]) are listed, when they were included in at least two of the top 5 apps. Theory-based stress management strategies that are already listed in the MARS-G strategies are not listed again (breathing [regenerative stress management strategy], relaxation [regenerative stress management strategy], mindfulness [regenerative stress management strategy], and acceptance [mental stress management strategy]).

[h]RSMS: regenerative stress management strategy.

[i]All studies were randomized controlled trials.

[j]App 4 was still available but only in the Google Play Store and was not available anymore in the Apple App Store, where it was found in 2020.

Discussion

Principal Findings

This systematic app search and standardized multidimensional assessment aimed to evaluate the general characteristics, quality, theory-based stress management strategies, evidence base, and long-term availability of SMAs. Furthermore, characteristics that might indicate high quality were derived from the 5 top-rated SMAs.

General Characteristics

Learning and maintaining stress management strategies requires regular engagement for not only changing the stress-enhancing cognitions and emotions, but also changing behavior [14]. Most of the included SMAs support this learning process by providing information on the background of the intervention and thus about stress management, tracking progress, or the use of reminder functions. Especially, the frequent presence of reminder functions was found to be similar in previous reviews [78]. A subgroup analysis of apps for mental health problems showed a moderate effect of reminder functions in reducing stress levels [79]. However, information on background, and tracking and reminder functions have been shown to improve long-term engagement within health apps [80], which might improve effectiveness through more intense and long-term usage. Similar to the results of Lau et al [34], most SMAs in this study were oriented toward self-help as merely 5 SMAs included the possibility to communicate synchronously (1/121, 0.8%) or asynchronously (4/121, 3.3%) with practitioners. This might be subject to change in future app development as a meta-analysis showed that professional guidance within mental health apps could significantly reduce stress levels compared with unguided apps (g=0.57 vs g=0.24) [79].

Support in terms of app communities was implemented in 16.5% (20/121) of the included SMAs. This is a positive trend compared with earlier findings showing only 4% of all included mindfulness-based apps providing this kind of support [78]. Since the availability of a community has beneficial effects on user engagement [80] and social support can positively influence the stress response [81], this seems to be a desirable trend.

Even though chronic stress is a major public health problem [1,9,12], only 3 SMAs were developed by institutions in the public sector (eg, universities or health authorities) and no SMA was officially certified (eg, according to the European Union Medical Device Regulation). This, together with the finding that only 5 SMAs were evaluated in randomized controlled trials, could indicate that thoroughly developed and evaluated apps might not find their way into the most popular app stores.

This study also focused on the declaration of the privacy and safety features within each identified SMA. The high percentage of SMAs providing a privacy policy (111/121, 91.7%) is promising. However, for 63.6% (77/121) of SMAs, no active confirmation of informed consent was required, and for 71.9% (87/121) of SMAs, there was no transparency regarding the right of withdrawal of informed consent. Data sharing with third parties was disclosed in the privacy policies of 17 (14.0%) SMAs. Regarding the actual practice of data security measures, Huckvale et al [40] showed that user data of health apps (depression and smoking cessation) have been shared with third parties, even without the necessary disclosure in the privacy policy. These results might be transferable to other health apps, including SMAs. Since the lack of data security is a common reason for user dissatisfaction with health apps and leads to the app being discontinued [82], improving data security measures may lead to increased engagement.

Quality

The 121 included SMAs showed an acceptable to good overall quality (mean score 3.59, SD 0.50). The scores of the dimensions functionality and esthetics were above the cutoff value of 3.5. The scores of the dimensions engagement (mean 3.05, SD 0.78) and information (mean 3.42, SD 0.46) did not exceed this cutoff score. Overall quality was similar to that of other (mental) health apps, such as mindfulness apps (mean score 3.66, SD 0.48 [37]), physical activity apps (mean score 3.60, SD 0.59 [83]), depression apps (mean score 3.01, SD 0.56 [53]), or apps for posttraumatic stress disorder (mean score 3.36, SD 0.65 [84]). The rating below the cutoff score in the information dimension was also consistent with the findings of previous systematic reviews [33,34,37,83]. One explanation is the limited evidence base of SMAs. Only 9% of the included SMAs were scientifically evaluated. The rating below the cutoff score in the dimension engagement implies that the content and functions of SMAs might currently not be sufficient to bind the users in the long term. Implementation of diverse content or the possibility of personalization could help as these aspects are particularly relevant for the users of mental health apps [82].

Theory-Based Stress Management Strategies

The examination of 3 types of theory-based stress management strategies resulted in 2.9 strategies per app. This is similar to the results in the study by Christmann et al [31], who reported 2.8 stress management strategies per app in their content analysis. Three of the four most implemented strategies are similar to the present results: meditation or mindfulness, breathing, and music (all categorized as regenerative stress management strategies). The results showed that instrumental and mental stress management strategies, which tend to be designed for prevention, are implemented less often than regenerative stress management strategies, which tend to be used for calming down after exposure to stress [13]. Therefore, the increased implementation of instrumental and mental strategies should be considered for a holistic approach to stress management in SMAs that seem to be relevant for effective prevention and coping with stress [13,15].

Evidence Base and Long-Term Availability

For 11 of the 121 (9.1%) SMAs, a scientific evaluation could be found. Moreover, 5 (4.1%) of the SMAs were evaluated in randomized controlled trials and 1 (0.8%) in a partially randomized controlled trial showing improvement in different outcomes such as stress, self-efficacy, mindfulness, anxiety, and depressive symptoms [59-65]. Previous reviews of mental health apps for other target groups included similar or even fewer efficacy studies [37,53,83-86]. This might be explained by the high rate of updates and the high volatility of apps [34,44,45], which could also be confirmed in this study. Of 163 apps, 13 (8.0%) became unavailable during the app rating period, and only 62.0% (75/121) of all SMAs were still available after 2 years, with most of them (64/75, 85.3%) being updated. This poses a great challenge for not only users and health professionals, but also researchers regarding the use, recommendation, and evaluation of SMAs or other health apps [44]. In addition, the trustworthiness of the information about the content and functions within app descriptions is questionable. In this study, 163 apps were included based on the information in the description. However, 24 (14.7%) apps had to be excluded after downloading because the actual content did not meet the previously described content. This confirms the findings of Coulon et al [33], who found that 33% of SMAs did not contain the content advertised in their descriptions. Potential users must check any eligible app for accuracy after overcoming the hurdles of downloading the app, installing the app, and, if required, registering an account. New evaluation frameworks are needed and do exist, but in a systematic review, it was concluded that none out of 45 evaluation frameworks for medical apps was rated as being fully suitable [87]. A different approach to deal with the fast-moving nature of apps has been proposed by a group of international and diverse stakeholders [88]. They harmonized elements of different frameworks into 5 priority levels (background info, data privacy and security, app effectiveness, user experience and adherence, and data integration) with the aim to enable informed app decision-making rather than to constantly evaluate the apps.

Overview of the Implications of the 5 Top-Rated SMAs

By presenting SMAs with top-rated MARS-G quality together with their characteristics in a comparative overview, a broad information and decision basis can be provided for researchers, health professionals, and users. The 5 top-rated SMAs showed both common characteristics and consistencies with existing evidence. Three of the 11 evidence-based apps were rated in the top 5 SMAs in terms of quality. Furthermore, some strategies, functions, and technical aspects previously shown to be effective in reducing stress or shown to improve engagement were found among the top-rated apps, such as providing psychoeducation [80], including breathing [89] and mindfulness [64], providing monitoring and tracking [80], using reminders [79], tailoring the content to the users' needs [90,91], and providing technical guidance and a privacy policy [82]. Across all SMAs and within the 5 top-rated SMAs, there were some aspects not covered by the MARS-G (eg, theory-based stress management strategies such as guided imagination, visualization, or music). This demonstrates the value of the overview of the top-rated SMAs (eg, compared with simple app rankings), in particular when special weight is given to certain aspects or app characteristics. In addition, the joint presentation of MARS-G content and additional uncovered aspects reveals certain revision potentials of the MARS-G.

Limitations

There were some limitations. First, owing to the rapid development of the app market and the short lifespan of apps, the content and quality of the reviewed SMAs may have already changed, some SMAs may no longer be available, or new SMAs may have been launched. However, this seems to be a challenge in general for health technology evaluation [44,87], and a screenshot of the current status might help to derive implications for improving SMA quality and effectiveness, and improving the evaluation frameworks for apps in the future. Second, the search results per search term were limited to 200 results and screening was based on the titles and descriptions of the apps. It is possible that apps that met the inclusion criteria were overlooked because relevant information was not provided or the word "stress" was not present in the titles or descriptions. Furthermore, only SMAs that were free of cost or provided a free basic version were evaluated. Further evaluation of paid (full version) SMAs could show whether there is a difference in quality, declaration of data privacy and security features, or access to professional support. Third, there was only a descriptive evaluation (not a technical evaluation; eg, for data privacy issues), and no conclusions about the overall effectiveness of the included SMAs could be drawn. Fourth, the number of SMAs including "breathing" as a strategy differed in the MARS rating and in the assessment of theory-based stress management strategies. Therefore, it should be emphasized that the discriminant differentiation of the related and partially overlapping concepts or strategies of "mindfulness" and "breathing" cannot be assessed conclusively (especially considering that "breathing" can also be practiced as a concrete strategy within the context of mindfulness). However, the fact that "breathing" and "mindfulness" were listed in the top 3 strategies remains unchanged. Finally, the review and evaluation of each app took an average of 30 minutes. It is possible that specific content could not be discovered owing to the limited amount of time spent evaluating each app.

Conclusion

In this comprehensive review including a systematic search and a standardized multidimensional assessment, the overall quality of 121 SMAs was rated as acceptable to good, with a rating below the cutoff score in the dimensions of information quality and engagement. The top-rated apps included psychoeducation, breathing and mindfulness, monitoring, reminder functions, tailoring, technical guidance, and a privacy policy. However, even though most SMAs provided a privacy policy, there is still a need for better personal data protection and transparency of data processing, such as the use of a password or information about data sharing with third parties. Theory-based strategies were mostly regenerative stress management strategies. For a holistic stress management approach, SMAs could benefit from the integration of more mental and instrumental stress management strategies. The evidence base for 11 (9.1%) of the 121 included apps showed that SMAs can reduce stress and improve further outcome variables such as self-efficacy, mindfulness, anxiety, and depressive symptoms. Moreover, SMAs have high scalability. Therefore, they have a high potential to reach and help a broad audience coping with increasing stress and demands in their work and daily living. However, the rather moderate information quality, the scarce evidence base of the included SMAs, and the fact that many SMAs changed or were unavailable after a 2-year period pose challenges for users and health professionals who are searching for high-quality apps that are effective and for long-term use. The common characteristics of SMAs with top-rated quality and evidence base of SMAs can be used as guidance for this search or even for SMA development. In addition, it is difficult for researchers to keep up to date with the latest research in this volatile field and provide potential users with helpful information. Enhanced evaluation frameworks are needed that might complement or even advance the idea of a continuous effectiveness and quality assessment to an approach that enables informed decision-making.

Acknowledgments

We want to thank Hannah Besel and Joshua Kannapin for assisting in the project. The study was self-funded (University of Freiburg). The article processing charge was partly funded by the Open Access Publication Fund of the University of Freiburg.

Data Availability

Data will be made available to researchers who provide a methodologically sound proposal, not already covered by other researchers. All requests should be directed to the corresponding author. Data requestors will need to sign a data access agreement. Provision of data is subject to data security regulations. Support depends on available resources.

Authors' Contributions

SP, EM, EMM, and HB designed the trial and initiated this study. SP, EM, RW, and VH conducted the German Mobile Application Rating Scale (MARS-G) ratings. SP supervised all MARS-G ratings. YT and DS conducted the statistical analyses. JS was involved as an expert in the field of stress research. MW was involved in the conceptualization and drafting of the manuscript and in the second assessment as part of an update in August 2022. SP wrote the first draft. All authors revised the manuscript, and read and approved the final report.

Conflicts of Interest

HB and EMM received payments for talks and workshops in the context of e-mental health. All other authors declare that they have no competing interests.

Multimedia Appendix 1

Frequency of declared data privacy and security features based on the German Mobile Application Rating Scale presented as a bar graph. The asterisk (*) indicates additional features that have been added to the original list.

[PNG File , 22 KB - mhealth_v11i1e42415_app1.png]

Multimedia Appendix 2

Frequency of the implemented strategies according to the German Mobile Application Rating Scale presented as a bar graph.

[PNG File , 14 KB - mhealth_v11i1e42415_app2.png]

Multimedia Appendix 3

Quality ratings of all included stress management apps.

[PDF File (Adobe PDF File), 183 KB - mhealth_v11i1e42415_app3.pdf]

Multimedia Appendix 4

Theory-based stress management strategies according to Kaluza [13] and Christmann et al [31] presented as a bar graph. Regenerative stress management strategies (RSMSs) are indicated with black, mental stress management strategies (MSMSs) are indicated with vertical lines, and instrumental stress management strategies (ISMSs) are indicated with horizontal lines. The asterisk (*) indicates theory-based stress management strategies that are also included in the German Mobile Application Rating Scale.

[PNG File , 48 KB - mhealth_v11i1e42415_app4.png]

Multimedia Appendix 5

Overview of stress management apps with an evidence base.
[PDF File (Adobe PDF File), 176 KB - mhealth_v11i1e42415_app5.pdf]

References

1. Hapke U, Maske UE, Scheidt-Nave C, Bode L, Schlack R, Busch MA. Chronischer Stress bei Erwachsenen in Deutschland [Chronic stress among adults in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. Bundesgesundheitsblatt 2013;56(5-6):749-754. [doi: 10.1007/s00103-013-1690-9] [Medline: 23703494]
2. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: A review. EXCLI J 2017;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]
3. Cohen S, Murphy MLM, Prather AA. Ten Surprising Facts About Stressful Life Events and Disease Risk. Annu Rev Psychol 2019;70:577-597 [FREE Full text] [doi: 10.1146/annurev-psych-010418-102857] [Medline: 29949726]
4. Harvey SB, Modini M, Joyce S, Milligan-Saville JS, Tan L, Mykletun A, et al. Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. Occup Environ Med 2017;74(4):301-310. [doi: 10.1136/oemed-2016-104015] [Medline: 28108676]
5. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
6. Stress in America: Paying with our health. American Psychological Association. Washington, DC: American Psychological Association; 2015. URL: https://www.apa.org/news/press/releases/stress/2014/stress-report.pdf [accessed 2023-07-11]
7. Everly GS, Lating JM. A Clinical Guide to the Treatment of the Human Stress Response. New York, NY: Springer; 2019.
8. Liu M, Li N, Li WA, Khan H. Association between psychosocial stress and hypertension: a systematic review and meta-analysis. Neurol Res 2017;39(6):573-580. [doi: 10.1080/01616412.2017.1317904] [Medline: 28415916]
9. Bergmann N, Gyntelberg F, Faber J. The appraisal of chronic stress and the development of the metabolic syndrome: a systematic review of prospective cohort studies. Endocr Connect 2014;3(2):R55-R80 [FREE Full text] [doi: 10.1530/EC-14-0031] [Medline: 24743684]
10. Hammen CL. Stress and depression: old questions, new approaches. Current Opinion in Psychology 2015;4:80-85. [doi: 10.1016/j.copsyc.2014.12.024]
11. Swanson SA, Zeng Y, Weeks M, Colman I. The contribution of stress to the comorbidity of migraine and major depression: results from a prospective cohort study. BMJ Open 2013;3(3):e002057 [FREE Full text] [doi: 10.1136/bmjopen-2012-002057] [Medline: 23474788]
12. Hassard J, Teoh KRH, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: A systematic review. J Occup Health Psychol 2018;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
13. Kaluza G. Calm and Confident Under Stress: The Stress Competence Book: Recognize, Understand, Manage Stress. Berlin, Heidelberg: Springer; 2022.
14. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, NY: Springer Publishing Company; 1984.
15. Kaluza G, Chevalier A. Stressbewältigungstrainings für Erwachsene [Stress management training for adults]. In: Fuchs R, Gerber M, editors. Handbuch Stressregulation und Sport. Berlin, Heidelberg: Springer; 2018:143-162.
16. Lazarus RS, Launier R. Stressbezogene Transaktionen zwischen Person und Umwelt [Stress-related transactions between person and environment]. In: Nitsch JR, editor. Stress. Theorien, Untersuchungen, Maßnahmen. Bern, Stuttgart, Wien: Verlag Hans Huber; 1981:213-259.
17. Ong L, Linden W, Young S. Stress management: what is it? J Psychosom Res 2004;56(1):133-137. [doi: 10.1016/S0022-3999(03)00128-4] [Medline: 14987975]
18. Murphy LR. Stress management in work settings: a critical review of the health effects. Am J Health Promot 1996;11(2):112-135. [doi: 10.4278/0890-1171-11.2.112] [Medline: 10163598]
19. Kaluza G. Gelassen und sicher im Stress: Das Stresskompetenz-Buch: Stress erkennen, verstehen, bewältigen [The Stress Competence Book: Recognize, Understand, Manage Stress]. Berlin, Heidelberg: Springer; 2018.
20. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
21. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
22. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. J Psychosom Res 2015;78(6):519-528. [doi: 10.1016/j.jpsychores.2015.03.009] [Medline: 25818837]
23. Sharma M, Rush SE. Mindfulness-based stress reduction as a stress management intervention for healthy individuals: a systematic review. J Evid Based Complementary Altern Med 2014 Oct;19(4):271-286. [doi: 10.1177/2156587214543143] [Medline: 25053754]
24. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
25. Rippe JM. Lifestyle Medicine: The Health Promoting Power of Daily Habits and Practices. Am J Lifestyle Med 2018;12(6):499-512 [FREE Full text] [doi: 10.1177/1559827618785554] [Medline: 30783405]

26. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. European Psychologist 2018;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
27. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2015;38(1):23-38. [doi: 10.1007/s13246-014-0315-4] [Medline: 25476753]
28. Anzahl der Downloads von mHealth-Apps weltweit in den Jahren 2013 bis 2018. Statista. 2018. URL: https://de.statista.com/statistik/daten/studie/695434/umfrage/anzahl-der-weltweiten-downloads-von-mhealth-apps/ [accessed 2021-08-02]
29. Percentage of U.S. adults that use apps for stress relief as of 2017, by gender. Statista. 2017. URL: https://www.statista.com/statistics/699523/us-adults-that-use-apps-to-relieve-stress-by-gender/ [accessed 2021-08-04]
30. Morrison LG, Geraghty AWA, Lloyd S, Goodman N, Michaelides DT, Hargood C, et al. Comparing usage of a web and app stress management intervention: An observational study. Internet Interv 2018;12:74-82 [FREE Full text] [doi: 10.1016/j.invent.2018.03.006] [Medline: 30135771]
31. Christmann CA, Hoffmann A, Bleser G. Stress Management Apps With Regard to Emotion-Focused Coping and Behavior Change Techniques: A Content Analysis. JMIR Mhealth Uhealth 2017;5(2):e22 [FREE Full text] [doi: 10.2196/mhealth.6471] [Medline: 28232299]
32. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and Controlling Stress Using mHealth: Systematic Search in App Stores. JMIR Mhealth Uhealth 2018;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
33. Coulon SM, Monroe CM, West DS. A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
34. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR Mhealth Uhealth 2020;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
35. Hoffmann A, Christmann CA, Bleser G. Gamification in Stress Management Apps: A Critical App Review. JMIR Serious Games 2017;5(2):e13 [FREE Full text] [doi: 10.2196/games.7216] [Medline: 28592397]
36. Alhasani M, Mulchandani D, Oyebode O, Baghaei N, Orji R. A Systematic and Comparative Review of Behavior Change Strategies in Stress Management Apps: Opportunities for Improvement. Front Public Health 2022;10:777567 [FREE Full text] [doi: 10.3389/fpubh.2022.777567] [Medline: 35284368]
37. Schultchen D, Terhorst Y, Holderied T, Stach M, Messner E, Baumeister H, et al. Stay Present with Your Phone: A Systematic Review and Standardized Rating of Mindfulness Apps in European App Stores. Int J Behav Med 2021;28(5):552-560 [FREE Full text] [doi: 10.1007/s12529-020-09944-y] [Medline: 33215348]
38. Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy Policies of Android Diabetes Apps and Sharing of Health Information. JAMA 2016;315(10):1051-1052. [doi: 10.1001/jama.2015.19426] [Medline: 26954415]
39. Grundy Q, Chiu K, Bero L. Commercialization of User Data by Developers of Medicines-Related Apps: a Content Analysis. J Gen Intern Med 2019;34(12):2833-2841 [FREE Full text] [doi: 10.1007/s11606-019-05214-0] [Medline: 31529374]
40. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
41. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
42. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German Version of the Mobile App Rating Scale (MARS-G): Development and Validation Study. JMIR Mhealth Uhealth 2020;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
43. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
44. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth 2016;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
45. McIlroy S, Ali N, Hassan AE. Fresh apps: an empirical study of frequently-updated mobile apps in the Google play store. Empir Software Eng 2015;21(3):1346-1370. [doi: 10.1007/s10664-015-9388-2]
46. Albrecht UV. Chancen und Risiken von Gesundheits-Apps (CHARISMHA); engl. Chances and Risks of Mobile Health Apps (CHARISMHA). Technische Universität Braunschweig. 2016. URL: https://leopard.tu-braunschweig.de/receive/dbbs_mods_00060000 [accessed 2023-07-11]
47. Paganini S, Meier P, Messner E, Sander L, Baumeister H, Terhorst Y. A Systematic Review of Content and Quality of Apps for Stress Management. OSF. 2019. URL: https://osf.io/8vg5c [accessed 2023-07-11]
48. Mobile Health App Database. URL: https://mhad.science/en/ [accessed 2023-07-15]

49. Stach M, Kraft R, Probst T, Messner E, Terhorst Y, Baumeister H, et al. Mobile Health App Database - A Repository for Quality Ratings of mHealth Apps. 2020 Presented at: 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN, USA. [doi: 10.1109/CBMS49503.2020.00087]
50. MARS - Mobile Anwendungen Rating Skala. YouTube. URL: https://www.youtube.com/watch?v=5vwMiCWC0Sc [accessed 2023-07-15]
51. Security tips. Android Developers. URL: https://developer.android.com/training/articles/security-tips [accessed 2023-05-07]
52. Orientierungshilfe zu den Datenschutzanforderungen an App-Entwickler und App-Anbieter [Guidance on data protection requirements for app developers and app providers]. Düsseldorfer Kreis - Bayerisches Landesamt für Datenschutzaufsicht. URL: https://datenschutz.hessen.de/sites/datenschutz.hessen.de/files/2022-08/oh_appentwicklung.pdf [accessed 2023-07-11]
53. Terhorst Y, Rathner E, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. Verhaltenstherapie 2018;28(2):101-112. [doi: 10.1159/000481692]
54. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
55. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
56. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Hoboken, NJ: Pearson/Prentice Hall; 2009.
57. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
58. Regulation 2017/745 of the European parliament and of the Council of 5 April 2017 on medical devices. Official Journal of the European Union. 2017. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745 [accessed 2023-07-11]
59. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. Intnl. J. Wellbeing 2018;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]
60. Nagel A, John D, Scheder A, Kohls N. Klassisches oder digitales Stressmanagement im Setting Hochschule? Präv Gesundheitsf 2018;14(2):138-145. [doi: 10.1007/s11553-018-0670-1]
61. Möltner H, Leve J, Esch T. [Burnout Prevention and Mobile Mindfulness: Evaluation of an App-Based Health Training Program for Employees]. Gesundheitswesen 2018;80(3):295-300 [FREE Full text] [doi: 10.1055/s-0043-114004] [Medline: 29017191]
62. Moberg C, Niles A, Beermann D. Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. J Med Internet Res 2019;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
63. Fuller-Tyszkiewicz M, Richardson B, Little K, Teague S, Hartley-Clark L, Capic T, et al. Efficacy of a Smartphone App Intervention for Reducing Caregiver Stress: Randomized Controlled Trial. JMIR Ment Health 2020;7(7):e17541 [FREE Full text] [doi: 10.2196/17541] [Medline: 32706716]
64. Flett JAM, Conner TS, Riordan BC, Patterson T, Hayne H. App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: a pragmatic, randomised, waitlist-controlled trial. Psychol Health 2020;35(9):1049-1074. [doi: 10.1080/08870446.2019.1711089] [Medline: 32046499]
65. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Australian Psychologist 2020;54(4):245-260. [doi: 10.1111/ap.12383]
66. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile Mindfulness Meditation: a Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health. Mindfulness 2018;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
67. Best NI, Durham CF, Woods-Giscombe C, Waldrop J. Combating Compassion Fatigue With Mindfulness Practice in Military Nurse Practitioners. The Journal for Nurse Practitioners 2020;16(5):e57-e60. [doi: 10.1016/j.nurpra.2020.02.023]
68. Lee RA, Jung ME. Evaluation of an mHealth App (DeStressify) on University Students' Mental Health: Pilot Trial. JMIR Ment Health 2018;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
69. Levin ME, Hicks ET, Krafft J. J Am Coll Health 2022 Jan;70(1):165-173. [doi: 10.1080/07448481.2020.1728281] [Medline: 32150519]
70. Hoffmann A, Faust-Christmann CA, Zolynski G, Bleser G. Gamification of a Stress Management App: Results of a User Study. In: Marcus A, Wang W, editors. Design, User Experience, and Usability. Application Domains. HCII 2019. Lecture Notes in Computer Science, vol 11585. Cham: Springer; 2019:303-313.
71. Yang E, Schamber E, Meyer R, Gold J. Happier Healers: Randomized Controlled Trial of Mobile Mindfulness for Stress Management. J Altern Complement Med 2018;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
72. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in Happiness: A Randomised Controlled Trial of a Smartphone-Based Mindfulness Intervention to Enhance Wellbeing. J Happiness Stud 2014;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
73. Wen L, Sweeney TE, Welton L, Trockel M, Katznelson L. Encouraging Mindfulness in Medical House Staff via Smartphone App: A Pilot Study. Acad Psychiatry 2017;41(5):646-650. [doi: 10.1007/s40596-017-0768-3] [Medline: 28795335]

74. Morrison Wylde C, Mahrer N, Meyer R, Gold J. Mindfulness for Novice Pediatric Nurses: Smartphone Application Versus Traditional Intervention. J Pediatr Nurs 2017;36:205-212. [doi: 10.1016/j.pedn.2017.06.008] [Medline: 28888505]
75. Bennike IH, Wieghorst A, Kirk U. Online-based Mindfulness Training Reduces Behavioral Markers of Mind Wandering. J Cogn Enhanc 2017;1(2):172-181. [doi: 10.1007/s41465-017-0020-9]
76. Economides M, Martman J, Bell M, Sanderson B. Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
77. Rosen K, Paniagua S, Kazanis W, Jones S, Potter J. Quality of life among women diagnosed with breast Cancer: A randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
78. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
79. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
80. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
81. Ditzen B, Heinrichs M. Psychobiology of social support: the social dimension of stress buffering. Restor Neurol Neurosci 2014;32(1):149-162. [doi: 10.3233/RNN-139008] [Medline: 23603443]
82. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
83. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler A, et al. Quality of Physical Activity Apps: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2021;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]
84. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
85. Terhorst Y, Messner E, Schultchen D, Paganini S, Portenhauser A, Eder A, et al. Systematic evaluation of content and quality of English and German pain apps in European app stores. Internet Interv 2021;24:100376 [FREE Full text] [doi: 10.1016/j.invent.2021.100376] [Medline: 33718002]
86. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
87. Moshi MR, Tooher R, Merlin T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: A systematic review. Int J Technol Assess Health Care 2018;34(5):464-475. [doi: 10.1017/S026646231800051X] [Medline: 30201060]
88. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health 2019;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
89. Hopper SI, Murray SL, Ferrara LR, Singleton JK. Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review. JBI Database System Rev Implement Rep 2019;17(9):1855-1876. [doi: 10.11124/JBISRIR-2017-003848] [Medline: 31436595]
90. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
91. Alqahtani F, Meier S, Orji R. Personality-based approach for tailoring persuasive mental health applications. User Model User-Adap Inter 2021;32(3):253-295. [doi: 10.1007/s11257-021-09289-5]

Abbreviations

ICC: interclass correlation
MARS: Mobile Application Rating Scale
MARS-G: German Mobile Application Rating Scale
SMA: stress management app

Edited by L Buis; submitted 02.09.22; peer-reviewed by J Lee, J Manning; comments to author 01.03.23; accepted 06.06.23; published 29.08.23.

Please cite as:
Paganini S, Meier E, Terhorst Y, Wurst R, Hohberg V, Schultchen D, Strahler J, Wursthorn M, Baumeister H, Messner EM
Stress Management Apps: Systematic Search and Multidimensional Assessment of Quality and Characteristics
JMIR Mhealth Uhealth 2023;11:e42415
URL: https://mhealth.jmir.org/2023/1/e42415
doi:10.2196/42415
PMID:37642999

Original Paper

Availability, Quality, and Evidence-Based Content of mHealth Apps for the Treatment of Nonspecific Low Back Pain in the German Language: Systematic Assessment

Lauro Ulrich[1], BSc; Phillip Thies[1], MSc; Annika Schwarz[1], Prof Dr

[1]Faculty of Social Sciences, City University of Applied Sciences Bremen, Bremen, Germany

Corresponding Author:
Annika Schwarz, Prof Dr
Faculty of Social Sciences, City University of Applied Sciences Bremen
Am Brill 2-4
Bremen, 28195
Germany
Phone: 49 176 151 40 227
Email: annika.schwarz@hs-bremen.de

Abstract

Background: Nonspecific low back pain (NSLBP) carries significant socioeconomic relevance and leads to substantial difficulties for those who are affected by it. The effectiveness of app-based treatments has been confirmed, and clinicians are recommended to use such interventions. As 88.8% of the German population uses smartphones, apps could support therapy. The available apps in mobile app stores are poorly regulated, and their quality can vary. Overviews of the availability and quality of mobile apps for Australia, Great Britain, and Spain have been compiled, but this has not yet been done for Germany.

Objective: We aimed to provide an overview of the availability and content-related quality of apps for the treatment of NSLBP in the German language.

Methods: A systematic search for apps on iOS and Android was conducted on July 6, 2022, in the Apple App Store and Google Play Store. The inclusion and exclusion criteria were defined before the search. Apps in the German language that were available in both stores were eligible. To check for evidence, the apps found were assessed using checklists based on the German national guideline for NSLBP and the British equivalent of the National Institute for Health and Care Excellence. The quality of the apps was measured using the Mobile Application Rating Scale. To control potential inaccuracies, a second reviewer resurveyed the outcomes for 30% (3/8) of the apps and checked the inclusion and exclusion criteria for these apps. The outcomes, measured using the assessment tools, are presented in tables with descriptive statistics. Furthermore, the characteristics of the included apps were summarized.

Results: In total, 8 apps were included for assessment. Features provided with different frequencies were exercise tracking of prefabricated or adaptable workout programs, educational aspects, artificial intelligence–based therapy or workout programs, and motion detection. All apps met some recommendations by the German national guideline and used forms of exercises as recommended by the National Institute for Health and Care Excellence guideline. The mean value of items rated as “Yes” was 5.75 (SD 2.71) out of 16. The best-rated app received an answer of “Yes” for 11 items. The mean Mobile Application Rating Scale quality score was 3.61 (SD 0.55). The highest mean score was obtained in “Section B–Functionality” (mean 3.81, SD 0.54).

Conclusions: Available apps in the German language meet guideline recommendations and are mostly of acceptable or good quality. Their use as a therapy supplement could help promote the implementation of home-based exercise protocols. A new assessment tool to obtain ratings on apps for the treatment of NSLBP, combining aspects of quality and evidence-based best practices, could be useful.

Trial Registration: Open Science Framework Registries sq435; https://osf.io/sq435

(JMIR Mhealth Uhealth 2023;11:e47502) doi:10.2196/47502

KEYWORDS

mobile health; mobile apps; smartphone; nonspecific low back pain; German language; intervention; digital health; home exercise; digital rehabilitation; workout; mobile phone

Introduction

Background

Low back pain (LBP) is a major global health concern affecting millions of people, with an estimated 7.5% of the population or 577 million people experiencing LBP in 2017 [1]. Furthermore, the condition was the leading cause for years lived with disability from 1990 to 2017, worldwide [1]. In Germany, LBP affects 59.4% of the population and results in decreased work performance and pain persistence, with an average cost of €1322 (US $1456.10) per patient per year [2-4]. Physical exercise and a healthy lifestyle are recommended by national and international guidelines for the management of nonspecific LBP (NSLBP) [3,4]. According to national guidelines, it should be emphasized that exercising does not cause harm but can help to alleviate symptoms in NSLBP [3]. In addition, an understanding of the biopsychosocial model of illness should be developed [3]. In this regard, several studies have shown promising evidence for the app, "Kaia Rückenschmerzen—Rückentraining für Zuhause," which provides a multidisciplinary pain treatment approach [5-7]. The app is based on 3 principles, which are education, physical exercising, and mindfulness and relaxation techniques [6]. This approach might even be superior to conventional physiotherapy [6]. Furthermore, a systematic review focused on the treatment of chronic pain with eHealth and mobile health (mHealth) interventions showed its significant efficacy on short- and medium-term outcomes on pain intensity and depression, as well as short-term reductions in pain-catastrophizing [8]. Due to their wide availability and low cost to patients, the authors of the systematic review encourage clinicians to use eHealth and mHealth interventions as an adjunct to their therapy [8]. Various sources report an increasing shortage of physiotherapists in Germany [9-11]. However, 88% of the population use smartphones [12]. Considering the prevailing lack of physiotherapists in Germany, mHealth and eHealth apps could be an addition to the management of patients with NSLBP [8]. Guideline-based apps could support therapy and help to close gaps in therapy or continue to support patients after they have completed physiotherapy. In Germany, the use of digital health apps (*Digitale Gesundheitsanwendungen* [DiGA]) is regulated by the Digital Health Care Act (*Digitale-Versorgung-Gesetz*) [13]. Health apps that are certified as a medical device of risk class I or IIa can be included in the so-called *DiGA directory* after a review process by the Federal Institute for Drugs and Medical Devices (*Bundesinstitut für Arzneimittel und Medizinprodukte*). In the review process, aspects of data protection, consumer protection, user-friendliness, and medical efficacy must be provided. Apps included in the DiGA directory can be reimbursed by health insurance companies after medical prescription [13]. The apps listed in the DiGA directory—and thus subjected to a review process—can be considered safe, especially considering the risk classification that has taken place. However, commercially, there are many other health apps available that are not institutionally reviewed. Their quality can therefore be highly variable [13].

Objective

Existing systematic reviews have evaluated the quality of apps in Australia [14,15], Spain, and the United Kingdom [16]; therefore, the included apps were restricted to the English and Spanish languages. Evaluated using the Mobile Application Rating Scale (MARS) [17], apps with good quality are available from app stores in Australia, Spain, and the United Kingdom [14-16]. Most Australian apps follow the recommendations of the UK guideline on LBP and sciatica by the National Institute for Health and Care Excellence (NICE) [4,14,15]. It has been shown that in-store user evaluations do not correlate with assessed quality [14,15]. Consequently, they are a poor indicator of app quality. To date, there is no comparable, objective analysis for the quality of apps in the German language that could help patients or clinicians to estimate the quality and guideline fidelity of the available apps. The objective of this assessment was to provide an overview of the availability and quality of apps for patients with NSLBP and to offer recommendations for clinicians in advising their patients with NSLBP.

Methods

Overview

The methods used in this study were adapted from the studies by Machado et al [14] and Didyk et al [15] and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [18]. A systematic search for smartphone apps on Apple iOS and Google Android app stores was conducted. To facilitate this process, we used the web scraping software Octoparse (version 8.5.2; Octoparse). All methods used were planned and publicly preregistered on the Open Science Framework Register before the searches were carried out (osf.io/sq435).

Inclusion and Exclusion Criteria

Previous research has shown that higher app price correlates with higher quality [14,15]. Therefore, no price limit was applied when including apps, and wherever available, the "pro" or "premium" version was considered for inclusion. The included apps were required to be available for download and use to the public, so that they could be accessed by the public and physiotherapists directly. According to the research objective, they should have been available in the German language. To ensure that the recommendations resulting from our research are as general as possible and applicable for use, regardless of the device and operating system, the identified apps should have been available for iOS and Android, as these are the most used smartphone systems in Germany [19]. Included apps had to be stand-alone and ready to use without accessories. The exceptions were a gym mat and resistance band. Apps had to be released or updated no later than 2021 to ensure technical support and compatibility with current software and devices [14]. Apps were required to be targeted at patients and consumers and physically and mentally engaging, as recommended by the German national guideline (GNG) for NSLBP and the NICE guideline [3,4]. For physical participation, we counted all forms of physical exercise. For mental participation, we counted interventions that incorporated mental or spiritual aspects, similar to the category of mind-body exercises in the NICE guideline.

Apps were excluded if they were designed only for diagnostic purposes (eg, the detection of risk factors). Finally, apps that explicitly addressed specific forms of LBP (eg, pregnancy-related LBP) and apps for general health promotion that did not address NSLBP were excluded.

Search

The search was performed on July 6, 2022. German synonyms for back pain were used as search terms: "Rückenschmerzen," "Rückenschmerz," "Kreuzschmerzen," and "Kreuzschmerz." A single search was performed for each search term. No search filters were used in either store. The metadata about the apps were collected from the browser-based view of the apps in each app store. These included the app name, developer name, last update of the app, app rating in the store, description of the app, and URL to the app. Duplicates were identified using these data. Once these were removed, a list was created for each store that contained all the apps available for the terms used.

Screening

The screening process can be divided into three phases: (1) identification of apps available in both stores; (2) screening of app names and descriptions from the stores according to the inclusion and exclusion criteria analogous to abstract screening [20]; and (3) screening of apps after installation. Apps that met the criteria and those for which it remained unclear whether they would meet the criteria were installed during the third screening phase. This screening was conducted by one rater (LU) according to the inclusion and exclusion criteria. A table of the screening and therefore excluded apps can be found in Multimedia Appendix 1. After installation on an iPhone SE (2020 model; Apple Inc), the apps were used and examined for at least 10 minutes. If the criteria were answered as "Unclear," the criterion was discussed with a second rater (PT and AS) until a consensus was reached to include or exclude the app.

Outcome Measures

Apps included in this study were assessed for evidence according to guidelines on the treatment of LBP and for quality using the MARS [3,4,17].

To assess the consistency with guidelines, a checklist was created along the GNG chapters *4.1—Principles of nonspecific low back pain therapy* and *4.2 Management of nonspecific low back pain* [3]. This resulted in a list of 16 items, each containing 1 recommendation. The checklist is presented in Multimedia Appendix 2 [3]. To evaluate apps, the question "Does the app meet the recommendation?" was asked for each recommendation or item. This could be answered with the response categories "Yes", "No", and "Unclear". In addition, the exercises used in the apps were classified according to the classification of exercises used in the UK NICE guideline. These categories were "biomechanical exercise" (BE), "aerobic exercise," "mind-body exercise," and "mixed modality exercise" [4]. Apps had to use at least one exercise that could be assigned along with this classification.

App quality was assessed using the MARS Tool [17]. It contains 23 items divided into five categories: 4 categories with objective quality criteria ("section A—engagement," "section B—functionality," "section C—aesthetics," and "section D—information quality") and 1 category with subjective quality criteria. Each item was rated on a 5-point scale (1=inadequate to 5=excellent). A full description of the categories is described elsewhere [17]. The MARS Tool has demonstrated excellent internal consistency (Cronbach α=.90) and interrater reliability (intraclass correlation coefficient [ICC]=0.79) [17]. Moreover, the tool has been used in methodologically related work [14-16]. Both raters (LU and PT) have been trained and proceeded according to the MARS training video [21].

To control for potential inaccuracies and check for reliability, around 30% (3/8) of the apps were rated by a second rater (PT) who used both instruments (MARS and guideline checklist). This approach follows the example of Machado et al [14], who checked a similar percentage of the MARS ratings. The second rater was trained by the first rater in the process and the use of the assessment instruments and also installed the apps on an iPhone SE (2020 model; Apple Inc). The control apps were randomly selected. For this purpose, a third person (AS), who was not involved in the evaluation process at the time, received a list of the included apps and created a randomization sequence using Research Randomizer [22]. As far as possible, the first and second raters reached a consensus for the identified differences between ratings. Where no consensus could be reached by the 2 raters, a third rater (AS) was consulted for a final verdict.

Data Analysis and Synthesis

App name, developer, models available, model used, date of last update or release, MARS quality mean score, and classification of exercises were compiled. The classification of exercises according to the NICE guideline were presented without further analyses. The results from the GNG checklist are presented with descriptive statistics (mean, median, SD, and range). Only the objective items 1 to 19 of the MARS were evaluated, as they are needed to calculate the app quality mean score [17]. In addition, an overview of the app characteristics is provided.

The agreement of the raters with the checklist was calculated using Cohen κ [23]. For this purpose, each item on the GNG was considered as a case that raters could answer "Yes", "No", or "Unclear". For agreement in the classifications according to the NICE guideline, each exercise class was considered a case in which raters could *accept* or *reject* each app. To calculate the interrater reliability of the MARS, the ICC was used as a 2-way mixed model with average measures and absolute agreement [24]. The mean values of the sections were used. SPSS (version 27; IBM Corp) was used to calculate the ICC. PSPP (version 3.0, 2007; GNU Project) was used for all other calculations.

Ethical Considerations

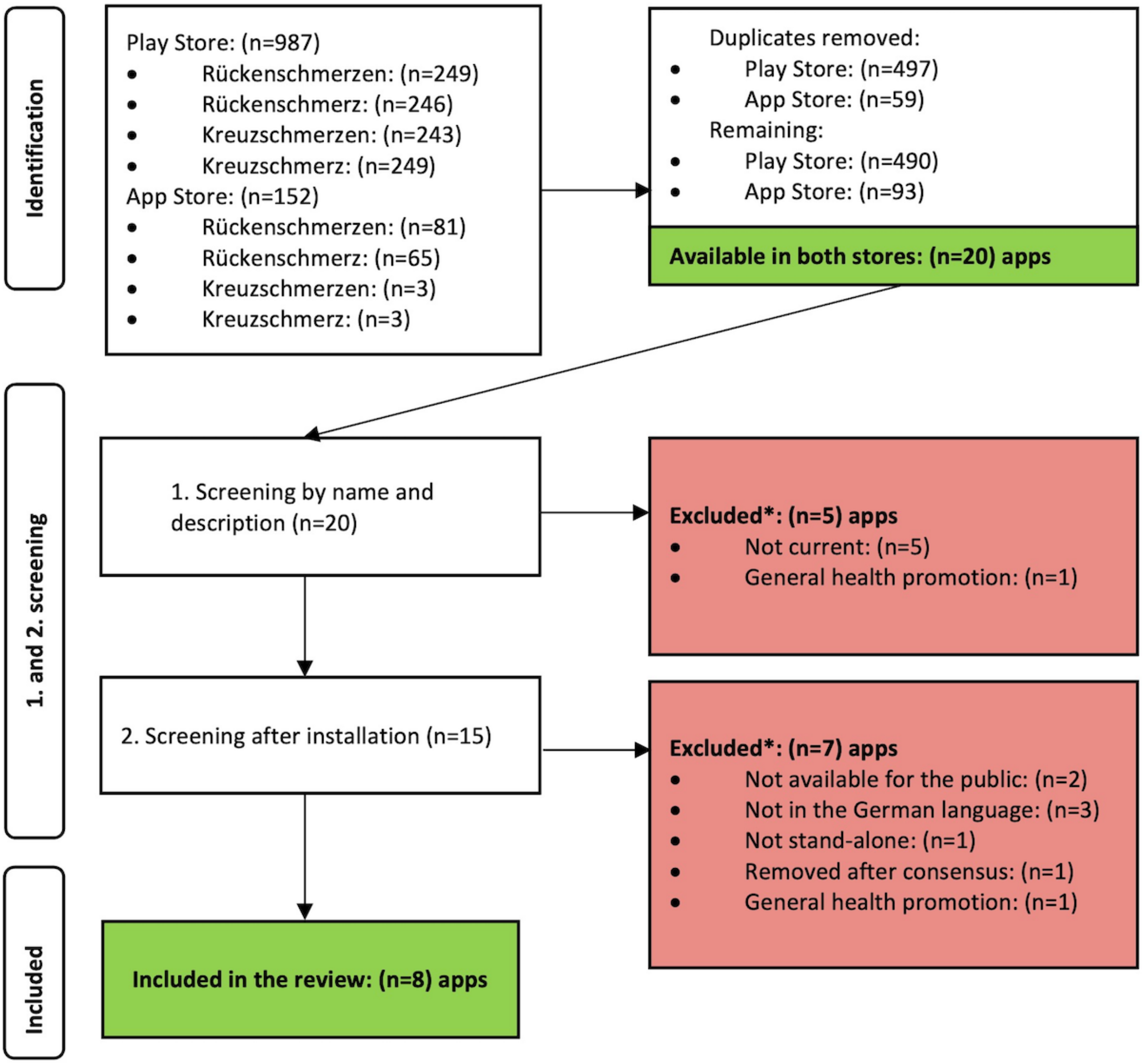
Ethical principles must be considered for medical research involving human subjects, including research on identifiable human material and data, according to Article 1 of the Preamble of the Declaration of Helsinki. As no patients were examined in this systematic assessment and only apps and data not requiring data protection were collected, no ethics vote is necessary according to the Declaration of Helsinki [25].

Results

App Selection

A total of 20 apps available in both stores were identified. After the initial screening for the inclusion and exclusion criteria based on the name and description in the stores, 5 apps were excluded because they received their last update before 2021. After the second screening of the remaining 15 apps, a further 7 apps were excluded. Eight apps were included in the assessment. The screening process is depicted in a flowchart diagram based on the PRISMA statement (Figure 1) [18].

Figure 1. Flowchart of the selection of apps. *Multiple criteria applicable.



App Characteristics

We detected a series of characteristic elements that were commonly used in combination with the assessed apps. Six apps delivered some form of educational content. One app used training videos to deliver the provided exercises. Three apps created individual exercise programs based on their algorithm. Seven apps provided an exercise tracking feature. Four apps suggested prefabricated workout plans that were customizable in 3 of these apps. One app provided motion detection of the exercising person using the camera of the smartphone. The characteristic elements and different combinations used for each app are listed in Table 1.

Table 1. Summary of characteristics of included apps.

App name (iOS; version)	App name (Android)	Developer	Used version	Published or last update at the time of assessment	Characteristics	MARS[a] score, mean (SD)	Classification of exercises according to NICE[b]
Dein Rückentraining (3.0)	Dein ganzheitliches Rückentraining	EBL Media Production OG	Purchase upon download (€24.99[c])	July 10, 2021	Education and train-along videos	2.7 (0.29)	BE[d] and MBE[e]
ViViRA bei Rückenschmerzen (2.41.0)	ViViRA bei Rückenschmerzen	Vivira Health Lab GmbH	Monthly subscription (€79.99)	July 6, 2022	Artificial intelligence–based program with education and tracked exercises	4.2 (0.19)	BE
Rückenschmerzen Übungen (1.0.99)	Rückenschmerzen Übungen	Vladimir Ratsev	"Pro"-version via in-app purchase (€2.99)	August 16, 2022	Exercise tracker with prefabricated workout plans and educational aspects	3.1 (0.59)	BE
ratiopharm Rückenschule (2.2.5)	ratiopharm Rückenschule für einen starken Rücken	ratiopharm GmbH	Free	October 13, 2021	Exercise tracker with prefabricated workout plans and educational aspects	3.1 (0.69)	BE
eCovery: Rücken, Hüfte & Knie (2.2.12)	eCovery: Rücken, Hüfte & Knie	eCovery GmbH	Free trial for 3 weeks	July 6, 2022	Artificial intelligence–based program with education and tracked exercises	4.1 (0.45)	BE and MBE
heyvie: Migräne & Resilienz (2.4.2)	heyvie: Resilienz & Migräne	HAIVE UG (haftungsbeschränkt)	Monthly subscription "Pro" (€9.99)	July 6, 2022	Artificial intelligence–based program with education and tracked exercises	4.2 (0.54)	BE and MBE
Rückentraining Gerade Haltung (1.2.1)	Rückentraining&Gerade Haltung	Nexoft Yazilim Limited Sirketi	Monthly subscription "Mitgliedschaft" (€4.49)	April 15, 2022	Exercise tracker with prefabricated plans	3.3 (0.25)	BE, AE[f], and MME[g]
AmbiCoach (1.1.27)	Dein Rückentraining: AmbiCoach	AmbiGate GmbH	Monthly subscription "Premium" (€49.99)	October 4, 2021	Exercise tracker with prefabricated plans and motion detection	3.6 (0.36)	BE

[a]MARS: Mobile Application Rating Scale.

[b]NICE: National Institute for Health and Care Excellence.

[c]A currency exchange rate of €1=US $1.02 is applicable.

[d]BE: biomechanical exercise.

[e]MBE: mind-body exercises.

[f]AE: aerobic exercises.

[g]MME: mixed modality exercises.

Consistency With Guidelines

All the included apps met some recommendations of the GNG. The mean value of items with the response "Yes" was 5.75 (SD 2.71). The mean value of items with the response "No" was 8.0 (SD 4.72). The mean value of items with the response "unclear" was 2.25 (SD 3.11). "Yes" was the most frequent response for item 14 (7/8, 88%), followed by item 2 (6/8, 75%) and then by items 8 and 13 (5/8, 62%). Items 6, 11, and 15 never received the response "Yes." No item was never answered "No". Items

2, 8, and 14 never received the response “Unclear”. Item 11 received the response “Unclear” most frequently (3/8, 38%). The results by item are shown in Table 2.

Table 2. German national guideline checklist items in the included apps (outcomes in total and per app).

Item	Yes, n (%)[a]	No, n (%)[b]	Unclear, n (%)[c]	Dein Rücken-training	ViViRa bei Rücken-schmerzen	Rücken-schmerzen Übungen	ratiopharm Rücken-schule	eCovery: Rücken, Hüfte & Knie	heyvie: Migräne & Re-silienz	Rücken-training Gerade Haltung	Ambi-Coach
1. Functional status	2 (25)	5 (62)	1 (12)	No	Yes	No	No	Yes	Unclear	No	No
2. Patient preferences	6 (75)	2 (25)	0 (0)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
3. Physical activity is safe	4 (50)	2 (25)	2 (25)	Yes	Yes	Yes	Unclear	Yes	No	No	No
4. Health-conscious behavior	3 (38)	4 (50)	1 (12)	Yes	Yes	No	Yes	Unclear	No	No	No
5. Promote understanding	3 (38)	4 (50)	1 (12)	Yes	Yes	No	No	Unclear	Yes	No	No
6. Education on healthy lifestyle	0 (0)	6 (75)	2 (25)	No	Unclear	No	No	Unclear	No	No	No
7. Maintaining activities	2 (25)	5 (62)	1 (12)	No	Yes	No	Yes	Unclear	No	No	No
8. Strength and endurance	5 (62)	3 (38)	0 (0)	Yes	Yes	Yes	Yes	Yes	No	No	No
9. Importance of activity	4 (50)	3 (38)	1 (12)	Yes	Yes	Unclear	No	Yes	Yes	No	No
10. Loading and resting	2 (25)	5 (64)	1 (12)	No	Yes	No	Yes	Unclear	No	No	No
11. Performance and pain	0 (0)	5 (62)	3 (38)	No	Unclear	No	No	Unclear	Unclear	No	No
12. Appropriate activities	2 (25)	5 (62)	1 (12)	No	Unclear	No	No	Yes	No	Yes	No
13. Iatrogenic fixations	5 (62)	2 (25)	1 (12)	Yes	Yes	Yes	Yes	Unclear	Yes	No	No
14. Preventing passive role	7 (88)	1 (12)	0 (0)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
15. Positive prognosis	0 (0)	6 (75)	2 (25)	No	No	No	No	Unclear	Unclear	No	No
16. Problematic patterns	1 (12)	6 (75)	1 (12)	No	No	No	No	Unclear	Yes	No	No

[a]Mean 5.75, SD 2.71; median 6.0; range 2.0-11.0.

[b]Mean 8.0, SD 3.11; median 9.0; range 1.0-14.0.

[c]Mean 2.25, SD 3.11; median 1.0; range 0.0-9.0.

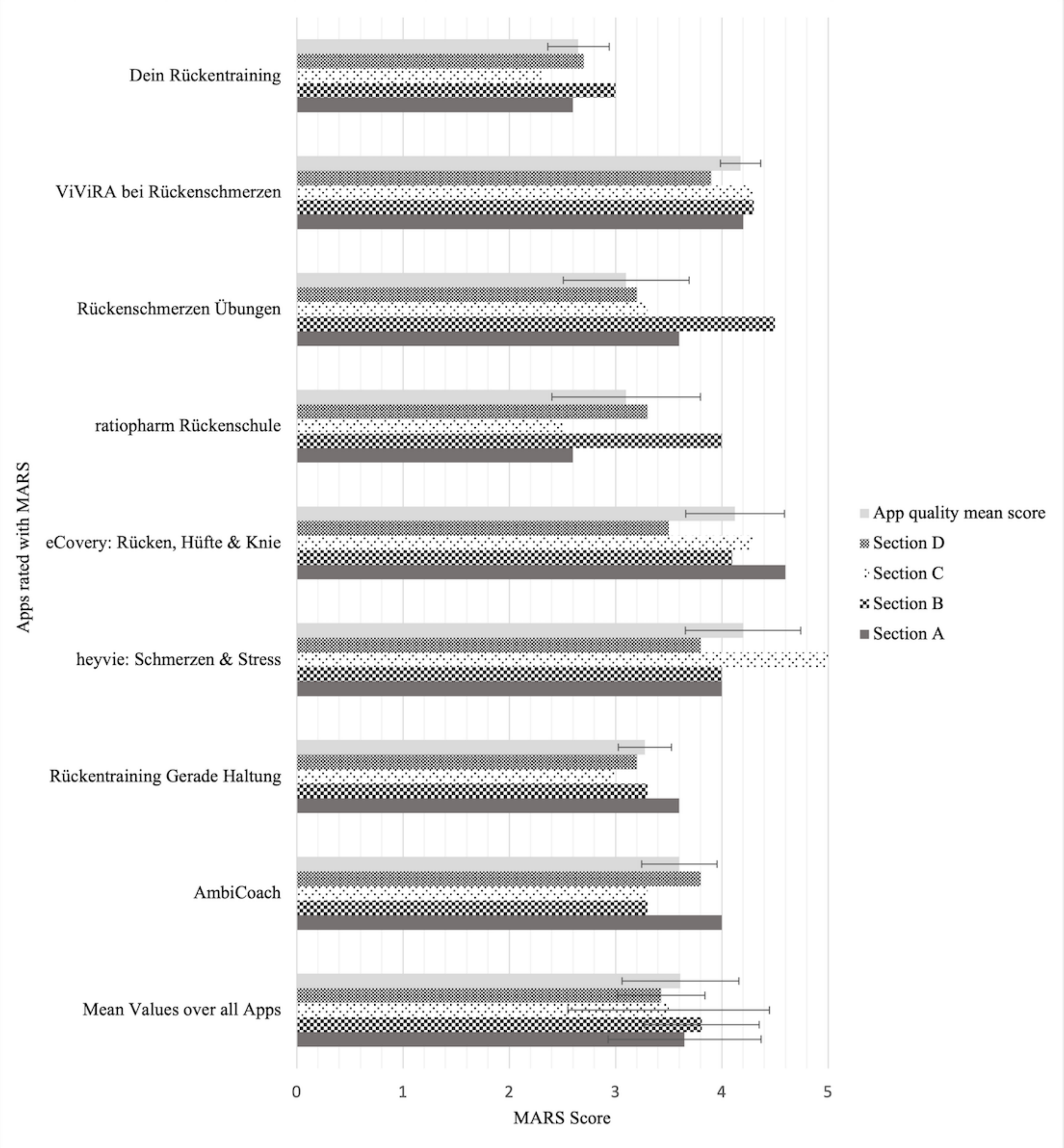
The app “ViViRa bei Rückenschmerzen” met the most recommendations (“Yes”: 11/16, 69%; “No”: 2/16, 12%; and “Unclear”: 3/16, 19%). For the app “eCovery: Rücken, Hüfte & Knie,” the most frequent response to recommendations was “Unclear” (“Yes”: 6/16, 38%; “No”: 1/16, 6%; and “Unclear”: 9/16, 56%). The app “AmbiCoach” met the fewest recommendations (“Yes”: 2/16, 12%; “No”: 14/16, 88%). The results by app are shown in Table 2.

All apps contained at least one form of exercise according to the NICE guideline [4]. All apps contained BE. Three apps also contained mind-body exercise. One app contained aerobic exercise and mixed modality exercise, in addition to BE. All forms of exercises are listed in Table 2.

App Quality

The overall mean MARS score for all apps included in the assessment was 3.61 (SD 0.55). “Section A–Engagement” surveyed whether apps were fun, engaging, and customizable in their use to increase users' engagement. The overall mean score for this section was 3.65 (SD 0.72). "Section B–Functionality" surveyed the functionality of the apps, in terms of usability, navigation, logical structure, and motor-gestural handling. The overall mean score for this section was 3.81 (SD 0.54). Therefore, the highest mean score was obtained in section B. "Section C–Aesthetics" surveyed the esthetics of the apps in terms of graphic design, visual stimuli, color design, and stylistic unity. The overall mean score for this section was 3.5 (SD 0.95). "Section D–Information" surveyed the quality of the apps' information and whether it was of high quality. The overall mean score for this section was 3.43 (SD 0.41), which was the lowest obtained mean score. The MARS scores for each section are shown in Figure 2.

Figure 2. Mobile Application Rating Scale (MARS) scores per section and mean scores over all Apps. Error bars indicate the SD of the mean values.



Discussion

Overview

In this systematic assessment, 8 apps for the treatment of NSLBP in the German language were identified. The recommendations of the GNG for NSLBP are partially fulfilled by these apps but often not. The quality of the apps, as measured by the MARS is acceptable (overall mean 3.61, SD 0.55). The section on information content had the lowest score among the apps. Each app contains at least one form of exercise that can be classified according to the NICE guideline. Therefore, all apps met the recommendations. Therefore, we can conclude that evidence-based and high-quality apps for the treatment of NSLBP in the German language are available. However, there is a big variance in how far the recommendations of the GNG are met. All apps provide at least one form of exercise that is classified and suggested by the NICE guideline [4].

Evidence According to the GNG Checklist

Many apps follow only a few recommendations of the GNG. The wide range of numbers of fulfilled recommendations is striking (range 2-11). For most apps, it was possible to clearly determine whether the recommendations were fulfilled. One exception was the app “eCovery Rücken, Hüfte & Knie.” For this app, many recommendations remained unclear. This was because of the design in which the app presented content or made it accessible.

Apps frequently met items 2, 8, 13, and 14 from the GNG checklist. Item 2 refers to whether the personal preferences of patients were considered [3]. However, the apps implemented this aspect in different ways. In some apps, single exercises could be rejected and were automatically replaced with new suggestions. Others allow the user to create their own individualized exercise plans from a selection of exercises. Taking personal preferences into consideration when treating patients is not only recommended by the GNG but also supported by evidence in the form of systematic reviews [3,26]. In addition, findings from qualitative research show that patients are more likely to adhere to exercise programs if they are created according to their personal preferences [27]. Therefore, it can be assumed that those apps with a more intense involvement of patient preferences are more likely to be used frequently. Item 8 asked whether education on improving strength and endurance was provided by the apps [3]. Systematic reviews and meta-analyses have concluded that strength or resistance training and endurance training have positive effects in treating patients with NSLBP [28]. Thus, education on the strength and endurance in apps is useful. Item 13 queried the recommendation of whether the risk of iatrogenic fixation of the patient is avoided, since iatrogenic fixation and early use of imaging techniques do not lead to improvement in symptoms [3]. Although the use of imaging leads to increased patient satisfaction, the outcomes of pain and function are not improved [29]. Consequently, the apps addressed to these patients should not encourage them to request more detailed examinations. Item 14 asked whether apps prevent medical procedures and applications that would push patients into passive coping [3]. A meta-analysis by Owen et al [30] showed that activities in the form of active interventions are superior to passive approaches. Accordingly, a large number of apps met items with clinically important suggestions.

Some recommendations of the GNG were not met by any app (items 6, 11, and 15). Item 6 asked whether continuous education and motivation for a healthy lifestyle with physical activity is provided by the apps [3]. In a systematic review on patients with chronic LBP, it was shown that pain and impairment can be reduced by health coaching in the sense of individual support for developing behavioral changes [31]. Thus, this seems to be an aspect of treatment that apps should support through education. Item 11 queried the recommended form of goal setting. According to the GNG, performance improvement without pain increase should be used as a goal definition instead of painlessness [3]. The included apps did not allow the individual selection of goals. For most apps, the set goal was pain reduction. An exception was “ViViRa bei Rückenschmerzen,” where the goal was performance enhancement. A systematic review by Haladay et al [32] shows that individual, patient-centered goals would be a useful addition to classic goals such as pain reduction. Apps should offer functions to formulate and track such goals. Item 15 queried whether apps provide guideline-compliant education on prognoses of the disease. Guideline-compliant information should include information about the frequency, the good prospects for recovery, and the self-limiting nature of the condition. In addition, it should be conveyed that pain does not necessarily mean actual tissue damage [3]. There were no apps that met this form of educational requirement. However, patient education is an important component of treatment [33]. Education can easily be provided by apps in the form of educational articles or videos, ideally with citations or links for further reading.

App Quality According to the MARS

There were 3 apps standing out in terms of quality with the rating “good” (“ViViRa bei Rückenschmerzen,” “eCovery: Rücken, Hüfte & Knie,” and “heyvie: Migräne & Resilienz”). There was 1 app that was rated “poor” (“Dein Rückentraining”). Most apps achieved the rating “acceptable.” In “Section A–Engagement,” most apps achieved a rating from “acceptable” to “good.” “Section B–Functionality” achieved the highest mean score. On average, the apps considered were good. Solitary apps such as “Dein Rückentraining,” “Rückentraining Gerade Haltung,” and “AmbiCoach” achieved the rating “acceptable.” The reasons were low ratings for the “navigation,” “ease of use,” and “performance” criteria. In “Section C–Aesthetics,” most apps achieved ratings from “acceptable” to “good.” Here, owing to its professional layout, high graphic quality, and unique design features, the app “heyvie: Migräne & Resilienz” achieved the best possible rating. In “Section D–Information,” the apps achieved the lowest scores. This could be because apps often do not claim to provide educational aspects but are rather intended as instructions and support for exercising. Only the store descriptions of 3 apps state that the app offers educational content (“ratiopharm Rückenschule,” “heyvie: Migräne & Resilienz,” and “eCovery: Rücken, Hüfte & Knie”). Because of the operational app design of “eCovery: Rücken, Hüfte & Knie,” educational information could hardly be considered. For the evaluation, the app was only used on 1 day. Most apps are presented in their store descriptions as instructions for exercises in the form of a home workout. As those apps did not aim to provide education, they lost points in the MARS quality rating but still might be useful apps to facilitate exercising.

Comparison With Prior Studies

Eight apps were identified in this systematic assessment. This is significantly fewer than that of previous studies, where 17 to 61 apps were identified [14-16]. This was because of the inclusion and exclusion criteria used. Our research was the first to include apps that were available in the Apple App Store as well as in the Google Play Store. The overall mean score of the quality of included apps (mean 3.61, SD 0.55) collected using the MARS was higher than that in the study by Machado et al [14] (mean 2.36, SD 0.83) but similar to those in the studies by Didyk et al [15] (mean 3.9, SD 0.5) and Escriche-Escuder et al [16] (mean 3.82). On the basis of this, it seems newer research tends to identify apps of higher quality. In the research by Machado et al [14] and Didyk et al [15], "Section A–Engagement" reached the lowest scores. In contrast, "Section D–Information" achieved the lowest score in our study. In all the aforementioned studies, including ours, "Section B–Functionality" achieved the highest mean score [14-16]. In our study, evidence was found for only 1 app. Evidence on any reviewed app was also rare or nonexistent in previous research [14-16]. All research, including ours, detected a maximum of 3 points for item 18, "credibility." Thus, the identified apps always originated from commercial businesses. In our research, all apps met the recommendations of the NICE guideline; this was also the case for almost all apps from the research by Machado et al [14] and for all apps from the research by Didyk et al [15]. However, in the latter case, this was a criterion for inclusion. Escriche-Escuder et al [16] did not collect this information. What was new in our research was the detailed evaluation along the GNG guideline checklist, which showed a wide range of recommendations met; 11 were met by the highest-rated app and only 2 were met by the lowest-rated app.

Use of Apps With Patients

Apps could be a useful addition to physiotherapeutic treatment. This is particularly conceivable for apps that have achieved high ratings. However, apps with lower ratings could also be useful if used appropriately. Palazzo et al [34] conducted a qualitative study on barriers to the implementation of home exercise programs in patients with chronic LBP. They found that the implementation of home exercise programs could be promoted through attractive designs and the provision of safety while exercising. Young patients were particularly interested in using new technologies [34]. Other studies showed improved adherence to home exercise programs when digital interventions were used [35,36]. In this context, apps could conceivably be used as a tool to support and implement home exercise programs. The positive effects of such programs on pain and function have been well studied [37]. The communication of educational aspects to patients via an app must be carefully considered, as only the app "ViViRa bei Rückenschmerzen" presented the sources used transparently. However, their descriptions were partly inaccurate. For example, in the educational text material on the development of pain, the term "nociception" was introduced very late and the term "pain stimulus" was used instead. This is not consistent with the terminology proposed by the International Association for the Study of Pain [38]. Other apps used negative and catastrophizing wording ("Dein Rückentraining" and "Rückenschule"). Such wording could have negative effects on the prognosis in the form of nocebo effects or fear-avoidance beliefs [39,40]. Such negative effects on patient prognosis are known from the presentation of magnetic resonance imaging results. Patients to whom results are explained as normal changes have more positive prognoses than those to whom presenting pathologies were explained in detail and without their clinical meaning [41]. Accordingly, the use of apps could consider patients' beliefs, knowledge, and fears. This is supported by the results of qualitative research, according to which the implementation of exercises is promoted when these aspects are considered in therapy [27,34]. However, patients also desire personalized advice and guidance from therapists [27,34,42]. Consequently, different versions and ways of working with apps may be appropriate for different patients. To ensure personalized advice and guidance, first contact with a professional remains crucial. This guidance, along with patient beliefs, knowledge, and fears, requires thorough clinical examinations including physical and psychosocial assessments, such as a stratification of patients based on their risk of chronification [3,43]. If a low risk is detected, patients can be treated with education and an exercise program [43]. Apps could be useful to deliver such education and facilitate exercises. High-risk patients should receive multimodal treatment guided by professionals [43]. It is advisable to closely involve the patient in the decision-making process regarding whether to use an app. Furthermore, we suggest using a screening tool with appropriate diagnostic properties to determine the patients' risk of chronicity before deciding to use an app.

Limitations

A few methodological limitations of this study can be noted. When apps were rated by 2 reviewers, there was no standardization of app installation. Although the raters used the same devices, technical differences appeared in the consensus process. The anamneses performed by 2 apps were not answered in a standardized way, so the raters were probably shown different content. These factors of individualization and technical aspects might have led to differences in ratings and thus to poor and moderate consensus. The evaluation of the apps took place during their use on 1 day. Some apps might meet more guideline aspects with longer use. A conceivable bias in the overall process would be that apps with good, professional, and appealing designs were perhaps also rated better in other criteria in terms of a primacy effect [44]. Since higher app prices correlate with higher quality [14,15], a biased rating of such apps in the sense of a confirmation bias is also conceivable [45]. Furthermore, the checklist used was not a validated instrument, and no criteria were formulated as to when a recommendation was or was not met by an app. This could be improved by precisely formulated conditions for the answer options. In addition, there was no weighting of the items on which ones are especially important to fulfill.

No apps with hybrid treatment approaches, such as web-based consultation with medical professionals, were investigated in this study. The focus was on the identification of apps and their evaluation by using rating tools.

This study was guided, among others, by the approach for systematic reviews according to the PRISMA statement [18]. However, app evaluation is very different from the evaluation of scientific literature. This was particularly noticeable in the process of reaching a consensus. The low level of consensus among raters in the guideline checklist may also reflect this. In contrast to the work of Didyk et al [15] and Machado et al [14], the raters in this study achieved a low ICC score. However, in both studies, the sample size was significantly larger [14,15]. In addition, there were no trial runs followed by consensus building, as recommended in the MARS web-based training by Stoyanov [21].

Conclusions

All apps considered in this systematic assessment met the recommendations of the GNG and included exercise forms classified and recommended by the NICE guideline. Most apps are of acceptable or good quality. There are apps with different designs: apps that create and guide home exercise programs and apps that create programs on their own or contain ready-made programs. Home exercise programs for the treatment of LBP are well researched. The use of apps as an adjunct to therapy could be useful if they succeed in getting patients to implement such programs or help in patient education. Whether health apps succeed in these matters should be the subject of research during the process of app development and publication, as it is required by the *Bundesinstitut für Arzneimittel und Medizinprodukte* to list such apps as a DiGA. The decision on whether and which app to use should be made in consideration of the preferences, knowledge, beliefs, and fears of patients in a joint exchange with a medical professional. Apps that create exercise programs should be tested for their effectiveness. Given the number of available apps for the treatment of NSLBP, an international checklist or assessment tool exclusively for the rating of such apps could be useful. By defining the requirements for safe and evidence-based treatment approaches for apps, such a tool could help to identify high-quality apps.

Data Availability

Additional material and information can be obtained from the corresponding author.

Authors' Contributions

All 3 authors were involved in the design of the study, whereas LU introduced the idea for the study. LU drafted the manuscript and analyzed the data. All 3 authors contributed to the draft of the manuscript and interpretation of the analyzed data. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion and exclusion of apps from screening phase 1 and 2.
[PDF File (Adobe PDF File), 76 KB - mhealth_v11i1e47502_app1.pdf]

Multimedia Appendix 2
Original German national guideline checklist items and English translations.
[PDF File (Adobe PDF File), 138 KB - mhealth_v11i1e47502_app2.pdf]

References

1. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299 [FREE Full text] [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
2. Raspe H. Themenheft 53 "Rückenschmerzen". Robert Koch-Institut. 2012. URL: https://edoc.rki.de/handle/176904/3239 [accessed 2022-03-09]
3. Nicht-spezifischer Kreuzschmerz. Programm für Nationale VersorgungsLeitlinien. 2017 Mar 02. URL: https://www.leitlinien.de/themen/kreuzschmerz [accessed 2022-07-06]
4. National Guideline Centre. Low Back Pain and Sciatica in Over 16s: Assessment and Management. London, UK: National Institute for Health and Care Excellence; 2016.
5. Huber S, Priebe JA, Baumann KM, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
6. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May 03;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
7. Priebe JA, Utpadel-Fischler D, Toelle TR. Less pain, better sleep? The effect of a multidisciplinary back pain app on sleep quality in individuals suffering from back pain - a secondary analysis of app user data. J Pain Res 2020 May 20;13:1121-1128 [FREE Full text] [doi: 10.2147/JPR.S232792] [Medline: 32547175]

8. Moman RN, Dvorkin J, Pollard EM, Wanderman R, Murad MH, Warner DO, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain-is it time to start prescribing electronic health applications? Pain Med 2019 Nov 01;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]
9. Engpassanalyse. Bundesagentur für Arbeit. 2021. URL: https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Fachkraeftebedarf/Engpassanalyse-Nav.html
10. Malin L, Hickmann H. Jahresrückblick- der Arbeitsmarkt 2021. Institut Der Deutschen Wirtschaft. 2021. URL: https://www.iwkoeln.de/studien/lydia-malin-helen-hickmann-jahresrueckblick-der-arbeitsmarkt-2021.html [accessed 2022-07-27]
11. Fachkräftemangel: keine Besserung in Sicht. Bundesverband selbstständiger Physiotherapeuten — IFK e. V. 2020. URL: https://ifk.de/artikel/fachkraeftemangel-keine-besserung-sicht
12. Tenzer F. Smartphone-Nutzung in Deutschland. Statista. URL: https://de.statista.com/statistik/studie/id/71707/dokument/smartphone-nutzung-in-deutschland/ [accessed 2022-03-18]
13. Bundesärztekammer, Kassenärztliche Bundesvereinigung. Gesundheits-Apps im klinischen Alltag: Handreichung für Ärztinnen und Ärzte. Ärztliches Zentrum für Qualität in der Medizin. 2020. URL: https://www.aezq.de/gesundheitsapps/ueberblick/# [accessed 2022-08-31]
14. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
15. Didyk C, Lewis LK, Lange B. Availability, content and quality of commercially available smartphone applications for the self-management of low back pain: a systematic assessment. Disabil Rehabil 2022 Dec;44(24):7600-7609. [doi: 10.1080/09638288.2021.1979664] [Medline: 34612133]
16. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (Apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Android vs. iOS: Smartphone OS sales market share evolution. Kantar Worldpanel. URL: https://www.kantarworldpanel.com/global/smartphone-os-market-share [accessed 2022-08-22]
20. Polanin JR, Pigott TD, Espelage DL, Grotpeter JK. Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. Res Syn Meth 2019 Jun 24;10(3):330-342 [FREE Full text] [doi: 10.1002/jrsm.1354]
21. MARS training video. Stoyan Stoyanov YouTube page. 2016 Jun 14. URL: https://www.youtube.com/watch?v=25vBwJQIOcE [accessed 2022-03-21]
22. Urbaniak GC, Plous S. Research randomizer. 2013. URL: https://www.randomizer.org/ [accessed 2022-08-22]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
25. WMA declaration of Helsinki – ethical principles for medical research involving human subjects. World Medical Association. 2022 Sep 06. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects [accessed 2023-08-17]
26. Karlsson M, Bergenheim A, Larsson ME, Nordeman L, van Tulder M, Bernhardsson S. Effects of exercise therapy in patients with acute low back pain: a systematic review of systematic reviews. Syst Rev 2020 Aug 14;9(1):182 [FREE Full text] [doi: 10.1186/s13643-020-01412-8] [Medline: 32795336]
27. Slade SC, Patel S, Underwood M, Keating JL. What are patient beliefs and perceptions about exercise for nonspecific chronic low back pain? A systematic review of qualitative studies. Clin J Pain 2014 Nov;30(11):995-1005. [doi: 10.1097/AJP.0000000000000044] [Medline: 24300225]
28. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Apr 24;4(4):CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
29. Kendrick D, Fielding K, Bentley E, Kerslake R, Miller P, Pringle M. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. BMJ 2001 Feb 17;322(7283):400-405 [FREE Full text] [doi: 10.1136/bmj.322.7283.400] [Medline: 11179160]
30. Owen PJ, Miller CT, Mundell NL, Verswijveren SJ, Tagliaferri SD, Brisby H, et al. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. Br J Sports Med 2020 Nov;54(21):1279-1287 [FREE Full text] [doi: 10.1136/bjsports-2019-100886] [Medline: 31666220]

31. Prior JL, Vesentini G, Michell De Gregorio JA, Ferreira PH, Hunter DJ, Ferreira ML. Health coaching for low back pain and hip and knee osteoarthritis: a systematic review with meta-analysis. Pain Med 2023 Jan 04;24(1):32-51 [FREE Full text] [doi: 10.1093/pm/pnac099] [Medline: 35775931]
32. Haladay D, Swisher L, Hardwick D. Goal attainment scaling for patients with low back pain in rehabilitation: a systematic review. Health Sci Rep 2021 Sep 22;4(3):e378 [FREE Full text] [doi: 10.1002/hsr2.378] [Medline: 34589616]
33. Traeger AC, Hübscher M, Henschke N, Moseley GL, Lee H, McAuley JH. Effect of primary care-based education on reassurance in patients with acute low back pain: systematic review and meta-analysis. JAMA Intern Med 2015 May;175(5):733-743. [doi: 10.1001/jamainternmed.2015.0217] [Medline: 25799308]
34. Palazzo C, Klinger E, Dorner V, Kadri A, Thierry O, Boumenir Y, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113 [FREE Full text] [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
35. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
36. Bennell KL, Marshall CJ, Dobson F, Kasza J, Lonsdale C, Hinman RS. Does a web-based exercise programming system improve home exercise adherence for people with musculoskeletal conditions?: A randomized controlled trial. Am J Phys Med Rehabil 2019 Oct;98(10):850-858. [doi: 10.1097/PHM.0000000000001204] [Medline: 31021823]
37. Quentin C, Bagheri R, Ugbolue UC, Coudeyre E, Pélissier C, Descatha A, et al. Effect of home exercise training in patients with nonspecific low-back pain: a systematic review and meta-analysis. Int J Environ Res Public Health 2021 Aug 10;18(16):8430 [FREE Full text] [doi: 10.3390/ijerph18168430] [Medline: 34444189]
38. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised international association for the study of pain definition of pain: concepts, challenges, and compromises. Pain 2020 Sep 01;161(9):1976-1982 [FREE Full text] [doi: 10.1097/j.pain.0000000000001939] [Medline: 32694387]
39. Corsi N, Emadi Andani M, Sometti D, Tinazzi M, Fiorio M. When words hurt: verbal suggestion prevails over conditioning in inducing the motor nocebo effect. Eur J Neurosci 2019 Oct;50(8):3311-3326. [doi: 10.1111/ejn.14489] [Medline: 31209960]
40. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. Spine J 2014 May 01;14(5):816-36.e4. [doi: 10.1016/j.spinee.2013.09.036] [Medline: 24412032]
41. Rajasekaran S, Dilip Chand Raja S, Pushpa BT, Ananda KB, Ajoy Prasad S, Rishi MK. The catastrophization effects of an MRI report on the patient and surgeon and the benefits of 'clinical reporting': results from an RCT and blinded trials. Eur Spine J 2021 Jul;30(7):2069-2081. [doi: 10.1007/s00586-021-06809-0] [Medline: 33748882]
42. Bernhardsson S, Samsson KS, Johansson K, Öberg B, Larsson ME. A preference for dialogue: exploring the influence of patient preferences on clinical decision making and treatment in primary care physiotherapy. Eur J Physiother 2019;21(2):107-114 [FREE Full text] [doi: 10.1080/21679169.2018.1496474]
43. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet 2011 Oct 29;378(9802):1560-1571 [FREE Full text] [doi: 10.1016/S0140-6736(11)60937-9] [Medline: 21963002]
44. Peterson CR, DuCharme WM. A primacy effect in subjective probability revision. J Exp Psychol 1967 Jan;73(1):61-65. [doi: 10.1037/h0024139] [Medline: 6031669]
45. Elston DM. Confirmation bias in medical decision-making. J Am Acad Dermatol 2020 Mar;82(3):572. [doi: 10.1016/j.jaad.2019.06.1286] [Medline: 31279036]

Abbreviations

BE: biomechanical exercise
DiGA: Digitale Gesundheitsanwendungen
GNG: German national guideline
ICC: intraclass correlation coefficient
LBP: low back pain
MARS: Mobile Application Rating Scale
mHealth: mobile health
NICE: National Institute for Health and Care Excellence
NSLBP: nonspecific low back pain
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 24.03.23; peer-reviewed by C Didyk, L Kerkemeyer; comments to author 27.04.23; revised version received 12.06.23; accepted 12.07.23; published 13.09.23.

Please cite as:
Ulrich L, Thies P, Schwarz A
Availability, Quality, and Evidence-Based Content of mHealth Apps for the Treatment of Nonspecific Low Back Pain in the German Language: Systematic Assessment
JMIR Mhealth Uhealth 2023;11:e47502
URL: https://mhealth.jmir.org/2023/1/e47502
doi:10.2196/47502
PMID:37703072

Original Paper

Association Between the Characteristics of mHealth Apps and User Input During Development and Testing: Secondary Analysis of App Assessment Data

Anna-Lena Frey[1], PhD; Rebecca Baines[1,2], PhD; Sophie Hunt[1], BSc; Rachael Kent[3], PhD; Tim Andrews[1,4], BA; Simon Leigh[1,4], PhD

[1]Organisation for the Review of Care and Health Apps, Daresbury, United Kingdom
[2]Centre for Health Technology, University of Plymouth, Plymouth, United Kingdom
[3]Department of Digital Humanities, King's College London, London, United Kingdom
[4]Warwick Medical School, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Anna-Lena Frey, PhD
Organisation for the Review of Care and Health Apps
V2, Sci-Tech Daresbury
Keckwick Lane
Daresbury, WA4 4FS
United Kingdom
Phone: 44 01925 606542
Email: anna.frey@orchahealth.com

Abstract

Background: User involvement is increasingly acknowledged as a central part of health care innovation. However, meaningful user involvement during the development and testing of mobile health apps is often not fully realized.

Objective: This study aims to examine in which areas user input is most prevalent and whether there is an association between user inclusion and compliance with best practices for mobile health apps.

Methods: A secondary analysis was conducted on an assessment data set of 1595 health apps. The data set contained information on whether the apps had been developed or tested with user input and whether they followed best practices across several domains. Background information was also available regarding the apps' country of origin, targeted condition areas, subjective user ratings, download numbers, and risk (as per the National Institute for Health and Care Excellence Evidence Standards Framework [ESF]). Descriptive statistics, Mann-Whitney *U* tests, and Pearson chi-square analyses were applied to the data.

Results: User involvement was reported by 8.71% (139/1595) of apps for only the development phase, by 33.67% (537/1595) of apps for only the testing phase, by 21.88% (349/1595) of apps for both phases, and by 35.74% (570/1595) of apps for neither phase. The highest percentage of health apps with reported user input during *development* was observed in Denmark (19/24, 79%); in the condition areas of diabetes (38/79, 48%), cardiology (15/32, 47%), pain management (20/43, 47%), and oncology (25/54, 46%); and for high app risk (ESF tier 3a; 105/263, 39.9%). The highest percentage of health apps with reported user input during *testing* was observed in Belgium (10/11, 91%), Sweden (29/34, 85%), and France (13/16, 81%); in the condition areas of neurodiversity (42/52, 81%), respiratory health (58/76, 76%), cardiology (23/32, 72%), and diabetes (56/79, 71%); and for high app risk (ESF tier 3a; 176/263, 66.9%). Notably, apps that reported seeking user input during testing demonstrated significantly more downloads than those that did not (P=.008), and user inclusion was associated with better compliance with best practices in clinical assurance, data privacy, risk management, and user experience.

Conclusions: The countries and condition areas in which the highest percentage of health apps with user involvement were observed tended to be those with higher digital maturity in health care and more funding availability, respectively. This suggests that there may be a trade-off between developers' willingness or ability to involve users and the need to meet challenges arising from infrastructure limitations and financial constraints. Moreover, the finding of a positive association between user inclusion and compliance with best practices indicates that, *where no other guidance is available*, users may benefit from prioritizing health apps developed with user input as the latter may be a proxy for broader app quality.

(JMIR Mhealth Uhealth 2023;11:e46937) doi:10.2196/46937

KEYWORDS

patient and public involvement; user involvement; mobile apps; digital health; mobile health; quality assessment

Introduction

Background

User involvement, also referred to as patient and public involvement, is increasingly being acknowledged as a central part of health care innovation [1]. In recent years, user involvement policies and strategies have been developed by organizations such as the National Institute for Health and Care Excellence (NICE) [2], the UK National Health Service [3,4], and the US Food and Drug Administration [5] to ensure that patient and public input is actively taken into account during the research and development of medical guidelines, products, and services.

A key social and political driver behind the increased emphasis placed on user involvement is the recognition that health care transformation is crucial to addressing current challenges and that, for such transformation to be efficient and successful, a people-centered approach needs to be taken [3,6]. For instance, the pressures faced by health care systems as a result of an increase in the number of older adult patients may be partially eased through more widespread use of health technologies and data. However, to achieve a positive impact, technologies have to address patient needs and must be easy to use, which can arguably only be accomplished by involving the public in the design of such innovations [6]. Another important challenge that has become particularly apparent during the COVID-19 pandemic is inequality in health care [7]. To ensure that new health technologies help improve rather than exacerbate such inequalities, it is critical to actively involve disadvantaged individuals to understand their needs and the barriers they may face to accessing health innovations [3]. In addition, shifts in demographics, such as an aging population and an increase in migration, highlight the need to seek input from diverse groups of individuals to ensure that their potential concerns related to new technologies are considered [6]. If such a participatory approach to health care innovation is taken, this can lead to more efficient use of resources and higher product uptake [3,8,9], as well as to the empowerment of patients to be active partners in decisions affecting their care [9,10].

However, despite these benefits, co-design and other forms of user involvement during the development and testing of mobile health apps are often not regarded as essential or are not fully realized [11-13]. Some developers express support for the idea of user involvement without implementing it, whereas others do not follow sound methodologies or do not meaningfully involve users throughout the app life cycle [11-13]. Moreover, remote user involvement, which has been on the rise since the beginning of the COVID-19 pandemic, brings with it additional implementation challenges related to technology access, digital literacy, data privacy, and difficulties in "claiming space" to speak out during web-based meetings that may not always be adequately addressed [14-16].

Given the reported benefits of meaningfully involving users, it is important to understand in which areas user inclusion is currently being implemented during the development and testing of mobile health apps. Such insights can help identify areas in which more education and support are needed to encourage app developers to seek user input. Moreover, an understanding of the possible relationships among user involvement, compliance with best practices, and app use and acceptance (eg, as indicated by user ratings and download numbers) can provide valuable guidance for future innovation practices.

Objectives

Therefore, the aims of this study were 3-fold: first, to examine how, if at all, the prevalence of user input during health app development and testing differs across countries, condition areas, and app risks; second, to determine whether user involvement is associated with higher user ratings or app download numbers; and, finally, to assess whether user input is associated with better compliance with best practices across the domains of clinical assurance, data privacy, risk management, and user experience. Exploring such associations may help strengthen the case for stakeholders, including patients and health care professionals, to prioritize mobile health apps that included users during their development and testing as this may be a proxy for broader app quality.

Methods

Data Provenance and Characteristics

A secondary analysis was conducted on a data set containing background and assessment information on 1595 mobile health apps. The data were collected between January 2021 and January 2022 as part of an app review conducted by the Organisation for the Review of Care and Health Apps (ORCHA), a digital health compliance company that specializes in the evaluation of mobile health apps and is currently working with National Health Service providers across 70% of regions in England [17].

During the ORCHA Baseline Review (OBR; version 6), each app was evaluated using approximately 300 objective (mostly binary yes or no) questions. The initial evaluation was performed by 1 of the ORCHA assessors, all of whom had undergone a thorough 6-month training course on standard operating procedures for the OBR and perform health app evaluations on a daily basis as part of their job. The OBR responses for all 1595 apps were reviewed and signed off by a second more senior assessor. Any disagreements between the 2 assessors were resolved by involving a third assessor with extensive review experience or, in the case of more difficult issues, were discussed by a panel of subject matter experts in the areas of assessment, clinical practice, and research who resolved the matter through a consensus decision. Notably, the OBR questions were based on standards, guidelines, and regulatory requirements such as the UK Medicines and Healthcare Products Regulatory Agency guidance on software as a medical device, the European Union General Data Protection Regulation (GDPR), the Web Content Accessibility Guidelines, the UK National Health Service DCB0129 Clinical Risk Management standard, and the NICE Evidence Standards Framework (ESF; see further details in the following paragraphs). As such, the OBR questions reflect widely accepted best practices.

For this study, a subset of 14 assessment questions from the OBR was selected for analysis based on their hypothesized association with user involvement. The questions covered best practices in the following domains: clinical assurance, data privacy, risk management, and user experience. The exact phrasing of the questions is noted in the *Results* section.

As part of the assessment, mobile health apps were classified into different tiers following the NICE ESF [18]. Tiers were assigned to the apps based on their functionality, which has implications for the risk and evidentiary requirements of the apps. The following ESF definitions were applied (note that the ESF has since been updated, with tiers 1, 2, and 3a or 3b having been replaced by tiers A, B, and C, respectively [19]):

1. Tier 1: apps that provide health and social care services with no measurable user outcomes
2. Tier 2: apps that provide 2-way communication between users and health care professionals, provide health information, or offer a health diary
3. Tier 3a: apps that support preventative behavior change aimed at health issues or allow users to self-manage a specific condition
4. Tier 3b: apps that provide or guide treatment for a condition; record and transmit data about this condition to a health care professional, carer, or third party without the user's input; contain a calculator that affects treatment; or guide diagnosis [18]

In addition to the ESF tier, the following background information was collected for each app: country of origin, app store user rating, downloads, and targeted condition areas. Note that download numbers were only available from the Google Play Store (for Android apps; n=777) and not from the iOS App Store. Moreover, a given app could cover more than one condition area.

Determination of User Input

To classify mobile health apps into those that did and did not report seeking user input during development or testing, 2 OBR assessment questions were used. These questions are listed below together with the conditions under which they were answered with *yes* (thus indicating user input).

Question 1: Is There a Statement Within the App or Store About User Feedback During Design or Development?

This question was answered with *yes* if information within the app or an associated policy or website stated that the app (1) was changed based on feedback received from users (eg, through suggestion forms provided on the associated website), (2) underwent a survey or pilot study and changes were made based on the outcome, or (3) was designed by the developer or publisher as a remedy to a problem that they (or someone they were caring for) were experiencing (that is to say, the developer was part of the intended user group and, therefore, had a firsthand understanding of user needs).

Question 2: Is There a Statement Within Either the App or Store About User Input During Testing?

This question was answered with "yes" if information within the app or an associated policy or website mentioned (1) a case study for the app, (2) that a beta version of the app was available before the app went live, (3) user feedback stating that the app is beneficial, (4) evidence of indicated user benefits, or (5) any other evidence of user testing.

Statistical Analysis

Descriptive statistics were used to examine the prevalence of user input for mobile health apps across different countries, condition areas, and ESF tiers. Note that only countries and condition areas with >10 apps were included in the relevant data summaries as samples of <10 apps are likely not representative of the larger app "population" in a given country or condition area.

Mann-Whitney *U* tests were conducted to assess differences in user ratings and download levels between mobile health apps that did and did not report seeking user input during development or testing. The Mann-Whitney *U* test is a nonparametric test used to compare ordinal or nonnormally distributed continuous dependent variables between 2 independent groups. Test statistics are calculated by placing the values of the dependent variable in ascending order (disregarding group membership); assigning a rank to each value; and then using the group sum of ranks and the group sample sizes to calculate the *U* value, the *Z* statistic, and an associated *P* value (for further details, see the book by Field [20]). We used Mann-Whitney tests as Shapiro-Wilks tests indicated that the user rating data were not normally distributed and as the download data were ordinal. Specifically, only download *ranges* were available from the app store (eg, 1-4, 5-9, 10-49, 50-99, and 100-499 up to 1 billion downloads), and each range was treated as a separate ordinal category in the analysis, referred to in the following sections as "download levels" (see Multimedia Appendix 1 for details).

Furthermore, Pearson chi-square analyses were performed to examine bivariate associations between user input and specific quality indicators of mobile health apps (as captured by individual dichotomous assessment questions), and odds ratios were reported. The Pearson chi-square test was used to examine relationships between 2 categorical variables, which can be represented in an *i* by *j* table, with *i* designating the number of categories in the first variable and *j* the number of categories in the second variable. The chi-square value represents the sum of squares of SDs between the *observed* frequencies within each cell of the table and the *expected* frequencies in each cell, with the latter being determined based on the total number of observations for the different categories. By comparing the calculated chi-square value against the critical values of the known chi-square distribution, considering the *df*, a *P* value can be obtained (for further details, see the book by Field [21]).

Unless otherwise indicated, the reported results remained significant after multiple-comparison correction for the 28 conducted association analyses (user involvement during testing and development examined for 14 questions) using the

Benjamini-Hochberg method with a false discovery rate of 10%. The Benjamini-Hochberg method is a multiple-comparison correction method that controls the false discovery rate. As part of this method, *P* values are placed in ascending order; a rank is assigned to each *P* value; and a Benjamini-Hochberg critical value is calculated for each *P* value using the rank, total number of tests, and selected false discovery rate. All *P* values above (but not below) the largest *P* value that is smaller than its critical value are considered significant (for further details, see the work by Benjamini and Hochberg [22]).

All analyses were conducted using SPSS Statistics (version 27.0; IBM Corp). Statistical significance was defined at the usual 5% level (ie, $P<.05$).

Ethical Considerations

As part of the OBR process, developers are informed of their assessment results and given the opportunity to contest the findings and request an amendment based on additional information. Moreover, the ORCHA privacy policy states that all reviews can be used for research purposes unless the developer asks for their app to be removed from the research database. Furthermore, outputs were anonymized in this paper, with no individual mobile health apps being named. As no data from human participants was used in this study, ethical approval was not required.

Results

App Characteristics

The number of mobile health apps within each country, condition area, and ESF tier can be found in Tables 1-3 and the distribution of user ratings and download levels across all apps is presented in Figure 1. Within the current data set, the largest number of health apps was developed in the United States (490/1595, 30.72%) and the United Kingdom (419/1595, 26.27%). The most covered condition areas were healthy living (563/1595, 35.3%), mental health (440/1595, 27.59%), and neurological conditions (135/1595, 8.46%).

Overall, user involvement was reported by 8.71% (139/1595) of the apps for only the development phase, by 33.67% (537/1595) of the apps for only the testing phase, by 21.88% (349/1595) of the apps for both phases, and by 35.74% (570/1595) of the apps for neither phase.

Table 1. Number and percentage of mobile health apps across countries that reported seeking user input during development and testing.

Country of origin	Apps with user input during development, n (%)	Apps with user input during testing, n (%)
Denmark (n=24)	19 (79.2)	17 (70.8)
United Kingdom (n=419)	185 (44.2)	255 (60.9)
France (n=16)	7 (43.8)	13 (81.3)
Turkey (n=14)	6 (42.9)	9 (64.3)
Canada (n=71)	30 (42.3)	38 (53.5)
Belgium (n=11)	4 (36.4)	10 (90.9)
Netherlands (n=30)	10 (33.3)	21 (70)
Israel (n=22)	7 (31.8)	16 (72.7)
China (n=10)	3 (30)	6 (60)
Spain (n=24)	7 (29.2)	11 (45.8)
Poland (n=11)	3 (27.3)	1 (9.1)
Sweden (n=34)	9 (26.5)	29 (85.3)
Germany (n=51)	12 (23.5)	33 (64.7)
United States (n=490)	113 (23.1)	274 (55.9)
Australia (n=36)	8 (22.2)	15 (41.7)
Ireland (n=19)	4 (21.1)	14 (73.7)
India (n=43)	2 (4.7)	14 (32.6)
Russia (n=16)	0 (0)	3 (18.8)
Singapore (n=20)	0 (0)	8 (40)

Table 2. Number and percentage of mobile health apps across condition areas that reported seeking user input during development and testing[a].

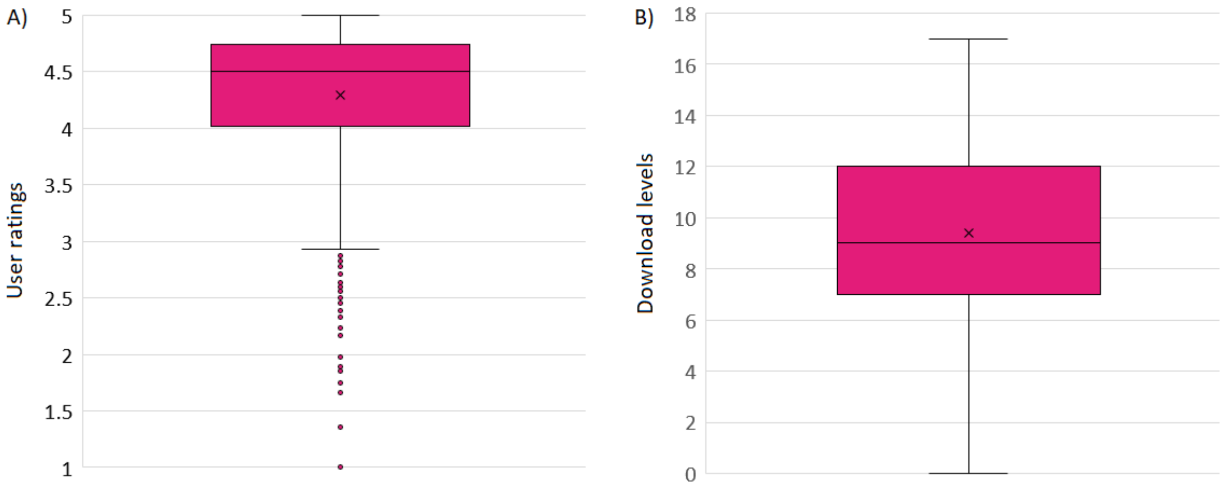
Targeted condition or group	Apps with user input during development, n (%)	Apps with user input during testing, n (%)
Diabetes (n=79)	38 (48.1)	56 (70.9)
Cardiology (n=32)	15 (46.9)	23 (71.9)
Pain management (n=43)	20 (46.5)	29 (67.4)
Cancer (n=54)	25 (46.3)	32 (59.3)
Neurological (n=135)	52 (38.5)	77 (57)
Respiratory (n=76)	29 (38.2)	58 (76.3)
Gastrointestinal (n=24)	9 (37.5)	12 (50)
Musculoskeletal (n=53)	19 (35.8)	35 (66)
Child health (n=71)	24 (33.8)	45 (63.4)
Pregnancy (n=81)	25 (30.9)	43 (53.1)
Allergy (n=13)	4 (30.8)	7 (53.8)
Neurodiverse (n=52)	15 (28.8)	42 (80.8)
Women’s health (n=66)	19 (28.8)	35 (53)
Mental health (n=440)	126 (28.6)	273 (62)
Dental (n=25)	7 (28)	13 (52)
Health living (n=563)	137 (24.3)	316 (56.1)
Dermatology (n=29)	7 (24.1)	12 (41.3)
Otorhinolaryngology (n=25)	5 (20)	9 (36)
Ophthalmology (n=56)	10 (17.9)	25 (44.6)
Sexual health (n=57)	9 (15.8)	29 (50.9)
First aid (n=13)	1 (7.7)	5 (38.5)
Urology (n=15)	1 (6.7)	10 (66.7)

[a]Note that a given app could cover more than 1 condition area.

Table 3. Number and percentage of mobile health apps that reported seeking user input during development or testing by National Institute for Health and Care Excellence Evidence Standards Framework (ESF) tier.

ESF tier	Apps with user input during development, n (%)	Apps with user input during testing, n (%)
Tier 1 (n=19)	3 (15.8)	5 (26.3)
Tier 2 (n=1207)	352 (29.2)	637 (52.8)
Tier 3a (n=263)	105 (39.9)	176 (66.9)
Tier 3b (n=105)	28 (26.7)	68 (64.8)

Figure 1. Distribution of (A) user ratings and (B) download levels across all mobile health apps. Apps with a higher download level have a higher number of downloads, as indicated by the range of downloads shown in the app store (see the Methods section for details).



User Input Across Countries, Condition Areas, and ESF Tiers

Examining user input across countries revealed that mobile health apps originating from Denmark had the highest percentage of reported user inclusion during development (19/24, 79%), followed by apps developed in the United Kingdom (185/419, 44.2%), France (7/16, 44%), Turkey (6/14, 43%), and Canada (30/71, 42%; Table 1). The percentage of mobile health apps reporting seeking user input during testing was highest in Belgium (10/11, 91%), Sweden (29/34, 85%), and France (13/16, 81%; Table 1). Across all countries except Denmark, the percentage of apps that included users during development was lower than the percentage of apps seeking user input during testing.

When considering apps across different health conditions, it was observed that the areas of diabetes (38/79, 48%), cardiology (15/32, 47%), pain management (20/43, 47%), and oncology (25/54, 46%) contained the highest percentage of apps that reported including users during development (Table 2). The percentage of apps reporting seeking user input during testing was highest for the condition areas of neurodiversity (42/52, 81%), respiratory health (58/76, 76%), cardiology (23/32, 72%), and diabetes (56/79, 71%; Table 2). Across all condition areas, the percentage of apps that included users during development was lower than the percentage of apps seeking user input during testing.

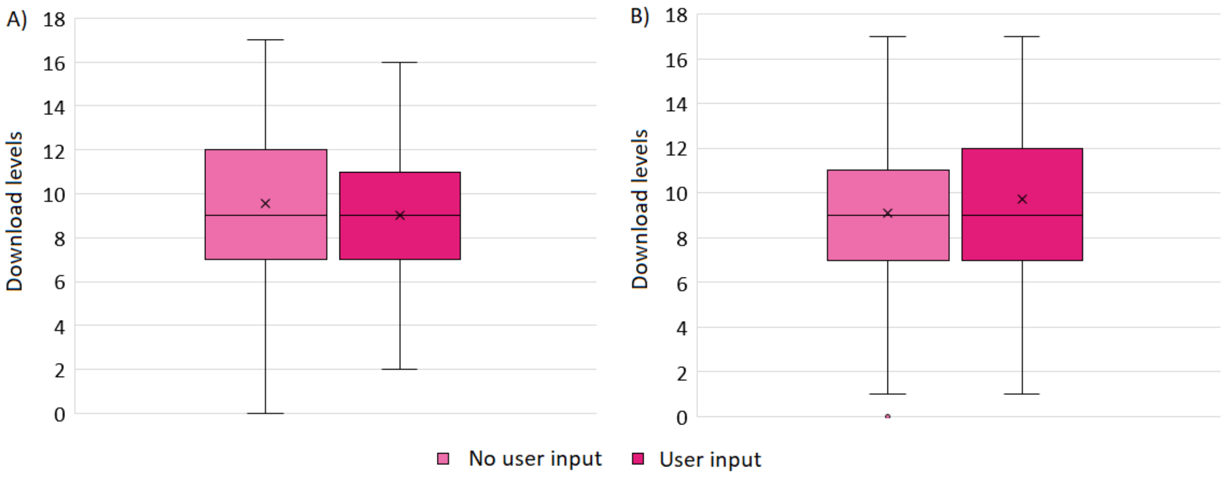
When examining mobile health apps across different ESF tiers, it was found that, with an increasing ESF tier (and, thus, with higher risk), an increasing percentage of apps reported seeking user input during development or testing. The exception to this pattern was tier 3b, in which a smaller percentage of apps reported including users during development compared with tiers 3a and 2. The percentage of apps reporting user input during testing was also slightly smaller for tier 3b than for tier 3a (Table 3).

User Input, User Ratings, and Download Levels

Mann-Whitney *U* tests indicated that the distribution of user ratings did not differ significantly between mobile health apps that did and did not report seeking user input during development (*P*=.45) or testing (*P*=.27).

By contrast, the distribution of download levels demonstrated a significant difference for mobile health apps that did and did not report including users during development (*P*=.02) or testing (*P*=.008). However, the direction of this effect differed between the 2 comparisons: apps that did not report including users *during development* showed a *larger* mean rank (and, thus, more downloads) than those that reported including users, whereas apps that did not report seeking user input *during testing* demonstrated a *smaller* mean rank (and, thus, fewer downloads) than those that did report seeking user input (Figure 2).

Figure 2. Download levels for mobile health apps that did and did not report seeking user input during (A) development or (B) testing. Apps with a higher download level have a higher number of downloads, as indicated by the range of downloads shown in the app store (see the Methods section for details).



User Input and App Compliance With Best Practices

Overview

Significant associations between compliance with best practices, as assessed through selected OBR questions, and reported user input were observed across all domains (clinical assurance, data privacy, risk management, and user experience).

The results of the Pearson chi-square analyses are described in detail in the following sections.

Clinical Assurance

Pearson chi-square analysis revealed a significant association between the answer to the question—"Does the app have appropriate evidence for its ESF tier?"—and the reported involvement of users during development (χ^2_1=24.8, *P*<.001) and testing (χ^2_1=218.2, *P*<.001). The odds ratio indicated that, if a mobile health app reported seeking user input during development or testing, the odds of it meeting its tier's evidence requirements were 1.75 (95% CI 1.40-2.19) or 6.02 (95% CI 4.68-7.75) times higher, respectively, than if it did not include users.

When only considering health apps for which an efficacy or effectiveness study had been conducted, no significant association was found between the significance of the study results (ie, the answer to the question "Does the p-value demonstrate significance [*P*<.05]?") and reported user input during development (χ^2_1=0.4, *P*=.55) or testing (χ^2_1=3.0, *P*=.09).

Data Privacy

In the data privacy domain, a significant association was observed between the answer to the question—"Is there a policy or statement that confirms the app's compliance with the GDPR?"—and reported user input during development (χ^2_1=22.1, *P*<.001) and testing (χ^2_1=67.0, *P*<.001). The odds ratio indicated that, if a mobile health app reported seeking user input during development or testing, the odds of it confirming compliance with the GDPR were 1.75 (95% CI 1.39-2.22) or 2.57 (95% CI 2.04-3.23) times higher, respectively, than if it did not involve users.

A significant association was also found between the answer to the question—"Does the developer provide users with details on all the purposes of processing user data?"—and the reported inclusion of users during development (χ^2_1=6.0, *P*=.01) and testing (χ^2_1=46.9, *P*<.001). The odds of explaining all data processing purposes were 1.84 (95% CI 1.12-3.01) or 4.29 (95% CI 2.74-6.72) times higher if a health app reported involving users during development or testing, respectively, than if it did not.

Risk Management

Regarding risk management, a significant association was revealed between the answer to the question—"Does the developer make clear risks associated with using the app?"—and the reported inclusion of users during testing (χ^2_1=14.6, *P*<.001) but not during development (χ^2_1=0.6, *P*=.45). If a health app reported seeking user input during testing, the odds of the developer clearly stating the risks associated with using the app were 2.42 (95% CI 1.52-3.86) times higher than if it did not include users.

Furthermore, a significant association was observed between the answer to the question—"Does the developer clearly identify who the app should or should not be used by?"—and reported user input during development (χ^2_1=38.7, *P*<.001) and testing (χ^2_1=17.9, *P*<.001). The odds ratio indicated that the odds of the developer clearly stating the target group for the health app were 2.30 (95% CI 1.76-3.00) or 1.61 (95% CI 1.29-2.00) times higher if the app reported including users during development and testing, respectively, than if it did not.

Moreover, a significant association was found between the answer to the question—"Is there a way for the user to confirm that the data input is accurate?"—and the reported inclusion of users during testing (χ^2_1=5.5, *P*=.02; note that this result does not survive multiple-comparison correction) but not during development (χ^2_1=0.4, *P*=.53). The odds of a health app incorporating data accuracy checks were 1.40 (95% CI 1.06-1.86) times higher if the app involved users during testing than if it did not.

User Experience and App Functionality

In the user experience and app functionality domain, a significant association was observed between the answer to the question—"Does the user have options to manage the settings for push or email notifications within the app for convenience and privacy?"—and reported user input during testing (χ^2_1=7.5, *P*=.01) but not development (χ^2_1=2.5, *P*=.12). The odds ratio indicated that, if an app reported including users during testing, the odds of it allowing users to manage notifications were 1.45 (95% CI 1.11-1.89) times higher than if it did not involve users.

A significant association was also found between the answer to the question—"Are any clinical or technical terms used explained clearly to the user?"—and the reported inclusion of users during development (χ^2_1=29.2, *P*<.001) and testing (χ^2_1=57.1, *P*<.001). The odds of clearly explaining technical terms were 2.21 (95% CI 1.65-2.96) and 2.51 (95% CI 1.97-3.20) times higher if a health app reported engaging users during development and testing, respectively, than if it did not.

Furthermore, a significant association was found between the answer to the question—"Is there any statement within the app about the developer's commitment to addressing problems reported to them (e.g., commitment to eradicate reported bugs and faults)?"—and the reported involvement of users during development (χ^2_1=37.1, *P*<.001) and testing (χ^2_1=60.5, *P*<.001). If a health app reported seeking user input during development or testing, the odds of the developer being committed to addressing reported issues were 2.19 (95% CI 1.70-2.82) or 3.02 (95% CI 2.27-4.02) times higher, respectively, than if it did not include users.

In addition, a significant association was revealed between the answer to the question—"Does the app provide gamification or goal setting features for the user?"—and the reported inclusion of users during testing (χ^2_1=74.5, *P*<.001) but not during development (χ^2_1=3.0, *P*=.08). The odds of a health app including gamification or goal setting were 2.53 (95% CI 2.04-3.13) times higher if the app involved users during testing than if it did not.

Moreover, a significant association was observed between the answer to the question—"Are there opportunities to link with other users (buddying, forums or group education)?"—and reported user input during development (χ^2_1=5.2, *P*=.02; note that this result does not survive multiple-comparison correction) or testing (χ^2_1=34.8, *P*<.001). If a health app reported including users during development or testing, the odds of the app allowing users to link to each other were 1.32 (95% CI 1.04-1.67) and 2.00 (95% CI 1.58-2.52) times higher, respectively, than if it did not involve users.

No significant association was found between the answer to the question—"Does the app allow the monitoring of key health information?"—and the reported inclusion of users during development (χ^2_1=0.0, *P*=.85) or testing (χ^2_1=2.5, *P*=.11).

Discussion

Principal Findings

This study aimed to examine the relationship between mobile health app characteristics and the inclusion of users during development and testing. For this purpose, a secondary analysis was conducted on an assessment data set of 1595 mobile health apps collected by ORCHA between January 2021 and January 2022. Descriptive statistics were used to explore the prevalence of user input for apps across different countries, condition areas, and ESF tiers. In addition, Mann-Whitney *U* tests and Pearson chi-square analyses were conducted to examine group differences and associations between health apps that did and did not report seeking user input and download numbers; user ratings; and assessment measures across the domains of clinical assurance, data privacy, risk management, user experience and app functionality.

Overall, user involvement was reported by 8.71% (139/1595) of the apps for only the development phase, by 33.67% (537/1595) of the apps for only the testing phase, and by 21.88% (349/1595) of the apps for both phases. The remaining 35.74% (570/1595) of the apps did not report including users during either phase. The highest percentage of mobile health apps with reported user input during *development* was observed in Denmark (19/24, 79%); in the condition areas of diabetes (38/79, 48%), cardiology (15/32, 47%), pain management (20/43, 47%), and oncology (25/54, 46%); and for high app risk (ESF tier 3a; 105/263, 39.9%). The highest percentage of health apps with reported user input during *testing* was observed in Belgium (10/11, 91%), Sweden (29/34, 85%), and France (13/16, 81%); in the condition areas of neurodiversity (42/52, 81%), respiratory health (58/76, 76%), cardiology (23/32, 72%), and diabetes (56/79, 71%); and for high app risk (ESF tier 3a; 176/263, 66.9%).

Moreover, health apps that reported seeking user input during testing demonstrated significantly more downloads than those that did not (*P*=.008), whereas the opposite was true for health apps that reported including users during development (*P*=.02). No significant group differences were observed in user ratings. Finally, reported user input was associated with improved compliance with best practices across all examined areas: clinical assurance (eg, meeting ESF evidence requirements), data privacy (eg, including a statement that confirms the app's compliance with the GDPR), risk management (eg, clearly stating who the app should or should not be used by), and user experience and app functionality (eg, allowing users to manage notification settings).

The interpretation of these findings in light of the previous literature is discussed in the following sections.

User Input Differs Across Countries, Condition Areas, and App Risks

The percentage of mobile health apps that reported user input differed across countries. Interestingly, the countries in which higher percentages of apps with reported user input were observed tended to be those with higher digital maturity levels in health care [23]. Digital maturity in this context describes the extent to which a country can derive value from technology in a health care setting and was evaluated across three main areas in the cited report [23]: (1) initiatives related to digital health, including policies and funding availability, as a foundation for the country's digital health care transformation; (2) infrastructure, including electronic health records, data standards, and interoperability, as a basis for interconnected systems and high-quality data; and (3) implementation efforts, including for telehealth, artificial intelligence, and internet-based studies, as a measure of the country's ability to make use of digitally collected data to improve population health management [23]. Notably, the country with the highest digital maturity score in health care across Europe, the Middle East, and Africa (Denmark [23]) also exhibited by far the highest percentage of reported user inclusion during development in our study. Similar relationships between high digital maturity and reported user input in our study were observed for England and Sweden and, to a lesser extent, France and Belgium [23]. Thus, it may be the case that a better digital health infrastructure and more established processes make it easier for app developers to involve users. For instance, countries with higher, compared with those with lower, digital maturity in health care may offer better access to funding or know-how to support user involvement activities or may allow developers to invest more resources into those activities because of more streamlined and, thus, less resource-intensive processes in other areas such as for establishing interoperability. If this is the case, user involvement activities may become more widespread in other countries as their digital maturity in health care improves.

The percentage of mobile health apps that reported seeking user input during development or testing also differed across condition areas, with the highest percentages of user input being observed in the areas of oncology, diabetes, and cardiology. This finding is in line with 2 recent reviews, which reported that most papers on patient and public involvement in health care innovation were published in the fields of oncology and diabetes (as well as mental health [11,12]). Notably, the condition areas in which the highest percentage of user inclusion was observed in this study were those with the largest number of mobile health apps on the market (apart from mental health [24]). Thus, facing high levels of competition may encourage developers to involve users in the development process. In addition, the uneven distribution of funding across condition areas may also play a role as the aforementioned clinical indications were among the top 5 most funded digital health areas in 2022 [25], which may increase both the number of apps on the market and the availability of resources for user involvement activities in those areas. This highlights the need for digital health funding in other condition areas, including to cover costs of user involvement activities as a crucial part of the app development and testing process.

The third dimension across which the percentage of mobile health apps with reported user input differed was the NICE ESF tier. Specifically, there was a tendency of an increasing ESF tier being associated with a greater percentage of apps that reported seeking user input during development or testing. This pattern suggests that developers are aware that user inclusion is especially important for health apps with high risk. Furthermore, the results may indicate that developers require more user input to understand how best to communicate with users and what their needs are for apps that provide tailored advice on diagnosis and treatment (tier 3) than for simpler health apps (tiers 1 and 2). Moreover, developers of riskier health apps may be more likely to seek funding from organizations such as the National Institute for Health and Care Research, which includes cocreation of the app with patients as a core requirement. However, it should be noted that, even in ESF tier 3, fewer than half (105/263, 39.9%) of the apps included users during development, which underscores the need for further education and support of developers regarding the importance and execution of user involvement during the early stages of the life cycle.

User Input During Testing Is Associated With More Downloads but Not With Higher User Ratings

When examining user ratings for mobile health apps that did and did not report seeking user input during development or testing, no significant differences were observed. This is somewhat surprising, especially in light of a recent study reporting that user experience was the most important determinant of positive health app user reviews [26]. However, the same study also observed that aspects such as payment problems and bugs or stability issues after updates were major factors that resulted in negative user reviews. Such issues are unlikely to arise during early user involvement (potentially working with free and low-fidelity app test versions), which may explain why no relationship between user ratings and user input during development was observed in this study. Moreover, the lack of relationship between user ratings and user input during testing could be partially due to user views only being taken into account during early testing and not at later stages as updates to the app are made [26]. Therefore, there may be a stronger relationship between user input and positive user ratings for earlier versions of the app, which may become weaker as updates are made without user contributions. We were unable to assess this suggestion using the available data, but if this was found to be the case, it would underline the importance of continued user involvement and consideration of user feedback (eg, from user reviews) at later life cycle stages.

Another surprising finding was that mobile health apps that did not report including users during development were downloaded more frequently than those that did. A potential explanation for this observation may be that there is a trade-off between resource investment for user involvement and marketing, with the latter being a stronger determinant of the number of downloads. In line with this suggestion, a recent review cited financial constraints as a barrier to patient and public involvement in digital health innovation [11]. In this context, it should also be noted that, although the number of initial downloads may not necessarily be related to user input, the *sustained* use, which ultimately determines health outcomes and the long-term success of the app, is likely to be linked more closely to user inclusion. Specifically, it has been argued that a lack of user involvement can give rise to barriers to sustained health app use, such as poor usability, failure to meet real-life user needs, and addition rather than elimination of effort in the health management process [27]. Moreover, it is worth noting that the hypothesized relationship between downloads and user input during *testing* was observed, with mobile health apps that did include users during testing demonstrating more downloads than those that did not. Notably, this study determined user input during testing based on publicly available information shown on app websites. Such information may be compelling to potential users, inducing them to download the app, which may explain the observed findings. This demonstrates the value to developers of conducting and showcasing user inclusion activities.

User Input Is Associated With Improved Compliance With Best Practices Regarding Clinical Assurance, Data Privacy, Risk Management, and User Experience

Examining the relationship between quality characteristics and user inclusion revealed a significant positive association between reported user input and alignment with ESF tier clinical evidence requirements. At first sight, it may be expected that this effect is due to user input resulting in more engaging apps, which, in turn, may lead to better adherence and, thus, higher clinical efficacy and effectiveness of the health apps. However, this hypothesis was not confirmed in this study as no significant association was observed between user input and reported app effectiveness or efficacy. Alternatively, the observed relationship may be due to a third-variable effect, with developers who are more aware of best practices for app development being more likely to include users *and* to conduct appropriate, high-quality research studies that meet ESF tier requirements. This could be a valuable area for future research.

Another quality indicator that was found to be associated with reported user input was data privacy best practice, as indicated by explicit compliance with the GDPR and transparency about data processing. Again, this may be a third-variable effect because of an association between awareness of best practices for both data privacy and user involvement. Alternatively, the observed association may be due to users stressing the importance of data privacy during user involvement activities. Consistent with this suggestion, a recent review of attitudes toward the use of health data found that, although public and patient participants generally support data sharing for research purposes, many raised concerns about confidentiality breaches and potential abuses of the data [28]. Along similar lines, a number of reviews and frameworks for user involvement in research emphasize the importance of transparency regarding data handling [29-31]. Therefore, it is possible that developers with more transparent data-handling practices find it easier to recruit users for involvement activities. This suggestion is in line with reports that concerns about data privacy are a barrier to patient and public involvement in digital health innovation [11].

Significant positive associations were also observed between reported user input and the presence of risk management practices, such as clearly indicating who the app should not be used by and what risks are associated with using the app. It is likely that the need to clearly inform users of potential risks may be highlighted during user involvement activities, which may provide insights into risks beyond the developers' initial expectations. This is likely the case as perceptions of risks and risk-benefit trade-offs can differ between developers and patients, who have firsthand experience of the condition being targeted by the app (which is why patient involvement in risk assessments is increasingly being considered an important factor for regulatory decisions regarding new health technologies [32-34]).

With regard to user experience, it was observed that reported user input was associated with a higher likelihood of the app allowing users to manage notifications, clinical terms being clearly explained, and the developer committing to addressing reported issues. Moreover, mobile health apps that involved users were also likely to include goal setting, gamification, or user connection (eg, forum) features. This finding is in line with previous research showing that aspects such as notification management and goal setting are important to users when they review health apps and decide whether to continue using them [26,35]. The observed association between the presence of these features and user input suggests that the decision to include these features may have arisen from user involvement activities, suggesting that user inclusion leads to improved adaption of features to user expectations.

Related Work

There are numerous previous studies that have evaluated health apps according to different quality dimensions. These reviews have revealed that many health apps do not follow best practices in areas such as data privacy and security [36-39], clinical safety [40], or efficacy and effectiveness evidence [37,38,41].

In addition, studies more specifically focused on user engagement have shown that aspects such as visual design and engaging presentation of app content are correlated with key mobile health app metrics, namely, use time and 30-day retention rates [42]. As design- and engagement-related aspects can likely be improved by seeking user input during development and testing, this finding underscores the importance of user involvement, which is in line with this study.

Relatedly, 2 recent papers have highlighted the usefulness of considering (automatically analyzed) app store user reviews in identifying and addressing the challenges that users experience with health apps. Key issues reported by users included compatibility and log-in difficulties, stability and accessibility problems, and privacy-related concerns, which contributed to low star ratings on the app store [26,43]. In line with these observations, a pilot study examining human-centric issues with health apps across different countries found that accessibility, usability, and data privacy issues were regarded as essential points to consider by different stakeholders [44]. These findings emphasize the importance of considering user views both during initial development and after app updates to ensure a positive user experience.

Summary of Implications for Practice and Recommendations for Future Work

Our research found that, even for high-risk apps, fewer than half of developers reported involving users during development. As mentioned previously, this highlights the need for further education and support for developers regarding the importance and implementation of user involvement during the early stages of the product life cycle. Building on existing work that revealed barriers to meaningful user involvement [11], a support program for developers could be devised. The effectiveness of this program could be evaluated through future research comparing user involvement practices between health app development companies (matched on various characteristics) that did and did not take part in the program.

Moreover, as discussed previously, there may be a relationship between funding availability and the willingness or ability of developers to involve users during development and testing. Future research (in collaboration with app developers) could examine this relationship more directly by assessing whether there is a significant association between investment in individual health apps and whether they seek user input or by determining what percentage of developers start involving users at which funding stage. If this research indicates that many developers are only willing to budget for user involvement activities once they have received a large amount of funding, this would suggest that it may be useful for investors to earmark some (earlier) funding for user involvement given the importance of considering user views to ensure the success of an app [26].

Furthermore, our findings indicate that seeking user input is associated with improved compliance with best practices across various domains. Therefore, *where no other guidance is available*, stakeholders such as patients and health care professionals looking for health apps may benefit from prioritizing apps that seek user input during development and testing as this may be a proxy for broader app quality.

Strengths and Limitations

The limitations of this study should be noted. First, only countries and condition areas with >10 mobile health apps were included in the relevant data summaries as samples of <10 mobile health apps are likely not representative of the larger app "population" in a given country or condition area. Therefore, our findings may not be generalizable to excluded countries and conditions with not enough observations. Relatedly, results from countries and condition areas for which a relatively small number of apps were included in our sample should be interpreted with caution.

Secondly, user involvement and input was defined somewhat more broadly in this study than is common in the previous literature (see the *Determination of User Input* section). This was because the study relied on a data set that had been collected for practical assessment purposes rather than specifically for research. Although this limitation should be kept in mind when interpreting the study findings, it is encouraging to see that there is convergence between the observations of this study and those of previous research (as described in the previous sections) that used a stricter definition of user or patient and public involvement.

Third, the determination of the presence of user input in this study relied on information from within the app and associated websites. It is possible that, in some cases, user involvement did take place without this being publicly indicated and, thus, without it being taken into account in this study. However, given that user involvement is generally regarded as a strength, it seems unlikely that (many) developers would forego the opportunity to publicly state that they took users' input into account.

Finally, it should be noted that no information was available regarding the quality and meaningfulness of the user involvement activities or who the involved users were (eg, whether they were representative of the actual users and whether they included minority groups or individuals from different socioeconomic backgrounds and with different levels of digital literacy—this will be examined in future versions of the ORCHA assessment). It was also not recorded what user involvement methods or levels of involvement were applied (eg, whether co-design or consultation was used). It is possible that findings may differ depending on the "depths" and quality of user involvement, which would be an interesting area of examination for future research.

Apart from these limitations, a major strength of this study is the use of a unique data set with detailed assessment information for >1500 mobile health apps across different countries and condition areas. This allowed the study to generate findings beyond those that could be gathered from the peer-reviewed literature, which is important given that likely not all user involvement activities are published in academic articles. Moreover, to our knowledge, this is the first study to examine the relationship between user input and compliance with best practices across the domains of clinical assurance, data privacy, risk management, and user experience.

Conclusions

In summary, this study found that the prevalence of user input during mobile health app development or testing differed across countries, condition areas, and ESF tiers. The countries and condition areas in which the highest percentage of health apps with user input were observed tended to be those with higher digital maturity in health care and more funding availability, respectively. This suggests that there may be a trade-off between developers' willingness or ability to involve users during development or testing and the need to meet challenges arising from infrastructure limitations and financial constraints. Moreover, the finding of a positive association between user input and compliance with best practices indicates that, *where no other guidance is available*, users may benefit from prioritizing mobile health apps that involved users during development and testing as this may be a proxy for broader app quality.

Acknowledgments

The data set and funding for this study were provided by the Organisation for the Review of Care and Health Apps.

Authors' Contributions

A-LF and SL conceptualized the study. A-LF conducted the data analysis and wrote the first draft of the manuscript. RB, SH, RK, TA, and SL reviewed and edited the manuscript draft. All authors have reviewed and approved the final manuscript for publication.

Conflicts of Interest

A-LF, RB, SH, TA, and SL are employed by the Organisation for the Review of Care and Health Apps. RK reports no conflicts of interest.

Multimedia Appendix 1
Conversion of download number ranges to download levels.
[DOCX File , 14 KB - mhealth_v11i1e46937_app1.docx]

References

1. Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 01;7(1):12 [FREE Full text] [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
2. Patient and public involvement policy. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/about/nice-communities/nice-and-the-public/public-involvement/public-involvement-programme/patient-public-involvement-policy#:~:text=NICE?s%20approach%20to%20patient%20and,and%20support%20their%20implementation%2C%20and [accessed 2023-02-08]
3. Accelerated access collaborative patient and public involvement strategy 2021-26. National Health Service. URL: https://www.england.nhs.uk/aac/wp-content/uploads/sites/50/2022/06/Accelerated-Access-Collaborative-patient-and-public-involvement-strategy.pdf [accessed 2023-02-08]
4. Patient and public participation policy. National Health Service England. 2017 Apr 6. URL: https://www.england.nhs.uk/wp-content/uploads/2017/04/ppp-policy.pdf [accessed 2023-02-08]
5. FDA patient-focused drug development guidance series for enhancing the incorporation of the patient’s voice in medical product development and regulatory decision making. U.S. Food & Drug Administration. URL: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical [accessed 2023-02-08]
6. Patrício L, Sangiorgi D, Mahr D, Čaić M, Kalantari S, Sundar S. Leveraging service design for healthcare transformation: toward people-centered, integrated, and technology-enabled healthcare systems. J Serv Manag 2020 Jun 08;31(5):889-909. [doi: 10.1108/josm-11-2019-0332]
7. Health inequalities annual report 2023. Northern Ireland Executive. 2023 Mar 29. URL: https://www.northernireland.gov.uk/news/health-inequalities-annual-report-2023#:~:text=The%20Health%20Inequalities%20Annual%20Report%20is%20one%20of%20a%20series,Social%20Care%20(HSC)%20Trust%20and [accessed 2023-10-23]
8. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
9. Pizzo E, Doyle C, Matthews R, Barlow J. Patient and public involvement: how much do we spend and what are the benefits? Health Expect 2015 Dec 12;18(6):1918-1926 [FREE Full text] [doi: 10.1111/hex.12204] [Medline: 24813243]
10. Locock L, Boaz A. Drawing straight lines along blurred boundaries: qualitative research, patient and public involvement in medical research, co-production and co-design. Evid Policy 2019 Aug 01;15(3):409-421. [doi: 10.1332/174426419X15552999451313]
11. Baines R, Bradwell H, Edwards K, Stevens S, Prime S, Tredinnick-Rowe J, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. Health Expect 2022 Aug 08;25(4):1232-1245 [FREE Full text] [doi: 10.1111/hex.13506] [Medline: 35526274]
12. Cluley V, Ziemann A, Feeley C, Olander EK, Shamah S, Stavropoulou C. Mapping the role of patient and public involvement during the different stages of healthcare innovation: a scoping review. Health Expect 2022 Jun 17;25(3):840-855 [FREE Full text] [doi: 10.1111/hex.13437] [Medline: 35174585]
13. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]

14. Jones E, Frith L, Gabbay M, Tahir N, Hossain M, Goodall M, et al. Remote working in public involvement: findings from a mixed methods study. Res Involv Engagem 2022 Nov 04;8(1):58 [FREE Full text] [doi: 10.1186/s40900-022-00396-0] [Medline: 36333757]
15. Lampa E, Sonnentheil B, Tökés A, Warner G. What has the COVID-19 pandemic taught us about conducting patient and public involvement remotely? insights from a series of digital meeting observations. Res Involv Engagem 2021 Oct 11;7(1):73 [FREE Full text] [doi: 10.1186/s40900-021-00315-9] [Medline: 34635179]
16. Martinez MG, Daleske CB, León Á, Rodelgo SM, Espinosa RO, López PR, et al. Empowering patients to co-design Covid-19 responses: the role of online health communities. R D Manage 2022 Mar;52(2):391-406. [doi: 10.1111/radm.12516]
17. Digital health in the UK - national attitudes and behaviour research. The Organisation for the Review of Care and Health Apps. 2021. URL: https://orchahealth.com/wp-content/uploads/2021/07/2107_ICS_Research_Report_2021_National_final.pdf [accessed 2023-02-24]
18. Evidence standards framework (ESF) for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2023-10-23]
19. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2022. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2023-02-24]
20. Field AP. Non-parametric tests. In: Discovering Statistics Using SPSS, 3rd Edition. Thousand Oaks, CA: SAGE Publications Ltd; 2009.
21. Field AP. Categorical data. In: Discovering Statistics Using SPSS, 3rd Edition. Thousand Oaks, CA: SAGE Publications; 2009.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
23. Arias A, Rickwood S. How can countries achieve digital maturity in healthcare? World Economic Forum. 2022 Aug 11. URL: https://www.weforum.org/agenda/2022/08/countries-achieve-digital-maturity-healthcare/ [accessed 2023-02-08]
24. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA. 2021 Jul 22. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2023-02-08]
25. Bryant K, Knowles M, Krasniansky A. 2022 year-end digital health funding: lessons at the end of a funding cycle. Rock Health. 2023 Jan 9. URL: https://rockhealth.com/insights/2022-year-end-digital-health-funding-lessons-at-the-end-of-a-funding-cycle/ [accessed 2023-02-08]
26. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196 [FREE Full text] [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
27. Wali S, Keshavjee K, Demers C. Moving towards sustainable electronic health applications. In: Heston TF, editor. eHealth - Making Health Care Smarter. London, UK: IntechOpen; 2018.
28. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
29. Greenhalgh T, Hinton L, Finlay T, Macfarlane A, Fahy N, Clyde B, et al. Frameworks for supporting patient and public involvement in research: systematic review and co-design pilot. Health Expect 2019 Aug;22(4):785-801 [FREE Full text] [doi: 10.1111/hex.12888] [Medline: 31012259]
30. Kent RC. From the UK welfare state to digital self-care: historical context of tracking public health and quantifying bodies. In: The Quantification of Bodies in Health Multidisciplinary Perspectives. Bingley, UK: Emerald Publishing Limited; 2021.
31. Kent R. Digital food tracking: combining traditional and digital ethnographic methods to identify the influence of social media sharing of health and foods upon users' everyday lives. In: Research Methods in Digital Food Studies. Milton Park, UK: Routledge; 2021.
32. Chachoua L, Dabbous M, François C, Dussart C, Aballéa S, Toumi M. Use of patient preference information in benefit-risk assessment, health technology assessment, and pricing and reimbursement decisions: a systematic literature review of attempts and initiatives. Front Med (Lausanne) 2020 Oct 26;7:543046 [FREE Full text] [doi: 10.3389/fmed.2020.543046] [Medline: 33195294]
33. Ho M, Saha A, McCleary KK, Levitan B, Christopher S, Zandlo K, et al. A framework for incorporating patient preferences regarding benefits and risks into regulatory assessment of medical technologies. Value Health 2016;19(6):746-750 [FREE Full text] [doi: 10.1016/j.jval.2016.02.019] [Medline: 27712701]
34. van Til JA, Ijzerman MJ. Why should regulators consider using patient preferences in benefit-risk assessment? Pharmacoeconomics 2014 Jan 28;32(1):1-4. [doi: 10.1007/s40273-013-0118-6] [Medline: 24288209]
35. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
36. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 06;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]

37. Melcher J, Torous J. Smartphone apps for college mental health: a concern for privacy and quality of current offerings. Psychiatr Serv 2020 Nov 01;71(11):1114-1119. [doi: 10.1176/appi.ps.202000098] [Medline: 32664822]
38. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' a systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020 Jan 09;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
39. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
40. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
41. Simon L, Reimann J, Steubl LS, Stach M, Spiegelhalder K, Sander LB, et al. Help for insomnia from the app store? a standardized rating of mobile health applications claiming to target insomnia. J Sleep Res 2023 Feb 27;32(1):e13642. [doi: 10.1111/jsr.13642] [Medline: 35624078]
42. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
43. Haggag O. Better identifying and addressing diverse issues in mHealth and emerging apps using user reviews. In: Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering. 2022 Presented at: EASE '22; June 13-15, 2022; Gothenburg, Sweden. [doi: 10.1145/3530019.3535346]
44. Shamsujjoha M, Grundy J, Li L, Khalajzadeh H, Lu Q. Human-centric issues in eHealth app development and usage: a preliminary assessment. In: Proceedings of the IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). 2021 Presented at: IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER); March 09-12, 2021; Honolulu, HI. [doi: 10.1109/saner50967.2021.00055]

Abbreviations

ESF: Evidence Standards Framework
GDPR: General Data Protection Regulation
NICE: National Institute for Health and Care Excellence
OBR: Organisation for the Review of Care and Health Apps Baseline Review
ORCHA: Organisation for the Review of Care and Health Apps

Edited by L Buis; submitted 03.03.23; peer-reviewed by J Grundy, B Merid; comments to author 10.05.23; revised version received 15.06.23; accepted 11.07.23; published 22.11.23.

Please cite as:
Frey AL, Baines R, Hunt S, Kent R, Andrews T, Leigh S
Association Between the Characteristics of mHealth Apps and User Input During Development and Testing: Secondary Analysis of App Assessment Data
JMIR Mhealth Uhealth 2023;11:e46937
URL: https://mhealth.jmir.org/2023/1/e46937
doi:10.2196/46937
PMID:37991822

Original Paper

Quality of Digital Health Interventions Across Different Health Care Domains: Secondary Data Analysis Study

Maciej Hyzy[1], BSc; Raymond Bond[1], PhD; Maurice Mulvenna[1], PhD; Lu Bai[2], PhD; Alan Dix[3], DPhil; Robert Daly[4], MSc; Anna-Lena Frey[4], PhD; Simon Leigh[4], PhD

[1]School of Computing, Ulster University, Belfast, United Kingdom

[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, United Kingdom

[3]Swansea University, Swansea, United Kingdom

[4]Organisation for the Review of Care and Health Apps (ORCHA), Daresbury, United Kingdom

Corresponding Author:
Maciej Hyzy, BSc
School of Computing
Ulster University
Ulster University, York St
Belfast, BT15 1ED
United Kingdom
Phone: 44 7526852505
Email: maciejmarekzych@gmail.com

Abstract

Background: There are more than 350,000 digital health interventions (DHIs) in the app stores. To ensure that they are effective and safe to use, they should be assessed for compliance with best practice standards.

Objective: The objective of this paper was to examine and compare the compliance of DHIs with best practice standards and adherence to user experience (UX), professional and clinical assurance (PCA), and data privacy (DP).

Methods: We collected assessment data from 1574 DHIs using the Organisation for the Review of Care and Health Apps Baseline Review (OBR) assessment tool. As part of the assessment, each DHI received a score out of 100 for each of the abovementioned areas (ie, UX, PCA, and DP). These 3 OBR scores are combined to make up the overall ORCHA score (a proxy for quality). Inferential statistics, probability distributions, Kruskal-Wallis, Wilcoxon rank sum test, Cliff delta, and Dunn tests were used to conduct the data analysis.

Results: We found that 57.3% (902/1574) of the DHIs had an Organisation for the Review of Care and Health Apps (ORCHA) score below the threshold of 65. The overall median OBR score (ORCHA score) for all DHIs was 61.5 (IQR 51.0-73.0) out of 100. A total of 46.2% (12/26) of DHI's health care domains had a median equal to or above the ORCHA threshold score of 65. For the 3 assessment areas (UX, DP, and PCA), DHIs scored the highest for the UX assessment 75.2 (IQR 70.0-79.6), followed by DP 65.1 (IQR 55.0-73.4) and PCA 49.6 (IQR 31.9-76.1). UX scores had the least variance (SD 13.9), while PCA scores had the most (SD 24.8). Respiratory and urology DHIs were consistently highly ranked in the National Institute for Health and Care Excellence Evidence Standards Framework tiers B and C based on their ORCHA score.

Conclusions: There is a high level of variability in the ORCHA scores of DHIs across different health care domains. This suggests that there is an urgent need to improve compliance with best practices in some health care areas. Possible explanations for the observed differences might include varied market maturity and commercial interests within the different health care domains. More investment to support the development of higher-quality DHIs in areas such as ophthalmology, allergy, women's health, sexual health, and dental care may be needed.

(JMIR Mhealth Uhealth 2023;11:e47043) doi:10.2196/47043

KEYWORDS

digital health interventions scoring; digital health interventions; digital health; mHealth assessment; mobile health; ORCHA assessment; Organisation for the Review of Care and Health Apps; quality assessment; quantifying DHIs

Introduction

According to a report from 2021 [1], there were more than 350,000 digital health interventions (DHIs) available in the app stores. And in 2020, more than 91,000 DHIs had been added to app stores, which amounts to 251 DHIs per day (on average). Moreover, searches for DHIs within app stores have also increased [2]. A potential catalyst for this could have been the COVID-19 pandemic and restricted access to incumbent services. Nevertheless, these findings clearly indicate that the public has a great interest in the use of DHIs.

However, some of these DHIs may contain harmful content. For example, a study from 2016 [3] conducted a systematic assessment of suicide prevention and deliberate self-harm mobile apps. The study found that some of the apps encouraged risky behaviors, such as the uptake of drugs. Similarly, reviews across different health care domains have demonstrated that many DHIs raise safety, security, or data privacy (DP) concerns [4-8], include incomplete or misleading medical information [9,10], or have not been supported by sufficient scientific evidence [5,6,11]. This indicates that the assessment of DHIs for adherence to best practice standards is critical to ensuring user safety and DHI's effectiveness, as well as allowing health care professionals to confidently recommend DHIs to clients or patients.

Previous systematic reviews have shown that there are numerous existing assessments for DHI evaluation by experts and users that encompass a large number of heterogenous assessment criteria [12-17]. However, many assessment frameworks demonstrate shortcomings, such as being limited to DHIs in a particular health care domain [12], not including important assessment areas such as DP [12], or being focused on health care professionals without providing meaningful insights to end users [18]. Assessment that addresses these issues, including disease-independent criteria across key areas and making assessment results easily accessible to end users, is needed.

The Organisation for the Review of Care and Health Apps (ORCHA) [19] is a UK-based digital health compliance company that specializes in the assessment of DHI quality in terms of compliance with best practice standards. The Organisation for the Review of Care and Health Apps Baseline Review (OBR) [20] provided by ORCHA is a proxy for DHI's compliance with best practice standards. ORCHA is currently working with 70% of National Health Service (NHS) organizations within England and provides DHI libraries, hosted by various health care organizations, that contain information about DHIs that have been assessed with the OBR. Specifically, the OBR results provide information (including an assessment score between 0 and 100) regarding DHIs' compliance with best practices in the domains of professional and clinical assurance (PCA), DP, and user experience (UX), allowing end users and clinical professionals to make informed decisions on whether to use or recommend these DHIs.

Notably, the OBR has been applied to thousands of DHIs, which provides a valuable data set for the investigation of best practice compliance across different types of DHIs. For instance, compliance may vary between DHIs in different health care domains and with different levels of risk (eg, as per the National Institute for Health and Care Excellence [NICE] Evidence Standard Framework [ESF] tier classification [21]). Gaining insights into such variations is important to determine what factors may drive high or low compliance with best practices and which health care domains require more effort and investment to improve the quality of DHIs. Moreover, an understanding of how compliance varies among different types of DHIs can serve as a future reference point for determining how particular DHIs compare to other similar DHIs.

This study aimed to explore these questions using a data set comprising OBR assessment results for 1574 DHIs. In this study, we explore OBR scores regarding 3 NICE tiers and compare the quality of DHIs across 26 different health care domains. We do this by establishing quantiles for each tier and for each health care domain. We want to determine if OBR scores are different across DHIs in different health care domains. This will allow us to identify health care domains that may require more investment to support their development. We hypothesize that the quality of DHIs is different across several health care domains.

Methods

The Data Set and Assessment

For this study, ORCHA provided a data set comprising raw data from 1574 DHIs, which were assessed using the OBR version 6 tool [20]. The OBR version 6 is the latest version of the "ORCHA assessment tool," which consists of almost 300 objective (mostly dichotomous "yes" or "no") assessment questions in 3 areas: PCA, DP, and UX. Each of the areas is scored individually on a scale from 0 to 100 and combined into an overall ORCHA score.

NICE tiers classify DHIs based on their functionality, risk, and regulatory status. Tier A indicates that the DHIs provide health and social care services with no measurable user outcome. Tier B denotes that the DHIs can provide 2-way communication between users and health care professionals and provide health care information or a health diary. Tier C indicates that DHIs provide preventative behavioral change aimed at health issues; they may allow users to self-manage a specific condition, indicate that DHIs provide or guide treatment for a condition, record and transmit data about this condition to a professional, caregiver, or third party without a user's input, contain a calculator that impacts treatment, provide diagnostics for a specific condition, or guide a diagnosis [21]. Since the NICE tiers are dependent upon functionality, risk, and regulatory status, it would be inappropriate to, for example, use the OBR score on a tier C DHI using the OBR tier B scoring.

An ORCHA threshold score of 65 is an NHS-accepted cutoff point that indicates compliance with best practice standards for DHIs, meaning that the DHI may be used or recommended by NHS staff. The score of 65 was established with NHS partners in 2020 and has since remained there. It represents the point at which (in the majority), excess risks are avoided; that is, you cannot possibly score above 65 while having no privacy policy, having no relevant evidence, or being a medical device that is not certified.

An ORCHA score of 65 is also an initial score for all the DHIs being assessed in all assessment areas (UX, PCA, and DP). Meaning that the initial score at the beginning of the assessment is 65 for each assessment area and overall ORCHA score. Then, based on answers to assessment questions, this score is altered through value and risk points (value points increase the score and risk points reduce the score) and assigned to a DHI. This process changes the initial score of 65 for each assessment area and is then combined to give an overall ORCHA score. For example, for apps that store personal or sensitive information, value points are assigned to such an app if they make their privacy policy immediately available when the user first uses the DHI. And risk points are assigned if a privacy policy is not clearly available when using the DHI. The amount of value and risk points assigned per question vary based on the NICE ESF tier that has been assigned to a DHI. If no value or risk points were assigned during the assessment, then the ORCHA score remains 65 [20]. Furthermore, to receive full points for appropriate evidence for its ESF tier, a tier B DHI (depending on its exact functionality) may only require a user benefits statement (eg, based on pilot results) and validation of the provided information by experts or references, while a tier C DHI will likely require a full-scale observational study or randomized controlled trial to meet the same evidence threshold. These differences in evidence requirements were introduced by the NICE ESF and adopted with slight amendments by the ORCHA assessment to ensure that standards are realistic and achievable for DHI companies without placing an undue burden on developers of low-risk DHIs, while at the same time setting expectations sufficiently high (especially for high-risk DHIs) to ensure safety and effectiveness and to provide users and health care providers with confidence in the DHIs. Some questions in the ORCHA assessment tool do not assign value or risk points but are there to provide information or context; for example, the question “When was the last Care Quality Commission [22] inspection completed?” does not assign value or risk points.

Each assessment of the 1574 apps has been carried out by at least 2 trained ORCHA reviewers as part of “business as usual” for ORCHA, where in the case of a dispute, a third ORCHA reviewer would resolve it during a discussion. All ORCHA reviewers have undergone the same training to use the OBR version 6 assessment tool.

It takes around 6 months for an ORCHA reviewer to be trained on how to use the OBR and considered ready to carry out live reviews using the tool. The training involves teaching the new reviewer about each area (UX, PCA, and DP) of the OBR. Training is carried out either in person or through web-based meetings.

The data set used included DHI assessments that were published between January 18, 2021, and January 6, 2022. All DHIs were assigned to 26 different health care domains and to 1 of the 3 NICE tiers, established by the NICE ESF [21].

Statistical Analysis

We carried out secondary data analyses of an ORCHA data set, which comprised the assessment of 1574 DHIs. The data analysis was carried out using R Studio (The R Foundation) and the R programming language (R Core Team). Descriptive statistics, including the minimum score, first quantile, median, mean (SD), third quantile, maximum score, and SE of the mean, were calculated for each of the OBR scores (ORCHA, PCA, DP, and UX).

Box plots were generated to study each score per NICE tier. DHIs were also grouped and analyzed across the different health care domains, with the sample size (number of DHIs) for each health care domain presented. Each OBR score (ORCHA, PCA, DP, and UX) per health care domain has been presented in quantiles from 0% to 100% in increments of 25%. Quantiles have been used so that an easy comparison could be made between different scores, NICE ESF tiers, and health care domains. Normality testing (the Shapiro-Wilk test [23]) was used to determine which hypothesis test was appropriate. The Kruskal-Wallis rank sum test [24] was used to compare the scores between the different NICE tiers, with a $P<.05$ considered statistically significant. The Kruskal-Wallis rank sum test was used to compare the scores across the health care domains with post hoc analysis using the Dunn test [25] and Holm’s [26] method for *P* value adjustment for multiple pairwise comparisons. A 2-sided unpaired Wilcoxon rank sum test has been used to determine if DHIs’ with International Organization for Standardization (ISO) 27001 certification [27] are statistically different from those without, regarding DP scores. The Wilcoxon rank sum test was also used to determine if DHIs classified as medical devices [28] are statistically significantly different than those that are not medical devices, regarding PCA scores. After the Wilcoxon rank sum test, Cliff delta has been used to indicate the magnitude of the difference between 2 compared samples of DHIs with a 95% CI. Cliff delta magnitude has been assessed using the thresholds $|d|<.147$ “negligible,” $|d|<.33$ “small,” $|d|<.474$ “medium,” otherwise “large” [29]. The above analyses have been conducted for all DHIs, separated by NICE tiers (n=number of DHIs), tier B (n=1155), and tier C (n=408). Tier A (n=11) was excluded due to sample size.

Ethical Considerations

This secondary data analysis study gained ethical approval (project number: CEBE_RE-22-002) by Ulster University (ethics filter committee, Faculty of Computing, Engineering, and the Built Environment). The process undertaken by ORCHA ensures that DHIs’ developers are aware of their score and are given time to contest the findings of the assessment, which may be amended if developers provide additional relevant information. All reviews, unless explicitly asked to be removed by the developer, are covered as suitable for research in ORCHA’s privacy policy [30].

Results

Overview

Table 1 presents a summary of the OBR scores for all DHIs. A Kruskal-Wallis test revealed that the distributions of the UX, PCA, and DP scores were statistically significantly different from each other (*P*<.001). Figure 1A shows that UX scores have the least variance of the 3 assessment areas, whereas PCA scores have the greatest variance. Table 1 shows that the SD for UX scores is 8.20, whereas the SD for PCA scores is 24.8, which is approximately 3 times greater than the SD of the UX scores. The UX scores are also typically higher than the other scores. A total of 57.3% (902/1574) of DHIs in the data set have an ORCHA score below the accepted ORCHA threshold of 65. Multimedia Appendix 1 contains the number of DHIs with varied ORCHA thresholds. Multimedia Appendix 2 contains the steps involved in selecting 1574 DHIs.

Table 1. Summary of scores for the 1574 digital health interventions (DHIs).

Score	0%	25%	50%	75%	100%	Mean (SD)	SEM[a]
ORCHA[b]	18.0	51.0	61.5	73.0	96.0	61.6 (13.9)	.350
UX[c]	27.4	70.0	75.2	79.6	94.2	74.3 (8.20)	.207
PCA[d]	7.14	31.9	49.6	76.1	98.5	52.2 (24.8)	.626
DP[e]	4.28	55.0	65.1	73.4	99.3	63.2 (14.9)	.375

[a]SEM: standard error of the mean.

[b]ORCHA: Organisation for the Review of Care and Health Apps.

[c]UX: user experience.

[d]PCA: professional and clinical assurance.

[e]DP: digital privacy.

Figure 1. (A) Overlaid density plot of the Organisation for the Review of Care and Health Apps (ORCHA) score, professional and clinical assurance (PCA) score, data privacy (DP) score, and user experience (UX) score for 1574 digital health interventions (DHIs). (B) Box plots of scores for 1574 DHIs. The Kruskal-Wallis test had *P*<.001, and the posthoc Dunn test on score distributions for UX, PCA, and DP, all had *P*<.001.

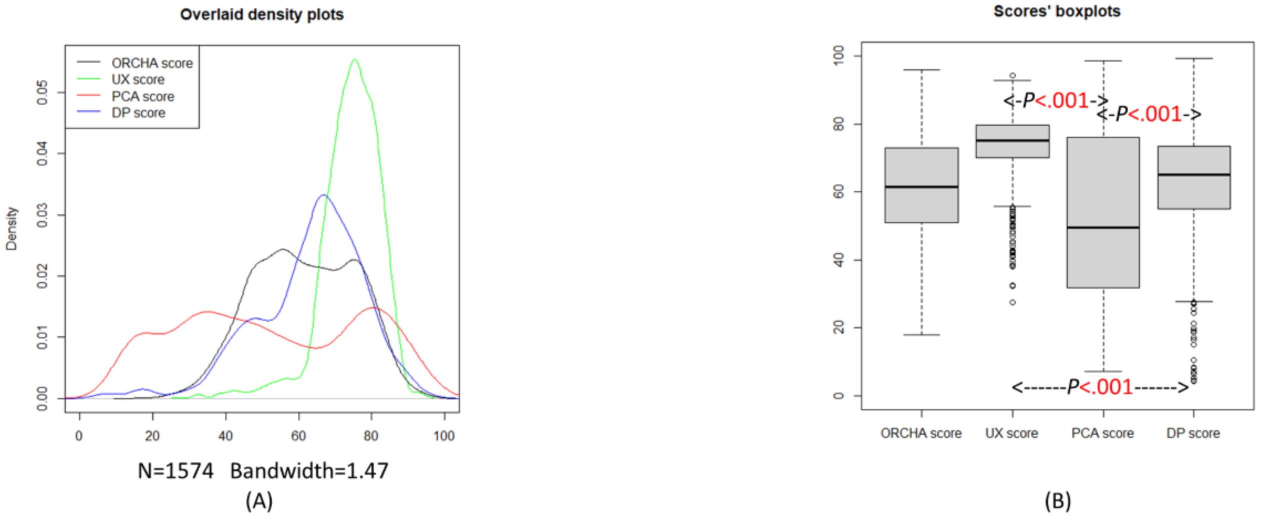
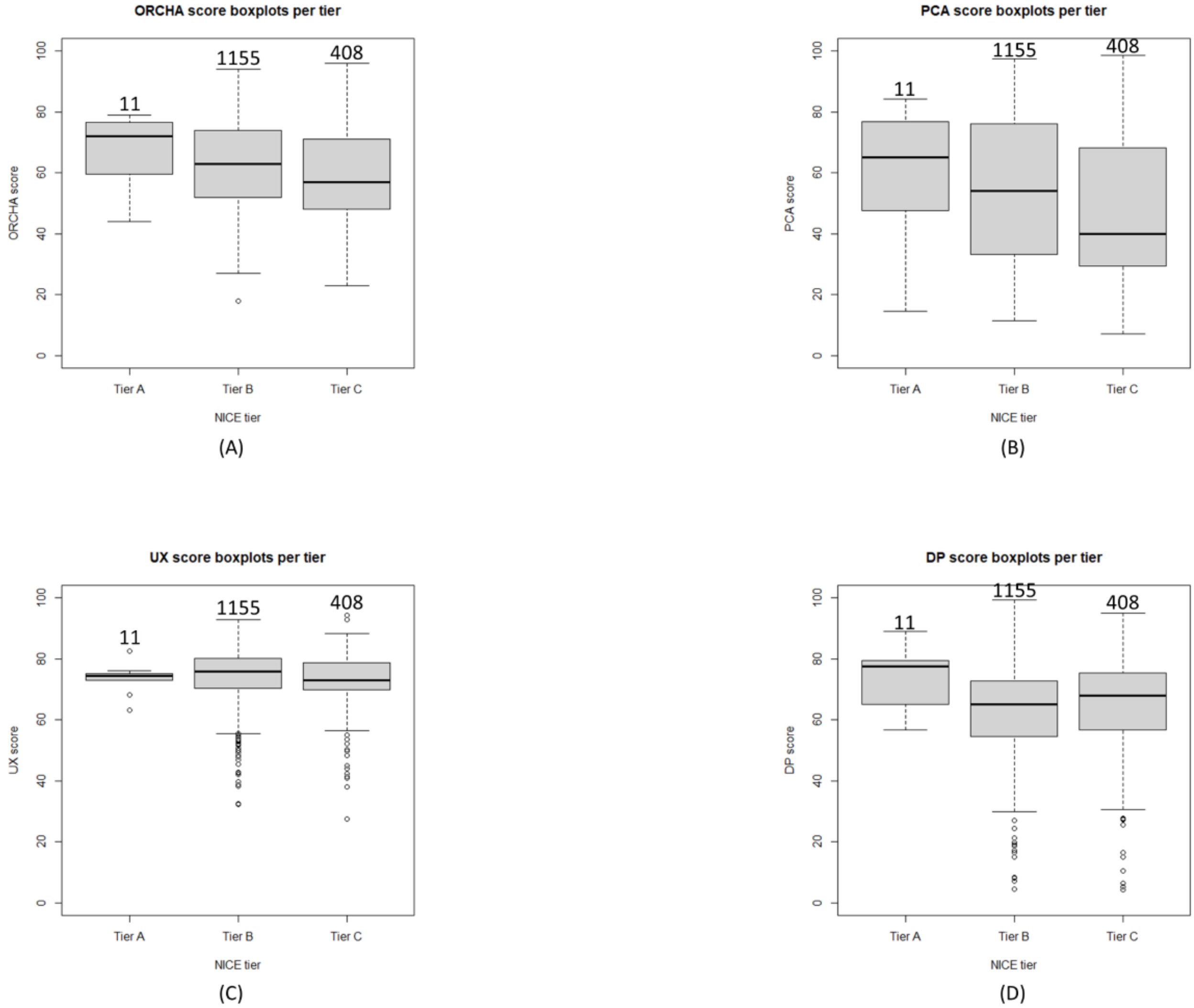


Table 1 shows the quantiles for each of the scores from 0% to 100% in increments of 25%. The ORCHA score for the 50% (median) quantile is 61.5, which is below ORCHA's threshold score of 65, meaning that most of the DHIs in the data set fail to adhere to the NHS cutoff for compliance with best practice standards. Median (IQR) for the OBR scores are ORCHA (61.5, IQR 51.0-73.0), UX (75.2, IQR 70.0-79.6), PCA (49.6, IQR 31.9-76.1), and DP (65.1, IQR 55.0-73.4).

Scores per NICE Tier

DHIs were distributed as follows across the different NICE tiers: tier A (11/1574, 0.699%), tier B (1155/1574, 73.4%), and tier C (408/1574, 25.9%). Figure 2 depicts box plots for the OBR scores within each tier. Further information is provided in Multimedia Appendix 3, which depicts quantiles for each score and NICE tier permutations from 0% to 100% in increments of 25%.

Figure 2. The number above the box plots is the sample size. (A) Organisation for the Review of Care and Health Apps (ORCHA) score box plots per National Institute for Health and Care Excellence (NICE) tier. (B) Professional and clinical assurance (PCA) score box plots per NICE tier. (C) User experience (UX) score box plots per NICE tier. (D) Data privacy (DP) score box plots per NICE tier.



Scores by Health Care Domains

The highest number of DHIs fell into the health care domains of healthy living (n=548) and mental health (n=436). Multimedia Appendix 4 provides a table of health care domains, including the number of DHIs within each health care domain (ie, the sample size) and the scores' quantiles from 0% to 100% in increments of 25%. Figures 3-6 show the distribution of scores within each health care domain as box plots in descending order of the median (except for the first box plot that shows overall performance). Further details regarding OBR scores (ORCHA, UX, PCA, and DP) for each health care domain can be found in Multimedia Appendices 4 and 5.

Figure 3. The 1574 digital health interventions (DHIs) are classified into categories (can be more than one), an overall Organisation for the Review of Care and Health Apps (ORCHA) score box plot (first from left), and ORCHA scores box plots per health care domain. Sample sizes are above box plots; the red line indicates an ORCHA threshold score of 65, and the blue line indicates an overall median score of 61.5. The Kruskal-Wallis rank sum test has *P*<.001. ENTM: ear, nose, throat, and mouth; MaCR: medicines and clinical reference; MD: musculoskeletal disorders; PM: pain management; SSN: social support network; UA: utilities or administration.

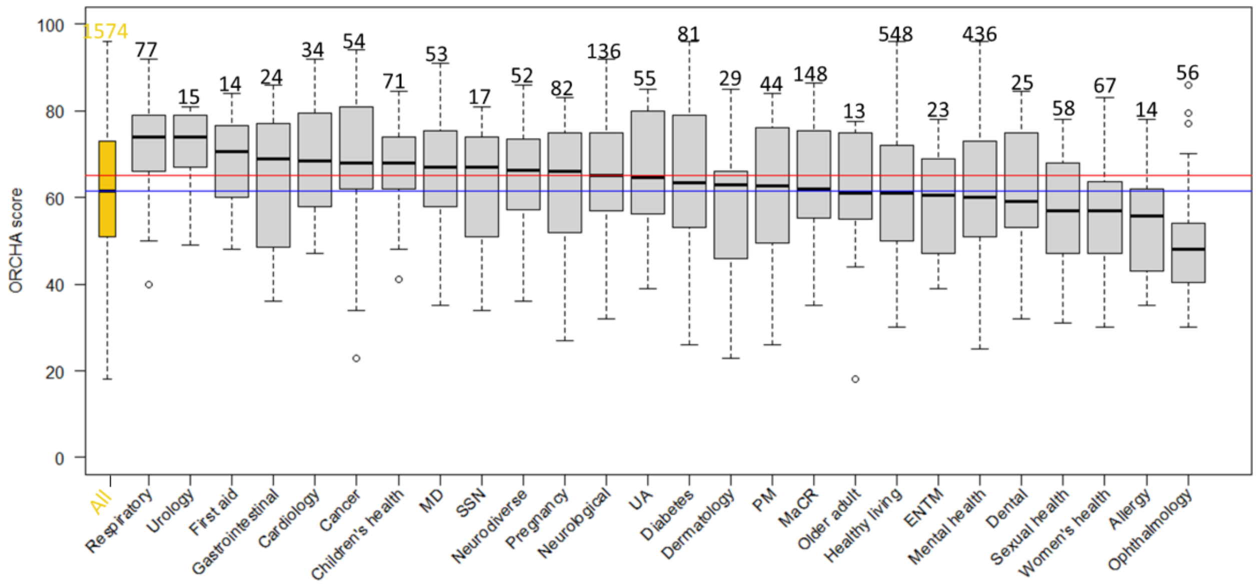


Figure 4. The 1574 digital health interventions (DHIs) are classified into categories (can be more than 1), an overall user experience (UX) score box plot (first from left), and UX scores box plots per health care domain. Sample sizes are above box plots; the blue line indicates an overall median score of 75.2. The Kruskal-Wallis rank sum test has *P*<.001. ENTM: ear, nose, throat, and mouth; MaCR: medicines and clinical reference; MD: musculoskeletal disorders; PM: pain management; SSN: social support network; UA: utilities or administration.

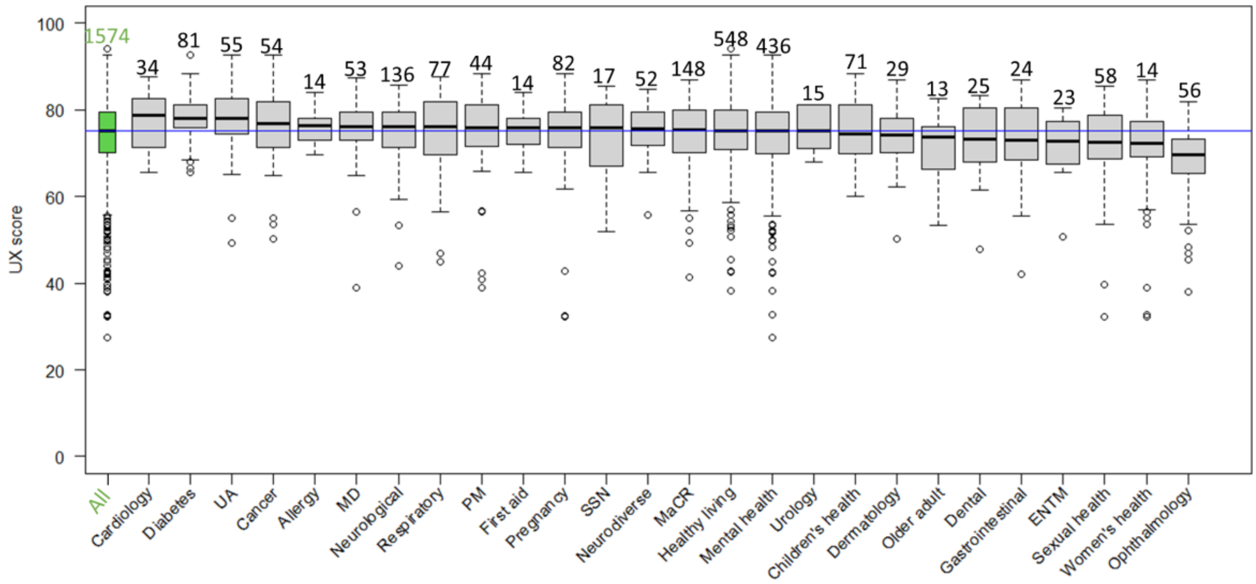


Figure 5. The 1574 digital health interventions (DHIs) are classified into categories (can be more than one), an overall professional and clinical assurance (PCA) score box plot (first from left), and PCA scores box plots per health care domain. Sample sizes are above box plots; the blue line indicates an overall median score of 49.6. The Kruskal-Wallis rank sum test has *P*<.001. ENTM: ear, nose, throat, and mouth; MaCR: medicines and clinical reference; MD: musculoskeletal disorders; PM: pain management; SSN: social support network; UA: utilities or administration.

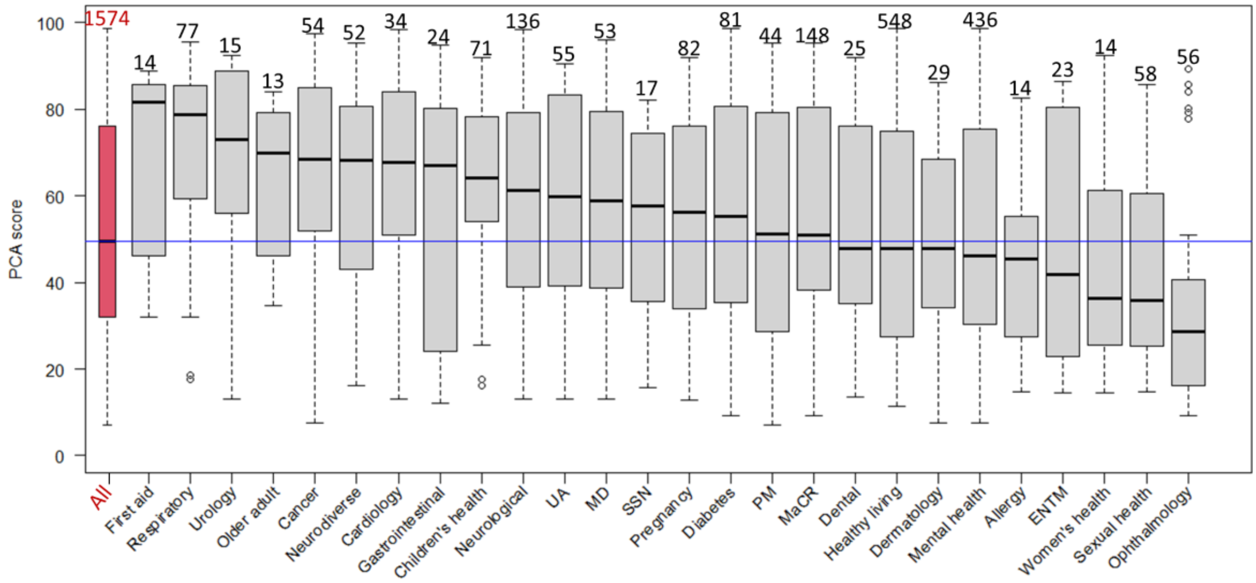
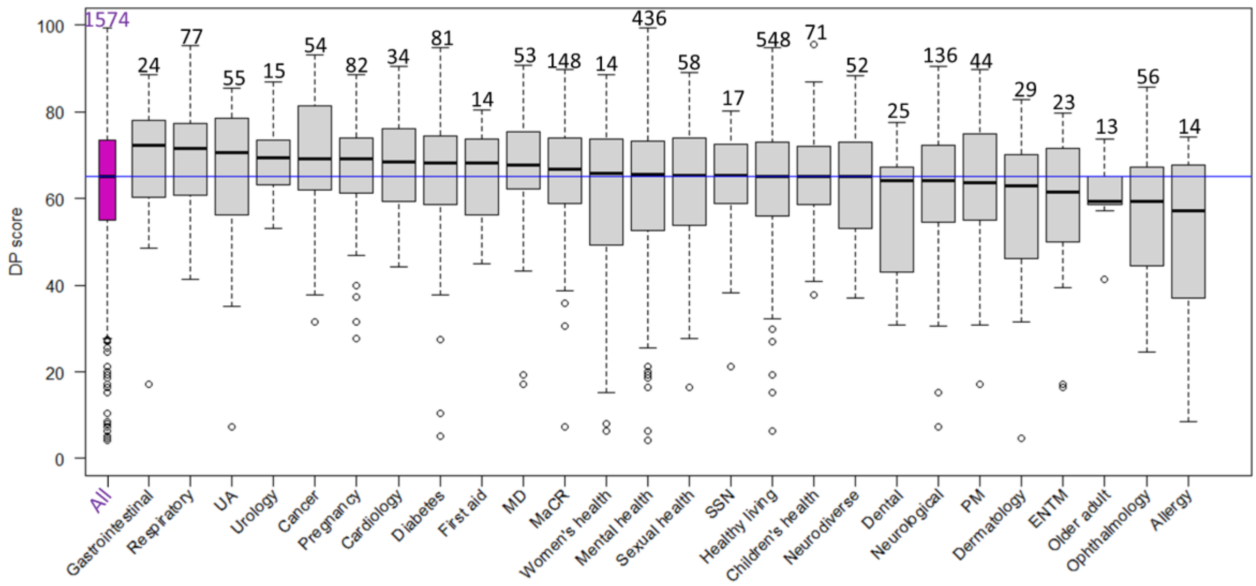


Figure 6. The 1574 digital health interventions (DHIs) are classified into categories (can be more than one), an overall digital privacy (DP) score box plot (first from left), and DP scores box plots per health care domain. Sample sizes are above box plots; the blue line indicates an overall median score of 65.1. The Kruskal-Wallis rank sum test has *P*<.001. ENTM: ear, nose, throat, and mouth; MaCR: medicines and clinical reference; MD: musculoskeletal disorders; PM: pain management; SSN: social support network; UA: utilities or administration.



Kruskal-Wallis rank sum tests were used to check for statistically significant differences between DHI categories. A statistically significant result (*P*<.001) was obtained for all OBR scores (ORCHA, UX, PCA, and DP) meaning that for all the scores at least 1 health care domain distribution is statistically significantly different from another. A post hoc analysis was conducted using the Dunn test to identify which categories are statistically different from each other (Multimedia Appendix 6).

For all DHIs, a total of 46.2% (12/26) health care domains had a median ORCHA score of 65 or more. The apps in each of the health care domains presented in descending order of quality (median ORCHA score; n) are as follows: respiratory (median 74.0; n=77), urology (median 74.0; n=15), first aid (median 70.5; n=14), gastrointestinal (median 69.0; n=24), cardiology (median 68.5; n=34), children's health (median 68.0; n=71), cancer (median 68.0; n=54), social support network (median 67.0; n=17), musculoskeletal disorders (median 67.0; n=53), neurodiverse (median 66.3; n=52), pregnancy (median 66.0; n=82), and neurological (median 65.0; n=136). A total of 53.8%

(14/26) health care domains had a median ORCHA score of less than 65. These, in descending order, are as follows: utilities or administration (median 64.5; n=55); diabetes (median 63.5; n=81); dermatology (median 63.0; n=29); pain management (median 62.8; n=44); medicines and clinical reference (median 62.0; n=148); healthy living (median 61.0; n=548); older adult (median 61.0; n=13); mental health (median 60.0; n=436); ear, nose, throat, and mouth (median 60.5; n=23); dental care (median 59.0; n=25); women's health (median 57.0; n=67); sexual health (median 57.0; n=58); allergy (median 55.8; n=14); and ophthalmology (median 48.0; n=56).

For tier B, a total of 57.7% (15/26) health care domains had a median ORCHA score of 65 or more. These, in descending order, are as follows: cancer (median 75.0; n=37), respiratory (median 73.0; n=49), urology (median 71.5; n=12), pregnancy (median 70.8; n=56), first aid (median 70.5; n=14), utilities or administration (median 70.0; n=41), children's health (median 68.0; n=66), social support network (median 68.0; n=16), neurological (median 68.0; n=104), medicines and clinical reference (median 68.0; n=95), neurodiverse (median 67.5; n=48), diabetes (median 67.5; n=33), musculoskeletal disorders (median 67.0; n=30), older adult (median 66.5; n=10), and cardiology (median 66.0; n=16). A total of 42.3% (11/26) health care domains had a median ORCHA score of less than 65. These, in descending order, are as follows: dermatology (median 64.3; n=16); pain management (median 63.5; n=35); mental health (median 62.0; n=332); sexual health (median 62.0; n=27); healthy living (median 61.0; n=436); dental care (median 59.5; n=20); women's health (median 58.3; n=36); allergy (median 58.0; n=9); gastrointestinal (median 56.0; n=13); ear, nose, throat, and mouth (median 55.0; n=17); and ophthalmology (median 50.0; n=30).

For tier C, a total of 24% (6/25; no "first aid" health care domain) health care domains had a median ORCHA score of 65 or more. These, in descending order, are as follows: urology (median 79.0; n=3), respiratory (median 74.5; n=28), cardiology (median 72.5; n=18), gastrointestinal (median 71.0; n=11), children's health (median 70.0; n=5), and musculoskeletal disorders (median 68.0; n=23). A total of 76% (19/25) health care domains had a median ORCHA score of less than 65. These, in descending order, are as follows: cancer (median 64.0; n=17); diabetes (median 63.0; n=48); ear, nose, throat, and mouth (median 61.8; n=6); healthy living (median 60.0; n=106); pain management (median 59.0; n=9); women's health (median 57.0; n=31); neurological (median 57.0; n=30); utilities or administration (median 57.0; n=14); medicines and clinical reference (median 55.5; n=49); mental health (median 55.0; n =102); pregnancy (median 55.0; n=26); older adult (median 55.0; n=2); sexual health (median 51.0; n=31); dermatology (median 46.0; n=13); neurodiverse (median 46.0; n=3); dental care (median 44.0; n=5); ophthalmology (median 43.3; n=26); allergy (median 42.0; n=5); and social support network (median 34.0; n=1). Multimedia Appendix 5 contains UX, PCA, and DP assessment areas ranked in order, and Multimedia Appendix 7 contains rank consistency. Multimedia Appendix 8 contains Distribution of DHIs across NICE Evidence Standards Framework (ESF) tiers by healthcare domain.

Partition of DHIs by ISO Certification and Medical Device Designation

Using median (IQR), the following difference has been found in DP scores among DHIs that received ISO 27001 certification [27] (79.4, IQR 73.6-85.3; n=77) and those that did not (65.0, IQR 54.1-72.4; n=1497), with a 2-sided unpaired Wilcoxon rank sum test with *P*<.001 and Cliff delta =.704 (95% CI 0.620-0.772). The following difference has been found in PCA scores among DHIs that have been designated as "medical device" [28] (58.8, IQR 33.7-84.4; n=162) and those that were not (49.3, IQR 31.9-76.1; n=1412), with a 2-sided unpaired Wilcoxon rank sum test with *P*=.003 and Cliff delta =.143 (95% CI 0.040-0.243).

For tier B, the following difference has been found in DP scores among DHIs that received ISO 27001 certification (78.5, IQR 71.4-81.8; n=42) and those that did not (65.0, IQR 53.8-72.2; n=1113), with a 2-sided unpaired Wilcoxon rank sum test with *P*<.001 and Cliff delta =.667 (95% CI 0.541-0.764). The following difference has been found in PCA scores among DHIs that have been designated as "medical device" (78.3, IQR 41.8-86.7; n=23) and those that were not (50.9, IQR 31.9-76.1; n=1132), with a 2-sided unpaired Wilcoxon rank sum test with *P*<.001 and Cliff delta =.644 (95% CI 0.470-0.769).

For tier C, the following difference has been found in DP scores among DHIs that received ISO 27001 certification (83.2, IQR 75.7-86.4; n=35) and those that did not (66.8, IQR 54.1-73.6; n=373), with a 2-sided unpaired Wilcoxon rank sum test with *P*<.001 and Cliff delta =.724 (95% CI 0.604-0.812). The following difference has been found in PCA scores among DHIs that have been designated as "medical device" (43.7, IQR 28.7-80.8; n=139) and those that were not (41.1, IQR 30.3-68.2; n=269), with a 2-sided unpaired Wilcoxon rank sum test with *P*=.002 and Cliff delta =.183 (95% CI 0.061-0.300).

Discussion

Principal Findings

A total of 57.3% (902/1574) DHIs in the data set failed to meet the ORCHA (a proxy for overall quality) threshold score of 65. The UX score was consistently the highest out of the 3 assessment areas (UX, PCA, and DP). The UX score also had the least variance when compared with other OBR scores. We found that scores differed widely between different health care domains. However, only some differences achieved statistical significance (Dunn test in Multimedia Appendix 6). The analysis revealed that the highest ORCHA scores were observed in the respiratory health care domain and the lowest in the ophthalmology health care domain (Figure 1 and Multimedia Appendix 4).

There have been several studies that suggest DHIs' quality could be further improved [4-8]. By identifying health care domains that have low OBR DHIs, this study indicates where a greater effort is needed to quality-assure these DHIs.

Table 1 shows that the largest variance has been observed in the PCA assessment area (PCA score IQR 44.2 and SD 24.8), which includes criteria related to the availability of scientific evidence to support the content and efficacy or effectiveness of the DHIs. This variation in clinical assurance across different DHIs is consistent with previous research. For instance, a paper from 2021 [31] found that evidence to support the claims made by health apps is often unavailable or of questionable quality. Similarly, a systematic review and exploratory meta-analysis from 2017 [32] with a focus on diagnostic apps found that the evidence for the diagnostic performance of health apps is limited. Additionally, a meta-analysis of randomized controlled trials from 2021 [27] concluded that, while there has been an increase in the rigorous evaluation of apps aimed at modifying behavior to promote health and manage disease, the evidence that such apps can improve health outcomes is weak.

Previous work has been done on benchmarking DHI system usability scores (SUS) across digital health apps [33] and for heart failure apps [34]. This study differs as it focuses on comparing DHIs using a broader selection of DHIs across health care domains and assessment areas (UX, PCA, and DP). Previous work from 2020 [35] has introduced an implementation framework called Technology Evaluation and Assessment Criteria for Health Apps. The aim of the framework is to enable users to make informed decisions regarding app use and increase app evaluation engagement by introducing a process to assist app implementation (Technology Evaluation and Assessment Criteria for Health Apps) across all DHIs. This study differs as it not only enables users to make informed decisions regarding app use but also enables the comparison of DHIs across health care domains. This study also identifies which health care domains may need more attention regarding their quality.

Compliance With Best Practices Across Health Care Domains

This study further observed differences in best practice compliance among health care domains. While DP and UX median scores were relatively similar across health care domains, large differences were observed between PCA scores (Figure 5 and Multimedia Appendices 4 and 5). A potential partial explanation for these findings may be that the proportion of DHIs within different tiers, and thus with different levels of evidence requirements (see above), may vary among health care domains. This suggestion is partially supported by the data, as a large proportion of DHIs in health care domains with high PCA scores fall into tiers A or B rather than C (Multimedia Appendices 5 and 8).

For all DHIs, a total of 12 of the 26 health care domains had a median ORCHA score of 65 or more. And a total of 14 of the 26 health care domains had a median ORCHA score of less than 65. For tier B, a total of 15 of the 26 health care domains had a median ORCHA score of 65 or more. And 11 of the 26 health care domains had a median ORCHA score of less than 65. For tier C, a total of 6 of the 25 (no "first aid" health care domain) health care domains had a median ORCHA score of 65 or more. And 19 of the 25 health care domains had a median ORCHA score of less than 65. Respiratory and urology DHIs were consistently highly ranked in NICE tiers B and C (Multimedia Appendices 4, 5, and 7).

The data indicate that DHIs that have received ISO 27001 certification (median 79.4, IQR 73.6-85.3; n=77) score higher regarding their DP score than those that have not (median 65.0, IQR 54.1-72.4; n=1497). The difference was statistically significant with a Wilcoxon rank sum test with $P<.001$ and Cliff delta =.704, indicating a large difference in DP scores. Similar results were obtained when the DHIs were partitioned by tiers B and C, as can be seen in the "Partition of DHIs by ISO Certification and Medical Device Designation" section.

DHIs that have been designated as medical device (median 58.8, IQR 33.7-84.4; n=162), scored higher on PCA than those that were not (median 49.3, IQR 31.9-76.1; n=1412). The difference was statistically significant with a Wilcoxon rank sum test with $P=.003$ and Cliff delta=.143, indicating a negligible difference in PCA scores. However, when partitioned by NICE tiers B and C, as can be seen in the "Partition of DHIs by ISO Certification and Medical Device Designation" section, results showed that for tier B DHIs that have been designated as medical device (median 78.3, IQR 41.8-86.7; n=23) scored higher than those that were not (median 50.9, IQR 31.9-76.1; n=1132). And had a Wilcoxon rank sum test with $P<.001$ and Cliff delta =.644, indicating a large difference in PCA scores. For tier C, DHIs that have been designated as medical device (median 43.7, IQR 28.7-80.8; n=139) and those that were not (median 41.1, IQR 30.3-68.2; n=269) had a Wilcoxon rank sum test with $P=.002$, but a much lower Cliff delta =.183, indicating a negligible difference in PCA scores.

Medical device DHIs seem to be outperforming nonmedical device DHIs regarding PCA scores. Especially medical device DHIs in tier B. Speculation can be made that since medical device DHIs have regulatory requirements [28], more is expected of them regarding PCA. This leads to low PCA scores among tier C apps, with a negligible difference in PCA score between medical device and nonmedical device DHIs, according to Cliff delta. However, since medical device DHIs are typically assigned to tier C, they outperform nonmedical device DHIs in tier B as developers attempt to meet regulatory demands. An alternative interpretation is that since medical device regulation is a gold standard where clinical evidence is evaluated, it would be expected to see higher PCA scores for DHIs designated as medical devices than nonmedical devices in tier C, similarly to DHIs in tier B. It could be that the PCA score for tier C is an inappropriate measure of clinical evidence. Meaning that the criteria for tier C DHIs are not ideal to differentiate between different levels of evidence.

A study from 2020 [36] focused on the value of mobile health (mHealth) for patients. Their analysis found that the highest level of clinical evidence for mHealth apps used for clinical scenarios is scarce. The analysis presented in this study identifies health care domains where DHIs may require improvements regarding their quality. Hence, this study may be helpful in mitigating the problem of scarce evidence regarding the quality of DHIs.

The current findings indicate that OBR scores differ among DHIs in different NICE tiers and health care domains. In the long term, the aim should be to elevate DHIs in lower-scoring categories to achieve an ORCHA threshold score of 65. The quantiles presented in Multimedia Appendices 3 and 4 can be used for the identification of low-quality DHIs as indicated by OBR scores.

After receiving OBR scores, a specific DHI can be compared with other DHIs in the same health care domain or NICE tier using quantiles. This will reveal how compliant the DHI is with best practice standards relative to similar DHIs. These comparisons can be conducted with ORCHA scores or for the separate assessment areas (UX, PCA, and DP; Multimedia Appendices 3 and 4).

Limitations

A few limitations of this study should be noted. There were uneven sample sizes for DHIs across NICE tiers (the sample size ranged from 11 to 1155 DHIs) and health care domains (the sample size ranged from 13 to 548 DHIs). When partitioning the data, lower samples in tiers and categories lead to less reliable results in those tiers and categories. Where the case was that the same DHIs were assessed twice, that is, the Android version and the iOS version (n=466 DHIs; Multimedia Appendix 2), the mean OBR scores were calculated using the Android and iOS assessments, and the result was included in the analysis. However, it is possible that if the names of the DHIs were somewhat different for the Android and iOS versions, both would have been included in the analysis as separate DHIs.

The OBR version 6 evolved from earlier versions of the OBR during the height of the COVID-19 pandemic. Originally, version 6 was created as a more stringent version of the OBR so that ORCHA could recommend the most compliant DHIs to members of the UK population with confidence. ORCHA tested version 6 on a selection of highly compliant DHIs (as determined by previous versions of the OBR). This set of 30 DHIs served as the pilot group, with the subsequent 2097 DHIs being assessed with ORCHA's typical assessment approach of categorizing DHIs into categories, ordering by number of downloads, and assessing the most downloaded DHI in each health care domain, followed by the second, and so forth.

Future Work

The concurrent validity can be performed on the ORCHA assessment tool by comparing ORCHA scores against other assessment frameworks (eg, Mobile Application Rating Scale [37]). The analysis conducted in this paper could be repeated with more DHIs in tier A.

Conclusion

This study examined assessment data for 1574 DHIs and found that 57.3% (902/1574) of the DHIs in the data set failed to meet the ORCHA threshold score of 65 (accepted by the NHS as a signal of compliance with best practice standards). This work also identified differences with regard to the OBRs of DHIs in different tiers and health care domains. Appropriate evidence and clinical assurance were especially lacking in DHIs with high risk (as per their tiers), which raises safety concerns and highlights the need for DHI assessments that support users in the selection of safe and effective DHIs. Interestingly, more stringent (tier C) clinical assurance and evidence requirements seemed more likely to be met in health care domains with high funding availability, such as diabetes and cardiology. This underscores the need for more investment in health care domains that currently demonstrate low compliance with best practices, such as women's health, ophthalmology, dental care, and allergy.

Additionally, this study produced quantiles across different health care domains and NICE tiers, which could be used to compare health care domain-specific DHIs in future studies.

Acknowledgments

This research is done in partnership with the Organisation for the Review of Care and Health Apps (ORCHA), which is a UK-based digital health compliance company. This work is supported by a Northern Ireland Department for the Economy (DfE) Co-operative Awards in Science and Technology (CAST) award PhD scholarship.

We would like to acknowledge the contribution of the many DHIs reviewers and developers who worked with ORCHA to allow for the review of DHIs and consent for their data to be used for research purposes. Without their contributions and consent, this research would not have been possible.

Data Availability

The data sets generated during and/or analyzed during this study are available in the ORCHA repository [38]. The data from this data set (and for more DHIs) are freely available to registered users (registration is free for a trial period). However, the data in the repository are presented for individual DHIs and would need to be extracted into a suitable format for analysis.

Conflicts of Interest

This study is funded by a DfE Cast award [38] and ORCHA [19]. SL, RD, and AF are or were employees at ORCHA.

Multimedia Appendix 1
Different ORCHA thresholds.
[DOCX File , 34 KB - mhealth_v11i1e47043_app1.docx]

Multimedia Appendix 2
Inclusion of digital health interventions (DHIs) in the Analysis.
[DOCX File , 23 KB - mhealth_v11i1e47043_app2.docx]

Multimedia Appendix 3
Organisation for the Review of Care and Health Apps Baseline Review (OBR) V6 score quantiles separated by National Institute for Health and Care Excellence (NICE) tiers.
[DOCX File , 15 KB - mhealth_v11i1e47043_app3.docx]

Multimedia Appendix 4
OBR V6 score quantiles classified into healthcare domains.
[DOCX File , 83 KB - mhealth_v11i1e47043_app4.docx]

Multimedia Appendix 5
Rank of DHIs.
[DOCX File , 36 KB - mhealth_v11i1e47043_app5.docx]

Multimedia Appendix 6
Post hoc analysis with Dunn test.
[DOCX File , 174 KB - mhealth_v11i1e47043_app6.docx]

Multimedia Appendix 7
DHI rank consistency.
[DOCX File , 34 KB - mhealth_v11i1e47043_app7.docx]

Multimedia Appendix 8
Distribution of DHIs across NICE Evidence Standards Framework (ESF) tiers by healthcare domain.
[DOCX File , 23 KB - mhealth_v11i1e47043_app8.docx]

References

1. Kern J, Skye A, Krupnick M, Pawley S, Pedersen A, Preciado K, et al. Written consent of IQVIA and the IQVIA Institute. Digital Health Trends. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2023-10-27]
2. Leigh S, Daly R, Stevens S, Lapajne L, Clayton C, Andrews T, et al. Web-based internet searches for digital health products in the United Kingdom before and during the COVID-19 pandemic: a time-series analysis using app libraries from the Organisation for the Review of Care and Health Applications (ORCHA). BMJ Open 2021;11(10):e053891 [FREE Full text] [doi: 10.1136/bmjopen-2021-053891] [Medline: 34635531]
3. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
4. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
5. Melcher J, Torous J. Smartphone apps for college mental health: a concern for privacy and quality of current offerings. Psychiatr Serv 2020;71(11):1114-1119 [FREE Full text] [doi: 10.1176/appi.ps.202000098] [Medline: 32664822]
6. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
7. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
8. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
9. Faessen JPM, Lucassen DA, Buso MEC, Camps G, Feskens EJM, Brouwer-Brolsma EM. Eating for 2: a systematic review of Dutch app stores for apps promoting a healthy diet during pregnancy. Curr Dev Nutr 2022;6(6):nzac087 [FREE Full text] [doi: 10.1093/cdn/nzac087] [Medline: 35711572]

10. van Galen LS, Xu X, Koh MJA, Thng S, Car J. Eczema apps conformance with clinical guidelines: a systematic assessment of functions, tools and content. Br J Dermatol 2020;182(2):444-453 [FREE Full text] [doi: 10.1111/bjd.18152] [Medline: 31179535]
11. Simon L, Reimann J, Steubl LS, Stach M, Spiegelhalder K, Sander LB, et al. Help for insomnia from the app store? A standardized rating of mobile health applications claiming to target insomnia. J Sleep Res 2023;32(1):e13642 [FREE Full text] [doi: 10.1111/jsr.13642] [Medline: 35624078]
12. Woulfe F, Fadahunsi KP, Smith S, Chirambo GB, Larsson E, Henn P, et al. Identification and evaluation of methodologies to assess the quality of mobile health apps in high-, low-, and middle-income countries: rapid review. JMIR Mhealth Uhealth 2021;9(10):e28384 [FREE Full text] [doi: 10.2196/28384] [Medline: 34636737]
13. Hongsanun W, Insuk S. Quality assessment criteria for mobile health apps: a systematic review. Walailak J Sci Tech 2020;17(8):745-759 [FREE Full text] [doi: 10.48048/wjst.2020.6482]
14. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
15. Hensher M, Cooper P, Dona SWA, Angeles MR, Nguyen D, Heynsbergh N, et al. Scoping review: development and assessment of evaluation frameworks of mobile health apps for recommendations to consumers. J Am Med Inform Assoc 2021;28(6):1318-1329 [FREE Full text] [doi: 10.1093/jamia/ocab041] [Medline: 33787894]
16. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
17. Moshi MR, Tooher R, Merlin T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: a systematic review. Int J Technol Assess Health Care 2018;34(5):464-475 [FREE Full text] [doi: 10.1017/S026646231800051X] [Medline: 30201060]
18. Azad-Khaneghah P, Neubauer N, Cruz AM, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
19. ORCHA. URL: https://orchahealth.com/ [accessed 2022-06-02]
20. Sophie H. Review documentation—review development and resources. External Share for Confluence. URL: https://confluence.external-share.com/content/b6055aac-83e4-4947-be0e-ebb8c39559ef [accessed 2022-03-13]
21. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2022-06-02]
22. Scope of registration: introduction. Care Quality Commission. URL: https://www.cqc.org.uk/guidance-providers/scope-registration [accessed 2022-07-04]
23. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965;52(3-4):591-611. [doi: 10.1093/biomet/52.3-4.591]
24. Ostertagová E, Ostertag O, Kováč J. Methodology and application of the Kruskal-Wallis test. Appl Mech Mater 2014;611:115-120. [doi: 10.4028/www.scientific.net/amm.611.115]
25. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. Stata J 2015;15(1):292-300 [FREE Full text] [doi: 10.1177/1536867x1501500117]
26. Holm S. A simple sequentially rejective multiple test procedure. Scand J Statist 1979;6(2):65-70. [doi: 10.1163/1872-9037_afco_asc_558]
27. ISO/IEC 27001:2022: information security, cybersecurity and privacy protection: information security management systems: requirements. International Organization for Standardization. URL: https://www.iso.org/standard/27001 [accessed 2023-06-27]
28. Regulating medical devices in the UK. Gov.uk. URL: https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk [accessed 2023-06-27]
29. Romano J, Kromrey JD. Appropriate statistics for ordinal level data: should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? ResearchGate. URL: https://tinyurl.com/222xcfux [accessed 2023-09-11]
30. Privacy policy. ORCHA. URL: https://appfinder.orcha.co.uk/privacy-policy/ [accessed 2022-08-11]
31. Morley J, Powell J, Floridi L. What's the evidence that health apps work? A scoping study of evidence of effectiveness for direct-to-consumer apps on the app store. SSRN J 2021:1-23 [FREE Full text] [doi: 10.2139/ssrn.3851242]
32. Buechi R, Faes L, Bachmann LM, Thiel MA, Bodmer NS, Schmid MK, et al. Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. BMJ Open 2017;7(12):e018280 [FREE Full text] [doi: 10.1136/bmjopen-2017-018280] [Medline: 29247099]
33. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
34. Cestari VRF, Florêncio RS, Garces TS, de Paula Pessoa VLM, Moreira TMM. Benchmarking of mobile apps on heart failure. Rev Bras Enferm 2021;75(1):e20201093 [FREE Full text] [doi: 10.1590/0034-7167-2020-1093] [Medline: 34614075]
35. Camacho E, Hoffman L, Lagan S, Rodriguez-Villa E, Rauseo-Ricupero N, Wisniewski H, et al. Technology Evaluation and Assessment Criteria for Health Apps (TEACH-Apps): pilot study. J Med Internet Res 2020;22(8):e18346 [FREE Full text] [doi: 10.2196/18346] [Medline: 32535548]

36. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
37. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
38. Library. ORCHA. URL: https://appfinder.orcha.co.uk/ [accessed 2022-08-16]
39. Co-operative awards in science and technology. nidirect. URL: https://www.nidirect.gov.uk/articles/co-operative-awards-science-and-technology [accessed 2022-08-16]

Abbreviations

DHI: digital health intervention
DP: data privacy
ESF: Evidence Standards Framework
ISO: International Organization for Standardization
mHealth: mobile health
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
OBR: Organisation for the Review of Care and Health Apps Baseline Review
ORCHA: Organisation for the Review of Care and Health Apps
PCA: professional and clinical assurance
SUS: system usability scores
UX: user experience

Edited by L Buis; submitted 06.03.23; peer-reviewed by A Maltz, M Rollwage, J Westheimer; comments to author 15.05.23; revised version received 06.07.23; accepted 18.10.23; published 23.11.23.

Please cite as:
Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Daly R, Frey AL, Leigh S
Quality of Digital Health Interventions Across Different Health Care Domains: Secondary Data Analysis Study
JMIR Mhealth Uhealth 2023;11:e47043
URL: https://mhealth.jmir.org/2023/1/e47043
doi:10.2196/47043
PMID:37995121

Original Paper

The Most Popular Commercial Weight Management Apps in the Chinese App Store: Analysis of Quality, Features, and Behavior Change Techniques

Lan Geng[1*], BSc; Genyan Jiang[1*], BSc; Lingling Yu[1], BSc; Yueming Xu[1], BSc; Wei Huang[1], BM; Zhiqi Chen[1], BSc; Xiaoyan Qi[1], PhD; Ting Zhang[1], MM; Mei Zhao[1], PhD

School of Nursing, Anhui Medical University, Hefei, China

[*]these authors contributed equally

Corresponding Author:
Mei Zhao, PhD
School of Nursing
Anhui Medical University
No.15 Feicui Road
Hefei, 230601
China
Phone: 86 551 63869167
Email: zhaomei@ahmu.edu.cn

Abstract

Background: Many smartphone apps designed to assist individuals in managing their weight are accessible, but the assessment of app quality and features has predominantly taken place in Western countries. Nevertheless, there is a scarcity of research evaluating weight management apps in China, which highlights the need for further investigation in this area.

Objective: This study aims to conduct a comprehensive search for the most popular commercial Chinese smartphone apps focused on weight management and assess their quality, behavior change techniques (BCTs), and content-related features using appropriate evaluation scales. Additionally, the study sought to investigate the associations between the quality of various domains within weight management apps and the number of incorporated BCTs and app features.

Methods: In April 2023, data on weight management apps from the iOS and Android app stores were downloaded from the Qimai Data platform. Subsequently, a total of 35 weight management apps were subjected to screening and analysis by 2 researchers. The features and quality of the apps were independently assessed by 6 professionals specializing in nutrition management and health behavioral change using the Mobile Application Rating Scale (MARS). Two registered dietitians, who had experience in app development and coding BCTs, applied the established 26-item BCT taxonomy to verify the presence of BCTs. Mean (SD) scores and their distributions were calculated for each section and item. Spearman correlations were used to assess the relationship between an app's quality and its technical features, as well as the number of incorporated BCTs.

Results: The data set included a total of 35 apps, with 8 available in the Android Store, 10 in the Apple Store, and 17 in both. The overall quality, with a mean MARS score of 3.44 (SD 0.44), showed that functionality was the highest scoring domain (mean 4.18, SD 0.37), followed by aesthetics (mean 3.43, SD 0.42), engagement (mean 3.26, SD 0.64), and information (mean 2.91, SD 0.52), which had the lowest score. The mean number of BCTs in the analyzed apps was 9.17 (range 2-18 BCTs/app). The most common BCTs were "prompt review of behavioral goals" and "provide instruction," present in 31 apps (89%). This was followed by "prompt self-monitoring of behavior" in 30 apps (86%), "prompt specific goal setting" in 29 apps (83%), and "provide feedback on performance" in 27 apps (77%). The most prevalent features in the analyzed apps were the need for web access (35/35, 100%), monitoring/tracking (30/35, 86%), goal setting (29/35, 83%), and sending alerts (28/35, 80%). The study also revealed strong positive correlations among the number of BCTs incorporated, app quality, and app features. This suggests that apps with a higher number of BCTs tend to have better overall quality and more features.

Conclusions: The study found that the overall quality of weight management apps in China is moderate, with a particular weakness in the quality of information provided. The most prevalent BCTs in these apps were reviewing behavioral goals, providing guidance, self-monitoring of behavior, goal setting, and offering performance feedback. The most common features were the need for web access, monitoring and tracking, goal setting, and sending alerts. Notably, higher-quality weight management apps in China tended to incorporate more BCTs and features. These findings can be valuable for developers looking to improve weight management apps and enhance their potential to drive behavioral change in weight management.

(JMIR Mhealth Uhealth 2023;11:e50226) doi:10.2196/50226

KEYWORDS

quality; mobile health; mobile apps; weight loss; weight management; apps; MARS; China

Introduction

Background

Obesity is becoming increasingly prevalent on a global scale, and this poses a significant challenge to public health worldwide. One of the primary contributing factors to this concerning trend is the expansion of global free trade, coupled with the economic growth and urbanization experienced by many low- and middle-income countries. These changes have given rise to environments that promote unhealthy dietary habits and a reduction in physical activity (PA) [1]. By the year 2025, it is anticipated that the worldwide prevalence of obesity will reach 18% among men and exceed 21% among women. Additionally, the prevalence of severe obesity is projected to surpass 6% for men and 9% for women [2]. Between 1980 and 2015, excess body weight has been linked to more than 4 million deaths worldwide [3]. Recent statistics indicate that over 50% of adults in China are overweight or obese, establishing China as the country with the highest number of individuals affected by overweight or obesity in the world [4]. Newly published studies have forecasted that by 2030, the combined prevalence of overweight and obesity among Chinese adults aged 18 years and older will reach 70.5% [5]. China's rate of overweight and obesity is significantly higher than that of the Middle East (21.17%) [6], France (37.3%), Portugal (43.3%), and Italy (44.0%) [7]. However, it is slightly lower than the rates in Romania (58.8%), Latvia (57.3%), and Bulgaria (56.9%) [7]. Nonetheless, owing to its vast population, China has the highest count of individuals with overweight or obesity globally [4]. This concerning statistic is expected to escalate to 810.65 million by 2030 [5], underscoring the gravity of the national health situation. Obesity has emerged as the leading cause of poor health, surpassing malnutrition and infectious diseases [8]. Excess body weight, particularly obesity, is the primary lifestyle-related risk factor for premature death. This increases the risk of morbidity and mortality from various diseases and conditions, such as cardiovascular disease, type 2 diabetes mellitus, nonalcoholic fatty liver disease, and certain types of cancer [8]. Hence, it holds significant importance and urgency to focus on behavioral interventions for individuals with overweight or obesity and bolster weight management. For decades, many countries have implemented policies aimed at promoting healthy lifestyles to prevent obesity [1], but these efforts have not succeeded in curbing the prevalence of obesity. At present, the scientific community is placing a priority on researching interventions that are accessible, cost-effective, and customized to individual needs. The goal of these interventions is to facilitate the translation of health knowledge into practical health behaviors and skills, a task that poses a significant challenge.

In the digital age, smartphones serve a purpose beyond making phone calls; they are now widely integrated into various aspects of life, including transportation, shopping, and social entertainment. Consequently, they have a profound impact on both people's work and personal lives. In China, smartphone ownership has witnessed exponential growth over the past decade. According to a statistical report on the development of the Chinese internet, by December 2020, there were 989 million Chinese internet users, of which a staggering 986 million (99.7%) used smartphones to access the internet [9].The widespread adoption of smartphones has spurred the rapid growth of the mobile app market. In an environment where the country and society actively promote national health, people's awareness of health and their expectations for health standards continue to rise. This has led to increased demands for personal health management, the development of healthy behaviors, and opportunities for healthy social engagement, among other aspects.According to available data, there are 660 million mobile health users in China [10]. Among these users, weight management is one of the most prominent demands, representing 41.9% of all requests [11]. Compared with traditional weight management methods, mobile apps are more cost-effective [12] and have the advantage of transcending time and space constraints. They enable users to monitor their personal health status and access essential health information whenever and wherever they desire. The convenience and practicality of weight management apps have made them highly popular among users, establishing them as a vital tool for effective weight management.

The weight management app market in China is experiencing rapid growth. Nevertheless, the content and quality of the available apps remain somewhat unclear. While app stores offer star ratings, user comments, and installation statistics, there is a notable absence of quality indicators specifically tailored to weight management apps. Furthermore, star ratings and comments can sometimes include inaccurate or subjective feedback, potentially causing confusion among users [13]. In the pursuit of offering users higher-quality mobile apps, an expert team has devised the Mobile Application Rating Scale (MARS). This scale serves as a straightforward, objective, and dependable tool for researchers, developers, and health professionals to assess app quality [14]. Since its inception in 2015, the MARS has been translated into multiple languages [15-17] and used to assess a wide range of mobile apps [18-26]. It comprises 23 items organized into distinct sections, covering engagement, functionality, aesthetics, information quality, and subjective quality [14]. Furthermore, mobile health (mHealth) apps that incorporate established behavior change techniques (BCTs) rooted in behavior change theories and designed to offer practical principles for altering determinants of behavior have the potential to facilitate behavioral change [27]. Nevertheless, not all commercially available weight management apps include BCTs [28]. Previous studies have evaluated English weight management apps from multiple angles, encompassing quality, features, and the presence of BCTs [28-30], but these studies cannot offer a comprehensive assessment of the quality of the currently available weight management apps in China.

Objectives

To fill this void, the objectives of this study were 4-fold: (1) to conduct a systematic assessment of the quality of well-received weight management apps; (2) to determine the count of BCTs and content-related features within these apps; (3) to evaluate the associations between various aspects of app quality and the inclusion of BCTs in these apps; and (4) to explore the connections between app quality and the number of app features.

Methods

Search Strategy

Qimai Data (Beijing Qimai Technology Co., Ltd.) is a specialized platform for analyzing mobile app data in China. It encompasses both the App Store (Apple Inc.) and Google Play Store (Alphabet Inc.) platforms, delivering comprehensive data insights for the iOS and Android app markets [31]. This platform constituted the primary data source for our study. In April 2023, 2 reviewers (LG and GYJ), 1 registered dietitian and 1 postgraduate in nutrition, performed a systematic search of mobile apps associated with weight management. This search encompassed 9 Chinese language mobile app stores and was facilitated using Qimai Data. The stores included the Apple iTunes Store for iOS and the Android system stores, including Tencent My App, Huawei, Xiaomi, VIVO, OPPO, Meizu, 360 Mobile Assistant and Baidu Mobile Assistant. The search was conducted using the keywords obesity, slimming, weight loss, weight, and fitness. Each keyword was entered into the general search bar of the 9 app stores mentioned above. Data extracted from Qimai Data were app names, subtitles, developers, rating scores, update time stamps, the number of ratings, and cumulative downloads. The inclusion or exclusion of the apps was documented following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [32].

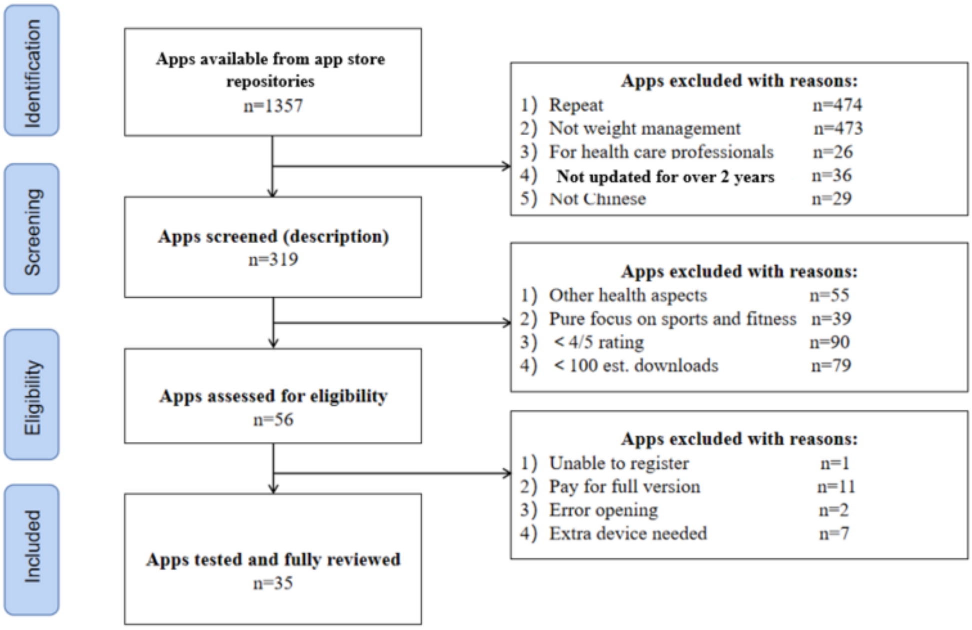
App Selection

The app filtering process involved several steps. Initially, information about weight management apps was retrieved and downloaded from the Qimai Data platform, with duplicate apps removed. Second, 2 reviewers (LG and GYJ) assessed app eligibility based on the app's title and introduction. An app was deemed eligible for inclusion in the study only if it was relevant to weight management, available in the Chinese language, designed for use by non–health care professionals, and had been updated within the preceding 2 years

Third, apps were excluded if they fell into the following categories: (1) they primarily addressed health aspects unrelated to weight management or related behaviors [28], such as smoking or sleeping; (2) they were purely focused on sports and fitness [28]; (3) they had a user rating below 4 on a scale ranging from 1 (inadequate) to 5 (excellent) [28]; and (4) they had accumulated less than 10,000 downloads.

Finally, apps requiring external devices, such as scales or wrist straps for proper functioning, and those that necessitated the purchase of a full version were excluded. We also removed apps that did not allow for registration or displayed errors upon opening. The app selection process is outlined in Figure 1. Following 4 rounds of screening, a collection of apps was identified and made ready for download and evaluation. This selection and evaluation process took place from April 2023 to May 2023.

Figure 1. Flow diagram of app search and selection.



Data Extraction

A team consisting of 6 raters (LG, GYJ, LLY, YMX, MZ, and TZ) downloaded the chosen apps and conducted independent assessments of their quality and features. The team comprised individuals with backgrounds in nursing and graduate students with research interests in clinical nutrition. Among the 6 reviewers, 2 were registered dietitians, and 4 were graduate nursing students. All members of the team possessed experience in nutrition management and health behavioral change, with 2 registered nutritionists having additional expertise in coding BCTs and app development. The raters used a variety of devices, including OPPO R15X, Huawei nova 11 Ultra, Redmi Note 8 Pro (Xiaomi), and OPPO R17, to download and evaluate apps from the Tencent App Store. For apps from the Apple App Store, iPhone 13 running iOS 15.6 and iPhone 12 running iOS 16.1.1 were used for downloading and testing. Each app was tested for 1 day. For all the apps included in the study, the following app characteristics were collected: app name; developer; version; date of the most recent update; average user rating (with a minimum of 4+); cumulative installation count; presence of BCTs; focus areas according to the MARS (such as enhancing happiness/well-being, behavioral change, entertainment, physical health); MARS theoretical underpinnings/strategies (including assessment, goal setting, information/education, advice/tips/strategies/skills training); and MARS technical features (eg, sharing capabilities, password protection, reminder notifications).

App Quality and Features Assessment

Six independent evaluators assessed app quality and features using the MARS [14]. Prior studies have demonstrated that the scale exhibits strong internal consistency and interrater reliability [33]. The MARS comprises 19 items, categorized into 4 dimensions: engagement (5 items, including entertainment, interest, customization, interactivity, and target group), functionality (4 items, encompassing performance, ease of use, navigation, and gestural design), aesthetics (3 items, covering layout, graphics, and visual appeal), and information (7 items, involving accuracy of app description, goals, quality of information, quantity of information, visual information, credibility, and evidence base). Each item was rated on a 5-point scale (1=inadequate; 2=poor; 3=acceptable; 4=good; and 5=excellent). The overall MARS score is calculated as the average of the 4 dimensions, providing an assessment of the app's overall quality. The MARS score ranges from 1 to 5, with higher scores indicating better app quality. A minimum acceptable score of 3.0 has been previously established as a cutoff point [34].

Beyond the 4 objective quality scales, the MARS also includes assessments of subjective app quality, which encompasses 4 items: recommendation, frequency of use, willingness to pay, and overall star rating. Additionally, it evaluates the likelihood of behavioral impact in weight management, considering 6 items: awareness, knowledge, attitudes, intention to change, help-seeking, and behavioral change. The subjective quality and likelihood of behavioral impact in weight management items were evaluated and scored independently.

Presence of BCTs Assessment

The presence of BCTs was evaluated by 2 registered dietitians using the taxonomy of BCTs [27]. Both registered dietitians possess experience in providing health behavioral change support, nutritional management, coding BCTs, and app development. The taxonomy of BCTs was developed by Abraham and Michie [35], and it was used to assess the presence or absence of BCTs in the included apps. This taxonomy comprises 26 BCTs, which can be used to gauge the effectiveness of behavior change interventions. It has also been used in previous studies to assess the presence of BCTs, similar to the approach taken in our study [36]. In this study, the 26 BCTs from the BCT taxonomy were used to assess the included apps, with 0 indicating absence and 1 indicating presence.

Quality Control

Initially, the 6 raters underwent training using the handbook to ensure their correct usage of the MARS and successfully completed a pilot test [14]. Before rating the apps included in the study, the raters randomly selected 6 weight management apps that were not part of the study, assessed them using the MARS, and discussed their findings. This step was taken to ensure that the raters were acquainted with the MARS program and the evaluation process. Any points of disagreement were deliberated until a consensus was ultimately reached. Subsequently, for all 35 eligible apps, the 6 raters initially downloaded and utilized all the functions of the apps on their own smartphones for 1 day to acquaint themselves with the apps. Lastly, the raters individually conducted the formal assessments of the 35 eligible apps using the MARS. For those apps that the raters could not access on their own smartphones due to compatibility issues, such as rater A with an Android smartphone system being unable to use an app on rater B's Apple smartphone system, the research team organized regular formal discussion meetings. On the first day, the raters swapped smartphones to become acquainted with the app interfaces and functions on each other's phones. Each app was used for a minimum of 20 minutes. On the second day, the app was used to gain further familiarity. On the third day, the app was used and formally evaluated using the MARS.

Statistical Analysis

Data analysis was conducted using SPSS version 26.0 (IBM Corp.). Descriptive statistics were applied to app quality and BCT data, and continuous variables were presented as mean and SD. If the data distribution was skewed, the median and IQR were reported. The interrater reliability between the raters for MARS scores and the number of BCTs present in the apps were assessed using 2-way mixed effects intraclass correlation coefficients. Agreement scores below 0.5 were considered indicative of poor agreement, while those within the range of 0.5-0.75 represented moderate reliability. Scores falling between 0.75 and 0.9 were indicative of good reliability, and scores exceeding 0.9 demonstrated excellent reliability [37]. The internal consistency of the MARS was evaluated using Cronbach α, while the internal consistency of BCTs was assessed using the Kuder-Richardson index 20 (KR-20). Spearman correlations were used to investigate the relationships between app quality, the number of technical app features, and the number of BCTs integrated into the apps. A correlation coefficient of >0.7 indicates a strong correlation, while a range of 0.3-0.7 suggests a moderate correlation, and <0.3 indicates a low correlation [38].

Ethical Considerations

This study was approved by the Ethics Committee of Anhui Medical University under the ethical number 84230017.

Results

Identification of Apps

A search was conducted in the Chinese app market for weight management apps available for iOS and Android systems, yielding 899 and 458 apps, respectively. The apps were subsequently screened according to the specified inclusion and exclusion criteria, ultimately resulting in the identification of 35 independent apps for analysis. Figure 1 presents an overview of the selection process and delineates the reasons for exclusion.

Characteristics and Features of the Selected Apps

Table 1 displays the characteristics and features of the apps included in this study. Of the 35 apps included, 10 were exclusive to iOS, 8 were exclusive to Android, and 17 apps were available on both platforms. Several apps were entirely free (12/35, 34%), but a larger number required payment for advanced services (23/35, 66%). The health behaviors targeted by the apps encompassed diet (28/35, 80%) and PA (21/35, 60%), with 16 apps addressing both of these behaviors. The majority of the apps necessitate a log-in (20/35, 57%), send reminders (28/35, 80%), and require internet access (35/35, 100%). Furthermore, they offer functions such as information/education (20/35, 57%), monitoring/tracking (30/35, 86%), goal setting (29/35, 83%), and advice/tips/strategies/skill training (27/35, 77%). Some apps provide the option for sharing (12/35, 34%), have an integrated app community (16/35, 46%), and offer password protection (14/35, 40%).

Table 1. Descriptive data of the included apps (N=35).

Descriptive data	Values			
	n (%)	Mean (SD)	Median (IQR)	Range (min-max)
App store				
Android	8 (23)	N/A[a]	N/A	N/A
Apple	10 (29)	N/A	N/A	N/A
Android and Apple	17 (49)	N/A	N/A	N/A
Cost				
Free	12 (34)	N/A	N/A	N/A
Paid (costs in RMB ￥[b])	23 (66)	29.41 (26.81)	16.50 (47.20)	75.00 (3.00-78.00)
User rating				
Average rating (4+)	35 (100)	4.63 (0.29)	4.80 (0.30)	1.00 (4.00-5.00)
The average number of user ratings (count)	35 (100)	114,684 (519,244)	2208 (17,548)	3,050,463 (25-3,050,488)
Health behavior				
Diet	28 (80)	N/A	N/A	N/A
Physical activity	21 (60)	N/A	N/A	N/A
Diet and physical activity	16 (46)	N/A	N/A	N/A
Age group				
Children (≤12 years)	22 (63)	N/A	N/A	N/A
Adolescents (13-17 years)	30 (86)	N/A	N/A	N/A
Young adults (18-25 years)	35 (100)	N/A	N/A	N/A
Adults (>25 years)	35 (100)	N/A	N/A	N/A
Number of app features (0-10)	N/A	6.00 (2.55)	6.00 (4)	8.00 (2.00-10.00)
App features				
Allows sharing	12 (34)	N/A	N/A	N/A
Has an app community	16 (46)	N/A	N/A	N/A
Allows password protection	14 (40)	N/A	N/A	N/A
Requires log-in	20 (57)	N/A	N/A	N/A
Sends reminders	28 (80)	N/A	N/A	N/A
Needs web access to function	35 (100)	N/A	N/A	N/A
Information/education	20 (57)	N/A	N/A	N/A
Monitoring/tracking	30 (86)	N/A	N/A	N/A
Goal setting	29 (83)	N/A	N/A	N/A
Advice/tips/strategies/skills training	27 (77)	N/A	N/A	N/A

[a]N/A: not applicable.

[b]RMB ¥1 = US $0.14.

App Quality

Table 2 presents the overall MARS scores and scores for each domain of the assessed apps. The interrater reliability for app quality assessment was high, with an intraclass correlation coefficient of 0.85 (95% CI 0.75-0.92) for the overall MARS score, 0.87 (95% CI 0.79-0.93) for engagement, 0.77 (95% CI 0.61-0.87) for functionality, 0.72 (95% CI 0.54-0.84) for aesthetics, 0.90 (95% CI 0.84-0.94) for information, 0.79 (95% CI 0.61-0.89) for subjective app quality, and 0.89 (95% CI 0.82-0.94) for app-specific quality. The internal consistency of the total MARS score and scores for each domain were considered excellent (Cronbach α>.70; Table 2).

The average total MARS score was 3.44 (out of 5), with scores ranging from 2.82 to 4.40, signifying a moderate overall quality. Functionality received the highest domain score, followed by aesthetics, engagement, and information. The average subjective quality score was 2.38, suggesting a low quality of the evaluated apps in terms of subjective assessments. The average app-specific quality score was 2.50, indicating a low quality concerning the potential to change awareness, knowledge, attitudes, intentions, help-seeking, and behavioral change (Table 2).

Table 3 presents the app quality scores and BCT scores separately for the 2 app platforms. The findings reveal that there were no significant differences in the scores for engagement (*P*=.93), functionality (*P*=.76), aesthetics (*P*=.77), information (*P*=.66), the overall MARS score (*P*=.85), subjective quality score (*P*=.88), app-specific quality score (*P*=.87), and BCT score between the 2 app platforms (*P*=.94).

Table 2. App quality and behavior change technique assessment scores.

Mobile Application Rating Scale domain	Mean (SD)	Median (range)	Cronbach α
Engagement (5 items)	3.26 (0.64)	3.20 (2.10-4.30)	.84
Functionality (4 items)	4.18 (0.37)	4.11 (3.70-4.90)	.74
Aesthetics (3 items)	3.43 (0.42)	3.42 (2.60-4.30)	.82
Information (7 items)	2.91 (0.52)	2.90 (2.10-4.30)	.83
App quality (overall mean)	3.44 (0.44)	3.34 (2.82-4.40)	.76
App subjective quality (4 items)	2.38 (0.75)	2.10 (1.30-4.30)	.82
App-specific quality (6 items)	2.50 (0.77)	2.20 (1.30-4.40)	.71
Number of behavior change techniques (26 items)	9.17 (4.07)	9.00 (2.00-18.0)	.87[a]

[a]The Kuder-Richardson index responses for all Mobile Application Rating Scale domains ranged from 1 to 5, with higher scores indicating a higher degree of app quality. The *P* value in the Kruskal-Wallis test between the 4 main domains of the Mobile Application Rating Scale was <.001.

Table 3. Scores for app quality and behavior change technique assessments, categorized by app platform.[a]

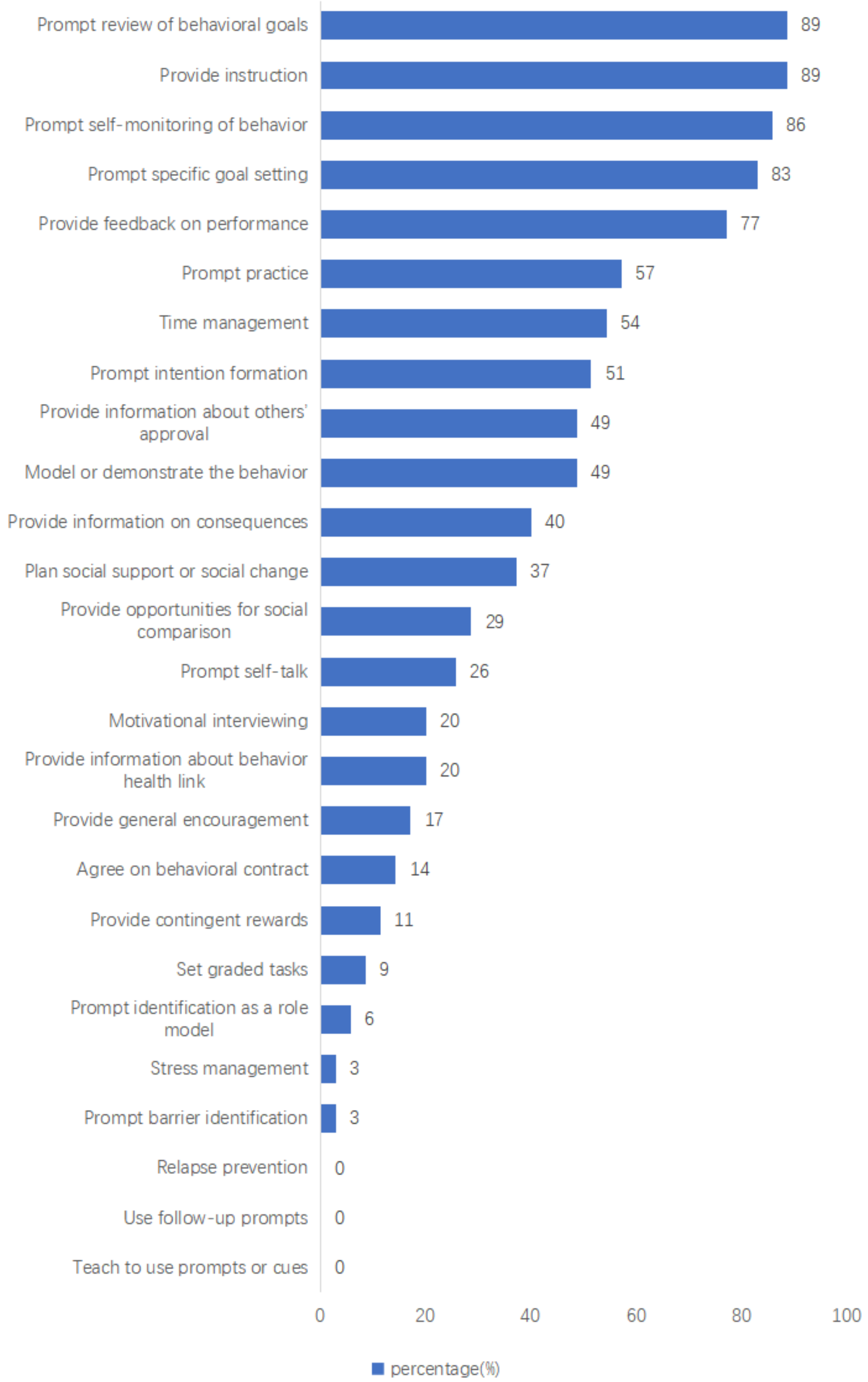
Mobile Application Rating Scale domain	App platform		*P* value
	Android (n=25), mean (SD)	Apple (n=27), mean (SD)	
Engagement	3.35 (0.89)	3.33 (0.88)	.93
Functionality	4.49 (0.31)	4.34 (0.49)	.76
Aesthetics	3.8 (0.42)	3.66 (0.58)	.77
Information	3.02 (0.83)	3.16 (0.63)	.66
App quality (overall mean)	3.67 (0.54)	3.61 (0.56)	.85
App subjective quality	2.71 (1.01)	2.65 (1.03)	.88
App-specific quality	2.73 (1.01)	2.83 (0.89)	.87
Number of behavior change techniques	9.24 (4.62)	9.85 (4.27)	.94

[a]All item responses in the Mobile Application Rating Scale ranged from 1 to 5, with higher scores signifying a higher level of app quality. Differences in scores and the number of behavior change techniques between platforms were evaluated using Mann-Whitney tests.

Presence of BCTs

The internal consistency in the evaluation of BCTs was likewise determined to be excellent, with the Kuder-Richardson value exceeding 0.70 (Table 2). The mean number of BCTs in the analyzed apps was 9.17 (range 2-18), indicating a moderate presence of BCTs in the evaluated apps. All apps contained a minimum of 2 BCTs, with 1 app (Boohee Health) having the highest recorded count of 18 BCTs. Figure 2 illustrates the proportion of apps that were evaluated to have the presence of BCTs. The most frequently observed BCTs were “prompt review of behavioral goals” and “provide instruction” (31/35, 89%), followed by “prompt self-monitoring of behavior” (30/35, 86%), “prompt-specific goal setting” (29/35, 83%), and “provide feedback on performance” (27/35, 77%). Three BCTs, namely, “relapse prevention,” “use follow-up prompts,” and “teach to use prompts or cues,” did not appear in any of the evaluated apps (Figure 2).

Figure 2. Presence of individual behavior change techniques in the analyzed apps..



Relationship Between App Quality and the Presence of BCTs/App Features

The number of BCTs exhibited a significant positive correlation with the MARS overall score (r=0.74; *P*<.01), MARS subjective quality score (r=0.56; *P*<.01), MARS-specific quality score (r=0.81; *P*<.01), MARS engagement score (r=0.84; *P*<.01), aesthetic quality score (r=0.55; *P*<.01), and information quality score (r=0.81; *P*<.01). There was no significant correlation between the functional score of the apps and the number of BCTs (r=0.30; *P*=.08). The number of app features displayed a strong and positive correlation with the objective quality of the app, subjective quality, and each domain. Furthermore, the number of app features was positively associated with the number of BCTs (Table 4).

Table 4. Correlations between app quality, number of app features, and behavior change techniques.

MARS[a] domain	Number of behavior change techniques	Number of app features
MARS engagement score	0.84[b]	0.77[b]
MARS functionality score	0.30	0.48[b]
MARS aesthetics score	0.55[b]	0.57[b]
MARS information quality score	0.81[b]	0.77[b]
MARS total score	0.74[b]	0.72[b]
App subjective quality	0.56[b]	0.49[b]
App-specific quality	0.81[b]	0.69[b]
Number of app features	0.69[b]	1

[a]MARS: Mobile App Rating Scale.

[b]*P*<.01.

Discussion

Principal Findings

After conducting a comprehensive search and screening of over 1000 apps, 35 were chosen for an in-depth evaluation in this study. This selection provided a snapshot of publicly available mobile apps for weight management in China. Through a comprehensive and systematic review, the study revealed that the overall quality of these apps was moderate, with the functionality domain achieving the highest score and the information domain attaining the lowest score based on the MARS evaluation. On average, each app included 6 features and 9.17 BCTs. The most common features were the requirement for web access to function, monitoring/tracking, goal setting, and sending reminders. The most common BCTs were prompting the review of behavioral goals, providing guidance, prompting self-monitoring of behavior, specific goal setting, and providing performance feedback. It was observed that apps with higher quality, as measured by the MARS, tended to include a greater number of features and BCTs. However, there were no significant differences in the quality or the presence of BCTs between different app platforms.

App Features

Of the 35 apps selected, 80% (28/35) provided only diet-related information, 60% (21/35) provided only PA-related information, and 46% (16/35) offered a combination of both types of information. The primary approach to weight loss is dietary calorie restriction [39], as PA alone tends to have only a moderate effect [40]. However, combining a healthy diet with PA is more beneficial for weight loss [41]. The Chinese Nutrition Society revised the 2022 edition of dietary guidelines for Chinese residents and also proposed the recommendation to “maintain a healthy weight by not overeating and being physically active every day.” As nutrition management and exercise are 2 major user demands for digital health management in China [11], and they are also essential and effective approaches for weight management, it is crucial that, in the future, when developing high-quality weight management apps, these 2 dimensions are integrated and emphasized. Furthermore, upon analyzing the app features, it was observed that options allowing for sharing and the presence of an app community were limited. Previous studies have indicated that engagement is crucial for enhancing the health of users, and apps that provide interactive features tend to score higher in the engagement quality domain [28,42].

Hence, in this study, the absence of interactive features such as app community sharing led to low engagement scores. Health-related social interactions are the primary trend among users of mobile health apps. Users’ main demands are tracking and recording their training results, enhancing enjoyment, sharing experiences, and boosting motivation. According to a previous study [11], a significant number of users (73.4%) were interested in health apps because they provide access to quality or professional users for communication [11]. With the onset of the 5G era, the social value in mobile health apps will be further enhanced and expanded. Hence, interactivity is an essential factor for the development of high-quality apps in the future.

App Quality

App quality was evaluated using the MARS scale, which assessed engagement, functionality, aesthetics, and information quality. We also conducted assessments for subjective quality and specific quality. The overall quality of the included apps was considered moderate, with an average MARS score of 3.44. The functionality domain received the highest score, followed by aesthetics and engagement, while the lowest-scoring domain was information quality. This finding aligns with previous studies that used MARS to evaluate the quality of health-related apps [27,43]. The high MARS functionality scores and the medium to high MARS aesthetics scores may be attributed to developers emphasizing a visually appealing and user-friendly app user experience [44]. This aligns with the concept that users typically prefer apps that are well-designed, engaging, user-friendly, and feature-rich. While these attributes attract users, they may not necessarily result in behavioral change [45].

The lowest score was observed in the information quality domain, indicating a general absence of high-quality, evidence-based content. The absence of evidence-based information could potentially expose end users to incorrect or misleading content, thereby increasing the risk of negative health outcomes among weight management users. Indeed, the credibility of the source of information and the scientific evidence presented to support the content play a significant role in establishing the trustworthiness of the app [46]. These findings indicate that developers of weight management apps should prioritize the improvement of information quality by offering clearly labeled and evidence-based scientific information to enhance the overall quality of the apps. Furthermore, government public health departments and network regulators should place a significant emphasis on ensuring the quality and safety of publicly available apps. They should also intensify their advocacy efforts aimed at promoting public health literacy and increasing awareness about the potential benefits and risks associated with the use of health apps. This approach will assist users in making informed decisions and selecting apps that are evidence based and promote their health [47].

Furthermore, the subjective quality of the analyzed apps was rated as low, with an average score of 2.38, which aligns with findings in existing studies [45,48]. One possible reason for this could be the absence of scientific evidence to substantiate the information content, leading to a low score in the information domain, which, in turn, diminishes the app's credibility [46]. The second reason could be the lack of an adequate information/education feature, with only 57% (20/35) of apps incorporating this feature. Insufficient information/education results in a knowledge gap among users regarding weight management, which hinders the enhancement of their awareness, motivation, and behavioral change. For users, understanding the reasons for engaging in weight management is crucial, as awareness of these benefits can serve as a motivating factor for them to adopt such practices. Our findings are consistent with the Information, Motivation, Behavioral Skills model of behavioral change [49]. According to this model, individuals require 3 components to initiate behavioral action: essential information (such as comprehensive knowledge related to weight management), motivation to pursue the goal (eg, willingness to embark on weight management), and behavioral skills required to make progress toward the goal (eg, self-guidance on managing nutrition and PA). The Information, Motivation, Behavioral Skills model illustrates that information can be transformed into actions that have the potential to motivate individuals and ultimately shape their attitudes and behaviors. Furthermore, "information" is regarded as the primary prerequisite for initiating behavior [49,50]. The absence of these features could diminish the user's subjective experience. The third reason may be the lack of interactive features such as an app community and sharing capabilities. These features allow users to document and share their experiences through photos and texts and exchange insights with other users, contributing to the enjoyment and motivation of using the app [11]. Thus, developers of weight management apps should regularly reassess the features of these apps to enhance both overall and subjective quality.

Behavior Change Techniques Included in Apps and App Features

Among the 26 BCT classifications developed by Abraham and Michie [35], there was a significant variation in the number of BCTs included in the analyzed apps, ranging from 2 to 18. In our study, each app contained 9.17 BCTs on average. This is higher than the findings of previous research, which reported an average of 4.2 [50,51] and 7.0 [52] BCTs for apps promoting PA in adults. The number of BCTs found in this review is slightly lower than that observed in a recent study that examined the number of BCTs in apps targeting PA, where an average of 11 BCTs per app was reported [53]. The inclusion of certain PA apps, which were developed with the involvement of health care professionals, may contribute to this difference in the number of BCTs. By contrast, the weight management apps examined in this study are commercially oriented. It is worth noting that apps developed with the involvement of health care professionals tend to have a stronger scientific foundation.

According to Michie et al [54], interventions aimed at improving diet and PA efficiency were associated with 5 specific BCTs, namely, prompt intention formation, prompt self-monitoring, providing feedback on performance, prompt review of behavioral goals, and prompt specific goal setting. In this study, the most commonly utilized BCTs were goal setting, reviewing behavioral goals, providing guidance, self-monitoring, and offering performance feedback, covering 4 out of the 5 valid BCTs. The 4 most commonly identified BCTs in apps targeting weight and PA in adults are goal setting, self-monitoring, providing instruction, and performance feedback [28,51]. From a psychological perspective, as specified by control theories, setting goals, monitoring behavior, receiving feedback, and reviewing relevant goals in light of feedback are central to self-management and behavioral control [12]. The apps analyzed in this study included a review of behavioral goals in addition to the 4 most effective BCTs. It is worth noting that interventions with a higher number of BCTs tend to be more effective [55]. However, it is essential to consider that incorporating too many BCTs can potentially reduce user engagement and app validity [43]. The optimal number and combination of BCTs for improving the effectiveness of weight management remain unclear. It is important to note that the impact of individual BCTs is often relatively small. To illustrate, according to control theory, setting goals and reviewing progress based on feedback are at the core of self-management, behavior control, and ultimately, behavioral change [12]. Participants can effectively self-monitor their health and make sustainable long-term changes [56]. As a result, combining techniques related to goal setting and goal review within an app may yield greater effectiveness. This insight has motivated developers to incorporate additional BCTs into future app designs and to experiment with different combinations of BCTs to determine which ones are most appealing and effective for specific populations.

Many Chinese apps offer a tracking feature (30/35, 86%). Users can record various parameters such as their daily diet, water intake, exercise, and weight. Additionally, these apps include reminders (28/35, 80%), which help users stay on track by sending prompts for meals, hydration, and exercise. They also set up goal achievement reminders to motivate users by celebrating each successful weight loss goal. Furthermore, these apps provide a range of functions, including advice, tips, strategies, skills training, goal setting, and information and education. These common features in apps support several BCTs, including goal setting, reviewing behavioral goals, providing guidance, self-monitoring, and performance feedback. Consequently, these BCTs were the most frequently found among the analyzed apps. An examination of PA and nutrition apps showed that approximately 55% of the analyzed apps included BCTs such as "plan social support" and "provide opportunities for social comparison" [57]. Based on our analysis, we found that only approximately 37% (13/35) of the apps included the BCT "plan social support," and 29% (10/35) of the apps included the BCT "provide opportunities for social comparison." Notably, social support is considered an important feature of web-based interventions [58]. As previously mentioned, the inclusion of sharing and app community features is less common among the apps we have examined. This reduced presence may result in decreased user interactivity and limited social support. User interaction enables individuals to share their weight loss progress and strategies within the app, facilitating feedback and suggestions from others. App users can join the community to connect with fellow users, exchange experiences and strategies, and receive valuable support and encouragement.

The fundamental attributes of mobile interventions encompass interactivity, adaptability, time sensitivity, and intraindividual dynamics [59]. Incorporating dynamic elements, such as the timing of information delivery, feedback, and reminders, and personalizing tasks and goals according to individual progress and capabilities align with persuasive techniques and are likely crucial components of an effective, targeted mobile intervention [59-63]. Hence, with the ever-expanding mobile app market, it is imperative to foster increased collaboration between app developers, nutritionists, health care professionals, and behavior change experts. This collaboration could further promote the integration of BCTs into apps, potentially unlocking new opportunities for health promotion.

Correlation

Our findings unveiled a connection between the quality of various domains within weight management apps and the number of BCTs as well as the quantity of app features. Notably, we observed a positive correlation between the overall app quality and the integration of both technical app features and BCTs. As the specific quality of a given weight management app, both subjectively and objectively, improves, there is a corresponding increase in the inclusion of BCTs and features within the app. Notably, the study revealed a positive correlation between the number of identified BCTs in the analyzed apps and the MARS score across all domains, with the exception of functionality. These findings align with previous studies conducted in various domains. For instance, a study focusing on diet, PA, and sedentary behavior in children and adolescents [43] revealed a positive association between the number of included BCTs and the total MARS score, MARS engagement score, and information quality score. Similarly, in the context of weight management [28], the number of incorporated techniques showed a positive correlation with the overall MARS score, engagement, and aesthetics. Furthermore, in the case of the Mediterranean diet [27], the number of BCTs was positively correlated with the app's information quality score, overall mean MARS score, subjective quality score, and app-specific quality score. Additionally, in our study, we observed a positive association between the number of features in the apps and app objective quality, subjective quality, and each of the domains. This aligns with findings from the China PA apps study [53]. We also found that the number of app features was positively linked to the number of BCTs, which is consistent with the study conducted on diet, PA, and sedentary behavior in children and adolescents [43].

Our research provides theoretical support for the notion that weight management apps with a greater number of functions and BCTs tend to exhibit higher objective quality. Moreover, it underscores the positive influence of these app features and BCTs on users' likelihood to recommend the app to others, frequency of use, willingness to pay, and their overall evaluation of the app's subjective quality. Additionally, the app's functions and BCTs enhance users' knowledge, awareness, and attitudes toward weight management. They also strengthen users' intentions for weight management, increase the likelihood of behavioral changes, and encourage users to seek further help for specific weight management needs. However, it is important to note that our study did not collect data on users' usage frequency; duration of engagement with weight management apps; willingness to pay; and users' knowledge, awareness, attitudes, intentions to enhance weight management, and actual behavioral changes resulting from app use. Therefore, caution must be exercised when interpreting our conclusions. This suggests that developers of weight management apps should consider incorporating more detailed, specific, and objective app usage statistics along with subjective user reviews in the future. These data can be invaluable in understanding how users engage with the app, their opinions, and the reasons behind their choices. Such insights can help developers address existing issues, tailor solutions to users' needs, and enhance overall app quality.

Strengths and Limitations

To ensure the representation and comprehensive coverage of the apps analyzed in this study, we selected 9 of the most popular Chinese mobile app markets from Qimai Data. China has the largest obese population in the world [4]. However, it is important to note that this study did not evaluate the quality of weight management apps. Instead, we used the validated MARS to assess the quality and features of apps with ratings higher than 4. Additionally, we identified BCTs using an established taxonomy.

This study has some limitations. First, given the rapid development of new apps and potential changes in the content of existing ones, the results presented can be seen as a snapshot of the current state of the included apps. Moreover, it is essential to acknowledge that the apps were assessed after relatively short-term use, which may have concealed certain characteristics, such as those BCTs requiring extended use (eg, follow-up prompts). Second, in alignment with prior research

[28], only apps with free content and those with a user score of ≥4 out of 5 were considered for this review. However, it is important to acknowledge that the exclusion of many apps from this review could be considered a limitation, as our findings may not be generalizable to apps that require an immediate paid subscription or apps with a user score of <4 out of 5. Third, it is essential to recognize that the MARS is subject to subjective influences. Factors such as apps requiring users to watch advertisements or the personal preferences of the rater, such as disliking PA while rating an app focused on PA for weight loss, could potentially have a negative impact on scores such as app subjective quality. Furthermore, it is important to note that in this study, weight management apps were assessed by professionals, and this evaluation may not entirely reflect the perspectives of real users of these apps. Therefore, additional research is warranted to obtain user evaluations of weight management apps. Fourth, it is worth noting that, as the Chinese version of the MARS is in the process of localization by other researchers, our study relied on the English version of the MARS for evaluation. This choice may have had an impact on the scoring results, even though the research team consisted of individuals with extensive experience in nutritional management, health behavioral change, and professional dietitians experienced in app development. Fifth, raters in this study used smartphones belonging to other raters for app evaluation, which could have resulted in some degree of unfamiliarity with the devices. Although efforts were made to acquaint the raters with the apps on these phones, there may still be some residual unfamiliarity.

Conclusions

Publicly available commercial Chinese weight management apps with a user rating of ≥4 exhibited a moderate overall quality. These apps ranked highest in functionality, had moderate scores for aesthetics and engagement, and scored lowest in information quality. The majority of the identified apps incorporated BCTs, with the most frequently used BCTs being goal setting, review of behavioral goals, providing guidance, self-monitoring, and performance feedback. Apps of higher quality typically included a greater number of technical app features and BCTs. There is a need for further efforts in the future to enhance these apps to engage users effectively. Efforts to boost user-app interaction and provide more comprehensive, evidence-based information within apps are crucial. Further research is necessary to identify the optimal number and combination of features and BCTs that can effectively influence behavioral change in weight management.

Acknowledgments

We were very grateful for the data provided by the Qimai Data Platform. This study was supported by the Major Foundation of Social Science and Humanity Anhui Province Department of Education (Grant No. SK2021ZD0030), Anhui Provincial Natural Science Foundation (Grant No. 2308085MH285), and School of Nursing, Anhui Medical University Graduate Student Seedling Cultivation Program (Grant No. hlqm12023005). The authors thank all the researchers who participated in this study. The authors also want to thank Stoyan R. Stoyanov, who provided helpful suggestions related to the use of the MARS and the strategies for ensuring the quality of the evaluation.

Authors' Contributions

LG, GJ, and MZ were responsible for research conception and design. LG and GJ performed initial searching and screening. LG, GJ, LY, YX, MZ, and TZ collected data on app ratings and performed data extraction. XQ and WH analyzed the data. LG, GJ, ZC, and MZ wrote the manuscript. LG and GJ are co-first authors.

Conflicts of Interest

None declared.

References

1. Malik VS, Willet WC, Hu FB. Nearly a decade on - trends, risk factors and policy implications in global obesity. Nat Rev Endocrinol 2020 Nov;16(11):615-616 [FREE Full text] [doi: 10.1038/s41574-020-00411-y] [Medline: 32873971]
2. Munir M, Zakaria ZA, Nisar H, Ahmed Z, Korma SA, Esatbeyoglu T. Global human obesity and global social index: relationship and clustering. Front Nutr 2023;10:1150403 [FREE Full text] [doi: 10.3389/fnut.2023.1150403] [Medline: 37063335]
3. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
4. Wang Y, Zhao L, Gao L, Pan A, Xue H. Health policy and public health implications of obesity in China. The Lancet Diabetes & Endocrinology 2021 Jul;9(7):446-461. [doi: 10.1016/s2213-8587(21)00118-2]
5. Sun X, Yan AF, Shi Z, Zhao B, Yan N, Li K, et al. Health consequences of obesity and projected future obesity health burden in China. Obesity (Silver Spring) 2022 Sep 24;30(9):1724-1751. [doi: 10.1002/oby.23472] [Medline: 36000246]
6. Okati-Aliabad H, Ansari-Moghaddam A, Kargar S, Jabbari N. Prevalence of obesity and overweight among adults in the Middle East countries from 2000 to 2020: a systematic review and meta-analysis. J Obes 2022;2022:8074837 [FREE Full text] [doi: 10.1155/2022/8074837] [Medline: 35154826]

7. Stival C, Lugo A, Odone A, van den Brandt PA, Fernandez E, Tigova O, TackSHS Project Investigators. Prevalence and correlates of overweight and obesity in 12 European countries in 2017-2018. Obes Facts 2022;15(5):655-665 [FREE Full text] [doi: 10.1159/000525792] [Medline: 35917801]
8. Kopelman PG. Obesity as a medical problem. Nature 2000 Apr 06;404(6778):635-643. [doi: 10.1038/35007508] [Medline: 10766250]
9. The 47th China statistical report on internet development. Office of the Central Cyberspace Affairs Commission. URL: http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm# [accessed 2023-05-13]
10. The number of mobile medical users in China reached 661 million by 2020. Central People's Government of the People's Republic of China. 2023. URL: http://www.gov.cn/xinwen/2021-04/21/content_5600947.htm [accessed 2023-05-13]
11. Research Report on the Development of China's Digital Health Management Industry in 2021. iiMedia. URL: https://www.iimedia.cn/c400/81977.html [accessed 2023-05-13]
12. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
13. Kuehnhausen M, Frost V. Trusting smartphone Apps? To install or not to install, that is the question. New York, NY: IEEE; 2013 Presented at: 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); February 25-28, 2013; San Diego, CA p. 30-37. [doi: 10.1109/cogsima.2013.6523820]
14. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
15. Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, et al. Promoting health via mHealth applications using a French version of the Mobile App Rating Scale: adaptation and validation study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480 [FREE Full text] [doi: 10.2196/30480] [Medline: 34463623]
16. Barzegari S, Sharifi Kia A, Bardus M, Stoyanov SR, GhaziSaeedi M, Rafizadeh M. The Persian version of the Mobile Application Rating Scale (MARS-Fa): translation and validation study. JMIR Form Res 2022 Dec 05;6(12):e42225 [FREE Full text] [doi: 10.2196/42225] [Medline: 36469402]
17. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 07;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
18. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
19. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):16 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
20. Raeesi A, Khajouei R, Ahmadian L. Evaluating and rating HIV/AIDS mobile apps using the feature-based application rating method and Mobile App Rating Scale. BMC Med Inform Decis Mak 2022 Oct 30;22(1):281 [FREE Full text] [doi: 10.1186/s12911-022-02029-8] [Medline: 36310157]
21. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, Kabuth B, Baltenneck N, Nusbaum F, et al. Mental health mobile apps in the French app store: assessment study of functionality and quality. JMIR Mhealth Uhealth 2022 Oct 12;10(10):e41282 [FREE Full text] [doi: 10.2196/41282] [Medline: 36223178]
22. Carrouel F, Bourgeois D, Clément C, Tardivo D, Martinon P, Guiral S, et al. Oral-hygiene-related mobile apps in the French app stores: assessment of functionality and quality. Int J Environ Res Public Health 2022 Jun 14;19(12):7293 [FREE Full text] [doi: 10.3390/ijerph19127293] [Medline: 35742541]
23. Martinon P, Saliasi I, Bourgeois D, Smentek C, Dussart C, Fraticelli L, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR Mhealth Uhealth 2022 Mar 14;10(3):e35879 [FREE Full text] [doi: 10.2196/35879] [Medline: 35285817]
24. Sharif MO, Alkadhimi A. Patient focused oral hygiene apps: an assessment of quality (using MARS) and knowledge content. Br Dent J 2019 Sep 13;227(5):383-386. [doi: 10.1038/s41415-019-0665-0] [Medline: 31520041]
25. Parker K, Bharmal RV, Sharif MO. The availability and characteristics of patient-focused oral hygiene apps. Br Dent J 2019 Apr;226(8):600-604. [doi: 10.1038/s41415-019-0197-7] [Medline: 31028329]
26. Kanoute A, Carrouel F, Gare J, Dieng SN, Dieng A, Diop M, et al. Evaluation of oral hygiene-related mobile apps for children in Sub-Saharan Africa. Int J Environ Res Public Health 2022 Oct 01;19(19):12565 [FREE Full text] [doi: 10.3390/ijerph191912565] [Medline: 36231862]
27. McAleese D, Linardakis M, Papadaki A. Quality and presence of behaviour change techniques in mobile apps for the Mediterranean diet: a content analysis of Android Google Play and Apple App Store apps. Nutrients 2022 Mar 18;14(6):1290 [FREE Full text] [doi: 10.3390/nu14061290] [Medline: 35334947]

28. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
29. Alshathri DM, Alhumaimeedy AS, Al-Hudhud G, Alsaleh A, Al-Musharaf S, Aljuraiban GS. Weight management apps in Saudi Arabia: evaluation of features and quality. JMIR Mhealth Uhealth 2020 Oct 26;8(10):e19844 [FREE Full text] [doi: 10.2196/19844] [Medline: 33104013]
30. Brammall BR, Garad RM, Boyle JA, Hayman MJ, de Jersey SJ, Teede HJ, et al. Assessing the content and quality of digital tools for managing gestational weight gain: systematic search and evaluation. J Med Internet Res 2022 Nov 25;24(11):e37552 [FREE Full text] [doi: 10.2196/37552] [Medline: 36427237]
31. Michie S, Abraham C. Reviews and product details of Qimai Data. SOQIFU. URL: https://www.soqifu.com/product/qimaishuju/ [accessed 2023-05-14]
32. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
33. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
34. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630 [FREE Full text] [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
35. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
36. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
37. Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e18495 [FREE Full text] [doi: 10.2196/18495] [Medline: 32706727]
38. Dawson LP, Biswas S, Lefkovits J, Stub D, Burchill L, Evans SM, et al. Characteristics and Quality of National Cardiac Registries: A Systematic Review. Circ Cardiovasc Qual Outcomes 2021 Sep;14(9):e007963 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.121.007963] [Medline: 34517724]
39. Kim S, Leng XI, Kritchevsky SB. Body composition and physical function in older adults with various comorbidities. Innov Aging 2017 Mar 01;1(1):igx008 [FREE Full text] [doi: 10.1093/geroni/igx008] [Medline: 30480107]
40. McNeil J, Liepert M, Brenner DR, Courneya KS, Friedenreich CM. Behavioral predictors of weight regain in postmenopausal women: exploratory results from the breast cancer and exercise trial in Alberta. Obesity (Silver Spring) 2019 Sep;27(9):1451-1463 [FREE Full text] [doi: 10.1002/oby.22569] [Medline: 31318492]
41. Wu T, Gao X, Chen M, van Dam RM. Long-term effectiveness of diet-plus-exercise interventions vs. diet-only interventions for weight loss: a meta-analysis. Obes Rev 2009 May;10(3):313-323. [doi: 10.1111/j.1467-789X.2008.00547.x] [Medline: 19175510]
42. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
43. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
44. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
45. Hayman M, Alfrey K, Cannon S, Alley S, Rebar AL, Williams S, et al. Quality, features, and presence of behavior change techniques in mobile apps designed to improve physical activity in pregnant women: systematic search and content analysis. JMIR Mhealth Uhealth 2021 Apr 07;9(4):e23649 [FREE Full text] [doi: 10.2196/23649] [Medline: 33825693]
46. Quintana Y, Feightner JW, Wathen CN, Sangster LM, Marshall JN. Preventive health information on the internet. Qualitative study of consumers' perspectives. Can Fam Physician 2001 Sep;47:1759-1765 [FREE Full text] [Medline: 11570301]
47. Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e14836 [FREE Full text] [doi: 10.2196/14836] [Medline: 32255432]
48. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jun 09;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]

49. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
50. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
51. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
52. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
53. Wang Y, Wang Y, Greene B, Sun L. An analysis and evaluation of quality and behavioral change techniques among physical activity apps in China. Int J Med Inform 2020 Jan;133:104029. [doi: 10.1016/j.ijmedinf.2019.104029] [Medline: 31759245]
54. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
55. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
56. Wang J, Hsieh R, Tung Y, Chen Y, Yang C, Chen YC. Evaluation of a technological image-based dietary assessment tool for children during pubertal growth: a pilot study. Nutrients 2019 Oct 20;11(10):2527 [FREE Full text] [doi: 10.3390/nu11102527] [Medline: 31635141]
57. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
58. Wang Z, Wang J, Maercker A. Program use and outcome change in a web-based trauma intervention: individual and social factors. J Med Internet Res 2016 Sep 09;18(9):e243 [FREE Full text] [doi: 10.2196/jmir.5839] [Medline: 27612932]
59. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 24;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
60. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
61. Fogg BJ. The Behavior Grid: 35 Ways Behavior Can Change. In: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology. New York, NY: ACM; 2009 Apr Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 1-5. [doi: 10.1145/1541948.1542001]
62. Fogg B. A behavior model for persuasive design. In: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology. New York, NY: ACM; 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 1-7. [doi: 10.1145/1541948.1541999]
63. Wahl DR, Villinger K, König LM, Ziesemer K, Schupp HT, Renner B. Healthy food choices are happy food choices: evidence from a real life sample using smartphone based assessments. Sci Rep 2017 Dec 06;7(1):17069 [FREE Full text] [doi: 10.1038/s41598-017-17262-9] [Medline: 29213109]

Abbreviations

BCT: behavior change technique
KR-20: Kuder-Richardson index 20
MARS: Mobile Application Rating Scale
mHealth: mobile health
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 23.06.23; peer-reviewed by A Carrandi, F Carrouel; comments to author 31.07.23; revised version received 24.09.23; accepted 06.10.23; published 24.11.23.

Please cite as:
Geng L, Jiang G, Yu L, Xu Y, Huang W, Chen Z, Qi X, Zhang T, Zhao M
The Most Popular Commercial Weight Management Apps in the Chinese App Store: Analysis of Quality, Features, and Behavior Change Techniques
JMIR Mhealth Uhealth 2023;11:e50226
URL: https://mhealth.jmir.org/2023/1/e50226
doi:10.2196/50226
PMID:37999950

Original Paper

Usage and Daily Attrition of a Smartphone-Based Health Behavior Intervention: Randomized Controlled Trial

Erlendur Egilsson[1], CandPsych; Ragnar Bjarnason[2,3], MD, PhD; Urdur Njardvik[1], PhD

[1]Department of Psychology, University of Iceland, Reykjavik, Iceland

[2]Department of Pediatrics, University of Iceland, Reykjavik, Iceland

[3]Faculty of Medicine, University of Iceland, Reykjavik, Iceland

Corresponding Author:
Erlendur Egilsson, CandPsych
Department of Psychology
University of Iceland
Saemundargata 12
Reykjavik, 102
Iceland
Phone: 354 5254240
Fax: 354 5255802
Email: erlendu@hi.is

Abstract

Background: Although most adolescents have access to smartphones, few of them use mobile health (mHealth) apps for health improvement, highlighting the apparent lack of interest in mHealth apps among adolescents. Adolescent mHealth interventions have been burdened with high attrition rates. Research on these interventions among adolescents has frequently lacked detailed time-related attrition data alongside analysis of attrition reasons through usage.

Objective: The objective was to obtain daily attrition rates among adolescents in an mHealth intervention to gain a deeper understanding of attrition patterns, including the role of motivational support, such as altruistic rewards, through analysis of app usage data.

Methods: A randomized controlled trial was conducted with 304 adolescent participants (152 boys and 152 girls) aged 13-15 years. Based on 3 participating schools, participants were randomly assigned to control, treatment as usual (TAU), and intervention groups. Measures were obtained at baseline, continuously throughout the 42-day trial period (research groups), and at the trial end. The mHealth app is called SidekickHealth and is a social health game with the following 3 main categories: nutrition, mental health, and physical health. Primary measures were attrition based on time from launch, and the type, frequency, and time of health behavior exercise usage. Outcome differences were obtained through comparison tests, while regression models and survival analyses were used for attrition measures.

Results: Attrition differed significantly between the intervention and TAU groups (44.4% vs 94.3%; χ^2_1=61.220; *P*<.001). The mean usage duration was 6.286 days in the TAU group and 24.975 days in the intervention group. In the intervention group, male participants were active significantly longer than female participants (29.155 vs 20.433 days; χ^2_1=6.574; *P*<.001). Participants in the intervention group completed a larger number of health exercises in all trial weeks, and a significant decrease in usage was observed from the first to second week in the TAU group (t_{105}=9.208; *P*<.001) but not in the intervention group. There was a significant increase in health exercises in the intervention group from the fifth to sixth week (t_{105}=3.446; *P*<.001). Such a significant increase in usage was not evident in the TAU group. The research group was significantly related to attrition time (hazard ratio 0.308, 95% CI 0.222-0.420), as well as the numbers of mental health exercises (*P*<.001) and nutrition exercises (*P*<.001).

Conclusions: Differences in attrition rates and usage between groups of adolescents were identified. Motivational support is a significant factor for lowering attrition in adolescent mHealth interventions. The results point to sensitivity periods in the completion of diverse health tasks, and emphasis on time-specific attrition, along with the type, frequency, and time of health behavior exercise usage, is likely a fruitful avenue for further research on mHealth interventions for adolescent populations, in which attrition rates remain excessive.

Trial Registration: ClinicalTrials.gov NCT05912439; https://clinicaltrials.gov/study/NCT05912439

(JMIR Mhealth Uhealth 2023;11:e45414) doi:10.2196/45414

KEYWORDS

mHealth intervention; mobile health; adolescent; attrition; mental health; physical activity

Introduction

Throughout the past decade, ownership of and access to smartphones and mobile devices have grown profoundly among adolescents and youth worldwide [1,2]. The growth has been such that smartphone ownership or access among US adolescents was 95% 4 years ago and had increased by 23% in the 4 years prior [1,2]. A similar development was observed in the majority of developed economies where adolescent smartphone access and ownership is above the 90th percentile [2]. Smartphones are so widely distributed and used that approximately 45% of adolescents spend nearly all waking hours online [3]. However, modest projections of daily usage indicate that many spend way less time online each day, though it is usually more than 4 hours [4-6].

Widespread smartphone usage in adolescent and youth populations has been extensively covered, but a more positive side to mobile usage is that a significant proportion of adolescents seek health information and clinical help online through their mobile devices, providing ample opportunities to reach at-risk adolescents with science-based methods focusing on health improvement [7-9]. Health problems (ie, mental health and lifestyle diseases) disproportionally burden lower socioeconomic status groups as well as diverse minority groups, and smartphones could become a vital tool for eliminating such disparities since smartphone access and ownership are not related to socioeconomic status, gender, or race in diverse economies [2,10,11]. The mobile health (mHealth) market is steadily becoming saturated with apps, and the yearly increase in the number of apps available has skyrocketed in recent years, with an estimated 350,000 mHealth apps currently on the market [12]. However, only 8% of adolescents seem to use health apps to improve their health, highlighting the apparent gap between easy access, extensive daily usage, and lack of interest in mHealth apps among adolescents [13].

Lack of physical activity has been labeled a global pandemic and has been reported as the fourth leading global cause of death [14]. Physical inactivity increases the risk of lifestyle diseases, such as heart disease, type 2 diabetes, and cancer, resulting in over 5 million annual global deaths [15]. Further, the estimated annual financial burden of physical inactivity is nearly USD 54 billion in health care costs around the world [16]. There seems to be a drop in physical activity in adolescence, and a large number of adolescents are under the recommended physical activity levels provided by the World Health Organization (WHO) [17-19]. Lack of sufficient physical activity tends to continue into adulthood, and research suggests that the majority of adolescents in the European Union do not even reach 30% of the recommended daily physical activity [19-21]. Further, adolescents seem to have the unhealthiest diet of all age groups, and they are particularly susceptible to weight gain [22]. Research has repeatedly revealed a significant relationship between nutritional behavior and physical activity in terms of weight management [23]. A tremendous increase in global adolescent obesity has been witnessed in the past decades, and the prevalence, for instance, has tripled since 1975 [24]. Cost-effective interventions to increase physical activity and improve nutritional behaviors in adolescent populations are therefore urgently needed.

Physical inactivity and inadequate nutritional habits are often interrelated to disabling emotional problems, and integrated strategies should include all 3 pillars to improve physical as well as mental well-being in adolescent populations. mHealth interventions targeting disabling emotional problems in adolescent populations have revealed encouraging outcomes, despite the fact that attrition rates in these interventions are generally high [11,25-30]. Varying definitions of attrition have complicated research on this topic, but attrition is defined as leaving treatment before obtaining a required level of improvement or completing intervention goals [31-33]. Research on mental mHealth interventions among adolescents has frequently lacked detailed time-related attrition data alongside accurate definitions and analysis of attrition reasons, though recent studies show promise in that regard [11,30,34]. Attrition is regularly reported at 2 distinct points of time, that is, intervention start and intervention end. A continuous measure of usage versus nonusage in mHealth interventions for adolescents while simultaneously obtaining detailed usage data to prevent or delay exact times of attrition in future interventions, would perhaps be an improved representation of attrition [35].

Increased knowledge on the actual attrition factors and patterns associated with mHealth interventions in adolescent populations is urgently needed. Obtaining a better understanding of how motivational support motivates adolescents to use mHealth apps and why adolescents maintain or lose interest in using these apps to improve their health is of vital importance. Motivational support in mHealth interventions, defined as strategies to enhance motivation and counter attrition to overcome behavior change barriers, often includes goal-setting, feedback, social support, and rewards [36,37]. Systematic reviews examining possible drivers behind usage point to group and task customization, localization, functional user support, gamification of health tasks, and immediate visual but simplified feedback on user action, while gender-related motivational support features could be contributing factors [36-38]. The timing of tailored motivational support, through just-in-time adaptive interventions, should be considered as well when implementing adolescent mHealth interventions, since time-based individualization could counter high attrition rates [35,39]. Given the magnitude of reported health problems among adolescents and lack of cost-effective health behavior interventions specifically developed for adolescent populations, the need for a better understanding of attrition reasons in adolescent mHealth interventions is large. The study aimed to (1) seek a richer understanding of continuous attrition rates for an mHealth intervention in an adolescent population and the effects motivational support has on attrition rates, and (2) examine the effectiveness of the intervention with the aim to increase daily mental, nutritional, and physical health behaviors.

Methods

Participants

The study included 304 individuals (152 girls and 152 boys) aged 13 to 15 years attending 1 of 3 public schools for children and adolescents in the greater capital area of Iceland. The mean age at baseline measurement was 13.70 (SD 0.83) years. All children attending the highest 3 classes (8th to 10th classes) in the 3 participating public elementary schools in Iceland were eligible to participate (n=661; male-to-female ratio of 313:348). All children in public schools in the municipality are equipped with an iPad from 10 years of age. The exclusion criterion was the diagnosis of a severe disorder of intellectual development or a physical, developmental, or mental illness significantly restricting the ability to use mobile apps. No participant was excluded from the study based on this exclusion criterion. Research specifications and an introduction to the app were sent via email to the parents and legal caretakers of all eligible participants through school officials, along with a confirmative survey link. If the link was answered, it provided confirmation for informed consent. Adolescents with informed consent from parents or legal caretakers were invited to take part in the study through a confirmative survey link.

Ethics Approval

The study was approved by the National Bioethics Committee of Iceland (license number: VSNb2015060065/03-01).

Measurements

The amount, time, and frequency of daily health activities measured through completion of in-app exercises, quality of sleep and energy levels, self-reported stress levels, and gratitude levels were primary outcome measures. The Cronbach α for the current sample was .920 for all self-reported health tasks within the app.

Anxiety and depressive symptoms were assessed using the Revised Children's Anxiety and Depression Scale (RCADS), a self-report assessment tool for children and youth. The scale involves a 4-point Likert scale, spans 47 questions, and is divided into 6 subscales (separation anxiety symptoms, general anxiety symptoms, obsessive-compulsion symptoms, social anxiety symptoms, panic symptoms, and depression symptoms). A *T*-score over 65 marks the clinical cutoff point. The inventory's psychometrics have been studied with acceptable findings in both US and Icelandic pediatric populations [40,41]. The Cronbach α for the current sample was .958.

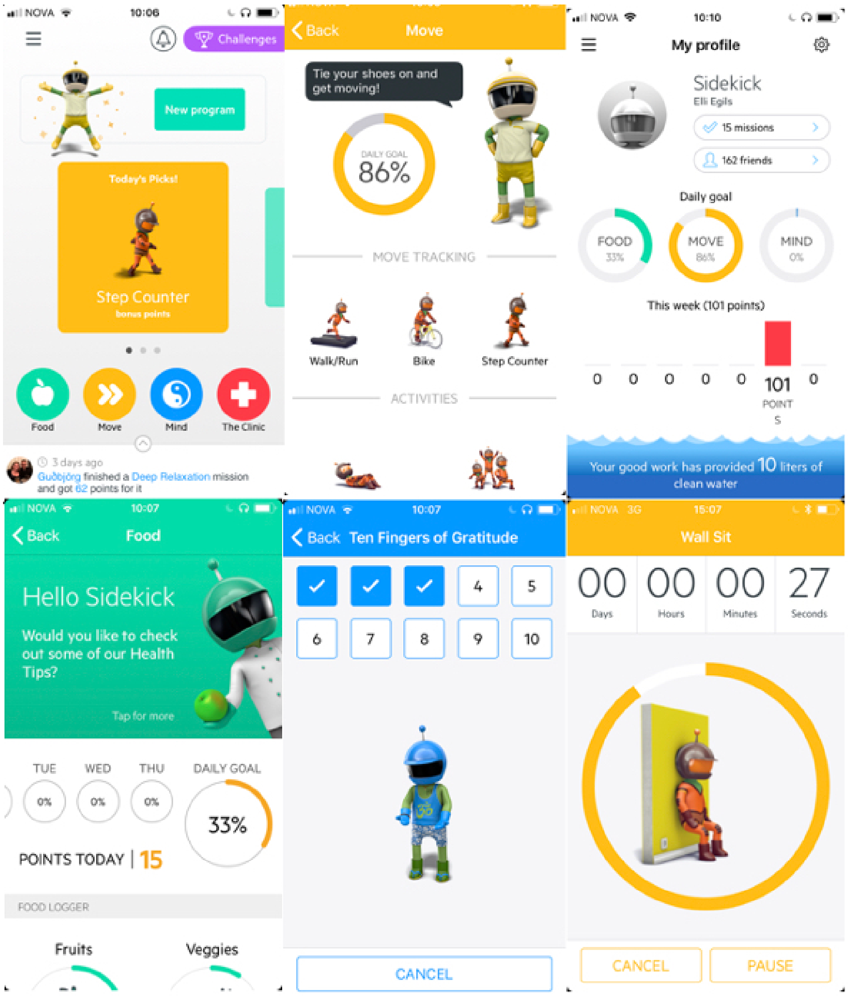
The General Self-Efficacy Scale (GSE), a 10-item self-report questionnaire with the total score ranging from 10 to 40, was used to measure self-efficacy levels, with a higher score indicating higher self-efficacy [42]. Acceptable psychometric properties for the questionnaire have been obtained, and it has been used globally in youth populations [43]. The Cronbach α for the current sample was .937.

The *BEARS* sleep screening algorithm was used to evaluate participants' sleep problems. It is a sleep screening instrument for children from 2 to 18 years old, and is divided into 5 sleep domains (bedtime problems, excessive daytime sleepiness, awakenings during night, regularity and duration of sleep, and snoring) [44]. The algorithm's psychometrics have been studied with acceptable findings in pediatric populations [45]. The Cronbach α for the current sample was .769.

mHealth App

The app is called SidekickHealth and has been described in the research group's previous work [35]. SidekickHealth was initially developed through multiple focus group studies among both Icelandic elementary school students and adolescents in the obesity clinic at the Landspitali University Hospital in Iceland to incorporate the target groups' needs and opinions. Based on results from focus group studies and design advisors, the app took the form of a social health game (Figure 1). Functionality of the app evolves around motivational support to help the user set goals and complete health tasks (gamification of tasks) in the following 3 main categories: food and drink (eg, vegetable and water intake, consumption of fruits, and avoiding sugary soda or energy drinks), physical activity (eg, body weight exercises, logged minutes of sports activity, GPS-based biking, walking, and running), and mental health exercises (eg, reducing stress, exercising gratitude, and improving sleep habits). By completing health tasks that are labeled missions and participating in friendly competitions with peers, users earn points (called "kicks") and badges providing altruistic rewards (eg, liters of water for children in need or polio vaccines that are sent in their name to children in need through UNICEF). A visual representation of the user's performance is provided in different categories. Keeping the app fun, entertaining, and easy to use is of integral importance and was a strong focus point throughout the developmental phases (Multimedia Appendix 1). The smartphone app operates on the Android and iOS platforms. The app's function focuses on education and simple health behavior changes through the benefits of increased physical activity and mental health exercises, as well as a healthy diet, portion sizing, and appetite awareness training. Appetite awareness training is a behavior modification tool that has, for instance, been used in obesity treatment and encourages overweight/obese children and youth to consume food and drink in response to internal appetite cues. It has shown promise for the treatment of overweight and obese children and teenagers, and has been visually developed as an individual mission in the app's nutrition category [46,47]. Participants in the intervention arm were randomly assigned to groups consisting of 8 individuals that collectively and individually competed in point collection through completion of in-app health tasks. In the beginning of each of the trial's 6 weeks, the intervention group received in-app messages where a new weekly competition (both individual and group levels) with altruistic rewards was introduced. In weeks 2 to 6, altruistic rewards for the past week's efforts were also handed out. Winners of competitions received confirmation that UNICEF had sent polio vaccines to children in need. Further, through completion of in-app health exercises, participants collected liters of water that were sent in their name to children in need through UNICEF. The total cost for the altruistic rewards, paid for by the first author, throughout the treatment period was roughly US $68 (56 US cents per participant).

Figure 1. Overview of app functions and categories.



Procedure

This study was a randomized controlled study. Group randomization was used to divide the 3 participating schools into control, treatment as usual (TAU), and intervention groups. Measures were obtained at baseline and 42 days later. Participants in both the TAU and intervention groups received an approximately 10-minute-long introduction regarding the study specifications and the app. The control group received no further contact, access to the app, or information until study-end questionnaire measures. Participants in the intervention group were randomly assigned to teams consisting of 8 individuals that collectively and individually competed in point collection through completion of in-app health tasks. Participation in the TAU and intervention groups was defined as downloading the SidekickHealth app and completing at least 3 health exercises within it. Exercise time was defined as the timestamp on completion of the exercise within any of the 3 types of exercise categories (physical activity, nutrition, and mental health) in the app. Exercise frequency referred to how often a given exercise was completed by a participant. Attrition time was defined as the timestamp of the last completed health exercise within the SidekickHealth app throughout the intervention period. The procedural difference between the TAU and intervention groups is related to motivational support. The intervention group received motivational support in the form of weekly individual and group feedback on usage, participation in friendly health task competitions, and weekly altruistic rewards for usage. Participants in the TAU group used the app individually throughout the trial period without any motivational support. A flowchart of participation is displayed in Multimedia Appendix 2.

Statistical Analysis

The descriptive characteristics of participants along with attrition reasons are reported. Pearson correlation coefficients, independent samples *t*-tests, repeated measures ANOVA with adjusted alpha levels, and χ^2 tests were used to measure mean differences in primary and secondary outcome measures from baseline to the trial end within and between research groups. Kaplan-Meier survival analysis plots and log-rank tests were used to assess the attrition time and possible significant differences between and within research groups [48,49]. The trial start was defined as the time of the first in-app health exercise completion, and the trial period was 6 weeks (42 days) from that moment. Attrition, or the event, was defined as the time of the participant's last completed health exercise in the SidekickHealth app. Participant cases were evaluated as censored when the app was still being used 42 days after the study start. Cox proportional hazard regression models with interacting covariables using research groups as clusters were used to examine attrition prediction based on usage of in-app health exercises for the time, type, and frequency of exercises, as well as sociodemographic variables (age and gender) [50]. Significance was defined as a *P* value <.05. Data were analyzed using IBM SPSS Statistics, Release Version 29 (IBM Corp).

Results

Among all invited participants with parental or caretaker consent to participate (N=451), 304 (67.41%) individuals took part in the study. Participants who did not answer questionnaires at the study end were excluded. Participant characteristics are presented in Table 1. Logged data revealed broad differences in app usage among participants as shown in Figure 2. There was a significant difference in the mean number of health exercises completed by participants, where individuals in the intervention group (mean 120.869, SD 32.434) completed on average roughly 6 times as many exercises as individuals in the TAU group (mean 18.341, SD 31.802) over the study period (t_{221}=−3.00; *P*<.001). When logged health exercises on the first day of the study period were examined, the results showed that the difference between the intervention group (mean 16.835, SD 21.820) and TAU group (mean 8.100, SD 7.237) was less extensive but still significant (t_{221}=−2.12; *P*=.04).

Forms of attrition over the 42-day study period are shown in Figure 3. Significant differences in completion rates were evident in log-rank tests between the intervention group and TAU group (χ^2_1=61.220; *P*<.001). Among participants in the TAU group, the mean survival time was 6.286 days (95% CI 4.304-8.277), while among participants in the intervention group, the mean survival time was 24.975 days (95% CI 21.452-28.518). Log-rank tests revealed significant differences in completion rates between male and female participants in the intervention group (χ^2_1=6.574; *P*<.001). In the intervention group, the mean survival time among male participants was 29.155 days (95% CI 24.519-33.812) and among female participants was 20.433 days (95% CI 15.301-25.558). Such differences were not evident in log-rank tests in the TAU group (χ^2_1=1.570; *P*=.21). Figure 4 presents Kaplan-Meier plots for gender-based attrition in the TAU and intervention groups.

There was a significant difference between groups in the average number of in-app health exercises in all weeks (Table 2). There was a significant mean drop (mean 12.347, SD 13.803) in usage in the TAU group from the first week to the second week (t_{105}=9.208; *P*<.001). Even though there was a drop in usage in the intervention group between the first and second weeks of the trial (mean 6.798, SD 37.481), the difference was not significant (t_{105}=1.959; *P*=.06). There was however a significant increase (mean 22.904, SD 71.721) in the average individual in-app health exercises completed by the intervention group from the fifth week to the sixth and last week of the trial (t_{105}=3.446; *P*<.001). Such a significant increase in usage was not evident in the TAU group.

No significant gender differences were found in the average weekly in-app exercise frequency within research groups (Table 3).

When exercise time was compared with the exercise category, the results revealed significant differences within both the intervention group (χ^2_6=2162.559; *P*<.001) and the TAU group (χ^2_6=69.372; *P*<.001). Differences in usage based on exercise time and exercise type are presented in Table 4 and Figure 5.

Results from the Cox proportional hazard regression models are shown in Multimedia Appendix 3. The research group that participants were assigned to (hazard ratio 0.308, 95% CI 0.222-0.420) was significantly related to attrition (*P*<.001), as well as the numbers of mental health exercises (*P*<.001) and nutrition exercises (*P*<.001) completed in the app by participants in both research groups. Participants in the TAU group (hazard ratio 0.387, 95% CI 0.201-0.748) who completed an in-app health exercise between day 2 and day 6 of the trial were found to be significantly more likely to finish (*P*=.03). Such significant differences were not found in the intervention group. Further, the numbers of health exercises completed in the app in the first week (*P*<.001), second week (*P*<.001), and last week (*P*<.001) of the trial were significantly related to survival rates in the TAU group. Similar significant differences were not evident in the intervention group. All types of health exercises completed in the app were also significantly related to attrition in the TAU group, although such differences were not found among intervention group members. Anxiety, depression, and self-efficacy measures between research groups are shown in Multimedia Appendix 4.

Table 1. Baseline participant characteristics.

Characteristic	Control group (n=81)	TAU[a] group (n=106)	Intervention group (n=117)
Age (years), mean (SD)	13.72 (0.45)	13.14 (0.51)	13.50 (0.63)
Male:female ratio	34:47	57:49	61:56
Disabling sleep problems, n (%)	32 (39.5)	23 (21.7)	38 (32.5)
Clinical anxiety symptoms, n (%)	18 (22.2)	10 (9.4)	14 (12.0)
Clinical depression symptoms, n (%)	12 (14.8)	7 (6.6)	7 (6.0)
General self-efficacy score, mean (SD)	17.90 (5.89)	16.17 (5.71)	18.14 (5.04)

[a]TAU: treatment as usual.

Figure 2. Mean weekly in-app exercises. TAU: treatment as usual.

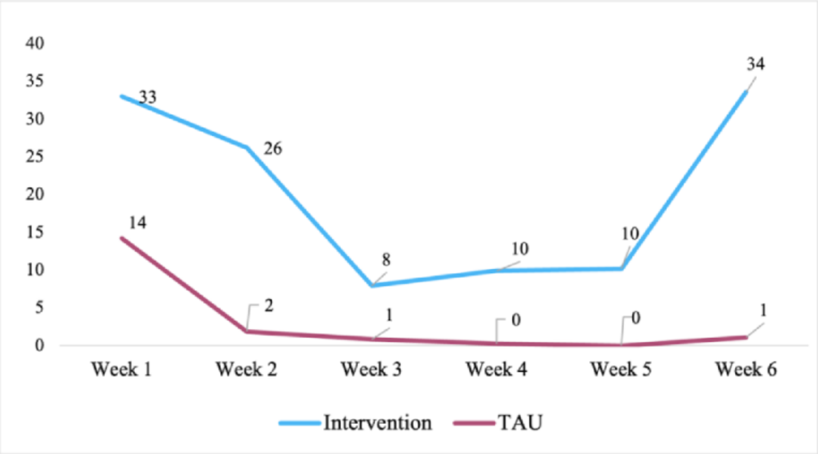


Figure 3. Forms of attrition. TAU: treatment as usual.

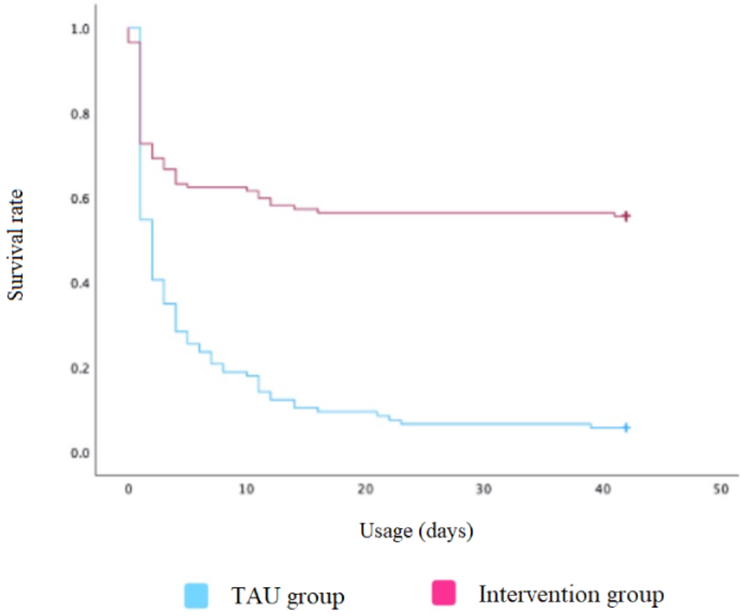


Figure 4. Gender-based attrition in the TAU and intervention groups. TAU: treatment as usual.

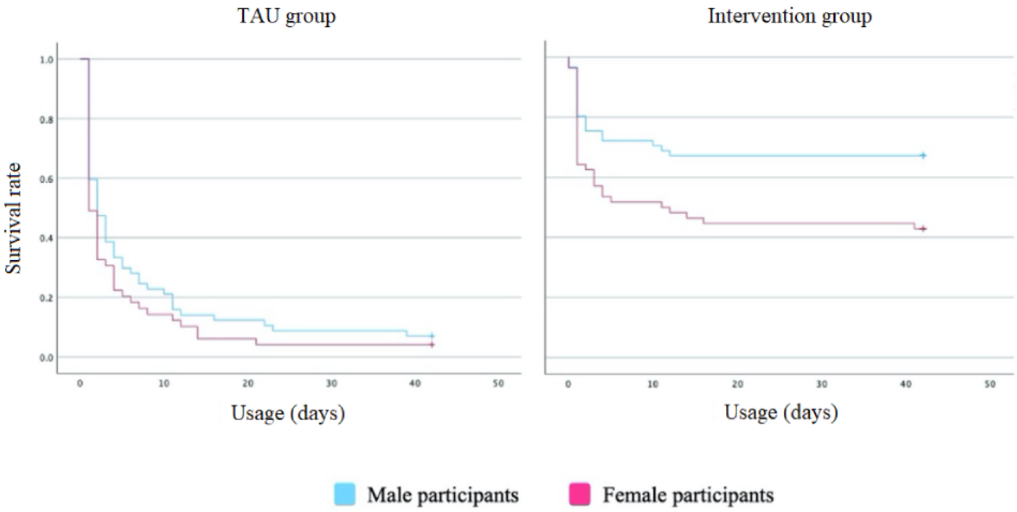


Table 2. Weekly comparison of usage and attrition.

Week	In-app health exercises completed, n			TAU[a] group usage, mean (SD)	Intervention group usage, mean (SD)	*P* value
	Overall	Male	Female			
1	5369	3531	1838	14.20 (17.34)	33.03 (72.38)	<.001
2	3265	2514	751	1.85 (8.71)	26.23 (90.61)	<.001
3	1016	548	468	0.84 (5.38)	7.92 (26.71)	<.001
4	1192	803	389	0.24 (1.28)	9.97 (34.08)	<.001
5	1223	891	332	0.00 (0.00)	10.45 (60.09)	<.001
6	4029	2889	1140	1.20 (6.99)	33.35 (110.99)	<.001
Overall	16,094	11,176	4918	18.32 (27.44)	120.96 (350.09)	<.001

[a]TAU: treatment as usual.

Table 3. Weekly average in-app health exercise frequency by gender.

Week and group	Male, mean (SD)	Female, mean (SD)	*P* value
Week 1			
TAU[a]	16.19 (19.36)	11.88 (14.50)	.20
Intervention	42.75 (97.75)	22.43 (20.12)	.13
Week 2			
TAU	3.02 (11.71)	0.49 (1.56)	.14
Intervention	38.89 (72.64)	12.98 (23.79)	.13
Week 3			
TAU	0.07 (0.42)	1.73 (7.84)	.11
Intervention	8.92 (22.29)	6.84 (30.99)	.68
Week 4			
TAU	0.37 (1.70)	0.08 (0.45)	.25
Intervention	12.84 (30.06)	6.88 (26.09)	.35
Week 5			
TAU	0.00 (0.00)	0.00 (0.00)	N/A[b]
Intervention	14.61 (62.43)	5.93 (12.73)	.44
Week 6			
TAU	1.02 (4.66)	1.41 (9.02)	.78
Intervention	46.41 (77.42)	19.13 (43.32)	.19
All weeks			
TAU	20.67 (33.60)	15.59 (17.85)	.35
Intervention	163.90 (271.22)	74.18 (109.50)	.08

[a]TAU: treatment as usual.

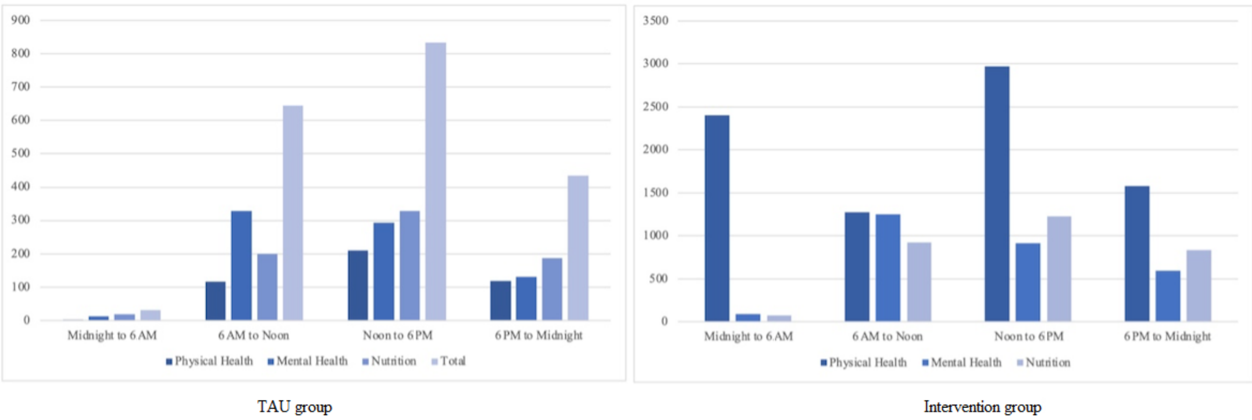
[b]N/A: not applicable.

Table 4. Frequency of exercise categories at different daily times.

Group and exercise category	Midnight to 6 AM, n (%)	6 AM to noon, n (%)	Noon to 6 PM, n (%)	6 PM to midnight, n (%)
TAU[a] group	32 (100)	644 (100)	833 (100)	435 (100)
Physical health	1 (3.1)	116 (18.0)	210 (25.2)	119 (27.4)
Mental health	12 (37.5)	329 (51.1)	294 (35.3)	130 (29.9)
Nutrition	19 (59.4)	199 (30.9)	329 (39.5)	186 (42.7)
Intervention group	2575 (100)	3446 (100)	5113 (100)	3008 (100)
Physical health	2407 (93.5)	1275 (37.0)	2970 (58.1)	1576 (52.4)
Mental health	90 (3.5)	1248 (36.2)	912 (17.8)	599 (19.9)
Nutrition	78 (3.0)	923 (26.8)	1231 (24.1)	833 (27.7)
Overall	2607 (100)	4090 (100)	5946 (100)	3443 (100)
Physical health	2408 (92.4)	1391 (34.0)	3180 (53.5)	1695 (49.2)
Mental health	102 (3.9)	1577 (38.6)	1206 (20.3)	729 (21.2)
Nutrition	97 (3.7)	1122 (27.4)	1560 (26.2)	1019 (29.6)

[a]TAU: treatment as usual.

Figure 5. Time-based exercise categories in the TAU and intervention groups. TAU: treatment as usual.



Discussion

The focus of this study was on time-specific attrition in an adolescent mHealth intervention. We hoped to build on previous work while focusing on the type, frequency, and time of usage in order to better understand why adolescent attrition from mHealth interventions is generally as excessive as it is, with a market saturated with roughly 350,000 mHealth apps adolescents seem reluctant to engage in [12,35]. The results showed that research groups are related to time of attrition (hazard ratio 0.308, 95% CI 0.222-0.420), and attrition differed between the TAU group (94.3%) and intervention group (44.4%). Attrition was somewhat higher than in our previous study, where attrition was in the 35th percentile. The difference in attrition rates between research groups was however vast, and the only distinction in program setup was motivational support in the form of weekly feedback on individual progress through the app, friendly health task competitions at the individual and group levels, and small altruistic rewards for completion of health tasks or active participation in competitions. The app was the same, but support through program setup differed. It is interesting that the altruistic reward cost per participant was only 56 US cents but still seemed to contribute strongly to increased usage and participation.

Completion rates differed greatly between research groups as did days of active usage since the number of usage days in the intervention group was nearly 25 (95% CI 21.452-28.518) and that in the TAU group was approximately 6 (95% CI 4.304-8.277). Differences in completion rates were therefore evident between the intervention group (55.6%) and TAU group (5.7%), and gender differences in completion rates were also observed in the intervention group but not in the TAU group. Male participants (95% CI 24.519-33.812) completed in-app health tasks longer than female participants (95% CI 15.301-25.558). This gender difference was not evident in our prior research but has been observed in adult populations and should be examined in future research hoping to explain attrition factors in adolescent mHealth interventions [38]. A deeper gender-based exploration into motivational support could be a promising avenue for further research on the matter, particularly how altruistic rewards and competitive intervention features facilitate motivational support between genders.

Broad differences in the completion of health exercises were evident between groups since the intervention group completed on average roughly 6 times as many health exercises as the TAU group throughout the trial period. It is somewhat interesting that this difference was only 2-fold on the first day of the trial, suggesting that participants in the TAU group did not lack usage motivation at the beginning of the trial but were simply not supported to keep on using the app. This was more evident when average weekly health exercise frequency was examined since there was a decrease in usage between the first and second weeks among participants in the TAU group, while such a difference was not evident among participants in the intervention group. In fact, intervention group members completed on average more health exercises in all 6 weeks of the trial than their peers in the TAU group. An interesting finding is that usage increased from the fifth week of the trial to the last one among intervention group members but not among TAU group members, which is thought to be related to motivational support features as well as the altruistic reward setup at the end of the trial's sixth week, and warrants further research.

There seemed to be different daily sensitivity periods for increased frequency of health task completion in different health categories. Adolescents in both research groups completed most of the physical activity exercises within the app from noon to 6 PM. The same applied to nutrition exercises in both groups. However, when it came to the frequency of mental health exercises, adolescents in both groups tended to do them from 6 AM to noon. Further, results from regression models indicated that the frequency of mental health exercises as well as nutrition exercises completed in the app by participants in both groups was related to delayed attrition. Physical activity exercises did not show such effects, possibly because those participants who did few exercises and were likely to drop out mainly used the physical activity category. Adolescents in the TAU group who completed an in-app health exercise between day 2 and day 6 of the trial were also found to be more likely to finish. This was not evident in the intervention group. These results imply that there are sensitivity usage periods that differ between the types of health behaviors adopted by adolescents in mHealth interventions and highlights the need for the development of just-in-time adaptive interventions in the future to hamper attrition and hopefully increase the frequency of health behavior exercises.

Taken together, the aim of this study was to examine time-specific attrition rates in an mHealth intervention for adolescents and hopefully increase our understanding of attrition in this group through a focus on the type, frequency, and time of health behavior exercise usage. Attrition between research groups was vastly different, and motivational support seems to be of vital importance to lower attrition in future mHealth interventions, specifically for adolescents in different age groups. Further research on how specific motivational support features in adolescent mHealth interventions function to lower attrition rates and how they affect usage patterns is evidently needed.

The limitations of this research include randomization factors, since randomization was between elementary schools rather than on an individual level to prevent contamination effects. Another limitation was related to the initial difference in usage between research groups. The data collection period was 6 weeks, and further research on the matter should include a prolonged research period with added randomization efforts to level usage between research groups, along with a 3-month follow-up to track usage and sustained gains from motivational support features. The generalizability of findings in adolescent mHealth studies to wider populations can be questionable, and this study is no exception (for instance, the function of altruistic reward schemes and competitive features in diverse cultural settings). The study's foremost strength lies in added knowledge to limited research on attrition rates and patterns in adolescent mHealth interventions. Additional strong points are related to the methodological approach. Continuous data collection throughout the trial period, efforts to accurately describe time-based attrition rates through survival analysis, and use of a relatively large sample size (n=304) are regarded as strengths.

Acknowledgments

The authors wish to thank the elementary school officials who helped in the study. The authors also wish to thank Icelandic Research Fund (IRF 141381051) for partially funding the study through a research grant.

Conflicts of Interest

EE is a minority shareholder in SidekickHealth AB and a former employee. The other authors have no conflicts to declare.

Editorial Notice

This randomized study was only retrospectively registered. The authors explained that the trial was originally registered in domestic registries through the University of Iceland and the Icelandic bioethics committee. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application [or other reasons for the exception, as argued by the authors]. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
App function.

[PNG File , 445 KB - mhealth_v11i1e45414_app1.png]

Multimedia Appendix 2
Study flowchart.
[PDF File (Adobe PDF File), 112 KB - mhealth_v11i1e45414_app2.pdf]

Multimedia Appendix 3
Regression model.
[PDF File (Adobe PDF File), 124 KB - mhealth_v11i1e45414_app3.pdf]

Multimedia Appendix 4
Anxiety, depression, and self-efficacy measures between research groups.
[PDF File (Adobe PDF File), 76 KB - mhealth_v11i1e45414_app4.pdf]

Multimedia Appendix 5
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1604 KB - mhealth_v11i1e45414_app5.pdf]

References

1. Smith A. U.S. Smartphone Use in 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2023-04-27]
2. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2023-04-27]
3. Anderson M, Jiang J. Teens, Social Media and Technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2023-04-27]
4. Jones EAK, Mitra AK, Bhuiyan AR. Impact of COVID-19 on Mental Health in Adolescents: A Systematic Review. Int J Environ Res Public Health 2021 Mar 03;18(5):2470 [FREE Full text] [doi: 10.3390/ijerph18052470] [Medline: 33802278]
5. Körmendi A. [Smartphone usage among adolescents]. Psychiatr Hung 2015;30(3):297-302. [Medline: 26471031]
6. Mobile marketing statistics. Smart Insights. 2016. URL: https://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/ [accessed 2023-04-27]
7. Birnbaum M, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295 [FREE Full text] [doi: 10.1111/eip.12237] [Medline: 25808317]
8. Lawlor A, Kirakowski J. Online support groups for mental health: A space for challenging self-stigma or a means of social avoidance? Computers in Human Behavior 2014 Mar;32:152-161 [FREE Full text] [doi: 10.1016/j.chb.2013.11.015]
9. Pretorius C, Chambers D, Coyle D. Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
10. McLaughlin K, Costello E, Leblanc W, Sampson N, Kessler R. Socioeconomic Status and Adolescent Mental Disorders. Am J Public Health 2012 Sep;102(9):1742-1750 [FREE Full text] [doi: 10.2105/ajph.2011.300477]
11. Radomski A, Wozney L, McGrath P, Huguet A, Hartling L, Dyson MP, et al. Design and Delivery Features That May Improve the Use of Internet-Based Cognitive Behavioral Therapy for Children and Adolescents With Anxiety: A Realist Literature Synthesis With a Persuasive Systems Design Perspective. J Med Internet Res 2019 Feb 05;21(2):e11128 [FREE Full text] [doi: 10.2196/11128] [Medline: 30720436]
12. Digital Health Trends 2021. IQVIA Institute. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2023-04-27]
13. Chan A, Kow R, Cheng J. Adolescents' Perceptions on Smartphone Applications (Apps) for Health Management. JournalMTM 2017 Aug;6(2):47-55 [FREE Full text] [doi: 10.7309/jmtm.6.2.6]
14. Kohl H, Craig C, Lambert E, Inoue S, Alkandari J, Leetongin G, et al. The pandemic of physical inactivity: global action for public health. The Lancet 2012 Jul;380(9838):294-305 [FREE Full text] [doi: 10.1016/s0140-6736(12)60898-8]
15. Lee I, Shiroma E, Lobelo F, Puska P, Blair S, Katzmarzyk P. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet 2012 Jul;380(9838):219-229 [FREE Full text] [doi: 10.1016/s0140-6736(12)61031-9]
16. Ding D, Lawson K, Kolbe-Alexander T, Finkelstein E, Katzmarzyk P, van Mechelen W, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. The Lancet 2016 Sep;388(10051):1311-1324 [FREE Full text] [doi: 10.1016/s0140-6736(16)30383-x]

17. Sember V, Jurak G, Kovač M, Đurić S, Starc G. Decline of physical activity in early adolescence: A 3-year cohort study. PLoS One 2020 Mar 11;15(3):e0229305 [FREE Full text] [doi: 10.1371/journal.pone.0229305] [Medline: 32160216]
18. Global recommendations on physical activity for health. World Health Organization. 2010. URL: https://www.who.int/publications-detail-redirect/9789241599979 [accessed 2023-04-27]
19. OECD/European Union. Health at a Glance: Europe 2016: State of Health in the EU Cycle. Paris, France: OECD Publishing; 2016.
20. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273 [FREE Full text] [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]
21. Craigie A, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. Maturitas 2011 Nov;70(3):266-284 [FREE Full text] [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
22. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A Systematic Review of Digital Interventions for Improving the Diet and Physical Activity Behaviors of Adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
23. Kouvari M, Karipidou M, Tsiampalis T, Mamalaki E, Poulimeneas D, Bathrellou E, et al. Digital Health Interventions for Weight Management in Children and Adolescents: Systematic Review and Meta-analysis. J Med Internet Res 2022 Feb 14;24(2):e30675 [FREE Full text] [doi: 10.2196/30675] [Medline: 35156934]
24. Obesity and Overweight. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight#cms [accessed 2023-04-27]
25. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
26. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
27. Topooco N, Byléhn S, Dahlström Nysäter E, Holmlund J, Lindegaard J, Johansson S, et al. Evaluating the Efficacy of Internet-Delivered Cognitive Behavioral Therapy Blended With Synchronous Chat Sessions to Treat Adolescent Depression: Randomized Controlled Trial. J Med Internet Res 2019 Nov 01;21(11):e13393 [FREE Full text] [doi: 10.2196/13393] [Medline: 31682572]
28. Adelman C, Panza K, Bartley C, Bontempo A, Bloch M. A Meta-Analysis of Computerized Cognitive-Behavioral Therapy for the Treatment of Anxiety Disorders. J. Clin. Psychiatry 2014 Jul 15;75(07):e695-e704 [FREE Full text] [doi: 10.4088/jcp.13r08894]
29. Mohr D, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
30. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage Attrition of Adolescents in an mHealth Promotion Intervention and the Role of Socioeconomic Status: Secondary Analysis of a 2-Arm Cluster-Controlled Trial. JMIR Mhealth Uhealth 2022 May 10;10(5):e36404 [FREE Full text] [doi: 10.2196/36404] [Medline: 35536640]
31. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450 [FREE Full text] [doi: 10.1111/bjc.12055] [Medline: 24831119]
32. Melville K, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471 [FREE Full text] [doi: 10.1348/014466509X472138] [Medline: 19799804]
33. Mitchell A, Selmes T. Why don't patients attend their appointments? Maintaining engagement with psychiatric services. Adv. psychiatr. treat 2018 Jan 02;13(6):423-434 [FREE Full text] [doi: 10.1192/apt.bp.106.003202]
34. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
35. Egilsson E, Bjarnason R, Njardvik U. Usage and Weekly Attrition in a Smartphone-Based Health Behavior Intervention for Adolescents: Pilot Randomized Controlled Trial. JMIR Form Res 2021 Feb 17;5(2):e21432 [FREE Full text] [doi: 10.2196/21432] [Medline: 33481750]
36. Jeminiwa R, Hohmann NS, Fox BI. Developing a Theoretical Framework for Evaluating the Quality of mHealth Apps for Adolescent Users: A Systematic Review. J Pediatr Pharmacol Ther 2019;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
37. Bear HA, Ayala Nunes L, DeJesus J, Liverpool S, Moltrecht B, Neelakantan L, et al. Determination of Markers of Successful Implementation of Mental Health Apps for Young People: Systematic Review. J Med Internet Res 2022 Nov 09;24(11):e40347 [FREE Full text] [doi: 10.2196/40347] [Medline: 36350704]

38. Palos-Sanchez P, Saura JR, Rios Martin M, Aguayo-Camacho M. Toward a Better Understanding of the Intention to Use mHealth Apps: Exploratory Study. JMIR Mhealth Uhealth 2021 Sep 09;9(9):e27021 [FREE Full text] [doi: 10.2196/27021] [Medline: 34499044]
39. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
40. Gísladóttir G, Njarðvík U, Hannesdóttir DK, Halldórsson F. Próffræðilegir eiginleikar íslenskrar þýðingar RCADS. Sálfræðiritið – Tímarit Sálfræðingafélags Íslands 2014:71-80 [FREE Full text]
41. Chorpita B, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. Behav Res Ther 2000 Aug;38(8):835-855 [FREE Full text] [doi: 10.1016/s0005-7967(99)00130-8] [Medline: 10937431]
42. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in health psychology: A user's portfolio. Causal and control beliefs. Windsor, UK: NFER-Nelson; 1995:35-37.
43. Luszczynska A, Diehl M, Gutiérrez-Doña B, Kuusinen P, Schwarzer R. Measuring one component of dispositional self-regulation: attention control in goal pursuit. Personality and Individual Differences 2004 Aug;37(3):555-566 [FREE Full text] [doi: 10.1016/j.paid.2003.09.026]
44. Owens J, Dalzell V. Use of the 'BEARS' sleep screening tool in a pediatric residents' continuity clinic: a pilot study. Sleep Med 2005 Jan;6(1):63-69 [FREE Full text] [doi: 10.1016/j.sleep.2004.07.015] [Medline: 15680298]
45. Ramírez-Vélez R, Huertas-Zamora L, Correa-Bautista JE, Cárdenas-Calderón EG. Confiabilidad y validez del cuestionario de trastornos de sueño BEARS en niños y adolescentes escolares de Bogotá, D.C., Colombia: Estudio FUPRECOL (Reliability and Validity of the BEARS Sleep Disorder Questionnaire in Children and Adolescent Students fr. Retos 2017 Nov 01(34):89-93. [doi: 10.47197/retos.v0i34.58585]
46. Bloom T, Sharpe L, Mullan B, Zucker N. A pilot evaluation of appetite-awareness training in the treatment of childhood overweight and obesity: a preliminary investigation. Int J Eat Disord 2013 Jan;46(1):47-51 [FREE Full text] [doi: 10.1002/eat.22041] [Medline: 22826019]
47. Gunnarsdottir T, Njardvik U, Olafsdottir AS, Craighead L, Bjarnason R. Childhood obesity and co-morbid problems: effects of Epstein's family-based behavioural treatment in an Icelandic sample. J Eval Clin Pract 2012 Apr;18(2):465-472 [FREE Full text] [doi: 10.1111/j.1365-2753.2010.01603.x] [Medline: 21210895]
48. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
49. Clark T, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. Br J Cancer 2003 Jul 21;89(2):232-238 [FREE Full text] [doi: 10.1038/sj.bjc.6601118] [Medline: 12865907]
50. Bradburn M, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. British Journal of Cancer 2003 Aug 04;89(3):431-436 [FREE Full text] [doi: 10.1038/sj.bjc.6601119] [Medline: 12888808]

Abbreviations

mHealth: mobile health
TAU: treatment as usual

Edited by L Buis; submitted 29.12.22; peer-reviewed by L Maenhout, H Bear, B Chaudhry, OR Patil; comments to author 18.01.23; revised version received 03.03.23; accepted 24.04.23; published 26.06.23.

Please cite as:
Egilsson E, Bjarnason R, Njardvik U
Usage and Daily Attrition of a Smartphone-Based Health Behavior Intervention: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e45414
URL: https://mhealth.jmir.org/2023/1/e45414
doi:10.2196/45414
PMID:37358888

Original Paper

Trajectories of Symptoms in Digital Interventions for Depression and Anxiety Using Routine Outcome Monitoring Data: Secondary Analysis Study

Diana Catalina Cumpanasoiu[1], PhD; Angel Enrique[1,2], PhD; Jorge E Palacios[1,2], MD, PhD; Daniel Duffy[1,2], PhD; Scott McNamara[1], MA; Derek Richards[1,2], PhD

[1]SilverCloud Science, SilverCloud Health, Dublin, Ireland

[2]E-Mental Health Group, School of Psychology, Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Diana Catalina Cumpanasoiu, PhD
SilverCloud Science
SilverCloud Health
One Stephen Street Upper
Dublin, D08 DR9P
Ireland
Phone: 353 6467031051
Email: catalina.cumpanasoiu@amwell.com

Abstract

Background: Research suggests there is heterogeneity in treatment response for internet-delivered cognitive behavioral therapy (iCBT) users, but few studies have investigated the trajectory of individual symptom change across iCBT treatment. Large patient data sets using routine outcome measures allows the investigation of treatment effects over time as well as the relationship between outcomes and platform use. Understanding trajectories of symptom change, as well as associated characteristics, may prove important for tailoring interventions or identifying patients who may not benefit from the intervention.

Objective: We aimed to identify latent trajectories of symptom change during the iCBT treatment course for depression and anxiety and to investigate the patients' characteristics and platform use for each of these classes.

Methods: This is a secondary analysis of data from a randomized controlled trial designed to examine the effectiveness of guided iCBT for anxiety and depression in the UK Improving Access to Psychological Therapies (IAPT) program. This study included patients from the intervention group (N=256) and followed a longitudinal retrospective design. As part of the IAPT's routine outcome monitoring system, patients were prompted to complete the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) after each supporter review during the treatment period. Latent class growth analysis was used to identify the underlying trajectories of symptom change across the treatment period for both depression and anxiety. Differences in patient characteristics were then evaluated between these trajectory classes, and the presence of a time-varying relationship between platform use and trajectory classes was investigated.

Results: Five-class models were identified as optimal for both PHQ-9 and GAD-7. Around two-thirds (PHQ-9: 155/221, 70.1%; GAD-7: 156/221, 70.6%) of the sample formed various trajectories of improvement classes that differed in baseline score, the pace of symptom change, and final clinical outcome score. The remaining patients were in 2 smaller groups: one that saw minimal to no gains and another with consistently high scores across the treatment journey. Baseline severity, medication status, and program assigned were significantly associated ($P<.001$) with different trajectories. Although we did not find a time-varying relationship between use and trajectory classes, we found an overall effect of time on platform use, suggesting that all participants used the intervention significantly more in the first 4 weeks ($P<.001$).

Conclusions: Most patients benefit from treatment, and the various patterns of improvement have implications for how the iCBT intervention is delivered. Identifying predictors of nonresponse or early response might inform the level of support and monitoring required for different types of patients. Further work is necessary to explore the differences between these trajectories to understand what works best for whom and to identify early on those patients who are less likely to benefit from treatment.

(JMIR Mhealth Uhealth 2023;11:e41815) doi:10.2196/41815

KEYWORDS

internet-delivered cognitive behavioral therapy; iCBT; depression; anxiety; trajectory of symptom change; routine outcome monitoring data

Introduction

Background

Depressive and anxiety disorders are 2 of the most common mental health difficulties, with epidemiological studies across countries suggesting that they are highly prevalent, can persist throughout lifetime, and are seriously impairing [1-4]. Cognitive behavioral therapy (CBT) is an effective psychological treatment regularly used to treat depression and anxiety symptoms [5,6]. In recent years, CBT has been adapted to an internet-delivered format (internet-delivered CBT [iCBT]) and has overcome some barriers associated with accessing traditional face-to-face psychological treatments [7,8]. Several meta-analyses have demonstrated the effectiveness of these interventions in treating depression and anxiety [9-13].

A key value in digitally delivered treatments is the ability to collect data on patient characteristics, routine outcome measures, and their engagement with the intervention. Given the volume of data collected, there is a potential new opportunity to understand treatment effects at an individual level. More than ever before, we can understand the trajectory of individual symptom changes and further explore the relationship between the use of the intervention and clinical outcomes. Understanding treatment effects over time and the relationship between use and outcomes may prove important for developing tailored interventions for different [14] patients. It may also be helpful in identifying patients who may be at risk of not responding, which can support clinical decision-making [15]. These efforts may have the potential to enhance digitally delivered treatments.

Most empirical evidence on trajectories of symptom change comes from face-to-face psychotherapy studies, which have found various classes of symptom courses, from early responders to late or delayed responders and steady or moderate responders [16,17]. Few studies have explored the evolution of symptoms during iCBT interventions. Some of them have consistently found a large group of users who show improvement and another group of users who show no or low symptom improvement [18,19]. Other studies have also found that most treatment responders experience the most clinical gains during the first weeks [20-22] and even before the treatment initiated [22]. Several of these studies also investigated the effects of individual baseline characteristics (eg, age and sex) on class membership; however, only symptom load has been consistently associated with class membership [20,21]. Similarly, studies have examined intervention use metrics and their relationship with class membership, with inconsistent findings reported [18,20,22]. While no differences were found between classes in terms of overall use time [18,22], two studies found differences between classes in the number of assessments, modules, and sessions completed [20,22].

In terms of intervention use and its relationship with outcomes from iCBT [23], it has been proposed that higher use (ie, better adherence or completion rate) predicts better outcomes [24]. However, other studies investigating the relationship between use metrics and outcomes have reported mixed results [25,26]. To date, many studies have been limited by their collection of outcomes at fixed time points. Evaluating use patterns in relation to continuous outcome monitoring may provide insight into how the temporal aspect of use is linked to changes in symptoms. Several studies [26,27] suggest that most use occurs at earlier stages of the intervention and that patients who improved have higher exposure levels to the intervention, especially in the first half [26]. A recent randomized controlled trial (RCT) presented an opportunity to examine the relationship between engagement and outcomes at different time points [28]. The results suggested that there was an association between use (completion rate and the frequency of items completed, but not time spent) and outcomes at 3 months but not earlier.

Objectives

Overall, the current literature on trajectories of symptom change in iCBT is at an early stage, and more research is needed to confirm whether the classes found in previous studies are also observed in diverse samples from different settings. Identifying individuals who benefit from iCBT and those at risk of not improving is key to offering tailored interventions that fit the needs of specific populations, which may ultimately lead to increased response rates. In addition, learning more about the stage of treatment at which change occurs and its association with intervention use will shed light on whether intervention use acts as a mechanism for change in these trajectories. On the basis of this, this study sought to use routine outcome monitoring (ROM) data gathered from a pragmatic RCT in a clinical service setting to (1) identify latent classes of responders and associated participant characteristics during an iCBT intervention for depression and anxiety treatment and (2) investigate the presence of a time-varying relationship between trajectories of change and intervention use.

Methods

Study Setting

This study is a secondary analysis of data collected [29] at the Berkshire Healthcare Foundation Trust, a provider within the National Health Service Improving Access to Psychological Therapies (IAPT) program. IAPT is a stepped-care model in which people with depression and anxiety are offered different intensities of treatment depending on their needs and symptom severity. At step 2, clients are recommended low-intensity CBT-based treatments, such as guided self-help, internet-delivered CBT, or group CBT, under the supervision of a psychological well-being practitioner (PWP). The PWPs are a specially trained cohort of psychology graduate students, with additional qualification in delivering low-intensity CBT. This study was conducted at step 2 of IAPT, with patients being assigned to iCBT as their preferred treatment option.

Design

The original RCT where these data were collected was designed to examine the effectiveness and cost-effectiveness of the SilverCloud programs for anxiety and depression [29]. The study used a parallel-group design, in which an intervention group was compared with a waitlist control group, and the results demonstrated the effectiveness of the intervention group compared to the waitlist control after treatment, and improvements were sustained over a 12-month period [29].

Between June 28, 2017, and April 30, 2018, a total of 464 participants were invited to the original RCT; however, this study followed a longitudinal, retrospective design and included only patients from the intervention group (N=256). It captured the clinical assessments and platform use that occurred during the first 12 weeks of intervention use. While the main RCT used outcome data collected at research time points, this study used ROM data. As part of their treatment journey in IAPT, clients were prompted to complete the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) when they received a review approximately every 2 weeks. Hence, for each participant, assessments were available at baseline and at weeks 2, 4, 6, 8, 10, and 12.

Participants

The eligibility criteria for this analysis mirrored that of the main RCT; therefore, to be included in the main RCT, a participant had to be aged >18 years, to present with mild to moderate symptoms of depression or anxiety, and to consent to engage with iCBT. In addition, comorbidity with psychotic illness, current psychological treatment, previous organic mental health disorder diagnosis, substance misuse, and suicidal risk (suicide-related thoughts, ideation, or active plans) were used as exclusion criteria.

Interventions

On the basis of their symptoms and needs, clients were offered one of the following SilverCloud programs: Space from depression; Space from anxiety (different programs for specific anxiety disorders—modules for phobia, social anxiety, or generalized anxiety disorders); and Space from depression and anxiety, with the possibility of customizing treatment. Each participant had a PWP monitoring their progress and providing asynchronous reviews through the SilverCloud platform. The supporters had access to clients' engagement and activity use through a dashboard interface, and they were encouraged to use this information to provide supportive and positive reviews to their clients.

All programs use evidence-based CBT principles and are delivered on the web (via a PC, tablet, or mobile device) on a Web 2.0 platform using media-rich interactive content (see Figures S1 and S2 in Multimedia Appendix 1, eg, for screenshots from the programs and prior publications [30] for more detailed descriptions). Each SilverCloud program has up to 8 modules, and it is recommended to complete 1 module per week. Each module incorporates quizzes, videos, informational content, interactive activities, homework assignments, and summaries. These interventions follow the National Institute for Health and Care Excellence guidelines [31] and have been tested and proved effective [30]. Participants were also assigned a PWP to support their progress through the program. These supporters monitor participants' activities through the platform, provide guidance, and tailor feedback based on patient needs. This feedback comes in the form of regular reviews of the participants' progress. In addition, supporters use these reviews to offer suggestions and guidance on how best to navigate the content and modules in each program to best fit an individual's needs.

Assessments

The data collected included demographic variables (age, sex, ethnicity, and employment), clinical measures for depression (PHQ-9) and anxiety (GAD-7), and use metrics.

The PHQ-9 is a brief, self-reported measure of depression [32,33] containing 9 items on a Likert scale (from 0 to 3). The score ranges from 0 to 27, with a cutoff score of ≥10 indicating the presence of depression and higher scores reflecting more severe symptoms. This assessment is widely used in clinical and research settings, and its validity, reliability (89%), sensitivity (88%), and specificity (88%) have been confirmed [32]. The GAD-7 is a brief, self-reported measure of anxiety [34] containing 7 items on a Likert scale (from 0 to 3). The score ranges from 0 to 21, with a cutoff score of ≥8 indicating the presence of anxiety and higher scores representing more severe symptoms. Similar to the PHQ-9, this assessment is widely used, and its validity and good internal consistency have been confirmed [34].

In terms of use, several objective metrics were obtained from the SilverCloud platform. Table 1 presents a full list of all use metrics and what each measured. To assess use at different time points in the treatment journey, all metrics were computed for each 2-week period between the start of treatment and week 12.

Table 1. Description of all usage metrics examined.

Use metrics	Descriptions
Number of log-ins	Number of log-ins for each participant adherence
Time spent	Length of time spent using the platform adherence
Number of reviews	Number of reviews each participant received from their psychological well-being practitioner engagement
Number of activities	Number of activities logged (eg, every time the participant used a tool or logged a journal entry)
Percentage of the program viewed	Percentage of new content viewed in each 2-week period engagement

Procedure

For the main RCT, participants were first screened and then invited to participate in the study, and the consented participants were assigned to active treatment or waitlist. Once assigned to the active treatment group, participants were offered 1 of 3 SilverCloud programs (Space from Depression, Space from Anxiety, and Space from Depression and Anxiety) based on their needs. As the participants worked through the programs, they were presented with assessments to evaluate their progress. ROM is used to trigger assessments of participants at various time intervals to provide regular measurements of their progress. These assessments correspond with predetermined research time intervals that allowed for a deeper understanding of everyone's journey through the program. All actions taken by the participants within the platform, such as module progress, content viewed, and activities completed, were collected through SilverCloud backend data collection.

Analysis Plan

First, differences between the included and excluded participants were established using descriptive statistics (mean and SD) and independent 2-tailed *t* tests. To answer the first research question and identify trajectories of change, latent class growth analysis (LCGA) was used to identify latent classes. LCGA is a type of growth mixture modeling that is used to identify latent classes with different trajectories of growth. Mixture modeling approaches such as LCGA have been used more broadly in recent years because they allow the identification of underlying clusters based on unobserved heterogeneity in the data [35]. Compared with other growth modeling approaches that describe all trajectories with a single growth estimate, LCGA allows the identification of latent classes that have different characteristics (eg, intercept and slope) and assumes that all individual trajectories in a class are homogeneous [36-38]. This is done by fixing the variance of the intercept and slope within a class to 0 and allowing them to vary only across classes [37,38]. LCGA models address missing data using maximum likelihood algorithms [36,37]. To determine the optimal number of classes, models with an increasing number of classes are estimated, and different fit indices are used to compare them. There are multiple considerations taken into account when choosing the optimal model, such as the model fit indices, theoretical framework, clinical interpretation, and other criteria such as the number of participants in each class [36]. After the model is chosen, the probability of each individual to belong to one of the classes is estimated using maximum posterior probabilities and thus each individual is assigned to one of the latent classes.

LCGA is commonly conducted using statistical software, such as MPlus and SAS; however, recent efforts have made it easier to conduct such analyses in open-source R software [39]. For this analysis, the *lcmm* package [40] in R was used following the steps in the tutorial provided by Wardenaar [38]. A single-class growth model with a fixed intercept and slope for the subjects was initially run to test whether a linear, quadratic, or cubic model would be more appropriate for capturing the overall observed pattern of the trajectory. The coefficients of the cubic and quadratic terms had a poor model fit; therefore, a linear LCGA model was fitted. Latent class models were then constructed by increasing the number of classes from 2 to 8 to identify the optimal number of classes. Once a model was selected, 1-way ANOVAs and chi-square tests were performed to evaluate differences between classes in individual characteristics (eg, age, sex, and baseline severity).

Before investigating the relationship between use and symptom change, descriptive methods were used to explore use data, and regressions were run to understand the predictive values of different patient characteristics (eg, age and baseline severity) on each use metric. Then, to understand how different trajectories relate to use, mixed (between-factors and within-factors) ANOVAs were conducted to examine the role of time and class (PHQ and GAD) membership in use metrics. For the number of log-ins, length of use, number of reviews, number of activities, and percentage of programs viewed, five 3×5 mixed ANOVAs were conducted, with time as within-factor (3 levels: use in the first 4 weeks, 4-8 weeks, and 8-12 weeks) and class as between-factor (5 levels: the 5 trajectories identified). A total of 10 ANOVAs were run, 5 using the depression trajectories and 5 using the anxiety trajectories. All regressions and mixed ANOVAs were run using the R platform.

Data Processing

As these were ROM data, it was decided to select the assignments completed on the closest date to the time points of interest (baseline, 2 weeks, 4 weeks, 6 weeks, 8 weeks, 10 weeks, and 12 weeks). The following criteria were used to do this: (1) an interval of –6 days to +6 days from each of the time points of interest was considered, and if there were more assignments done in that period, the one on the closest date to the time point of interest was selected; (2) if in the –6 days to +6 days interval there were 2 assignments equally close to the time point (eg, 1 assignment done 4 days before the time point and 1 assignment done 4 days after the time point), the second one was selected (ie, the one after the time point); and (3) if no assignments were completed in that time interval, a check was done to see if any assignments were done on the seventh day before or after the time point. If neither of these conditions was met, no assignment was selected for that time point.

As ROM data were used for outcomes with the criteria explained earlier, there were a number of missing PHQ-9 and GAD-7 assessments at each time point (for more details regarding missing data, see Table S1 in Multimedia Appendix 1). On average, each individual had 2 missed measurements out of 7 possible. There were also no significant results (all $P>.05$) from the logistic regression evaluating the predictive power of age, sex, baseline PHQ score, baseline GAD score, presence of a long-term condition (LTC), psychiatric medication status, and employment status on having more (defined as 3+) or less (defined as 0-2) missing assessments.

LCGA uses the maximum likelihood algorithm to handle both participants with full and missing data [36]. Missing data patterns for outcomes were evaluated and based on the Little test. After studying the different patterns of missing data, there was no evidence to disconfirm that data were missing at random; thus, analyses proceeded under this assumption. After identifying the latent classes, the pattern of missing data in each class was evaluated, and similar proportions of missing assessments were found in each class (Table S2 in Multimedia Appendix 1). For use, there were no missing data. Box plots were constructed to identify outliers, and the Winsorization method was applied. As use data are expected to be skewed, Winsorization was chosen over other methods (eg, truncating) to preserve all the data points and limit the influence of the extreme outliers. Several values were identified as potential outliers for the number of log-ins, length of use, and number of activities. All values determined as extreme (ie, 3 SDs or more away from the mean) were Winsorized to reduce the impact of those data points without removing them.

Ethics Approval

This trial was approved by the National Health Service England Research Ethics Committee (reference: 17/NW/0311). The trial was prospectively registered at Current Controlled Trials (ISRCTN91967124).

Results

Overview and Sample Characteristics

Of the 256 participants in the intervention arm, 23 were excluded because they did not have a start date, and another 12 were excluded because of having only 1 assessment (the baseline). The analyses were conducted on the remaining 221 individuals.

Participants were aged 18 to 74 (mean 33, SD 12.68) years, had an overall baseline PHQ-9 score of 13.82, and had an overall baseline GAD-7 score of 12.26. Table 2 presents further descriptive information on the participants' demographic and clinical characteristics.

Table 2. Patient characteristics (n=221).

Characteristics	Values
Age (years), mean (SD)	32.9 (12.68)
Baseline Patient Health Questionnaire-9 score, mean (SD)	13.82 (5.36)
Baseline Generalized Anxiety Disorder-7 score, mean (SD)	12.26 (4.97)
Sex, n (%)	
Male	63 (29)
Female	158 (71)
Religion, n (%)	
No religious group or secular	143 (65)
Other	72 (33)
N/A[a]	6 (2)
Sexual orientation, n (%)	
Heterosexual	193 (87)
Other	20 (9)
N/A	8 (4)
Employment, n (%)	
Employed full time	166 (75)
Other	55 (25)
Psychiatric medication, n (%)	
Prescribed and taking	87 (39)
Other	134 (61)
L ong-term condition , n (%)	
No	176 (80)
Yes	41 (19)
N/A	4 (1)
Program assigned, n (%)	
Comorbid	104 (47)
Depression	50 (23)
Anxiety	67 (30)

[a]N/A: not available.

The *t* test and Chi-square test comparisons showed no significant differences (all P>.05) between the 35 excluded participants and those included in terms of baseline PHQ-9 and GAD-7 scores and demographics, such as age, sex, religion, sexual orientation, employment status, psychiatric medication status, LTCs, and type of program.

Latent Class Growth Analysis

Latent class models were constructed with an increasing number of classes, from 2 to 8 (Tables S3 and S4). The goodness-of-fit was assessed using Bayesian information criterion for each model to determine the optimal number of classes. In addition, the interpretability of the identified trajectories as well as their clinical meaningfulness were considered when choosing a model. On the basis of Bayesian information criterion index and theoretical considerations, the models with 5 classes were chosen for both the PHQ-9 and GAD-7.

PHQ-9 Classes Description

A graphical representation of the 5 classes of depression trajectories can be found in Figure 1, where the individual patient trajectories and the mean trajectory for each class are shown. The characteristics of each class are summarized in Table 3. The 5 classes were as follows: "stable high" symptoms (10/221, 4.5%), "improving high" symptoms (17/221, 7.7%), "improving moderate" symptoms (77/221, 34.8%), "stable moderate" symptoms (56/221, 25.3%), and "improving low" symptoms (61/221, 27.6%). At baseline, the mean PHQ-9 scores for classes 1 ("stable high") and 2 ("improving high") were similar, but the patients in class 2 had a sharp decrease in symptoms across the treatment journey, whereas those in class 1 retained consistently high scores throughout treatment throughout the 12 weeks. Patients in class 3, "improving moderate," showed a similar but slower decrease in symptoms to class 2 ("improving high"). Class 2 was the largest class, and the 77 individuals in it started with moderate PHQ-9 scores and consistently improved across the treatment journey, reaching a mean PHQ-9 score of 6.64 (SD 3.89) at 12 weeks. The patients in class 4, "Stable moderate," started with moderate levels of depression and had a slight decrease in scores, remaining in the moderate range at the end of the intervention. Patients in class 5, "improving low," started with subclinical levels at baseline and slowly but consistently decreased across the treatment journey.

Figure 1. Trajectory classes for depression (left) and anxiety (right). GAD: Generalized Anxiety Disorder; PHQ: Patient Health Questionnaire.

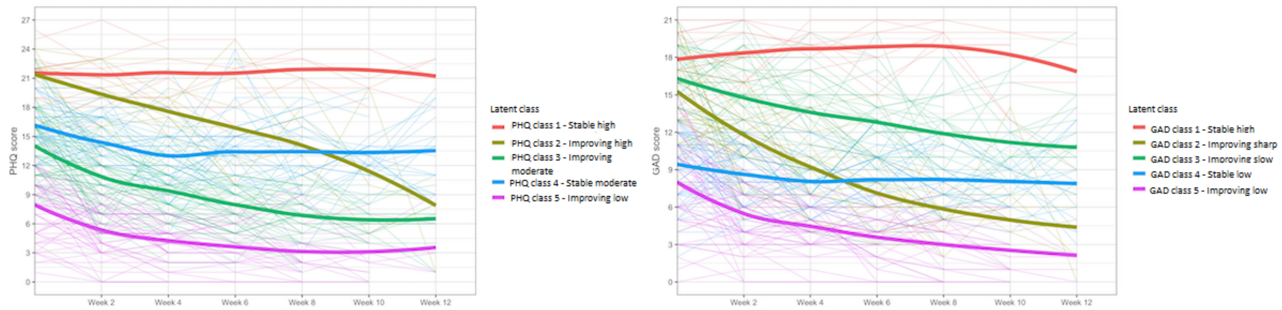


Table 3. Patient Health Questionnaire-9 (PHQ-9) class characteristics.

	PHQ class 1 (“stable high”; n=10)	PHQ class 2 (“improving high”; n=17)	PHQ class 3 (“improving moderate”; n=77)	PHQ class 4 (“stable moderate”; n=56)	PHQ class 5 (“improving low”; n=61)
Participants (n=221), n (%)	10 (4.52)	17 (7.69)	77 (34.84)	56 (25.34)	61 (27.6)
Age (years), mean (SD)	30.1 (13.92)	34.47 (13.31)	31.31 (11.43)	32.38 (13.48)	35.41 (13.01)
Baseline PHQ-9 score, mean (SD)	21.6 (2.95)	21.41 (2.58)	14.12 (3.52)	16.05 (2.94)	8 (3.61)
Week 12 PHQ-9 score, mean (SD)	21.4 (2.07)	7.88 (3.64)	6.64 (3.89)	13.64 (3.55)	3.64 (2.10)
Baseline GAD-7[a] score, mean (SD)	17.4 (3.47)	16.65 (3.00)	11.96 (5.03)	13.55 (4.32)	9.39 (4.14)
Week 12 GAD-7 score, mean (SD)	14.2 (5.31)	5.75 (3.11)	5.56 (3.80)	11.42 (4.46)	3.72 (3.58)
Sex (male), n (%)	1 (10)	6 (35)	25 (32)	16 (29)	15 (25)
Employment status (employed full time), n (%)	6 (60)	14 (82)	60 (78)	36 (64)	50 (82)
Psychiatric medication status (prescribed and taking), n (%)	7 (70)	12 (71)	29 (38)	25 (45)	14 (23)
Long-term condition (no), n (%)	8 (80)	12 (71)	61 (79)	45 (80)	50 (82)
Program type, n (%)					
Comorbid	6 (60)	10 (59)	35 (45)	32 (57)	21 (34)
Depression	1 (10)	6 (35)	21 (27)	14 (25)	8 (13)
Anxiety	3 (30)	1 (6)	21 (27)	10 (18)	32 (52)

[a]GAD-7: Generalized Anxiety Disorder-7.

GAD-7 Classes Description

A graphical representation of the 5 classes of anxiety trajectories can be found in Figure 1 (also Figure S3 in Multimedia Appendix 1), where the individual patient trajectories and the mean trajectory for each class are shown. The characteristics of each class are summarized in Table 4. The 5 classes were as follows: “stable high” symptoms (16/221, 7.2%), “improving sharp” symptoms (36/221, 16.3%), “improving slow” symptoms (54/221, 24.4%), “stable low” symptoms (49/221, 22.2%), and “improving low” symptoms (66/221, 29.9%). Class 1 (“stable high”) was the smallest, consisting of only 16 individuals whose trajectories of change were marked by consistently high scores from the baseline to 12 weeks. Patients in class 2 (“improving sharp”) and class 3 (“improving slow”) both started with severe anxiety, but the former had a much steeper decline in symptoms across the treatment journey. Patients in class 3, “improving slow,” started with severe levels of anxiety and had a slow and consistent decrease in scores. Class 4, “stable low,” consisted of participants who started and finished with mild levels with minimal changes in symptoms. Finally, patients in class 5, “improving low,” started with mild symptoms at baseline and slowly but consistently transitioned to minimal symptoms.

One-way ANOVAs and Tukey post hoc comparisons revealed significant differences between depression classes in baseline PHQ-9 ($F_{4,216}$=89.21; P<.001) and baseline GAD-7 ($F_{4,216}$=15.21; P<.001) scores. A similar difference was found between anxiety classes for both the baseline PHQ-9 ($F_{4,216}$=23.14; P<.001) and GAD-7 ($F_{4,216}$=82.59; P<.001) scores. Chi-square tests and Bonferroni-corrected pairwise comparisons also revealed significant differences between some of the depression classes regarding medication status (n=221, χ^2_4=18.5, P<.001) and the type of program (n=221, χ^2_8=25.7, P=.001). Depression class 5 (“improving low”) had a significantly lower percentage (14/61, 23%) of members with “prescribed and taking” medication compared with classes 2 (12/17, 70.6% “improving high”; P=.003) and 1 (7/10, 70.0% “stable high”; P=.003). Class 5 was also significantly different from classes 4 (“stable moderate”; P=.003) and 2 (“improving high”; P=.02) in terms of program type, with a much higher percentage (32/61, 52%) of members of class 5 being in the anxiety program compared with the other 2 classes where anxiety seemed to be the least common program (17.9% in the stable moderate class, and 5.9% in the improving high class). These results suggest that there are differences in the severity of depression and anxiety at baseline, with some of the identified PHQ and GAD classes starting with higher levels. Moreover, depression class 5 (“improving low”) seemed to distinguish itself from other depression classes by having more members in the anxiety program and a lower number of members taking medication. No other significant differences (all P>.05) were found between depression or anxiety classes in terms of age, sex, employment status, and the presence of LTCs.

Table 4. Generalized Anxiety Disorder-7 (GAD-7) class characteristics.

	GAD class 1 ("stable high"; n=16)	GAD class 2 ("improving sharp"; n=36)	GAD class 3 ("improving slow"; n=54)	GAD class 4 ("stable low"; n=49)	GAD class 5 ("improving low"; n=66)
Patients (n=221), n (%)	16 (7.24)	36 (16.29)	54 (24.43)	49 (22.17)	66 (29.86)
Age (years), mean (SD)	33.06 (13.16)	31.22 (10.78)	31.46 (12.48)	30.86 (11.57)	36.47 (13.74)
Baseline PHQ-9[a] score, mean (SD)	19.69 (4.03)	15.08 (5.47)	16.65 (4.65)	12.43 (3.94)	10.42 (4.37)
Week 12 PHQ-9 score, mean (SD)	15.38 (5.34)	5.73 (2.99)	12.1 (4.60)	10.41 (5.92)	4.19 (3.31)
Baseline GAD-7 score, mean (SD)	17.88 (2.16)	15.28 (3.22)	16.31 (2.85)	9.41 (2.91)	8.06 (3.67)
Week 12 GAD-7 score, mean (SD)	17.13 (3.83)	4 (1.96)	10.97 (3.54)	8 (3.33)	2.08 (1.78)
Sex (male), n (%)	3 (19)	9 (25)	14 (26)	9 (18)	18 (27)
Employment status (employed full time), n (%)	12 (75)	28 (78)	39 (72)	37 (76)	50 (76)
Psychiatric medication status (prescribed and taking), n (%)	9 (56)	13 (36)	28 (52)	17 (35)	20 (30)
Long-term condition (no), n (%)	14 (88)	31 (86)	40 (74)	38 (78)	53 (80)
Program type, n (%)					
Comorbid	11 (69)	15 (42)	28 (52)	27 (55)	23 (35)
Depression	1 (6)	5 (14)	11 (20)	13 (27)	20 (30)
Anxiety	4 (25)	16 (44)	15 (28)	9 (18)	23 (35)

[a]PHQ-9: Patient Health Questionnaire-9.

Platform Use

Overall, participants spent an average of 312 minutes on the platform, logged in approximately 18 times, viewed 57% of the total program, completed 150 activities, and received 4.26 reviews from their supporters (Table 5 provides descriptive information on platform use). It is noteworthy that there is large variability in all these use metrics, indicating considerable individual differences.

The results of the regressions indicate that age was a significant predictor of the number of log-ins (β=.13; *P*=.04), length of time spent on the internet (β=166.62; *P*=.03), and number of reviews (β=−.02; *P*=.009), whereas the presence of an LTC was a significant predictor of the number of activities (β=−43.5; *P*=.02), and the type of program was a significant predictor of the percentage of programs viewed (β=.07; *P*<.001). Overall, older people had a higher number of log-ins and spent more time on the platform but fewer reviews. The evaluation of box plots and descriptive summaries further showed that patients with an LTC had a higher number of activities compared with those without an LTC, and patients in the comorbid program had a smaller percentage of programs viewed compared with those in the depression or anxiety programs. No other baseline characteristics or demographic variables were found to be significant predictors of use (all *P*>.05).

Table 5. Platform use.

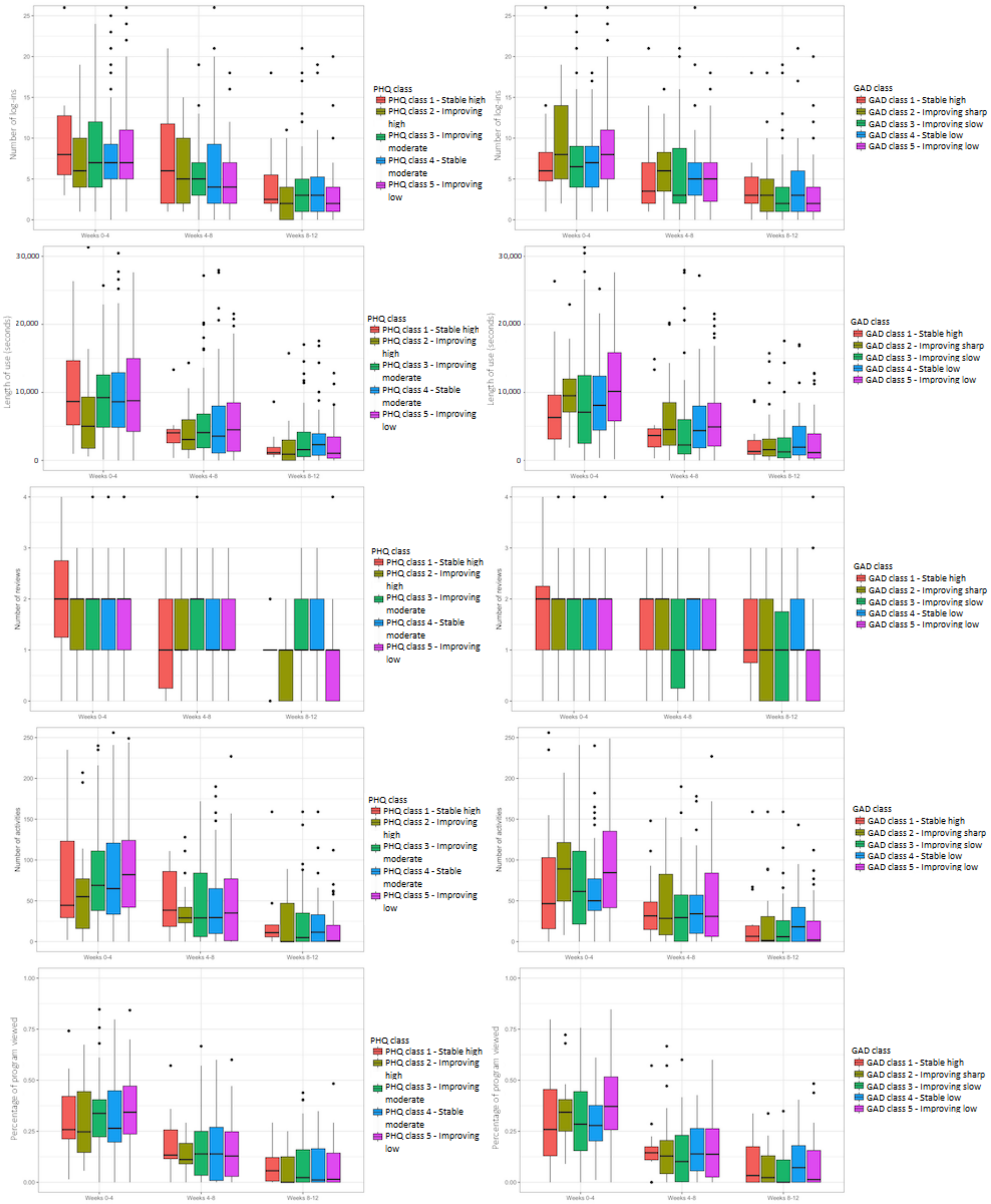
	Values, mean (SD)	Values, median (range)
Number of log-ins (Winsorized)	17.67 (11.13)	15 (1-70)
Length of use (Winsorized; minutes)	312.16 (228.83)	274.92 (2.47-1376.55)
Number of reviews	4.26 (1.63)	5 (0-8)
Number of activities (Winsorized)	150.48 (108.47)	126 (0-646)
Percentage of program viewed	0.57 (0.25)	0.58 (0.01-1)

Relationship Between Outcomes and Use

Mixed ANOVAs were conducted to identify the effect of depressive symptom trajectory and time on use. There were no significant interactions of time by trajectory (all *P*>.05). For all 5 use metrics: the number of log-ins ($F_{2,432}$=53.37; *P*<.001), length of use ($F_{2,432}$=70.51; *P*<.001), number of reviews ($F_{2,432}$=28.13; *P*<.001), number of activities ($F_{2,432}$=56.20; *P*<.001), and percentage of program viewed ($F_{2,432}$=86.34; *P*<.001), the ANOVAs revealed a significant effect of time. Paired *t* tests with *P*-adjusted Bonferroni post hoc analyses indicated that there were significant differences between all time points for all 5 use metrics (all *P*<.001), with most use occurring in the first 4 weeks, followed by weeks 4 to 8, and the least use occurring in the last 4 weeks (8-12). A detailed set of comparisons has been included as supplementary material for anyone who wants to delve deeper into the data (Tables S5 and S6 in Multimedia Appendix 1).

Mixed ANOVAs similar to the ones mentioned earlier were conducted to explore the effect of anxiety trajectory and time on usage. The Mauchly test demonstrated that the sphericity assumption had been violated, so Greenhouse-Geisser corrections were applied. There were significant interaction effects of time by trajectory for the number of log-ins ($F_{8,432}$=1.99; P=.045) and number of activities ($F_{8,432}$=1.98; P=.047); however, these effects were no longer significant after applying Greenhouse-Geisser corrections (P>.05). Interaction effects were not statistically significant for other use metrics (all P>.05). For the number of logins ($F_{2,432}$=65.43; P<.001), length of use ($F_{2,432}$=91.46; P<.001), number of reviews ($F_{2,432}$=30.57; P<.001), number of activities ($F_{2,432}$=85.62; P<.001), and percentage of programs viewed ($F_{2,432}$=121.85; P<.001), the ANOVAs revealed a significant effect of time. . For all use metrics where a main effect of time was observed in the ANOVA, paired *t* tests with *P*-adjusted Bonferroni post hoc analyses indicated that there were significant differences between all time points (all P<.001), with most use occurring in the first 4 weeks, followed by 8 weeks, and the least at 12 weeks. Moreover, in the graphs for both depression and anxiety (Figure 2), it can be observed that there is substantial variability in use between the PHQ and GAD classes, where most of the variability between classes seems to occur in the early stages of treatment (first 4 weeks), and the differences between classes in the other 2 periods (weeks 4-8 and weeks 8-12) are less obvious. In addition, even after the Winsorization of the use metrics, the graphs allow us to see a high number of outliers.

Figure 2. Differences in usage between trajectory classes at different time points. GAD: Generalized Anxiety Disorder; PHQ: Patient Health Questionnaire.



Discussion

Principal Findings

This study aimed to identify different trajectories of symptom change across supported iCBT treatment for depression and anxiety using continuous outcomes data from patients treated in a mental health service setting. In addition, this study aimed to explore the relationship between the identified trajectories and use data. Overall, we found high heterogeneity in treatment response, with 5 latent classes emerging from data for both depression and anxiety. Across the 5 classes for depression and anxiety, we identified 3 improving classes and 2 classes that showed little to no change in symptoms. Across all classes, we also found an effect of time, with most use occurring in the first 4 weeks.

Understanding the different types of responses to an intervention as well as factors that may influence those responses could help improve the quality and delivery of iCBT interventions. This study found that approximately 70% of the sample improved across treatments; however, these individuals experienced different trajectories of change based on 3 characteristics: their baseline score, the pace of the improvement, and their posttreatment symptom scores. Some of the improver classes identified in this study are consistent with other iCBT studies that found individuals who started with moderate or moderate-severe scores and progressed toward recovery at a steady, moderate pace [19-22] and individuals who started with milder baseline symptoms and improved at a slower pace [20,21]. However, in general, our results showed steady improvements across improver classes instead of the early improvement classes observed in iCBT research [22]. These differences could be partially attributed to the nature of the ROM assessments linked to supporter reviews, as opposed to fixed time points. Of importance, we also found a class with severe anxiety (anxiety class 3) who, despite showing steady improvements, were still within the clinical definition of anxiety after treatment. Patients showing these trajectories could benefit from extending treatment or even adding high-intensity clinical interventions, such as face-to-face therapy, to support continued improvement that could help them achieve recovery [41].

It is perhaps even more important to understand the class characteristics of the individuals that see smaller to no gains so that in future, we could identify them early on, monitor their trajectory of symptoms, and make any necessary treatment decisions earlier to maximize treatment benefits. In particular, the 2 classes of nonresponders that start in the moderate severity range and see limited improvements (depression class 4 and anxiety class 4) could be ripe candidates for monitoring more closely symptom change trajectories and perhaps identify symptom change thresholds, whereby, from week to week, if thresholds are not met for symptom change, it could result in treatment decisions being made. Although some prior iCBT studies [19,22] have identified a class of limited to no improvement akin to depression class 4 and anxiety class 4 mentioned earlier, the nonresponder severe class found here has not been reported in other studies, perhaps because not many studies have included baseline scores on the severe end. It is possible that individuals in this severe nonresponder class need more support or to be stepped up in their care or that the interventions are unsuitable because they are primarily developed for mild to moderate ranges of symptom presentation. Therefore, the early identification of patients with high symptoms at baseline and an unchanged trajectory of symptom change early in treatment may also support better clinical decision-making.

It is also worth mentioning that we did not find a deterioration class similar to others [22], which could be a result of the intervention used here, or because of differences in the sample and methods used. In our primary RCT, where the current sample data originated, among 8-week measure completers, 5.2% (10/194) of participants in the intervention arm deteriorated (ie, increases in PHQ-9 score ≥6 or GAD-7 score ≥4) [42], which is in line with a recent individual patient data meta-analysis [43] that suggested only 5.8% of individuals in the intervention groups showed deterioration. Therefore, individuals with deteriorating trajectories in the current sample could be too few to create a subgroup of their own and may be mixed across the nonimproving classes.

A better understanding of the attributes of the classes identified has implications for tailoring, intervention delivery, and the early identification of individuals who are not on an improving path. We found no significant associations between the classes and baseline sociodemographic variables (ie, age, sex, employment status, and LTC), which is consistent with some studies [18,21], although other studies found that female individuals were more likely to be in the high-severity class [20]. Overall, the small classes with high interpersonal variability may have made it more difficult to identify differences between the groups, but future work could investigate other moderators related to clinical variables instead of demographic variables.

The results did not show a time-varying relationship between use and the various trajectories of change, indicating that the effect of platform use did not result in immediate clinical gains. This is consistent with findings from a study by Zeng et al [28], who only found this association by the end of treatment and therefore calls into question the potential role of use as a mechanism of change in iCBT. Studies with large observational samples should be conducted to detect these effects because the association between use and outcomes is more consistently found in large samples with aggregated data [12,24], whereas small-scale studies lead to more inconsistent findings [25,26]. However, we did find a consistent effect of time on the overall use of the intervention across classes and outcomes, suggesting that all participants used the intervention significantly more in the first 4 weeks. This is an important finding, as it replicates work from our group from a previous RCT [26], and the finding is consistent with other published literature [26,27,44]. Attention could be given to help maximize patients' use of the intervention early in treatment. This can be achieved through frontloading key content or by increasing the schedule of support and guidance in the first 4 weeks.

The predictors of use were also investigated to better understand overall platform use. Age was a significant predictor for some of the use metrics, with older clients logging in more frequently and for longer periods, which is consistent with previous findings [27,45,46]. The content of the intervention may be better suited for older clients or older clients may need or have more time to read the material [45,46]. Alternatively, there may be a need to tailor the content in terms of cognitive load, delivery mode, and time commitment across different age groups.

Limitations

A limitation of this study is the observational nature of this substudy, with no manipulation of the variables related to use, which makes it impossible to establish causal relationships between use metrics and outcomes. Moreover, using routine continuous outcome data compared with regular time point assignments comes with challenges, such as many missing questionnaires or complex data processing required to retrieve the relevant assignments. The small size of some of the classes presents another limitation, and future studies may benefit from larger sample sizes that could allow the detection of smaller effects and could provide an opportunity to apply other methods (eg, growth mixture modeling) to allow for both between- and within-class variability. Another limitation was the lack of access to baseline clinical information, which could have been useful for investigating differences between classes. For instance, variables that have been previously linked to treatment response include previous episodes of depression and anxiety [47], previous treatment or medication [48], client expectations [49], and treatment credibility [48,50]. Moreover, some of these clinical variables could be comorbidities and have a significant effect on the clusters found here. Further work is necessary to investigate and understand the role of these other clinical characteristics on the classes identified here.

Conclusions

This study identified 5 distinct classes of symptom trajectories for depression and anxiety over the course of iCBT treatment. The results showed that although iCBT works for the majority, the way improvement occurs varies, which may have implications for how iCBT is delivered. The absence of effects on the time-varying relationship between platform use and trajectories calls into question the role of use as a mechanism for change. Other contextual information and larger sample sizes may need to be presented to explore these effects better. Further work is necessary to better understand these patterns of change, as well as factors impacting them as insights gained, which may be useful in tailoring treatments for different patient groups and in identifying and monitoring patient groups to enable earlier and enhanced treatment decisions.

Acknowledgments

The authors would like to thank the staff at Berkshire National Health Service Foundation Trust for providing assistance that allowed us to successfully complete the execution of the original study, which is the basis of this secondary analysis.

Conflicts of Interest

DCC, AE, JEP, DD, SM, and DR are employees of SilverCloud Health. DR is a shareholder in Amwell, a company that SilverCloud Health has been a subsidiary of since 2021. SM was an employee at Amwell while this work was in progress.

Multimedia Appendix 1

Supplementary tables and figures with additional analysis information.

[DOCX File , 1060 KB - mhealth_v11i1e41815_app1.docx]

References

1. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004(420):21-27. [doi: 10.1111/j.1600-0047.2004.00327.x] [Medline: 15128384]
2. Kessler RC, Aguilar-Gaxiola S, Alonso J, Chatterji S, Lee S, Ormel J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. Epidemiol Psichiatr Soc 2009 Jan;18(1):23-33 [FREE Full text] [doi: 10.1017/s1121189x00001421] [Medline: 19378696]
3. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013;34:119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
4. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 2011 Nov;31(7):1117-1125. [doi: 10.1016/j.cpr.2011.07.004] [Medline: 21820991]
5. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
6. Nathan PE, Gorman JM. A Guide to Treatments That Work. 4th edition. Oxford, UK: Oxford University Press; Jul 2015.
7. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]

8. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010 Dec 07;8:82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]
9. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
10. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
11. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
12. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):18r12188 [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
13. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: 10.2196/18100] [Medline: 32865497]
14. Beintner I, Vollert B, Zarski AC, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
15. Müller VN, Boyle K, Zimmermann D, Weinmann-Lutz B, Rubel JA, Lutz RW. What is individually tailored mental health care? Rev Argentina de Clin Psicol 2018;27(2):157-181 [FREE Full text] [doi: 10.24205/03276716.2018.1062]
16. Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. J Affect Disord 2009 Nov;118(1-3):60-68. [doi: 10.1016/j.jad.2009.01.019] [Medline: 19217669]
17. Skriner LC, Chu BC, Kaplan M, Bodden DH, Bögels SM, Kendall PC, et al. Trajectories and predictors of response in youth anxiety CBT: integrative data analysis. J Consult Clin Psychol 2019 Feb;87(2):198-211. [doi: 10.1037/ccp0000367] [Medline: 30570308]
18. Sunderland M, Wong N, Hilvert-Bruce Z, Andrews G. Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. Behav Res Ther 2012 Jun;50(6):374-380. [doi: 10.1016/j.brat.2012.03.005] [Medline: 22498311]
19. Batterham PJ, van Spijker BA, Mackinnon AJ, Calear AL, Wong Q, Christensen H. Consistency of trajectories of suicidal ideation and depression symptoms: evidence from a randomized controlled trial. Depress Anxiety 2019 Apr;36(4):321-329. [doi: 10.1002/da.22863] [Medline: 30466148]
20. March S, Batterham PJ, Rowe A, Donovan C, Calear AL, Spence SH. Trajectories of change in an open-access internet-based cognitive behavior program for childhood and adolescent anxiety: open trial. JMIR Ment Health 2021 Jun 18;8(6):e27981 [FREE Full text] [doi: 10.2196/27981] [Medline: 34142971]
21. Zagorscak P, Heinrich M, Bohn J, Stein J, Knaevelsrud C. How individuals change during internet-based interventions for depression: a randomized controlled trial comparing standardized and individualized feedback. Brain Behav 2020 Jan;10(1):e01484 [FREE Full text] [doi: 10.1002/brb3.1484] [Medline: 31777204]
22. Lutz W, Arndt A, Rubel J, Berger T, Schröder J, Späth C, et al. Defining and predicting patterns of early response in a web-based intervention for depression. J Med Internet Res 2017 Jun 09;19(6):e206 [FREE Full text] [doi: 10.2196/jmir.7367] [Medline: 28600278]
23. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
24. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
25. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
26. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
27. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
28. Zeng Y, Guo Y, Li L, Hong YA, Li Y, Zhu M, et al. Relationship between patient engagement and depressive symptoms among people living with HIV in a mobile health intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e20847 [FREE Full text] [doi: 10.2196/20847] [Medline: 33118956]

29. Richards D, Duffy D, Blackburn B, Earley C, Enrique A, Palacios J, et al. Digital IAPT: the effectiveness and cost-effectiveness of internet-delivered interventions for depression and anxiety disorders in the improving access to psychological therapies programme: study protocol for a randomised control trial. BMC Psychiatry 2018 Mar 02;18(1):1-3 [FREE Full text] [doi: 10.1186/s12888-018-1639-5]
30. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
31. Depression in adults: treatment and management. National Institute for HealthCare Excellence. 2022 Jun 29. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2021-05-17]
32. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
34. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
35. Schmiege SJ, Meek P, Bryan AD, Petersen H. Latent variable mixture modeling: a flexible statistical approach for identifying and classifying heterogeneity. Nurs Res 2012 May;61(3):204-212. [doi: 10.1097/NNR.0b013e3182539f4c] [Medline: 22551995]
36. Jung T, Wickrama KA. An introduction to latent class growth analysis and growth mixture modeling. Soc Personal Psychol Compass 2008 Jan;2(1):302-317 [FREE Full text] [doi: 10.1111/j.1751-9004.2007.00054.x]
37. Berlin KS, Williams NA, Parra GR. An introduction to latent variable mixture modeling (part 1): overview and cross-sectional latent class and latent profile analyses. J Pediatr Psychol 2014 Mar;39(2):174-187. [doi: 10.1093/jpepsy/jst084] [Medline: 24277769]
38. Wardenaar KJ. Latent class growth analysis and growth mixture modeling using R: a tutorial for two R-packages and a comparison with Mplus. PsyArXiv Preprint posted online on April 7, 2020. [FREE Full text] [doi: 10.31234/osf.io/m58wx]
39. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2020. URL: https://www.r-project.org/ [accessed 2021-03-21]
40. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R Package 1cmm. J Stat Soft 2017;78(2):1-56 [FREE Full text] [doi: 10.18637/jss.v078.i02]
41. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. Front Psychiatry 2020 Jan 09;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
42. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020 Jun 15;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
43. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. Psychol Med 2018 Nov;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
44. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
45. Castro A, López-Del-Hoyo Y, Peake C, Mayoral F, Botella C, García-Campayo J, et al. Adherence predictors in an internet-based intervention program for depression. Cogn Behav Ther 2018 May;47(3):246-261. [doi: 10.1080/16506073.2017.1366546] [Medline: 28871896]
46. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an internet intervention for depression. J Affect Disord 2018 Mar 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
47. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017 May;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
48. El Alaoui S, Ljótsson B, Hedman E, Svanborg C, Kaldo V, Lindefors N. Predicting outcome in internet-based cognitive behaviour therapy for major depression: a large cohort study of adult patients in routine psychiatric care. PLoS One 2016 Sep 12;11(9):e0161191 [FREE Full text] [doi: 10.1371/journal.pone.0161191] [Medline: 27618548]
49. Hedman E, Andersson E, Ljótsson B, Andersson G, Andersson E, Schalling M, et al. Clinical and genetic outcome determinants of internet- and group-based cognitive behavior therapy for social anxiety disorder. Acta Psychiatr Scand 2012 Aug;126(2):126-136. [doi: 10.1111/j.1600-0447.2012.01834.x] [Medline: 22320999]

50. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. PLoS One 2015 Apr 20;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder-7
IAPT: Improving Access to Psychological Therapies
iCBT: internet-delivered cognitive behavioral therapy
LCGA: latent class growth analysis
LTC: long-term condition
PHQ-9: Patient Health Questionnaire-9
PWP: psychological well-being practitioner
RCT: randomized controlled trial
ROM: routine outcome monitoring

Edited by L Buis; submitted 09.08.22; peer-reviewed by J Wielgosz, C Webb, B Chu; comments to author 06.10.22; revised version received 23.11.22; accepted 30.11.22; published 12.07.23.

Please cite as:
Cumpanasoiu DC, Enrique A, Palacios JE, Duffy D, McNamara S, Richards D
Trajectories of Symptoms in Digital Interventions for Depression and Anxiety Using Routine Outcome Monitoring Data: Secondary Analysis Study
JMIR Mhealth Uhealth 2023;11:e41815
URL: https://mhealth.jmir.org/2023/1/e41815
doi:10.2196/41815
PMID:37436812

Original Paper

Longer-Term Effects of Cardiac Telerehabilitation on Patients With Coronary Artery Disease: Systematic Review and Meta-Analysis

Wen Zhong[1,2*], BSc; Rui Liu[1,2*], BSc; Hongxin Cheng[1,2], MD; Lin Xu[1,2], BSc; Lu Wang[1,2], MD; Chengqi He[1,2], MD, PhD; Quan Wei[1,2], MD, PhD

[1]Department of Rehabilitation Medicine and Institute of Rehabilitation Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China

[2]Key Laboratory of Rehabilitation Medicine in Sichuan Province, Chengdu, Sichuan, China

[*]these authors contributed equally

Corresponding Author:
Quan Wei, MD, PhD
Department of Rehabilitation Medicine and Institute of Rehabilitation Medicine
West China Hospital
Sichuan University
No 37 Guoxue Alley
Wuhou District, Sichuan Province
Chengdu, Sichuan, 610041
China
Phone: 86 2885422847
Fax: 86 2828423819
Email: weiquan@scu.edu.cn

Abstract

Background: Cardiac telerehabilitation offers a flexible and accessible model for patients with coronary artery disease (CAD), effectively transforming the traditional cardiac rehabilitation (CR) approach.

Objective: This systematic review and meta-analysis aimed to evaluate the long-term effectiveness of cardiac telerehabilitation.

Methods: We searched randomized controlled trials (RCTs) in 7 electronic databases: PubMed, Web of Science, EMBASE, the Cochrane Central Register of Controlled Trials, ClinicalTrials.gov, the China National Knowledge Infrastructure, and WANFANG. The primary outcome focused on cardiopulmonary fitness. For secondary outcomes, we examined cardiovascular risk factors (blood pressure, BMI, and serum lipids), psychological scales of depression and anxiety, quality of life (QoL), cardiac telerehabilitation adherence, and adverse events.

Results: In total, 10 RCTs fulfilled the predefined criteria, which were reviewed in our meta-analysis. The results showed that after cardiac telerehabilitation, there was a significant difference in the improvement in long-term peak oxygen uptake compared to center-based CR (mean difference [MD] 1.61, 95% CI 0.38-2.85, *P*=.01), particularly after 6-month rehabilitation training (MD 1.87, 95% CI 0.34-3.39, *P*=.02). The pooled effect size of the meta-analysis indicated that there were no significant differences in the reduction in cardiovascular risk factor control. There was also no practical demonstration of anxiety scores or depression scores. However, cardiac telerehabilitation demonstrated an improvement in the long-term QoL of patients (MD 0.92, 95% CI 0.06-1.78, *P*=.04). In addition, the study reported a high completion rate (80%) for cardiac telerehabilitation interventions. The incidence of adverse events was also low during long-term follow-up.

Conclusions: Cardiac telerehabilitation proves to be more effective in improving cardiopulmonary fitness and QoL during the long-term follow-up for patients with CAD. Our study highlights monitoring-enabled and patient-centered telerehabilitation programs, which play a vital role in the recovery and development of CAD and in the long-term prognosis of patients.

(JMIR Mhealth Uhealth 2023;11:e46359) doi:10.2196/46359

KEYWORDS

cardiac telerehabilitation; coronary artery disease; CAD; cardiac rehabilitation; CR; long-term effect; meta-analysis

Introduction

Cardiovascular diseases (CVDs) are the leading cause of a large proportion of deaths worldwide, accounting for approximately 1/3 of all deaths [1]. The global burden of CVDs is rising, increasing from 271 million in 1990 to 523 million in 2019, and the total number of cases of CVDs have almost doubled, especially coronary artery disease (CAD) [2]. CAD occurs mainly due to the progression of atherosclerosis of coronary arteries to narrowing or occlusion, leading to blood flow limitation, which causes cardiomyocyte or myocardial necrosis [3]. Population growth and aging are the main drivers, and the prevention and control of CVDs face significant challenges.

Cardiac rehabilitation (CR) is a complex, multidisciplinary intervention aimed at comprehensive rehabilitation assessment of the patient's condition to meet the needs of patients with CVD, preventing disease recurrence and further progression. CR involves various components, such as exercise training, drug counseling, diet, nutrition, psychological regulation, and risk factor management [4]. To better serve patients with CAD, this multidisciplinary treatment method can be combined with well-validated strategies to assess physical function and risk factors in order to personalize treatment strategies [5]. Studies have shown that CR can reduce the risk of recurrent heart attacks by 47%, heart disease mortality by 36%, and all-cause mortality by 26% [6]. Furthermore, secondary prevention recommendations emphasize CR's importance in patients with CAD, which has been recognized as comprehensive medical monitoring to reduce CAD mortality, morbidity, disability, and high expense [7,8]. Although participation in CR is a class IA recommendation for CAD [9], rates of referral and usage remain low [8]. Aragam et al [10] demonstrated that approximately >40% of patients are not referred for CR after percutaneous coronary intervention (PCI) by the time of hospital discharge. Participation enrollment in CR ranges from only 20% to 30% in the United States [8]. Low participation and adherence to CR programs may be attributed to multifactorial conditions, such as comorbidities, living farther from medical organizations, high medical costs, and time commitment [11,12].

To alleviate these barriers and improve the uptake rates of CR, cardiac telerehabilitation, a targeted approach, is used to effectively shift the traditional rehabilitation mode to a high-value overall strategy. Telerehabilitation is defined as a telemedicine platform, including telediagnosis, teletreatment, and remote monitoring [13]. These technologies are conducive to the joint participation of doctors and patients and medical departments in patients' health management work. For patients with CVD, doctors can monitor the patients' vital signs and cardiac telerehabilitation progress through a remote monitoring system and adjust the CR treatment plan according to the patients' condition. Cardiac telerehabilitation uses internet information technology to allow patients to receive rehabilitation treatment at home or in other nonhospital settings regardless of time and geographical restrictions. This can motivate and help supervise patients, improving patient compliance. These factors have promoted the popularity of cardiac telerehabilitation, and the comprehensive promotion of telemedicine construction and development is gradually focusing on this aspect. Most patients eligible for cardiac telerehabilitation have a low rate of adverse events during exercise training if previously adequately evaluated. A review by Stefanakis et al [14], which included 5 studies on adverse event rates in home telerehabilitation, estimated the incidence of adverse events in the sample to be 1 in 23,823 patient-hours of exercise.

As a more accessible and flexible model of CR, cardiac telerehabilitation has been developed based on new communication technologies and advanced telemedical devices, such as smartphones, web-based apps, wearable sensors, and virtual reality [15]. This is supported by recent meta-analyses [16,17] that have shown that cardiac telerehabilitation as an alternative rehabilitation delivery model achieves an equivalent effect on physical exercise capacity, behavior change, reduction in risk factors, and improvement in the quality of life (QoL) of patients with CAD compared to the traditional rehabilitation model. At present, in addition to the traditional outpatient rehabilitation model, the main approaches to the CR model include home-based CR and cardiac telerehabilitation. Both cardiac telerehabilitation and home-based CR refer to rehabilitation in a nonhospital setting and have their advantages. Telerehabilitation is delivered and implemented through telemedical equipment, while home-based CR refers to rehabilitation carried out in the patient's home. Both approaches can serve as important supplements or alternatives to traditional in-hospital rehabilitation models. However, the main difference between the 2 approaches is that home-based CR patients rely on outpatient or community follow-up guidance, take subjective initiatives at home for rehabilitation, and lack uninterrupted supervision and guidance from doctors; cardiac telerehabilitation makes up for these disadvantages of home-based CR. Therefore, the control of patients participating in telerehabilitation is strict. In a 2019 joint statement, the American Association of Cardiovascular and Pulmonary Rehabilitation (AACVPR), the American Heart Association (AHA), and the American College of Cardiology (ACC) [5] suggested that cardiac telerehabilitation could be a reasonable alternative for patients with clinically stable, low-to-moderate-risk CVD. However, high-risk patients, such as patients with unstable angina, CVD with heart failure or symptomatic arrhythmias, and other hemodynamic instabilities, usually require careful evaluation, outpatient CR under supervision, and reassessment for stabilization during the convalescent phase before participating in cardiac telerehabilitation.

Cardiac telerehabilitation has been found to be effective, as evidenced by improvements in the condition of patients with CVD. Still, current telerehabilitation studies have aimed to assess whether telerehabilitation affects short-term (about 3-month follow-up) or medium-term (about 6-month follow-up) effectiveness [18-20]. There are limited data describing the long-term (more than 1-year follow-up) effects of telerehabilitation. A self-regulation lifestyle program [21] reported that motivation for lifestyle changes tends to diminish. At the same time, patients with CAD feel healed, which influences long-term beneficial changes in lifestyle and risk factors. Hence, evaluating the long-term effectiveness of telerehabilitation has practical significance for implementing CR. Considering the well-established association between telerehabilitation and the potential beneficial effects in CAD, in this study, we hypothesized that cardiac telerehabilitation could maintain the results for longer-term consequences.

Methods

Study Design

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement.

Ethical Considerations

All analyses were based on previously published studies, so no ethical approval or patient consent was needed.

Literature Search

A literature search was performed to identify relevant studies in the following 7 electronic databases: PubMed, Web of Science, EMBASE, the Cochrane Central Register of Controlled Trials (CENTRAL), ClinicalTrials.gov, the China National Knowledge Infrastructure (CNKI), and WANFANG.

The search proceedings used the following different keywords without a time or language limit: (coronary artery disease or left main coronary artery disease or coronary arteriosclerosis) AND (telerehabilitation or tele-rehabilitation or tele rehabilitation or remote rehabilitation or virtual rehabilitation or telemedicine or mobile health or telehealth or eHealth or internet or online or web or sensor or wearable or smartphone or App or WeChat or QQ). Multimedia Appendix 1 provides the complete search strategy. To retrieve a comprehensive list of eligible papers, we manually screened reference lists, relevant conference lists, and even gray literature.

Inclusion and Exclusion Criteria

We defined and applied explicit inclusion criteria to select eligible studies for this meta-analysis as follows:

- Population: adults (≥18 years old) diagnosed with CVD (stable angina pectoris, acute coronary syndrome, myocardial infarction, or postcoronary revascularization, such as PCI or coronary artery bypass grafting [CABG]).
- Intervention: We focused on telerehabilitation as an emerging model of rehabilitation service delivery based on smartphones, wearable monitoring portable devices, virtual reality, or other internet interventions. Modern telecommunications technology combined with rehabilitation is designed to complete functional exercise training sessions, achieve self-management, and improve physical function. Patients are available at home with remote monitoring and remote health care consultation.
- Comparison: A control group was randomly assigned to usual-care or center-based CR.
- Outcomes: The primary outcome focused on cardiopulmonary function assessed with the peak oxygen uptake (peak VO_2) using the cardiopulmonary exercise test (CPET). For secondary outcomes, we focused on changes in cardiovascular risk factors, psychological scales of depression and anxiety, QoL, cardiac telerehabilitation adherence, and adverse events.
- Study design: Randomized controlled trials (RCTs) that compared the cardiac telerehabilitation group with the control group and evaluated the longer-term effects during least 12 months' follow-up were included in the review.

We also set the following exclusion criteria: (1) patients with severe heart failure with New York Heart Association (NYHA) functional class III or IV, malignant cardiomyopathy, valvular disease, heart failure; (2) papers with an unreasonable literature research design, non-RCTs, nonhuman or animal studies, or a follow-up time of <12 months; (3) repeated publication; (4) unavailable full text, incomplete information and data, or an inability to extract and compare data; and (5) conference papers, abstracts, reviews, letters to the editor, and case reports.

Selection of Studies

Potentially relevant papers meeting the abovementioned search strategy were imported into the EndNote X9.2 tool (Clarivate). Initially, 2 reviewers each independently screened the titles and abstracts of all studies on the finalized list. Next, they conducted full-text screening according to the inclusion and exclusion criteria to determine the final eligibility. During the overall flow of the process, if there were any different views, a third reviewer provided an opinion and resolved the disagreement via consensus.

Data Extraction

Two authors collaborated on the final decision of data extraction, which was summarized in Microsoft Word 2019 and Microsoft Excel 2019: (1) study design (eg, first author, year of publication, country, study design, follow-up time), (2) participants (eg, sample size, sex, age, diagnosis), (3) intervention (eg, telerehabilitation group vs control group), (4) change in our protocol-specified outcomes, and (5) risk of bias. Although relevant details were insufficiently reported in the included studies, we contacted the authors via email for further information.

Risk-of-Bias and Quality Assessment of Studies

We evaluated each study's eligibility using the Cochrane risk-of-bias tool [22] to assess the risk of all types of bias (selection bias, performance bias, attrition bias, reporting bias, and other sources of bias). Furthermore, we also used the Physiotherapy Evidence Database (PEDro) scale [23] to perform a quality assessment of the studies included. The PEDro scale comprises 11 items that correspond to a maximum of 10 points, except for item 1. A study with a PEDro score of 9-10 points is considered excellent, a score of 6-8 points is considered good, and a score of ≤5 is considered poor (low level of quality). The quality of the studies was independently assessed by 2 authors, and any dissent was settled through discussion or via consultation with a third reviewer.

Statistical Analysis

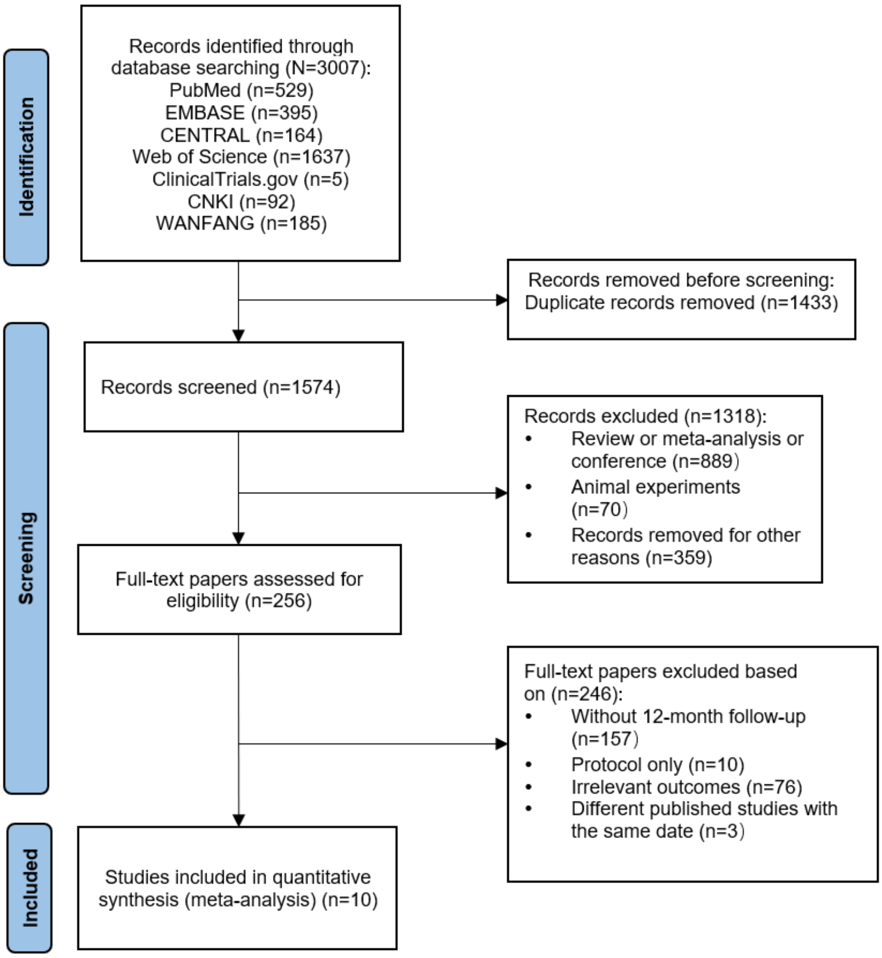
Data analysis and synthesis were performed using Cochrane Review Manager (RevMan) version 5.2 for Windows. Although all RCTs follow the principle of randomization and most baseline characteristics have no significant difference, we calculated the change from initial to final follow-up treatment difference values with a correlation coefficient of 0.5 [24] to obtain a more accurate comparison of changes in the outcomes. For continuous data, we presented the outcomes using the mean difference (MD) with 95% CIs, and $P<.05$ was considered statistically significant. If studies used different units or measurement scales, we used the standardized mean difference (SMD) with 95% CIs. As a basis for assessing heterogeneity, an inconsistency of I^2 test values of more than 50% was considered indicative of substantially high heterogeneity. If we observed statistical heterogeneity with a threshold of >50%, random-effect models were used; otherwise, fixed-effect models were applied. In addition, if this threshold was exceeded, we performed a leave-one-out sensitivity analysis to ascertain whether our findings were driven by a single study, and checked the potential reasons for heterogeneity.

Results

Search and Study Selection

A comprehensive overview of the study screening and selection process is presented using a PRISMA 2020 flow diagram in Figure 1. The diagram provides a visual representation of the selection process, while the exclusion and inclusion reasons are explained in detail later. A total of 3007 citations were identified, and 1433 (47.7%) duplicates were excluded. Of the remaining 1574 (52.3%) papers that underwent title and abstract screening, 256 (16.3%) were found to be potentially relevant to the research topic. For further assessment, full-text screening was performed, and 10 (3.9%) RCTs fulfilled the predefined criteria and were incorporated into our systematic review and meta-analysis. Importantly, a significant number of intervention-related RCTs were excluded from the analysis due to having a follow-up period of less than 12 months.

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the selection of studies. CENTRAL: Central Register of Controlled Trials; CNKI: China National Knowledge Infrastructure.



Study Characteristics

Descriptive characteristics of the 10 studies [25-34] were captured and are summarized in Multimedia Appendix 2, including study design, main follow-up time, population, intervention, control setting, and measured outcomes. Of these 10 studies, 9 (90%) [26-34] were designed as 2-arm prospective RCTs with a parallel design, and the remaining study [25] was a 3-arm RCT. In addition, 2 (20%) studies each were performed in the Netherlands [30,33], China [28,34], Canada [31,32], and Belgium [25,29] and 1 (10%) each in the Czech Republic [26] and Spain [27]. In total, 1417 participants completed the RCTs, with 709 (50%; n=123, 17.3%, women and n=586, 82.7%, men) participants in the intervention group and 708 (50%; n=119, 16.8%, women and n=589, 83.2%, men) participants in the control group. The mean age of the intervention and control groups at baseline was 59.1 (SD 9.5) years and 59.9 (SD 9.8) years, respectively (P=.16).

Intervention Programs

In our systematic review, CR involved multiple components, mainly focusing on exercise intervention, risk factor management, medical evaluation, reasonable dietary combinations, and psychosocial counseling [35]. In the papers included, these telemonitoring and telerehabilitation-delivered CR interventions were implemented using different remote application equipment, as summarized in Multimedia Appendix 2. Depending on how a network (the internet) is accessed and the devices used, there are different design options available for the presentation of cardiac telerehabilitation. For example, Kraal et al [30] and Snoek et al [33] used wearable monitoring devices to enable real-time monitoring of patients' personalized exercise training. Avila et al [25], Batalik et al [26], Frederix et al [29], Kraal et al [30], and Snoek et al [33] focused mainly on exercise-based cardiac telerehabilitation, with a predominantly moderate exercise intensity, a training heart rate equivalent to 70%-80% of the heart rate reserve, and individualized control through electronic monitoring. Among them, only Snoek et al [33] used an exercise intensity not only above the first ventilation threshold but also in the moderate exercise intensity range, approximately equal to 70%-80% of the heart rate reserve. When constructing personalized exercise sessions, the purpose of exercise also needs to be considered, and different exercise methods can have different health benefits. We included studies that mainly aimed to enhance cardiorespiratory fitness, choosing walking, jogging, and cycling, while Blasco et al [27], Reid et al [32], and Wang et al [34] performed some aerobic exercises in a home environment according to their preferences. However, Batalik et al [26], Kraal et al [30], Snoek et al [33], and particularly Frederix et al [29] and Dorje et al [28] recorded patients' walking to assess their exercise condition. In addition to enhancing cardiorespiratory function, it is also vitally essential to increase muscle strength and endurance, and most of the exercises involved in these studies are aerobic exercises. Only Avila et al [25] used strength exercises, such as arm ergometry and rowing, but did not report whether strength training, such as weightlifting, sit-ups, and push-ups, was used. However, there was a lack of exercise to improve flexibility and body coordination. In 5 (50%) studies, Blasco et al [27], Dorje et al [28], Lear et al [31], Reid et al [32], and Wang et al [34], the focus was on structured comprehensive CR, including risk factor management for CVD, emotional management, dietary management, and medication management, and the studies also involved the core component of exercise coaching. Although these studies did not have detailed exercise prescriptions, using social platforms, such as WeChat, to provide more comprehensive CR, with motivational feedback about progress, can lead to better effects.

Moreover, Avila et al [25], Frederix et al [29], and Reid et al [32] supervised home exercises and uploaded web-based reports to motivate patients to improve adherence and self-management enthusiasm. Batalik et al [26], Blasco et al [27], Dorje et al [28], Lear et al [31], and Wang et al [34], mainly used educational videos or electronic pamphlets supported by WeChat and other apps, which enabled medical staff to communicate online with patients. Patients could carry out remote health consultations to improve their QoL and control cardiac risk factors.

Risk of Bias

All the 10 (100%) studies analyzed in this review using the PEDro scale had acceptable methodological quality (score≥6); see Table 1. The results of the risk-of-bias assessment for the included studies are graphically displayed in Figure 2. We first used the Cochrane risk-of-bias tool, including selection, performance, attrition, and reporting biases. The 10 studies described specific randomization methods, 9 (90%) [25,26,28-34] reported allocation concealment, while 1 (10%) [27] had no allocation concealment. All studies had no subject blinding, and most studies had no therapist blinding due to regular supervision and timely feedback in the rehabilitation environments. Moreover, 8 (80%) studies had no reporting bias, 2 (20%) were unclear, and all studies had no clear descriptions of other biases.

Table 1. The PEDro[a] scale to assess the included RCTs'[b] methodological quality.

Quality metric	Author									
	Avila et al [25]	Batalik et al [26]	Blasco et al [27]	Dorje et al [28]	Frederix et al [29]	Kraal et al [30]	Lear et al [31]	Reid et al [32]	Snoek et al [33]	Wang et al [34]
Eligibility criteria[c]	1[d]	1	1	1	1	1	1	1	1	1
Random allocation	1	1	1	1	1	1	1	1	1	1
Concealed allocation	1	1	0[e]	1	1	1	1	1	1	1
Baseline comparability	1	1	1	1	1	1	1	1	1	1
Blinded subjects	0	0	0	0	0	0	0	0	0	0
Blinded therapists	0	0	0	0	0	0	1	1	1	0
Blinded assessors	0	0	1	1	1	0	1	1	0	0
Adequate follow-up	1	1	1	1	1	1	1	1	1	1
Intention-to-treat analysis	0	0	0	1	1	0	0	0	1	1
Between-group comparisons	1	1	1	1	1	1	1	1	1	1
Point estimates and variability	1	1	1	1	1	1	1	1	1	1
Total score	6	6	6	8	8	6	8	8	8	7

[a]PEDro: Physiotherapy Evidence Database.

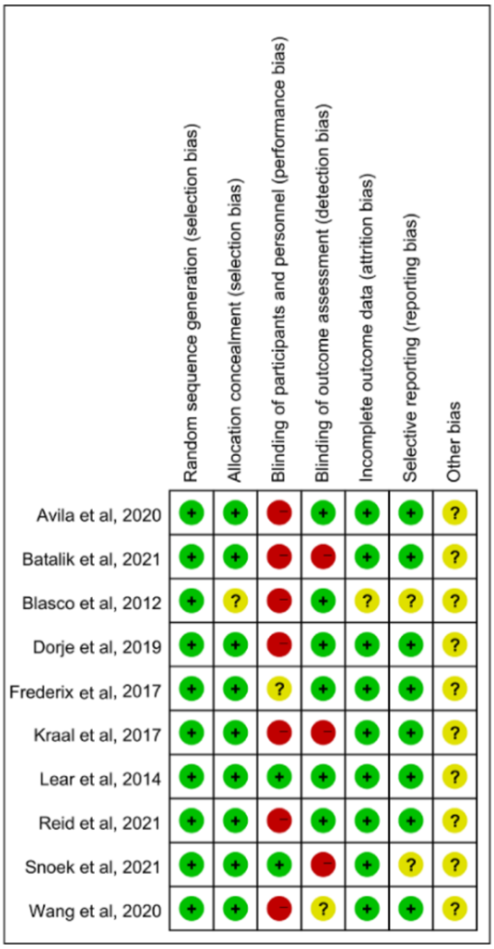
[b]RCT: randomized controlled trial.

[c]Eligibility criteria did not contribute to the total score.

[d]1=yes (reported in the study).

[e]0=no (not met).

Figure 2. Risk-of-bias summary: the authors' judgments about each risk-of-bias item for each included study were reviewed. Red, green, and yellow colors indicate high, low, and unclear risk of bias, respectively.

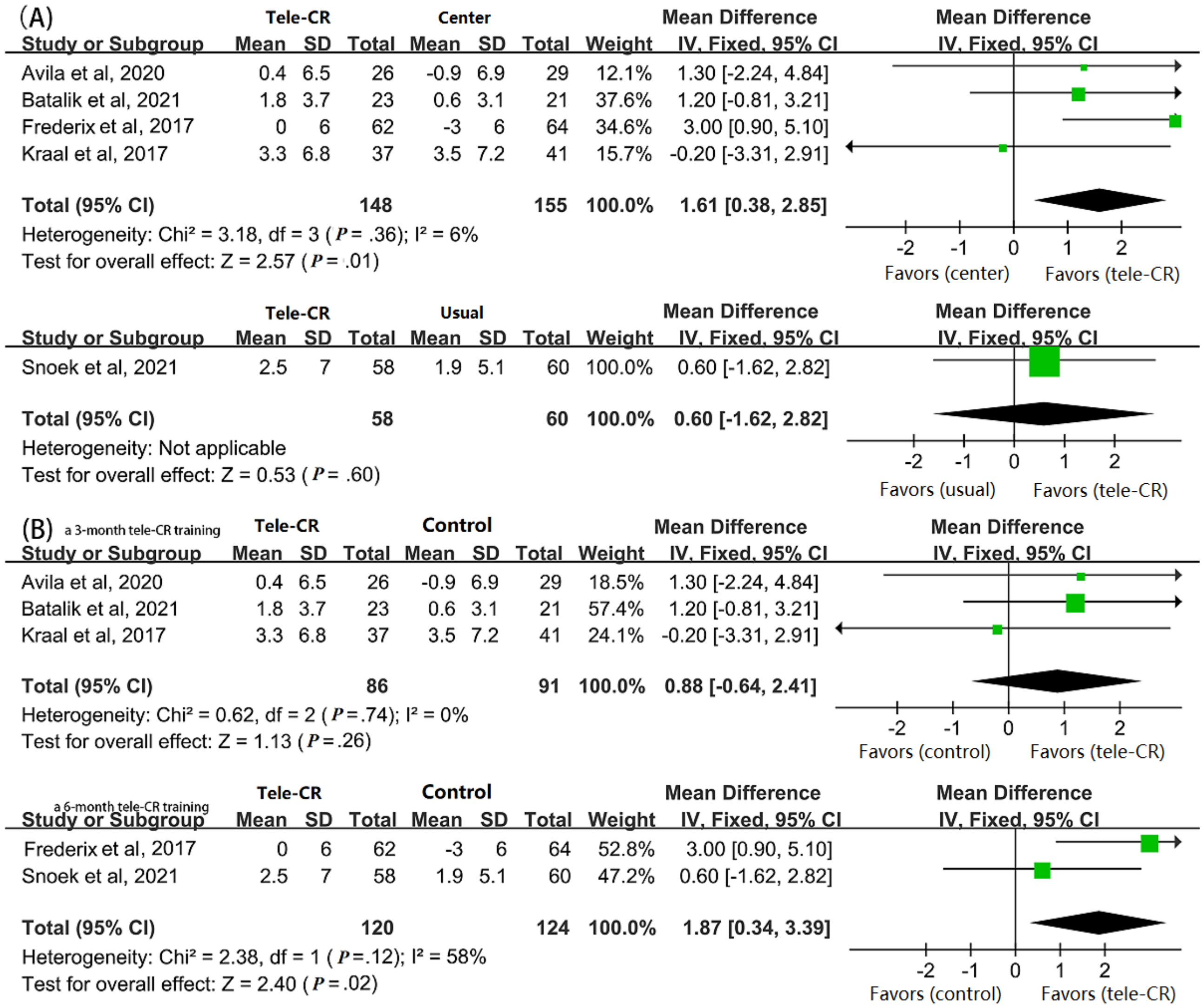


Assessment of Outcomes

Cardiorespiratory Fitness

We included 10 studies on people with CVD that evaluated long-impact cardiac telerehabilitation interventions, 5 (50%) [25,26,29,30,33] of the 10 RCTs reported peak VO_2, and a total of 421 participants who had at least 12 months of follow-up were included in the analysis. To exclude the effect of the type of control group intervention, stratified based on center-based CR or usual care, on outcomes, subgroup analyses of the 5 studies were performed, of which the control group of only 1 (20%) study, by Snoke et al [33], received usual care. Therefore, the combined meta-results of the analyses of the remaining 4 (80%) studies showed that real-time monitored exercise-based cardiac telerehabilitation significantly improves long-term peak VO_2 compared to center-based CR (MD 1.61, 95% CI 0.38-2.85, P=.01), as shown in Figure 3A. Subgroup analyses were also performed for patients considering the different effects of intervention durations in exercise protocols. As shown in Figure 3B, peak VO_2 improvement in the intervention group was significantly greater than that in the control group after 6-month telerehabilitation training (MD 1.87, 95% CI 0.34-3.39, P=.02), but there was no significant difference after 3-month telerehabilitation training.

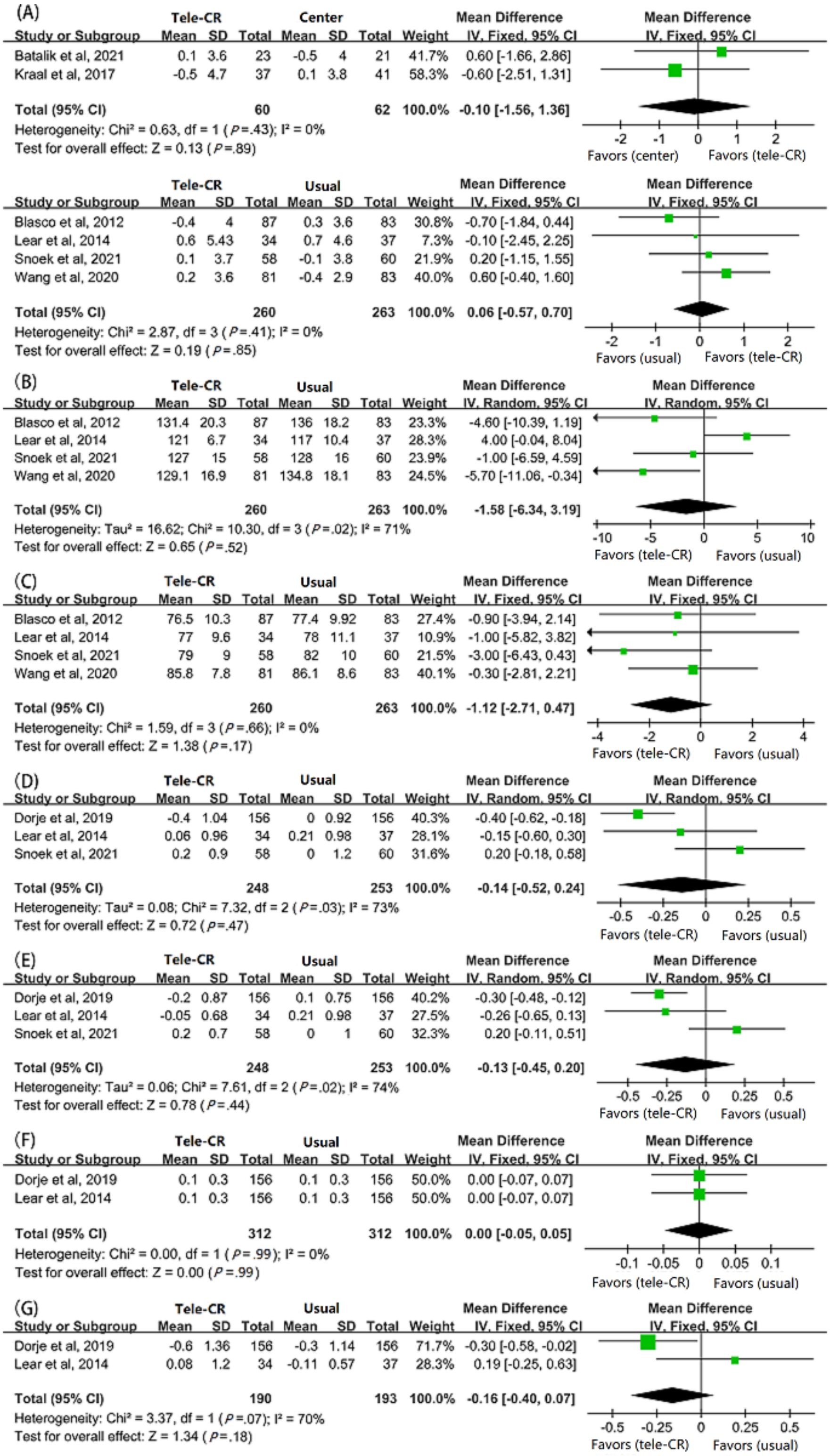
Figure 3. Pooled MD between cardiac telerehabilitation and control groups in terms of peak VO2 in long-term follow-up, divided into (A) cardiac telerehabilitation vs center-based CR or usual care and (B) a 3- or 6-month telerehabilitation program in the intervention group. CR: cardiac rehabilitation; MD: mean difference; peak VO2: peak oxygen uptake.



Cardiovascular Risk Factors

In our study, cardiovascular risk factors mainly included the BMI, blood pressure, and lipid profile. Of the 10 studies, 6 (60%) [26,27,30,31,33,34] reported the BMI after long-term follow-up, showing no significant difference between cardiac telerehabilitation and center-based CR (MD –0.10, 95% CI –1.56 to 1.36, *P*=.89; I^2=0%) or between cardiac telerehabilitation and usual care (MD –0.60, 95% CI –0.57 to 0.70, *P*=.85; I^2=0%); see Figure 4A. In terms of blood pressure, there was no significant difference in systolic blood pressure (MD –1.58, 95% CI –6.34 to 3.19, *P*=0.52; I^2=0%; Figure 4B) and diastolic blood pressure (MD –1.12, 95% CI –2.71 to 0.47, *P*=0.17; I^2=0%; Figure 4C) compared to the usual-care combined effect size. Only 4 (40%) studies [28,29,31,33] included the blood lipid index as an outcome measure, of which 3 (75%) studies [28,31,33] used usual care for the control group, and only Frederix et al [29] used center-based CR. Therefore, to maintain the consistency of the control group and reduce bias, the 3 (75%) studies [28,31,33] were finally included to analyze the improvement in blood lipids. The results showed that there was no more significance than the usual-care group in improving total cholesterol (TC; MD –0.14, 95% CI –0.52 to 0.24, *P*=.47; I^2=73%; Figure 4D), low-density lipoprotein cholesterol (LDL-C; MD –0.13, 95% CI –0.45 to 0.20, *P*=.44; I^2=74%; Figure 4E), high-density lipoprotein cholesterol (HDL-C; MD 0.00, 95% CI –0.05 to 0.05, *P*=.99; I^2=0%; Figure 4F), and triglycerides (TGs; MD –0.16, 95% CI –0.40 to 0.07, *P*=.18; I^2=70%; Figure 4G).

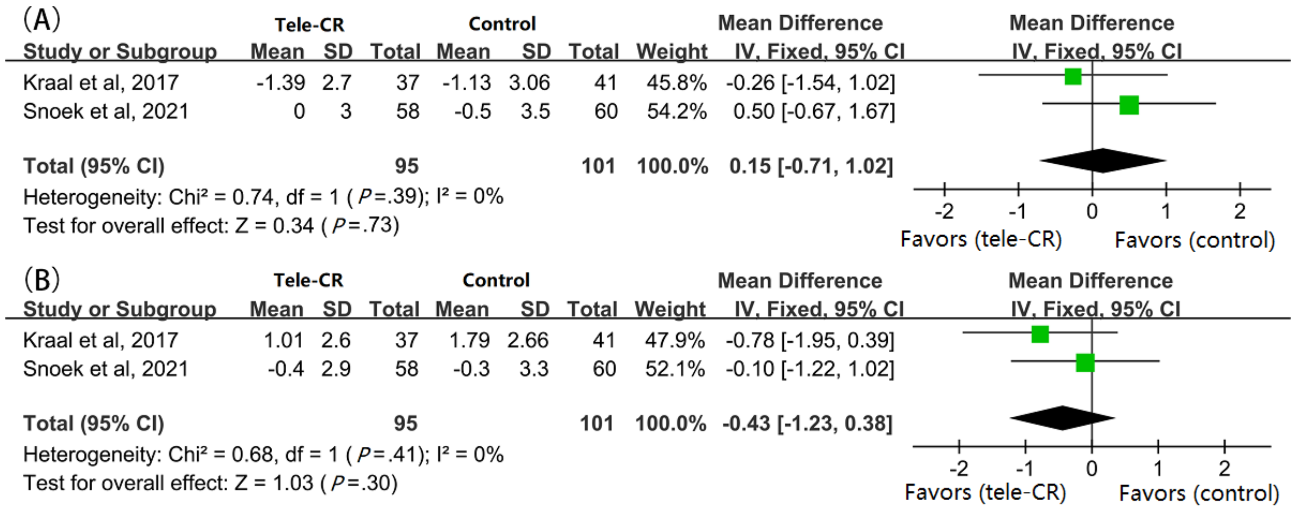
Figure 4. Pooled MD between the cardiac telerehabilitation and control groups in terms of (A) BMI, (B) systolic blood pressure, (C) diastolic blood pressure, (D) total cholesterol (TC), (E) low-density lipoprotein cholesterol (LDL-C), (F) high-density lipoprotein cholesterol (HDL-C), and (G) triglycerides (TGs) in long-term follow-up. CR: cardiac rehabilitation; MD: mean difference.



Depression and Anxiety

Of the 10 studies included, 2 (20%) [30,33] applied fixed-effect meta-analysis and showed no significant long-term improvement in anxiety scores (MD 0.15, 95% CI –0.71 to 1.02, *P*=.73; I^2=0%; Figure 5A) or depression scores (MD –0.43, 95% CI –1.23 to 0.38, *P*=.30; I^2=0%; Figure 5B).

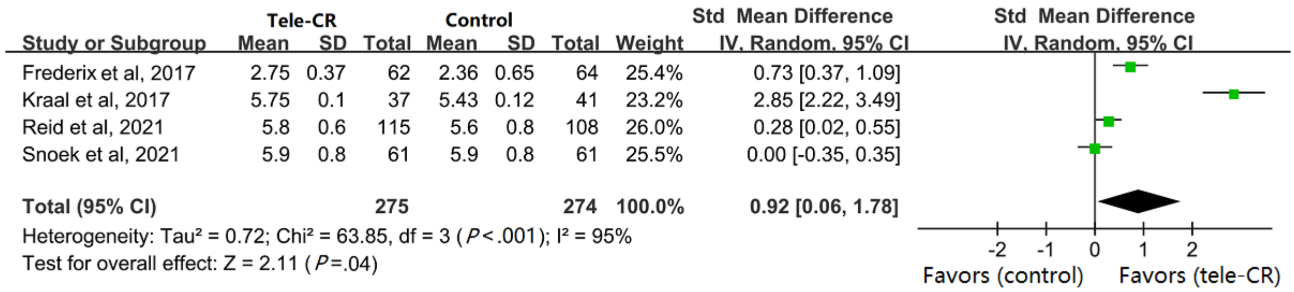
Figure 5. Pooled MD between the cardiac telerehabilitation and control groups in terms of (A) anxiety score change and (B) depression score change in long-term follow-up. CR: cardiac rehabilitation; MD: mean difference.



Quality of Life

Due to the different scales used to measure QoL, we used the SMD as a practical measure. The MacNew Heart Disease Health-related Quality of Life (MacNew) questionnaire was used to explore the effects of cardiac telerehabilitation on QoL in people with CVD in 3 (30%) of the 10 studies included [30,32,33], and 1 (10%) study [29] used the HeartQoL questionnaire. However, the combined results showed high heterogeneity (I^2=95%; Figure 6), and I^2 still fluctuated between 70% and 95% after sensitivity analysis, so the random-effect model was used for analysis. The results showed that cardiac telerehabilitation could improve the long-term QoL of patients with CVD (MD 0.92, 95% CI 0.06-1.78, *P*=.04).

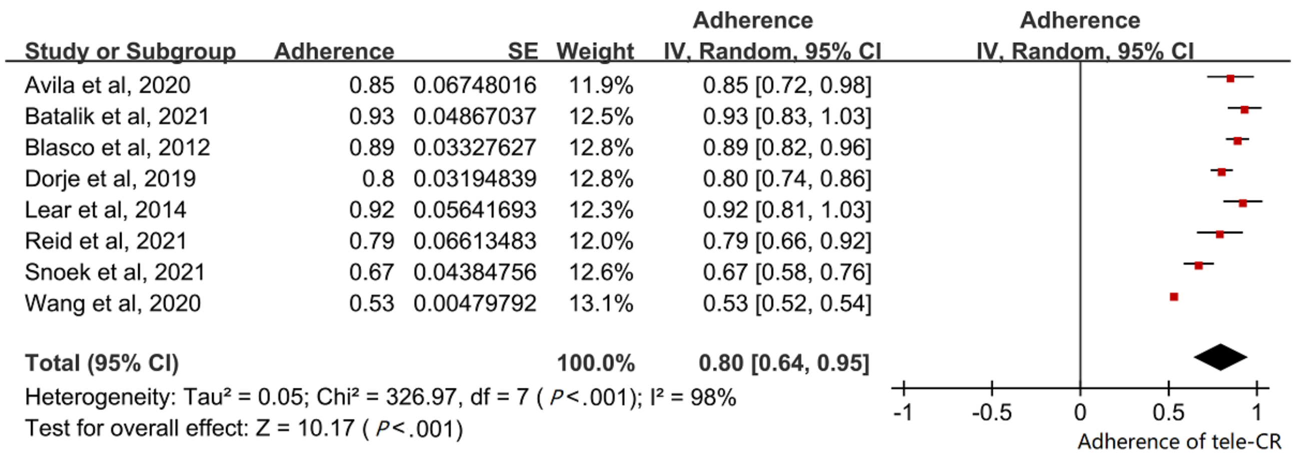
Figure 6. Pooled SMD in terms of QoL. CR: cardiac rehabilitation; QoL: quality of life; SMD: standardized mean difference.



Adherence to the Telerehabilitation Program

Completion rates for cardiac telerehabilitation were reported in 8 (80%) of the 10 studies (MD 0.80, 95% CI 0.64-0.95; Figure 7) [25-28,31-34], with high heterogeneity (I^2=98%) based on our pooled meta-analysis.

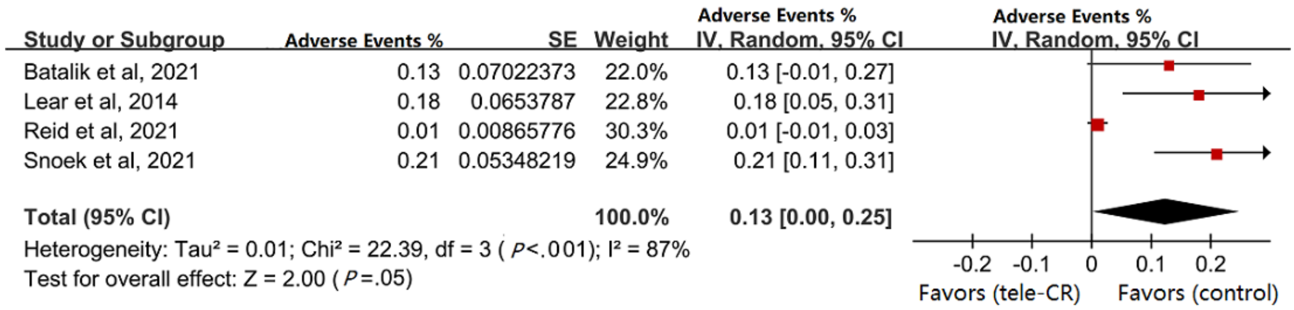
Figure 7. Pooled completion rate. CR: cardiac rehabilitation.



Adverse Events

Of the 10 studies, in 4 (40%) [26,31-33], 13% of patients reported all-cause adverse events (MD 0.13, 95% CI 0.00-0.25; see Figure 8).

Figure 8. Pooled adverse event rate.



Discussion

Principal Findings

Compared to center-based CR, cardiac telerehabilitation was effective in improving cardiorespiratory fitness and exercise capacity in patients with CAD, particularly in terms of peak VO_2, during long-term follow-up. However, technology-based cardiac telerehabilitation had no long-term benefits in terms of risk factor management. Based on the RCTs included, we also found that telehealth interventions do not result in significant improvements in depression and anxiety scores in the long term. Nevertheless, there was evidence of improved long-term QoL with cardiac telerehabilitation. Our study also revealed positive adherence to cardiac telerehabilitation interventions, and the incidence of adverse events during long-term follow-up was low.

Cardiorespiratory fitness is a crucial determinant of CR and a strong predictor of all-cause mortality and cardiovascular mortality [36]. The peak VO_2 in CPET is 1 of the most critical and gold-standard indicators of cardiorespiratory fitness in patients with CVD, representing the maximum oxygen intake by the human body per unit body weight (mL/kg/minute). Peak VO_2 reflects cardiopulmonary function in transporting oxygen and carbon dioxide around the body [37], the maximum aerobic metabolic capacity [38], and the skeletal muscles' ability to absorb and use oxygen [39]. Our study confirms that compared to center-based CR training, cardiac telerehabilitation can significantly improve long-term peak VO_2 in patients. Furthermore, we found that the duration of exercise training has a positive impact on the improvement in peak VO_2. There is debate over whether ambulatory cardiac telerehabilitation is superior to traditional in-hospital or center-based CR, and this has also been widely discussed in recent years. Our primary finding, consistent with previous studies and related reviews [40-42], showed that exercise-based cardiac telerehabilitation with remote monitoring is equivalent or superior to center-based CR in terms of exercise capacity in cardiac disease. This can help patients overcome barriers such as transportation limitations and conflicts with work schedules, thereby expanding the implementation of rehabilitation programs for a wider range of patients. However, it is worth noting that these previous studies have primarily focused on short- or medium-term follow-up, and there is a lack of long-term evaluation. Therefore, our results further support the notion that compared to center-based CR, cardiac telerehabilitation can better sustain patients' cardiorespiratory endurance over an extended period. Our finding is consistent with a retrospective study performed by Ramadi et al [43], who provided evidence that an extensive multidisciplinary and structured CR program retains the improvement in exercise capacity until 1-year follow-up. Interestingly, Aamot et al [44] and Smith et al [45] indicated that the monitoring strategy in CR enables the therapist to help patients sustain long-term exercise adherence so that peak VO_2 significantly increases compared to baseline values. To strengthen our results, we also focused on other variables of CPET. Unexpectedly, we did not observe a significant difference. For example, only 2 RCTs [25,46] reported oxygen consumption at the anaerobic threshold (VO_2 AT, mL/kg/minute) and showed that the improvement in VO_2 AT does not persist for long. These indicators, especially VO_2 AT, have a strong correlation with the clinical symptoms of patients with CVD [47], so further studies can evaluate whether cardiac telerehabilitation protocols can improve long-term clinical outcomes in order to strengthen the evidence that cardiac telerehabilitation can maintain cardiopulmonary fitness levels and exercise capacity.

Although our study and previous studies have shown that cardiac telerehabilitation shows significant long-term effectiveness in increasing peak VO_2, the current intervention methods for cardiac telerehabilitation show considerable variability in their design. For example, in the 10 studies included in this review, Avila et al [25], Batalik et al [26], Kraal et al [30], and Snoek et al [33] used heart rate monitors, such as heart rate belts and sports bracelets, to detect exercise intensity and record health data in order to promptly synchronize patients' exercise status. Blasco et al [27], Frederix et al [29], Lear et al [31], and Reid et al [32] used self-health management systems, such as smartphones or computers, to transmit patients' blood sugar levels, blood pressure, smoking status, and other daily life conditions to the online platform to manage CVD risk factors; Frederix et al [29] also used activity trackers to monitor patients' exercise. Furthermore, with continuous software and hardware optimization, Dorje et al [28] and Wang et al [34] used social software, such as the WeChat group mode, for education, management, and follow-up, as well as emotional and nutrition management. Although remote methods of intervention vary, at the core, they are the same; that is, using various types of remote equipment, they mobilize the subjective initiative of patients for rehabilitation, under the effective communication of medical care, for them to implement rehabilitation treatments, such as exercise prescription, drug prescription, psychological prescription, and risk factor management. Therefore, future remote CR content and the operation process will be more standardized.

Our subgroup analysis revealed that an extended intervention duration of 6 months in exercise protocols significantly improves and maintains cardiorespiratory fitness. However, compared to the control group, no significant difference was observed after the 3-month rehabilitation training. This could be attributed to the reinforcement of patient self-awareness of CR through extended telehealth guidance and monitoring. Slovinec et al [48] found that motivational orientation and self-efficacy of exercise behavior affect exercise maintenance and physical activity levels. Self-regulation theory [21] is supported by physical and mental aspects to motivate patients to realize their potential, actively face diseases and adverse reactions, and improve physical and mental health. Janssen et al [49] showed that a theory-based lifestyle program could stimulate and sustain improvements in exercise adherence. Therefore, cardiac telerehabilitation delivered through modern network technology has the potential to enhance long-term effectiveness by tailoring, coaching, monitoring, and providing objective feedback programs that enhance patients' self-efficacy and subjective initiative. Maddison et al [50] indicated that mobile health interventions have a positive therapeutic effect on leisure-time physical activity and walking, which may be moderated by changes in self-efficacy, which also strengthens our conclusions.

We found no strongly favorable evidence of a difference in cardiovascular risk factor management. For cardiovascular risk factors, this meta-analysis indicated no significant differences in reduction in blood pressure, BMI, and blood lipid analysis changes over 12 months. Evaluation and management of risk factors are crucial for the prognosis of patients with CVD [5,51]. Furthermore, with the younger age of patients with CVD and the increase in human life expectancy worldwide, long-lasting beneficial changes are needed for targeted preventive activities, so future research projects in this field should focus on extending the efficacy of risk factor management and maintenance effects.

No effectiveness was demonstrated in anxiety and depression score changes compared to the control group. Negative emotions and the occurrence and development of CVDs are 2-way causes [52]. Penninx et al [53] carried out a 4-year follow-up of 2397 patients with undiagnosed CVD and found that patients with negative emotions are more likely to suffer from CVD than patients without mood disorders. Another study [54] found that CR can help people with anxiety and depression shift their attention, better vent their emotions, and effectively alleviate mood disorders. Internet-based cardiac telerehabilitation may enhance communication and feedback between patients and medical staff and achieve emotional problem solving on time [55]. There are few studies on telerehabilitation and emotions of patients with CVD. In addition to improving cardiopulmonary fitness and cardiovascular risk factor management, 1 of the ultimate goals of telerehabilitation is to improve the long-term QoL of people with CVD. There is a significant statistical difference between groups in long-term follow-up results. However, the effect of cardiac telerehabilitation on the QoL of patients with CVD may be influenced by different assessment tools, and there is no consensus at present. The relevant pathogenesis and treatment guidelines are imperfect, so further research is needed.

We found high participation rates in CR during long-term security follow-up in our study. Compliance with rehabilitation training is a key factor in improving the rehabilitation effect [56], due to the patients' need to go to the hospital regularly for rehabilitation training, resulting in a poor participation rate and compliance, and medium- to long-term recovery rates after discharge are low [57]. Telerehabilitation compliance is high, which effectively improves the efficiency of the CR of patients. Ivers et al [58] showed that large multicenter RCT telehealth interventions can improve the completion rate of CR in patients with myocardial infarction, albeit only using simple and convenient remote methods, such as Short Messaging Service (SMS) and email, which was also verified in the summary results of the review by Santiago et al [59]. At the same time, cardiac telerehabilitation has a low incidence of adverse events if it is fully evaluated before implementation. For example, Piotrowicz et al [60] found no significant difference in the incidence of adverse events between 2 groups (12.5% vs 12.4%) during the 12- to 24-month follow-up after 9 weeks of cardiac telerehabilitation intervention in patients with heart failure compared to usual care. The most effective way to improve the safety of telerehabilitation is to fully assess the patient's status, such as using CPET, noninvasive cardiac output, physical assessment, etc.

CR is a vital part of the rehabilitation process of patients with CVD. Based on the development and application of the internet and tele-equipment, cardiac telerehabilitation, as a new means of rehabilitation, can effectively carry out CR in the home-based environment, improve the participation and compliance of patients with CVD to undergo CR, and improve their functional status. Our study adds evidence to the advancement of telerehabilitation, in the hope that the popularization of telerehabilitation and the improvements in CR treatment for patients with CVD can improve the quality of rehabilitation, save medical time and medical costs, and solve the problem of some young patients being unable to participate in conventional center-based CR due to work. It is necessary for future studies to promote the "internet + cardiac rehabilitation" model to overcome the clinical problems of actual patients seeking medical treatment and effectively implement the model for every patient who needs CR.

Limitations

There are some limitations of the study. First is the large variability and complexity of the interventions due to cardiac telerehabilitation delivering exercise intervention details according to the frequency, intensity, timing, and type (FITT) principle. Because relatively few trials are investigating the long-term outcomes of cardiac telerehabilitation, it is difficult to unify specifically detailed research protocols; this requires future investigations to verify the long-term effectiveness of cardiac telerehabilitation under different administration conditions and to develop a uniform personalized plan. Second, our study did not include as outcome measures long-term improvements in major adverse cardiovascular events (MACEs), all-cause mortality, or all-cause hospitalization in people with CVD. Cardiopulmonary fitness indicates that peak VO_2 corresponds to a 13% reduction in all-cause mortality and a 15% reduction in cardiovascular mortality [61]. Therefore, future studies are required to include this indicator based on a sufficient sample size. Finally, we did not consider economic cost-effectiveness. Collecting data on patients' medical cost burden and rehabilitation cycle during cardiac telerehabilitation and conducting financial analysis from a societal perspective are constructive. For example, Batalik et al [62] mentioned that cost-benefit analysis is essential for policy makers, systematic review of exercise-based telehealth CR is cost-effective, and the 12 studies included in their research showed no clear difference between telerehabilitation and center-based CR. Short- and long-term clinical-economic analyses are still required to facilitate the implementation of telerehabilitation interventions in the clinic.

Conclusion

Cardiac telerehabilitation, as a promising treatment method, plays a crucial role in the comprehensive rehabilitation of patients with CVD. It addresses the diverse rehabilitation needs of patients and helps enhance their recovery. Our results demonstrated a significant difference in peak VO_2 and QoL in terms of long-term improvements but no significant differences in changes in cardiovascular risk factor management and the psychological scales of depression and anxiety. Our results provide initial evidence supporting the use of cardiac telerehabilitation as an alternative model to center-based CR. By extending the benefits of cardiorespiratory effectiveness, cardiac telerehabilitation can promote patients' long-term awareness of rehabilitation, thereby maximizing the prognosis for each patient.

Acknowledgments

This work was supported by the National Key R&D Program of China (grant numbers 2020YFC2008500 and 2020YFC2008502), the National Natural Science Foundation of China (grant numbers 81572231 and 82172534), and the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (grant number ZYJC21038).

Authors' Contributions

WZ and QW designed the study. WZ, HC, and QW performed the research and analyzed the data. LX and RL provided help and advice on the tables and figures. CH provided resources. WZ, LW, and QW wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete search strategy.
[DOCX File , 21 KB - mhealth_v11i1e46359_app1.docx]

Multimedia Appendix 2
Descriptive characteristics of the 10 studies.
[DOCX File , 27 KB - mhealth_v11i1e46359_app2.docx]

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
2. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021 [FREE Full text] [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
3. Wang H, Dou K, Wang Y, Yin D, Xu B, Gao R. Benefit-risk profile of DAPT continuation beyond 1year after PCI in patients with high thrombotic risk features as endorsed by 2018 ESC/EACTS Myocardial Revascularization Guideline. Cardiovasc Drugs Ther 2020 Oct;34(5):663-675. [doi: 10.1007/s10557-020-07030-9] [Medline: 32601780]
4. Pelliccia A, Sharma S, Gati S, Bäck M, Börjesson M, Caselli S, ESC Scientific Document Group. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J 2021 Jan 01;42(1):17-96 [FREE Full text] [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]

5. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019 Jul 09;74(1):133-153 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
6. Fang J, Ayala C, Luncheon C, Ritchey M, Loustalot F. Use of outpatient cardiac rehabilitation among heart attack survivors - 20 states and the District of Columbia, 2013 and four states, 2015. MMWR Morb Mortal Wkly Rep 2017 Aug 25;66(33):869-873 [FREE Full text] [doi: 10.15585/mmwr.mm6633a1] [Medline: 28837549]
7. Ambrosetti M, Abreu A, Corrà U, Davos CH, Hansen D, Frederix I, et al. Secondary prevention through comprehensive cardiovascular rehabilitation: from knowledge to implementation. 2020 update. A position paper from the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2021 May 14;28(5):460-495. [doi: 10.1177/2047487320913379] [Medline: 33611446]
8. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the Million Hearts Cardiac Rehabilitation Collaborative. Mayo Clin Proc 2017 Feb;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
9. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
10. Aragam KG, Dai D, Neely ML, Bhatt DL, Roe MT, Rumsfeld JS, et al. Gaps in referral to cardiac rehabilitation of patients undergoing percutaneous coronary intervention in the United States. J Am Coll Cardiol 2015 May 19;65(19):2079-2088 [FREE Full text] [doi: 10.1016/j.jacc.2015.02.063] [Medline: 25975470]
11. Ruano-Ravina A, Pena-Gil C, Abu-Assi E, Raposeiras S, van 't Hof A, Meindersma E, et al. Participation and adherence to cardiac rehabilitation programs. A systematic review. Int J Cardiol 2016 Nov 15;223:436-443. [doi: 10.1016/j.ijcard.2016.08.120] [Medline: 27557484]
12. Bakhshayeh S, Sarbaz M, Kimiafar K, Vakilian F, Eslami S. Barriers to participation in center-based cardiac rehabilitation programs and patients' attitude toward home-based cardiac rehabilitation programs. Physiother Theory Pract 2021 Jan;37(1):158-168. [doi: 10.1080/09593985.2019.1620388] [Medline: 31155986]
13. Patil R, Shrivastava R, Juvekar S, McKinstry B, Fairhurst K. Specialist to non-specialist teleconsultations in chronic respiratory disease management: a systematic review. J Glob Health 2021 Mar 27;11:04019 [FREE Full text] [doi: 10.7189/jogh.11.04019] [Medline: 34326988]
14. Stefanakis M, Batalik L, Antoniou V, Pepera G. Safety of home-based cardiac rehabilitation: a systematic review. Heart Lung 2022;55:117-126 [FREE Full text] [doi: 10.1016/j.hrtlng.2022.04.016] [Medline: 35533492]
15. Subedi N, Rawstorn JC, Gao L, Koorts H, Maddison R. Implementation of telerehabilitation interventions for the self-management of cardiovascular disease: systematic review. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e17957 [FREE Full text] [doi: 10.2196/17957] [Medline: 33245286]
16. Jin Choo Y, Chang MC. Effects of telecardiac rehabilitation on coronary heart disease: A PRISMA-compliant systematic review and meta-analysis. Medicine (Baltimore) 2022 Jul 15;101(28):e29459 [FREE Full text] [doi: 10.1097/MD.0000000000029459] [Medline: 35839029]
17. Ramachandran HJ, Jiang Y, Tam WWS, Yeo TJ, Wang W. Effectiveness of home-based cardiac telerehabilitation as an alternative to phase 2 cardiac rehabilitation of coronary heart disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2022 May 25;29(7):1017-1043 [FREE Full text] [doi: 10.1093/eurjpc/zwab106] [Medline: 34254118]
18. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart Randomized Controlled Trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
19. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-term effectiveness of a comprehensive internet-based and patient-specific telerehabilitation program with text messaging support for cardiac patients: randomized controlled trial. J Med Internet Res 2015 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
20. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014 Mar 11;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
21. Janssen V, De Gucht V, van Exel H, Maes S. Beyond resolutions? A randomized controlled trial of a self-regulation lifestyle programme for post-cardiac rehabilitation patients. Eur J Prev Cardiol 2013 Jun;20(3):431-441. [doi: 10.1177/2047487312441728] [Medline: 22396248]
22. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 03;10(10):ED000142 [FREE Full text] [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
23. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [Medline: 12882612]
24. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol 1992 Jul;45(7):769-773. [doi: 10.1016/0895-4356(92)90054-q] [Medline: 1619456]

25. Avila A, Claes J, Buys R, Azzawi M, Vanhees L, Cornelissen V. Home-based exercise with telemonitoring guidance in patients with coronary artery disease: does it improve long-term physical fitness? Eur J Prev Cardiol 2020 Mar;27(4):367-377. [doi: 10.1177/2047487319892201] [Medline: 31787026]
26. Batalik L, Dosbaba F, Hartman M, Konecny V, Batalikova K, Spinar J. Long-term exercise effects after cardiac telerehabilitation in patients with coronary artery disease: 1-year follow-up results of the randomized study. Eur J Phys Rehabil Med 2021 Oct;57(5):807-814 [FREE Full text] [doi: 10.23736/S1973-9087.21.06653-3] [Medline: 33619944]
27. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
28. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
29. Frederix I, Solmi F, Piepoli MF, Dendale P. Cardiac telerehabilitation: a novel cost-efficient care delivery strategy that can induce long-term health benefits. Eur J Prev Cardiol 2017 Nov;24(16):1708-1717. [doi: 10.1177/2047487317732274] [Medline: 28925749]
30. Kraal JJ, Van den Akker-Van Marle ME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273 [FREE Full text] [doi: 10.1177/2047487317710803] [Medline: 28534417]
31. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Randomized trial of a virtual cardiac rehabilitation program delivered at a distance via the internet. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):952-959. [doi: 10.1161/CIRCOUTCOMES.114.001230] [Medline: 25271050]
32. Reid RD, Morrin LI, Beaton LJ, Papadakis S, Kocourek J, McDonnell L, et al. Randomized trial of an internet-based computer-tailored expert system for physical activity in patients with heart disease. Eur J Prev Cardiol 2012 Dec;19(6):1357-1364. [doi: 10.1177/1741826711422988] [Medline: 21903744]
33. Snoek JA, Meindersma EP, Prins LF, Van't Hof AW, de Boer M, Hopman MT, et al. The sustained effects of extending cardiac rehabilitation with a six-month telemonitoring and telecoaching programme on fitness, quality of life, cardiovascular risk factors and care utilisation in CAD patients: the TeleCaRe study. J Telemed Telecare 2021 Sep 23;27(8):473-483. [doi: 10.1177/1357633X19885793] [Medline: 31760855]
34. Wang J, Zeng Z, Dong R, Sheng J, Lai Y, Yu J, et al. Efficacy of a WeChat-based intervention for adherence to secondary prevention therapies in patients undergoing coronary artery bypass graft in China: a randomized controlled trial. J Telemed Telecare 2020 Sep 30;28(9):653-661. [doi: 10.1177/1357633x20960639]
35. Guo Y, Ledesma RA, Peng R, Liu Q, Xu D. The beneficial effects of cardiac rehabilitation on the function and levels of endothelial progenitor cells. Heart Lung Circ 2017 Jan;26(1):10-17. [doi: 10.1016/j.hlc.2016.06.1210] [Medline: 27614559]
36. Lee D, Sui X, Ortega FB, Kim Y, Church TS, Winett RA, et al. Comparisons of leisure-time physical activity and cardiorespiratory fitness as predictors of all-cause mortality in men and women. Br J Sports Med 2011 May;45(6):504-510. [doi: 10.1136/bjsm.2009.066209] [Medline: 20418526]
37. Taylor JL, Medina-Inojosa JR, Chacin-Suarez A, Smith JR, Squires RW, Thomas RJ, et al. Age-related differences for cardiorespiratory fitness improvement in patients undergoing cardiac rehabilitation. Front Cardiovasc Med 2022;9:872757 [FREE Full text] [doi: 10.3389/fcvm.2022.872757] [Medline: 35498026]
38. Costache A, Roca M, Honceriu C, Costache I, Leon-Constantin M, Mitu O, et al. Cardiopulmonary exercise testing and cardiac biomarker measurements in young football players: a pilot study. J Clin Med 2022 May 14;11(10):2772 [FREE Full text] [doi: 10.3390/jcm11102772] [Medline: 35628899]
39. Banydeen R, Monfort A, Inamo J, Neviere R. Diagnostic and prognostic values of cardiopulmonary exercise testing in cardiac amyloidosis. Front Cardiovasc Med 2022;9:898033 [FREE Full text] [doi: 10.3389/fcvm.2022.898033] [Medline: 35734274]
40. Maddison R, Rawstorn JC, Stewart RAH, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
41. Taylor RS, Dalal H, Jolly K, Zawada A, Dean SG, Cowie A, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2015 Aug 18;8(8):CD007130. [doi: 10.1002/14651858.CD007130.pub3] [Medline: 26282071]
42. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6(6):CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
43. Ramadi A, Haennel RG, Stone JA, Arena R, Threlfall TG, Hitt E, et al. The sustainability of exercise capacity changes in home versus center-based cardiac rehabilitation. J Cardiopulm Rehabil Prev 2015;35(1):21-28. [doi: 10.1097/HCR.0000000000000084] [Medline: 25313452]
44. Aamot I, Karlsen T, Dalen H, Støylen A. Long-term exercise adherence after high-intensity interval training in cardiac rehabilitation: a randomized study. Physiother Res Int 2016 Mar;21(1):54-64. [doi: 10.1002/pri.1619] [Medline: 25689059]

45. Smith KM, McKelvie RS, Thorpe KE, Arthur HM. Six-year follow-up of a randomised controlled trial examining hospital versus home-based exercise training after coronary artery bypass graft surgery. Heart 2011 Jul;97(14):1169-1174. [doi: 10.1136/hrt.2010.202036] [Medline: 21561899]
46. Ma J, Ge C, Shi Y, Xu Y, Zhao C, Gao L, et al. Chinese home-based cardiac rehabilitation model delivered by smartphone interaction improves clinical outcomes in patients with coronary heart disease. Front Cardiovasc Med 2021;8:731557 [FREE Full text] [doi: 10.3389/fcvm.2021.731557] [Medline: 34676252]
47. Corrà U, Agostoni PG, Anker SD, Coats AJS, Crespo Leiro MG, de Boer RA, et al. Role of cardiopulmonary exercise testing in clinical stratification in heart failure. A position paper from the Committee on Exercise Physiology and Training of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2018 Jan;20(1):3-15 [FREE Full text] [doi: 10.1002/ejhf.979] [Medline: 28925073]
48. Slovinec D'Angelo ME, Pelletier LG, Reid RD, Huta V. The roles of self-efficacy and motivation in the prediction of short- and long-term adherence to exercise among patients with coronary heart disease. Health Psychol 2014 Nov;33(11):1344-1353. [doi: 10.1037/hea0000094] [Medline: 25133848]
49. Janssen V, De Gucht V, van Exel H, Maes S. A self-regulation lifestyle program for post-cardiac rehabilitation patients has long-term effects on exercise adherence. J Behav Med 2014 Apr;37(2):308-321. [doi: 10.1007/s10865-012-9489-y] [Medline: 23334387]
50. Maddison R, Pfaeffli L, Stewart R, Kerr A, Jiang Y, Rawstorn J, et al. The HEART Mobile Phone Trial: the partial mediating effects of self-efficacy on physical activity among cardiac patients. Front Public Health 2014;2:56 [FREE Full text] [doi: 10.3389/fpubh.2014.00056] [Medline: 24904918]
51. Gharibzadeh A, Shahsanaei F, Rahimi Petrudi N. Clinical and cardiovascular characteristics of patients suffering ST-segment elevation myocardial infarction after Covid-19: a systematic review and meta-analysis. Curr Probl Cardiol 2023 Jan;48(1):101045 [FREE Full text] [doi: 10.1016/j.cpcardiol.2021.101045] [Medline: 34780870]
52. Ikeda A, Steptoe A, Shipley M, Wilkinson IB, McEniery CM, Tanigawa T, et al. Psychological wellbeing and aortic stiffness: longitudinal study. Hypertension 2020 Sep;76(3):675-682 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.119.14284] [Medline: 32654561]
53. Penninx BW, Beekman AT, Honig A, Deeg DJ, Schoevers RA, van Eijk JT, et al. Depression and cardiac mortality: results from a community-based longitudinal study. Arch Gen Psychiatry 2001 Mar;58(3):221-227. [doi: 10.1001/archpsyc.58.3.221] [Medline: 11231827]
54. Chauvet-Gelinier J, Bonin B. Stress, anxiety and depression in heart disease patients: a major challenge for cardiac rehabilitation. Ann Phys Rehabil Med 2017 Jan;60(1):6-12 [FREE Full text] [doi: 10.1016/j.rehab.2016.09.002] [Medline: 27771272]
55. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity eHealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
56. Ge C, Ma J, Xu Y, Shi Y, Zhao C, Gao L, et al. Predictors of adherence to home-based cardiac rehabilitation program among coronary artery disease outpatients in China. J Geriatr Cardiol 2019 Oct;16(10):749-755 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2019.10.003] [Medline: 31700514]
57. Hutchinson P, Meyer A, Marshall B. Factors influencing outpatient cardiac rehabilitation attendance. Rehabil Nurs 2015;40(6):360-367. [doi: 10.1002/rnj.202] [Medline: 25771985]
58. Ivers NM, Schwalm J, Bouck Z, McCready T, Taljaard M, Grace SL, et al. Interventions supporting long term adherence and decreasing cardiovascular events after myocardial infarction (ISLAND): pragmatic randomised controlled trial. BMJ 2020 Jun 10;369:m1731 [FREE Full text] [doi: 10.1136/bmj.m1731] [Medline: 32522811]
59. Santiago de Araújo Pio C, Chaves GS, Davies P, Taylor RS, Grace SL. Interventions to promote patient utilisation of cardiac rehabilitation. Cochrane Database Syst Rev 2019 Feb 01;2(2):CD007131 [FREE Full text] [doi: 10.1002/14651858.CD007131.pub4] [Medline: 30706942]
60. Piotrowicz E, Pencina MJ, Opolski G, Zareba W, Banach M, Kowalik I, et al. Effects of a 9-week hybrid comprehensive telerehabilitation program on long-term outcomes in patients with heart failure: the Telerehabilitation in Heart Failure Patients (TELEREH-HF) randomized clinical trial. JAMA Cardiol 2020 Mar 01;5(3):300-308 [FREE Full text] [doi: 10.1001/jamacardio.2019.5006] [Medline: 31734701]
61. Soga Y, Yokoi H, Ando K, Shirai S, Sakai K, Kondo K, et al. Safety of early exercise training after elective coronary stenting in patients with stable coronary artery disease. Eur J Cardiovasc Prev Rehabil 2010 Apr;17(2):230-234. [doi: 10.1097/HJR.0b013e3283359c4e] [Medline: 20051870]
62. Batalik L, Filakova K, Sladeckova M, Dosbaba F, Su J, Pepera G. The cost-effectiveness of exercise-based cardiac telerehabilitation intervention: a systematic review. Eur J Phys Rehabil Med 2023 Apr;59(2):248-258 [FREE Full text] [doi: 10.23736/S1973-9087.23.07773-0] [Medline: 36692413]

Abbreviations

AT: anaerobic threshold

CAD: coronary artery disease
CPET: cardiopulmonary exercise test
CR: cardiac rehabilitation
CVD: cardiovascular disease
MD: mean difference
PCI: percutaneous coronary intervention
peak VO2: peak oxygen uptake
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by L Buis; submitted 08.02.23; peer-reviewed by L Batalik, R Thomas; comments to author 10.05.23; revised version received 19.06.23; accepted 03.07.23; published 28.07.23.

Please cite as:
Zhong W, Liu R, Cheng H, Xu L, Wang L, He C, Wei Q
Longer-Term Effects of Cardiac Telerehabilitation on Patients With Coronary Artery Disease: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2023;11:e46359
URL: https://mhealth.jmir.org/2023/1/e46359
doi:10.2196/46359
PMID:37505803

Review

Smartphone-Assisted Medical Care for Vestibular Dysfunction as a Telehealth Strategy for Digital Therapy Beyond COVID-19: Scoping Review

Masao Noda[1,2], MD, PhD; Tatsuaki Kuroda[2,3], MD; Akihiro Nomura[4], MD, PhD; Makoto Ito[1], MD, PhD; Tomokazu Yoshizaki[5], MD, PhD; Hiroaki Fushiki[2], MD, PhD

[1]Department of Otolaryngology, Jichi Medical University, Shimotsuke, Japan

[2]Mejiro University Ear Institute Clinic, Saitama, Japan

[3]Kuroda Ear, Nose and Throat Clinic, Kumamoto, Japan

[4]College of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan

[5]Department of Otolaryngology, Kanazawa University, Kanazawa, Japan

Corresponding Author:
Hiroaki Fushiki, MD, PhD
Mejiro University Ear Institute Clinic
320 Ukiya, Iwatsukiku, Saitama-shi
Saitama, 339-8501
Japan
Phone: 81 48 797 3341
Email: fushiki@mejiro.ac.jp

Abstract

Background: Dizziness and vertigo can be caused by various factors, such as peripheral vestibular and central disorders. Although consultations with specialists are advisable when necessary, patients with severe vertigo symptoms may have limited mobility, which may interfere with hospital visits. The spread of COVID-19 has further limited the number of hospital visits for patients with dizziness; therefore, a method of medical care that enables more accurate treatment under time and geographical constraints is needed. Telemedicine has become widespread, owing to the popularity of smartphone and tablet devices in recent years, and the use of devices and systems has made it possible to provide efficient medical care. However, no previous scoping review has mapped existing studies on telemedicine for vertigo and dizziness, and no recommendations have been made regarding which devices and systems should be used for specific diseases.

Objective: The aim of this review was to map and assess previous studies on the use of information communications technology, smartphones, and apps for treating patients with vertigo and discuss the added value of introducing telemedicine to improve the quality of medical care and create an environment that builds security and trust among patients.

Methods: A scoping review was conducted with the methodological framework of Arksey and O'Malley and in accordance with the of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews) guidelines. The PubMed, MEDLINE, and Cochrane Library databases were searched to retrieve previous reports on smartphone-assisted telemedicine treatment for vertigo published between January 2000 and May 2023. Two authors independently assessed eligibility and extracted data.

Results: This review included 20 papers that reported devices or systems for telemedicine for vestibular dysfunction. Among studies that reported the use of a device or app, 2 were related to anamnesis and subjective symptoms, 12 were related to objective examination, 7 were related to remote diagnosis, and 7 were related to treatment and rehabilitation.

Conclusions: With the advancement of technology, the use of telemedicine in patients with dizziness may be feasible. In the future, it will be necessary to consider how telemedicine can be used in dizziness treatment and develop an effective treatment system combining in-person medical care and the effective use of devices for the management of severe vertigo and related diseases. The smooth introduction of telemedicine in vertigo treatment is expected to improve the quality of treatment, increase opportunities for patients to receive medical care, and reduce time and travel costs, leading to a sense of security and trust among patients.

(JMIR Mhealth Uhealth 2023;11:e48638) doi:10.2196/48638

KEYWORDS

dizziness; vertigo; telemedicine; smartphone; digital therapy; telehealth; COVID-19; information technology; scoping review; health device; remote diagnosis; medical care

Introduction

Dizziness and vertigo are common symptoms experienced by people of all ages, with lifetime prevalence estimates of 17%-30% for dizziness and 3%-10% for vertigo [1]. Dizziness and vertigo can be caused by various factors, including peripheral vestibular disorders, central disorders, circulatory dysfunction, headache-related vertigo, and psychogenic vertigo, which transcend the framework of medical departments. Among various cases of dizziness, differentiating between peripheral vertigo and stroke is clinically important for the early detection of a life-threatening condition. However, patients with severe vertigo symptoms may have limited mobility, which may interfere with hospital visits. In addition, the diagnosis may be delayed or prevented as the dizziness symptoms and significant clinical findings have subsided or are no longer available at the time of the hospital visit. Thus, a system to inform patients of their symptoms and findings outside the premises of the hospital is necessary to improve the accuracy of medical care and make an early diagnosis.

In recent years, telemedicine has become widespread owing to the popularity of smartphone and tablet devices. Moreover, the development of data communication technologies and the use of devices and systems for medical care has made it possible to provide efficient medical care [2-5]. Establishing a way to acquire observations remotely in cases of medical emergencies, such as stroke or arrhythmia during the acute phase, will lead to effective medical treatment methods in the future. The COVID-19 pandemic has significantly impacted the diagnosis and treatment of dizziness or vertigo. The spread of COVID-19 has limited the number of hospital visits for patients with dizziness, and many physicians are searching for newer methods to treat dizziness during the pandemic. Interestingly, Barreto et al [5] have set a new direction for vestibular evaluation using smartphones for the diagnosis of vestibular hypofunction and mentioned technical conditions and limitations for patient privacy. However, no guidelines have been established for treatment strategies for vertigo and dizziness, and no recommendations have been made regarding which devices and systems should be used for which diseases. Furthermore, a strategy for treating vertigo that works in a combined approach must be developed, rather than just an isolated examination or consultation method using smartphones.

With the rapid spread of telemedicine, there is a need to conduct an updated and broader literature review to develop an overview of the body of knowledge within this field. To our knowledge, no scoping review has mapped existing studies on telemedicine for vertigo and dizziness, and no recommendations have been made regarding specific devices and systems that should be used for various diseases.

Herein, we aim to map and assess previous studies on the use of information communication technology (ICT), smartphones, and apps for treating vertigo patients by reviewing previous reports on telemedicine for treating patients with vertigo using smartphones or apps.

Methods

Overall Study Design

A scoping review was performed using PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping review) guidelines [6].

The review subsequently proceeded with the stages using the framework given by Arksey and O’Malley [7]: identifying the research question and relevant studies; selecting studies; charting the data; and collating, summarizing, and reporting the results. The protocol for this scoping review has not been registered or published.

Search Strategy

We conducted a search of medical literature databases (MEDLINE via PubMed, Cochrane Library) to identify relevant reports for this scoping review on smartphone-assisted treatment for vertigo telemedicine. The search was conducted between January 1, 2000, and May 30, 2023.

The search strategy was developed by 3 authors (TK, MN, and HF) and applied to each database. A set of controlled variables (“dizziness” or “vertigo” and “telemedicine” or “teletreatment” or “Internet” or “diagnosis” and “smartphone” or “cellphone” or “application” or “Digital Therapeutics”) were used to identify relevant studies. In addition, a manual search was performed to screen the reference lists of the included papers.

All identified studies were reviewed to determine whether they answered the following questions: “How can a smartphone be used in dizziness telemedicine?” and “What types of devices and applications have been reported to be related to dizziness, especially for monitoring, testing, teleconsultation, and therapy?” We applied specific inclusion and exclusion criteria based on study type, period, type of materials, target population, and type of medical act (Textbox 1). The authors independently screened titles, abstracts, and full-text papers for inclusion in the study.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Qualitative, quantitative, and mixed methods studies published in peer-reviewed journals
- From January 1, 2000, to May 30, 2023
- All languages
- Use of web apps, smartphones, and devices for participants
- Use of materials related to dizziness and vertigo for patients
- Use of materials for participants at the stages of anamnesis, symptoms, objectives, diagnosis, and treatments

Exclusion criteria

- Conference abstracts, doctoral theses, and reviews
- Before January 1, 2000
- No use of devices or smartphones or the internet
- Use of materials for diseases not related to dizziness
- Introduction of devices and functional test of devices

Charting Data

Two investigators (MN and HF) independently extracted study-specific data using a standardized collection form with the following information: authors, year of publication, title, aim, telehealth device or app, design and methods, timing of material use, and summary.

Summarizing and Reporting Results

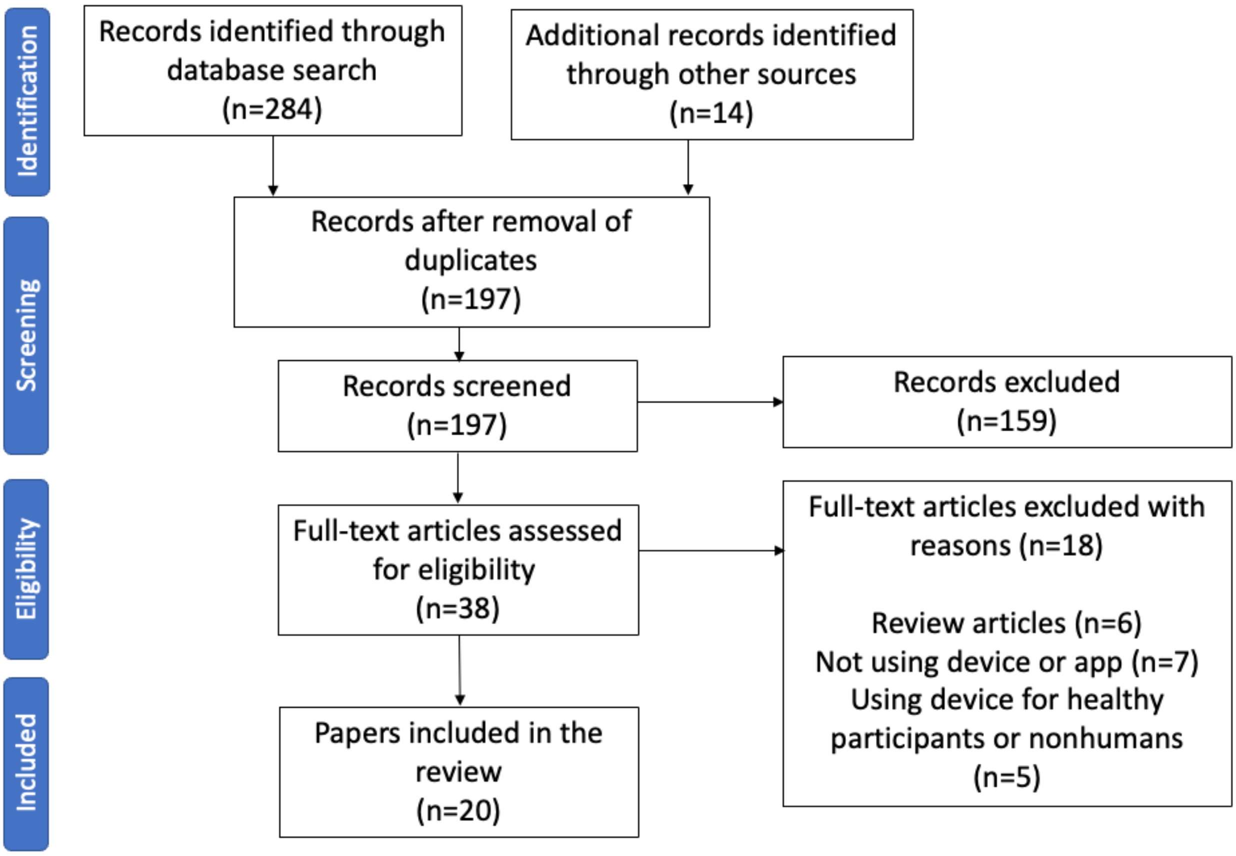
We used an inductive approach to address the research question, focusing on themes identified from the included papers. The results were summarized accordingly. The types of medical acts were divided into 4 thematic groups: anamnesis and subjective symptoms, objective tests, diagnosis, and treatment.

Results

Overview

The database and manual searches yielded 356 publications. After the removal of duplicates, the titles and abstracts of 289 publications were screened. Based on the inclusion and exclusion criteria, the full texts of 38 publications were read, 18 publications were excluded, and 20 publications were included in this review (Figure 1, Multimedia Appendix 1 [3,8-26]).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the systematic search strategy.



Among the 20 studies that reported the use of a device or app, 2 were related to anamnesis and subjective symptoms [3,8], 12 were related to objective examination [8-19], 7 were related to remote diagnosis [9,11,15,16,18,20], and 7 were related to treatment and rehabilitation [20-26] (Table 1).

Table 1. Articles included in thematic groupings.

Theme	Study	Articles, n
Used for anamnesis and subjective symptoms	• Jaensson et al [3] • Martin et al [8]	2
Used for objective tests	• Martin et al [8] • Kiroglu et al [9] • Jacob et al [10] • Gold et al [11] • Phillips et al [12] • Parker et al [13] • Ippisch et al [14] • van Bonn et al [15] • Wagle et al [16] • Riera-Tur et al [17] • Shah et al [18] • Wengier et al [19]	12
Used for diagnosis	• Kiroglu et al [9] • Gold et al [11] • Phillips et al [12] • van Bonn et al [15] • Wagle et al [16] • Shah et al [18] • Bamiou et al [20]	7
Used for treatment	• Bamiou et al [20] • Tabanfar et al [21] • van Vugt et al [22] • Kanyılmaz et al [23] • Guerrero et al [24] • Soto-Varela et al [25] • Lubetzky et al [26]	7

All papers included adult patients. Seven papers were randomized controlled trials [3,9,20-23,25]. Ten papers used a mobile phone, among which 5 included an app for the subjective visual vertical (SVV) test [10,17,19], head impulse test (HIT) [13], and hearing test [8]. Three papers used virtual reality (VR) for treatment with the Epley maneuver [21] and vestibular rehabilitation [23,26]. Eight papers included apps or devices with more than one category of medical act performed; all papers were about sequential acts [8,9,11,12,15,16,18,20].

To answer the research question, “What types of devices and applications have been reported to be related to dizziness, especially for monitoring, testing, teleconsultation, and therapy?” the results of this review are presented in 4 thematic groupings: anamnesis and subjective symptoms, objective test, diagnosis, and treatment.

Anamnesis and Subjective Symptoms

Two papers reported the use of mobile phone apps for assessing anamnesis and subjective symptoms. Both systems involved the input of daily symptoms by patients, which were then reviewed by a medical provider who monitored the postoperative symptoms and severity of symptoms related to dizziness [3,8]. Postoperative patient follow-up led to reassurance and positive feedback, and postoperative dizziness symptoms improved significantly [3]. Subjective symptoms and self-hearing tests were recorded in combination with a hearing test [8]. The severity of self-reported tinnitus increased significantly with the increase in the severity of self-reported hearing loss in the affected (P=.01) and unaffected ears (P<.001) and was useful in determining the severity of Ménière disease.

Objective Test

Twelve papers reported the use of mobile phone apps for objective tests. Seven of the reports used examination data to diagnose various diseases, including acute and chronic vertigo, in the emergency department [9,11,18]. In contrast, only 1 study combined these with subjective symptoms (Martin et al [8]) and used Dizzy Quest and an iPad-based hearing test.

Various types of tests are conducted in vertigo treatment, including balance function, nystagmus, and visual stimulation tests, such as SVV. The types of tests in the report included HIT [13], SVV [10,17,19], nystagmus [9,11,12,15,16,18], gait analysis [14], and hearing tests [8].

For nystagmus, some researchers used the camera functionality of smart devices and directly shared their videos, while others used deep learning to classify the video data [16].

In a report comparing remote and existing examinations, Parker et al [13] reported a video HIT app (iPhone) that uses the iPhone’s Augmented Reality Kit system. Shah et al [18] reported that the remote screening of benign paroxysmal positional vertigo (BPPV) is possible with high specificity. The sensitivity for diagnosing BPPV via a smartphone recording of eye movements of the Dix Hallpike test was 92.86%, with a specificity of 100% and a negative predictive value of 97.87%.

Diagnosis

Among the included papers, 7 reported the use of mobile phone apps for the diagnosis of dizziness. These apps were used after conducting objective tests, and they focused on identifying central vertigo, BPPV, and Ménière disease [9,11,18]. All of these studies used smartphone apps as diagnostic tools.

Another paper reported the EMBalance decision support system (DSS), a multilanguage platform, for diagnosing vestibular disorders in primary care settings [20]. The EMBalance DSS serves as a supportive tool by providing a structured and detailed diagnostic and management plan for various vestibular disorders. In terms of diagnostic accuracy, the study found that nonspecialist physicians using the EMBalance DSS had a higher percentage of correct diagnoses compared with those without the DSS (54% vs 41.5% correct diagnoses, respectively). As a standalone tool, the DSS demonstrated better sensitivity for first- and second-line diagnostic decisions compared with the nonspecialist group without the DSS (odds ratio 3). However, it is important to note that the specificity of the DSS was somewhat weakened.

Treatment

Seven papers reported the use of mobile phone apps or the internet for the treatment of dizziness. Physical therapy, such as head position therapy and rehabilitation, was identified as an effective treatment for vertigo. Among these 6 papers, 2 used VR technology, the Epley maneuver [21], and vestibular rehabilitation [23,26]. Both studies reported long-term therapeutic effects over a 6-month period. In addition, van Vugt et al [22] reported more effective results when the mobile phone app was incorporated into regular rehabilitation. Guerrero et al [24] examined the effects of a virtual exercise program on balance in adults with Down syndrome and reported a significant improvement compared with the pretesting level.

Discussion

Principal Findings

This scoping review mapped and assessed previous studies on the use of ICT, smartphones, and apps for treating patients with vertigo. The results demonstrated that smartphones and apps were feasible for use in dizziness telemedicine and enabled high-quality dizziness care at home and other remote locations, resulting in a practice that benefits both physicians and patients. Dizziness is a symptom perceived by people of various ages. Approximately 30% of adults aged >65 years report experiencing dizziness that interferes with their daily life [27]. Patients with dizziness are at an increased risk of falling, which interferes with daily life and limits their participation in social activities [27,28]. In addition, the time available for in-person medical care is limited in the clinic, and there are many geographical constraints. To use telemedicine and teleconsultation for patients with dizziness, considering a variety of symptoms and tests that feature in dizziness treatment and an effective combination of individualized interviews, examination, and treatment methods could lead to the development of an effective treatment strategy.

In this review of articles published since 2000, most were published after 2015, indicating an increasing trend in the need for telemedicine over the years.

Many reports on telemedicine for vertigo were published around 1990. Wolf et al [29] reported "telemetric" electronystagmography, which involved conducting nystagmus testing outside of the clinic setting. Viirre et al [30,31] reported on efforts to use a head mount device for nystagmus evaluation and VR technologies in rehabilitation. As many mechanical tests, such as nystagmus, balance function, and hearing tests, are used in clinical practice for vertigo, the medical care approach is influenced by the evolution of telecommunications and devices. Furthermore, the demand for telemedicine has been increasing since the late 2010s due to the widespread use of smartphones, internet connectivity, and the global COVID-19 pandemic. With the advent of innovative methods, there is a need to discuss the effectiveness and challenges associated with devices in vertigo management and the optimal use of these tools.

Regarding the current theoretical and practical research gaps, the use of smartphones and apps allows for a variety of medical treatments equivalent to existing dizziness treatments. Studies have been conducted on several diseases, their use in acute and chronic phases, and their application to examination, as well as diagnosis, symptoms, and rehabilitation. However, there is a lack of comprehensive use of each app and device in combination and a lack of examples of use for diagnosis and interviewing. These limitations may be attributed to the following reasons: multiple systems are not yet well coordinated, which makes medical treatment time consuming; the IT literacy of users; the development of an environment in which the systems can be used; and the medical care system, including insurance coverage.

Regarding the use of devices in the papers reviewed, most of the cases used the device for objective examination rather than other medical acts. There were a few cases of device use for multiple medical acts; however, there were no reports of device use for multiple acts, such as anamnesis-to-diagnosis or examination-to-treatment.

Dizziness is caused by various pathological conditions in the inner ear, nervous system, and musculoskeletal system. Consequently, a combination of tests, including nystagmus evaluation, balance function assessment, and imaging techniques (such as computed tomography and magnetic resonance imaging), is required for thorough examination and diagnosis. Therefore, smartphones and other devices are being increasingly used as substitutes for conventional testing methods, facilitating the diagnostic process for patients with vertigo. By combining various devices and apps, it is feasible to build a comprehensive system for vertigo treatment, and the future holds great promise.

VestAid is an innovative tablet-based system for vestibulo-ocular reflex exercises [32]. VestAid validated eye-gaze accuracy as part of the study and detected eye movement abnormalities in the participants with directed energy exposure, concussion, and vestibular neuritis [33]. Although not included in the review as this paper is in the initial evaluation stage of the system, this system is capable of recognizing facial movements and providing rehabilitation for vertigo through games. VestAid can also perform objective and subjective data collection and is expected to be used for multiple medical acts in various apps in the clinical practice of vertigo.

In the management of dizziness, subjective symptoms and objective findings are used for diagnosis, and in some cases, the final diagnosis depends on the response after treatment. Thus, comprehensive dizziness care leads to improvement in the quality of care. For chronic dizziness, diagnosis may be confirmed by excluding other diseases, depending on the accompanying symptoms and their severity. Among individuals under the age of 20 years, migraine-associated dizziness, including vestibular migraine and benign paroxysmal vertigo of childhood or orthostatic dysregulation, are the most common types of vertigo; however, there are cases in which the two overlap, and the diagnosis is unclear. There have been several reports of anxiety and fear, anemia, menstruation, hypotension, insomnia, headache, stiff shoulders, and neck, with coexisting or overlapping menopausal symptoms in women with dizziness. It is essential to monitor dizziness and lightheadedness during daily activities in such cases.

For diagnosis using a smart device, diseases related to the diagnosis or exclusion were central vertigo identification, BPPV, and Ménière disease [9,11,18]. The dizziness experienced by the patient is mild in many cases of recurrent dizziness, such as Ménière disease, delayed endolymphatic hydrops, vestibular migraine, and BPPV, and there are no significant findings at the time of presentation, making a definitive diagnosis difficult. Therefore, eye movement findings during a vertigo attack are crucial for a definitive diagnosis. Kiroglu et al [9] reported the use of a cell phone camera for diagnosing Ménière disease. Compared with the conventional method of diagnosis, no significant differences were observed in the rate of diagnosis of Ménière disease when recording eye movement during a vertigo attack [9].

There have also been several instances of artificial intelligence implementation in nystagmus and vestibular rehabilitation using VR. The accuracy of video-oculography using a smartphone has shown steady improvement. Parker et al [34] and Friedrich et al [35] reported, respectively, that ARkit-based apps and ConVNG are developing methods ranging from pupil detection to nystagmus analysis, which are expected to further improve the accuracy of diagnosis in smartphone vertigo practice [34,35].

Examination findings in vertigo treatment are often judged by referring to waveforms and images, which are also quantified, making it a suitable field for machine learning and deep learning. In addition, medical devices that monitor eye movements, such as Frenzel glasses, are similar to VR goggles and could be introduced to both physicians and patients without any significant modifications. With technological advancements, the quality and accuracy of vertigo medical care are expected to improve further.

Concerning the target users of the system, most apps were used by patients, and the data were confirmed by the medical staff. The current style of medical care mainly involves interaction between physicians and patients. There is an effort to change this style, as enhancing collaboration between patients and physicians as well as between physicians is expected to further expand the range of medical care that can be provided. Physician-to-patient telemedicine has many advantages, such as the reduction of time and travel costs, reduction in the waiting period for medical care, increased opportunities to see patients, avoidance of risk of spread of infection during pandemics, and complementary medical care in the event of a patient becoming a high-risk contact.

The EMBalance DSS developed by Bamiou et al [20] describes a system in which primary care physicians consult with specialists. EMBalance is effective both as a supporting tool and a standalone tool, creating an opportunity for appropriate medical care in the absence of specialists. Balance disorders can occur due to various causes, including peripheral vertigo caused by the inner ear and vestibular nerves, problems in the cerebellum and brainstem, circulatory dysregulation (ie, arrhythmia), and dizziness related to migraines, anxiety, and depression. Although circumstances may vary by region, this type of consultation between general physicians and specialists may be necessary. Nevertheless, specialists in dizziness are unevenly distributed, and there is a possibility of disparities in the medical care they can provide. Telemedicine enables collaboration between physicians and patients, as well as among physical therapists, pharmacists, and patients' families. Regarding the execution of rehabilitation and sharing of medical conditions, the geographical limitations of conventional in-person consultations will be eliminated.

There are challenges and limitations in the dissemination of dizziness telemedicine, such as privacy, communication environment, information leakage, and ICT literacy. Although it is technically possible to provide medical care equivalent to in-person consultation, it is necessary to clarify the safety and legal issues related to apps and devices. Another challenge is differentiating between acute and emergency dizziness.

Serious conditions must be ruled out in the emergency department as telemedicine is not always able to respond to emergencies for which contact is required. It is important to differentiate between BPPV, Ménière disease, first attacks (ie, vestibular migraine), and peripheral dizziness (ie, vestibular neuritis and stroke) in cases of acute dizziness. Currently, the Head Impulse, Nystagmus, and Test of Skew are recommended to differentiate vertigo from stroke [36]. Some studies have used telemedicine in emergency settings for medical care coordination of acute vertigo [37,38]. This study will contribute to organizing the existing practice and clarifying the possibilities and challenges of combining the two in considering a system of telemedicine using smartphones and devices in vertigo practice.

No previous study has reviewed the use of smartphones for vertigo telemedicine. We believe that the expansion of vertigo telemedicine based on this study will lead to increased opportunities to see patients, reduced treatment and travel time, reduced risk of infection, and improved quality of care. It also clarifies issues such as limitations in the acute setting and app combinations, which are crucial for achieving an appropriate comprehensive vertigo treatment system. A strength of this review was that we used an acknowledged framework for conducting scoping reviews, in addition to the PRISMA-ScR for guiding the reporting of the review. In addition, a broad, comprehensive, and systematic search was performed to identify published studies. To our knowledge, this review is the first to examine the characteristics of smartphones and smart devices to support dizziness telemedicine inclusively. Nevertheless, this review has some limitations. Searching the literature for "smartphone-assisted dizziness therapy" was difficult due to the diversity in the language used to describe such materials and interventions. Second, some articles may have been missed. Although multiple reviewers extracted data from included articles, unclear descriptions may have arisen.

Conclusions

This review evaluated the use and benefits of digital technologies, such as smartphones and apps, in the treatment of vertigo, highlighting the potential of telemedicine in enhancing care quality and building patient trust. Despite the increased adoption of these technologies, further improvements in individual and combined systems are needed for better accuracy.

Authors' Contributions

MN, TK, and HF conceived the project and wrote the manuscript. MN, TK, AN, and HF designed and conducted the experiments and summarized the data. MN coordinated the project. MI and TY contributed substantially to drafting the manuscript. All authors critically reviewed and revised the manuscript draft for intellectual content. All authors approved the final version of the manuscript to be published.

Conflicts of Interest

Akihiro Nomura received consulting fees and research grants from CureApp, Inc.

Multimedia Appendix 1
The aim and summary of reviewed studies.
[DOCX File , 2227 KB - mhealth_v11i1e48638_app1.docx]

References

1. Murdin L, Schilder AGM. Epidemiology of balance symptoms and disorders in the community: a systematic review. Otol Neurotol 2015;36(3):387-392. [doi: 10.1097/MAO.0000000000000691] [Medline: 25548891]
2. Nomura A, Tanigawa T, Muto T, Oga T, Fukushima Y, Kiyosue A, et al. Clinical efficacy of telemedicine compared to face-to-face clinic visits for smoking cessation: multicenter open-label randomized controlled noninferiority trial. J Med Internet Res 2019;21(4):e13520 [FREE Full text] [doi: 10.2196/13520] [Medline: 30982776]
3. Jaensson M, Dahlberg K, Eriksson M, Nilsson U. Evaluation of postoperative recovery in day surgery patients using a mobile phone application: a multicentre randomized trial. Br J Anaesth 2017;119(5):1030-1038 [FREE Full text] [doi: 10.1093/bja/aex331] [Medline: 29077818]
4. Aida A, Svensson T, Svensson AK, Urushiyama H, Okushin K, Oguri G, et al. Using mHealth to provide mobile app users with visualization of health checkup data and educational videos on lifestyle-related diseases: methodological framework for content development. JMIR Mhealth Uhealth 2020;8(10):e20982 [FREE Full text] [doi: 10.2196/20982] [Medline: 33084586]
5. Barreto RG, Yacovino DA, Cherchi M, Nader SN, Teixeira LJ, da Silva DA, et al. The role of the smartphone in the diagnosis of vestibular hypofunction: a clinical strategy for teleconsultation during the COVID-19 pandemic and beyond. Int Arch Otorhinolaryngol 2021;25(4):e602-e609 [FREE Full text] [doi: 10.1055/s-0041-1736340] [Medline: 34777592]
6. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
8. Martin EC, Verkaik R, Stultiens JJA, van de Berg MR, Janssen AML, Leue C, et al. The DizzyQuest: relation between self-reported hearing loss, tinnitus and objective hearing thresholds in patients with Meniere's disease. J Neurol 2022;269(10):5239-5248 [FREE Full text] [doi: 10.1007/s00415-021-10909-8] [Medline: 34894282]
9. Kıroğlu M, Dağkıran M. The role of mobile phone camera recordings in the diagnosis of Meniere's disease and pathophysiological implications. J Int Adv Otol 2020;16(1):18-23 [FREE Full text] [doi: 10.5152/iao.2019.6605] [Medline: 32066548]
10. Brodsky JR, Cusick BA, Kawai K, Kenna M, Zhou G. Peripheral vestibular loss detected in pediatric patients using a smartphone-based test of the subjective visual vertical. Int J Pediatr Otorhinolaryngol 2015;79(12):2094-2098 [FREE Full text] [doi: 10.1016/j.ijporl.2015.09.020] [Medline: 26434547]

11. Gold D, Tourkevich R, Shemesh A, Brune A, Choi W, Peterson S, et al. A novel tele-dizzy consultation program in the emergency department using portable video-oculography to improve peripheral vestibular and stroke diagnosis. Neurology 2019;92(15 Supplement):S28.002.
12. Phillips JS, Newman JL, FitzGerald JE, Cox SJ. Implications of vestibular telemetry for the management of Ménière's disease—our experience with three adults. Clin Otolaryngol 2020;46(2):425-429 [FREE Full text] [doi: 10.1111/coa.13676]
13. Parker TM, Farrell N, Otero-Millan J, Kheradmand A, McClenney A, Newman-Toker DE. Proof of concept for an "eyePhone" app to measure video head impulses. Digit Biomark 2021;5(1):1-8 [FREE Full text] [doi: 10.1159/000511287] [Medline: 33615116]
14. Ippisch R, Jelusic A, Bertram J, Schniepp R, Wuehr M. mVEGAS—mobile smartphone-based spatiotemporal gait analysis in healthy and ataxic gait disorders. Gait Posture 2022;97:80-85 [FREE Full text] [doi: 10.1016/j.gaitpost.2022.07.256] [Medline: 35914387]
15. van Bonn SM, Behrendt SP, Pawar BL, Schraven SP, Mlynski R, Schuldt T. Smartphone-based nystagmus diagnostics: development of an innovative app for the targeted detection of vertigo. Eur Arch Otorhinolaryngol 2022;279(12):5565-5571 [FREE Full text] [doi: 10.1007/s00405-022-07385-9] [Medline: 35451613]
16. Wagle N, Morkos J, Liu J, Reith H, Greenstein J, Gong K, et al. aEYE: a deep learning system for video nystagmus detection. Front Neurol 2022;13:963968 [FREE Full text] [doi: 10.3389/fneur.2022.963968] [Medline: 36034311]
17. Riera-Tur L, Caballero-Garcia A, Martin-Mateos AJ, Lechuga-Sancho AM. Efficacy of the subjective visual vertical test performed using a mobile application to detect vestibular pathology. J Vestib Res 2022;32(1):21-27. [doi: 10.3233/VES-201526] [Medline: 34180441]
18. Shah MU, Lotterman S, Roberts D, Eisen M. Smartphone telemedical emergency department consults for screening of nonacute dizziness. Laryngoscope 2019;129(2):466-469. [doi: 10.1002/lary.27424] [Medline: 30478898]
19. Wengier A, Ungar OJ, Handzel O, Cavel O, Oron Y. Subjective visual vertical evaluation by a smartphone-based test-taking the phone out of the bucket. Otol Neurotol 2021;42(3):455-460. [doi: 10.1097/MAO.0000000000002944] [Medline: 33555757]
20. Bamiou DE, Kikidis D, Bibas T, Koohi N, Macdonald N, Maurer C, et al. Diagnostic accuracy and usability of the EMBalance decision support system for vestibular disorders in primary care: proof of concept randomised controlled study results. J Neurol 2022;269(5):2584-2598 [FREE Full text] [doi: 10.1007/s00415-021-10829-7] [Medline: 34669009]
21. Tabanfar R, Chan HHL, Lin V, Le T, Irish JC. Development and face validation of a Virtual Reality Epley Maneuver System (VREMS) for home epley treatment of benign paroxysmal positional vertigo: a randomized, controlled trial. Am J Otolaryngol 2018;39(2):184-191 [FREE Full text] [doi: 10.1016/j.amjoto.2017.11.006] [Medline: 29169952]
22. van Vugt VA, van der Wouden JC, Essery R, Yardley L, Twisk JWR, van der Horst HE, et al. Internet based vestibular rehabilitation with and without physiotherapy support for adults aged 50 and older with a chronic vestibular syndrome in general practice: three armed randomised controlled trial. BMJ 2019;367:l5922 [FREE Full text] [doi: 10.1136/bmj.l5922] [Medline: 31690561]
23. Kanyılmaz T, Topuz O, Ardıç FN, Alkan H, Öztekin SNS, Topuz B, et al. Effectiveness of conventional versus virtual reality-based vestibular rehabilitation exercises in elderly patients with dizziness: a randomized controlled study with 6-month follow-up. Braz J Otorhinolaryngol 2022;88 Suppl 3(Suppl 3):S41-S49 [FREE Full text] [doi: 10.1016/j.bjorl.2021.08.010] [Medline: 34799265]
24. Guerrero K, Umagat A, Barton M, Martinez A, Ho KY, Mann S, et al. The effect of a telehealth exercise intervention on balance in adults with down syndrome. J Appl Res Intellect Disabil 2023;36(2):385-393. [doi: 10.1111/jar.13068] [Medline: 36585748]
25. Soto-Varela A, Rossi-Izquierdo M, Del-Río-Valeiras M, Faraldo-García A, Vaamonde-Sánchez-Andrade I, Lirola-Delgado A, et al. Vestibular rehabilitation with mobile posturography as a "low-cost" alternative to vestibular rehabilitation with computerized dynamic posturography, in old people with imbalance: a randomized clinical trial. Aging Clin Exp Res 2021;33(10):2807-2819. [doi: 10.1007/s40520-021-01813-2] [Medline: 33677737]
26. Lubetzky AV, Kelly J, Wang Z, Gospodarek M, Fu G, Sutera J, et al. Contextual sensory integration training via head mounted display for individuals with vestibular disorders: a feasibility study. Disabil Rehabil Assist Technol 2022;17(1):74-84. [doi: 10.1080/17483107.2020.1765419] [Medline: 32421374]
27. Neuhauser HK. Epidemiology of vertigo. Curr Opin Neurol 2007;20(1):40-46. [doi: 10.1097/WCO.0b013e328013f432] [Medline: 17215687]
28. Yin M, Ishikawa K, Wong WH, Shibata Y. A clinical epidemiological study in 2169 patients with vertigo. Auris Nasus Larynx 2009;36(1):30-35. [doi: 10.1016/j.anl.2008.03.006] [Medline: 18486378]
29. Wolf SR, Christ P, Haid CT. "Telemetric" electronystagmography: a new method for examination of nystagmus outside the clinic. Acta Otolaryngol Suppl 1991;481:374-381. [doi: 10.3109/00016489109131426] [Medline: 1927422]
30. Viirre E. Vestibular telemedicine and rehabilitation. Applications for virtual reality. Stud Health Technol Inform 1996;29:299-305. [Medline: 10163763]
31. Viirre E, Warner D, Balch D, Nelson JR. Remote medical consultation for vestibular disorders: technological solutions and case report. Telemed J 1997;3(1):53-58. [doi: 10.1089/tmj.1.1997.3.53] [Medline: 10166445]

32. Hovareshti P, Roeder S, Holt LS, Gao P, Xiao L, Zalkin C, et al. VestAid: a tablet-based technology for objective exercise monitoring in vestibular rehabilitation. Sensors (Basel) 2021;21(24):8388 [FREE Full text] [doi: 10.3390/s21248388] [Medline: 34960480]
33. Whitney SL, Ou V, Hovareshti P, Costa CM, Cassidy AR, Dunlap PM, et al. Utility of VestAid to detect eye-gaze accuracy in a participant exposed to directed energy. Mil Med 2022:usac294 [FREE Full text] [doi: 10.1093/milmed/usac294] [Medline: 36208334]
34. Parker TM, Badihian S, Hassoon A, Tehrani ASS, Farrell N, Newman-Toker DE, et al. Eye and head movement recordings using smartphones for telemedicine applications: measurements of accuracy and precision. Front Neurol 2022;13:789581 [FREE Full text] [doi: 10.3389/fneur.2022.789581] [Medline: 35370913]
35. Friedrich MU, Schneider E, Buerklein M, Taeger J, Hartig J, Volkmann J, et al. Smartphone video nystagmography using convolutional neural networks: ConVNG. J Neurol 2023;270(5):2518-2530 [FREE Full text] [doi: 10.1007/s00415-022-11493-1] [Medline: 36422668]
36. Newman-Toker DE, Kerber KA, Hsieh YH, Pula JH, Omron R, Tehrani ASS, et al. HINTS outperforms ABCD2 to screen for stroke in acute continuous vertigo and dizziness. Acad Emerg Med 2013;20(10):986-996 [FREE Full text] [doi: 10.1111/acem.12223] [Medline: 24127701]
37. Müller-Barna P, Hubert ND, Bergner C, Schütt-Becker N, Rambold H, Haberl RL, et al. TeleVertigo: diagnosing stroke in acute dizziness: a telemedicine-supported approach. Stroke 2019;50(11):3293-3298 [FREE Full text] [doi: 10.1161/STROKEAHA.119.026505] [Medline: 31607244]
38. Green KE, Pogson JM, Otero-Millan J, Gold DR, Tevzadze N, Tehrani ASS, et al. Opinion and special articles: remote evaluation of acute vertigo: strategies and technological considerations. Neurol 2021;96(1):34-38 [FREE Full text]

Abbreviations

DSS: decision support system
HIT: head impulse test
ICT: information communications technology
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SVV: subjective visual vertical
VR: virtual reality

Edited by M Dorsch; submitted 02.05.23; peer-reviewed by B Pawar; comments to author 31.05.23; revised version received 13.06.23; accepted 08.08.23; published 11.09.23.

Please cite as:
Noda M, Kuroda T, Nomura A, Ito M, Yoshizaki T, Fushiki H
Smartphone-Assisted Medical Care for Vestibular Dysfunction as a Telehealth Strategy for Digital Therapy Beyond COVID-19: Scoping Review
JMIR Mhealth Uhealth 2023;11:e48638
URL: https://mhealth.jmir.org/2023/1/e48638
doi:10.2196/48638
PMID:37695671

Original Paper

Effects of a Mobile-Based Intervention for Parents of Children With Crying, Sleeping, and Feeding Problems: Randomized Controlled Trial

Michaela Augustin[1], MSc; Maria Licata-Dandel[1,2], PhD; Linda D Breeman[3], PhD; Mathias Harrer[4,5], MSc; Ayten Bilgin[6], PhD; Dieter Wolke[7,8], PhD; Volker Mall[1,2], Dr med; Margret Ziegler[2], Dr med; David Daniel Ebert[4], PhD; Anna Friedmann[1], PhD

[1]Social Pediatrics, TUM School of Medicine, Technical University of Munich, Munich, Germany
[2]kbo-Kinderzentrum Munich, Munich, Germany
[3]Health, Medical, and Neuropsychology Unit, Leiden University, Leiden, Netherlands
[4]Psychology & Digital Mental Health Care, Department of Sports and Health Sciences, Technical University of Munich, Munich, Germany
[5]Clinical Psychology and Psychotherapy, Institute for Psychology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
[6]School of Psychology, University of Kent, Canterbury, United Kingdom
[7]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom
[8]Department of Psychology, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Michaela Augustin, MSc
Social Pediatrics
TUM School of Medicine
Technical University of Munich
Heiglhofstr.65
Munich, 81377
Germany
Phone: 49 8971009149
Email: michaela.augustin@tum.de

Abstract

Background: Excessive crying, sleeping, and feeding problems in early childhood are major stressors that can result in parents feeling socially isolated and having low self-efficacy. Affected children are a risk group for being maltreated and developing emotional and behavioral problems. Thus, the development of an innovative and interactive psychoeducational app for parents of children with crying, sleeping, and feeding problems may provide low-threshold access to scientifically based information and reduce negative outcomes in parents and children.

Objective: We aimed to investigate whether following the use of a newly developed psychoeducational app, the parents of children with crying, sleeping, or feeding problems experienced less parenting stress; gained more knowledge about crying, sleeping, and feeding problems; and perceived themselves as more self-effective and as better socially supported and whether their children's symptoms decreased more than those of the parents who did not use the app.

Methods: Our clinical sample consisted of 136 parents of children (aged 0-24 months) who contacted a cry baby outpatient clinic in Bavaria (Southern Germany) for an initial consultation. Using a randomized controlled design, families were randomly allocated to either an intervention group (IG; 73/136, 53.7%) or a waitlist control group (WCG; 63/136, 46.3%) during the usual waiting time until consultation. The IG was given a psychoeducational app that included evidence-based information via text and videos, a child behavior diary function, a parent chat forum and experience report, tips on relaxation, an emergency plan, and a regional directory of specialized counseling centers. Outcome variables were assessed using validated questionnaires at baseline test and posttest. Both groups were compared at posttest regarding changes in parenting stress (primary outcome) and secondary outcomes, namely knowledge about crying, sleeping, and feeding problems; perceived self-efficacy; perceived social support; and child symptoms.

Results: The mean individual study duration was 23.41 (SD 10.42) days. The IG reported significantly lower levels of parenting stress (mean 83.18, SD 19.94) after app use compared with the WCG (mean 87.46, SD 16.67; P=.03; Cohen d=0.23). Furthermore, parents in the IG reported a higher level of knowledge about crying, sleeping, and feeding (mean 62.91, SD 4.30) than those in the WCG (mean 61.15, SD 4.46; *P*<.001; Cohen *d*=0.38). No differences at posttest were found between groups in terms of parental efficacy (*P*=.34; Cohen *d*=0.05), perceived social support (*P*=.66; Cohen *d*=0.04), and child symptoms (*P*=.35; Cohen *d*=0.10).

Conclusions: This study provides initial evidence of the efficacy of a psychoeducational app for parents with child crying, sleeping, and feeding problems. By reducing parental stress and increasing knowledge of children's symptoms, the app has the potential to serve as an effective secondary preventive measure. Additional large-scale studies are needed to investigate long-term benefits.

Trial Registration: German Clinical Trials Register DRKS00019001; https://drks.de/search/en/trial/DRKS00019001

(JMIR Mhealth Uhealth 2023;11:e41804) doi:10.2196/41804

KEYWORDS

children; crying problems; sleeping problems; feeding problems; feeding; regulatory problems; intervention study; Mobile Health Care; health app; mobile app; patient education; psychoeducation; eHealth; mobile health; mHealth; parenting; baby; babies; sleep; crying; newborn; mobile phone

Introduction

Definition

Excessive crying, sleeping, and feeding problems are common sources of stress for parents during early childhood [1,2]. They are characterized by self-regulation difficulties [2], which may manifest as fussiness, prolonged crying, difficulties transitioning to sleep or maintaining sleep, or refusal of new or selective foods [3]. Some children may experience a single problem, whereas others experience multiple problems simultaneously or persistently from early childhood to toddlerhood [2,4].

Prevalence

The prevalence rates of crying, sleeping, and feeding problems vary widely depending on the exact definition of the problems, diagnostic procedures, and the study population. For excessive crying, the incidence rates range from 5% to 26% in the first 18 months of life [5-7]. Sleeping problems in early childhood are evident in 10% to 33.3% of children aged <24 months [7-11], whereas mild to moderate feeding difficulties occur in 20% to 43% of children in the first year of life [7,10,12,13]. Two or more simultaneous problems of feeding, sleeping, or excessive crying emerge in approximately 1% to 26.6% of children in their first 2 years of life [7,10,14]. Although in many cases early crying, sleeping, and feeding problems are transient [14-16], persistent problems are associated with short- and long-term negative consequences for child health and development.

Associated Short- and Long-term Risks

Several factors are associated with the development of excessive crying, sleeping, and feeding problems, and they include neurodevelopmental vulnerability, high levels of parenting stress, parental pre- and postnatal psychopathology such as depression and anxiety, and impaired parent-child relationships [10,14,17-20]. In addition to enormous exhaustion and dejection, parents often feel helpless and incompetent in their parenting skills, which can be reinforced by persistent crying despite all their efforts to calm and satisfy their babies [2,21-23]. Furthermore, affected families often report that they feel increasingly socially isolated and lack support from their social environment [2,24,25]. Affected families bear a high risk of stressful experiences in which childcare demands outbalance parental resources. Thus, children who cry excessively are considered a specific risk group for child maltreatment, for example, by shaking the child, which can have severe long-term and even life-threatening consequences (shaken baby syndrome) [26-29]. Impairment of the parent-child relationship may persist for several years [14,30-32]. Other long-term consequences include the persistence of sleeping and feeding problems beyond the first months of life up to preschool and even elementary school age [15,16,33] and an increased risk of developing other mental health problems, such as emotional and behavioral problems, as well as cognitive deficits [1,4,33-37]. There is emerging evidence that multiple or persistent crying, sleeping, and feeding problems may have long-lasting effects on attention and avoidant personality symptoms in adulthood [4,27,38]. In summary, early crying, sleeping, and feeding problems are associated with high family stress and increase the risk of impaired child development.

Psychoeducation as a Suitable Intervention Tool for Affected Families

Thus, it is necessary to provide support to affected families as early as possible to reduce parenting stress. This can contribute to preventing crisis situations as well as child and parent mental health problems in the short and long term [39-41]. Psychoeducation is generally an essential tool for the prevention and treatment of mental illness and stress. In addition to increasing knowledge about the child's problem or disorder [42,43], psychoeducational interventions can promote empowerment by improving subjective coping strategies [44-46]. This is also relevant for families of children with early crying, sleeping, and feeding problems, as parents consider corresponding reliable information helpful for understanding children's symptom patterns and signals [24,47,48]. Moreover, Gilkerson et al [49] found that an early psychoeducational intervention program for parents of excessively crying children improved parenting self-efficacy. Findings suggest that psychoeducational interventions not only have a positive impact on parental parameters but can also reduce early child symptoms such as sleeping problems [50]. Furthermore, Hiscock et al [47] demonstrated the positive effects of a psychoeducational intervention on sleeping and crying problems in a subgroup of children who were fed very frequently.

Need for a Low-Threshold Intervention

Although early intervention is necessary and helpful for affected families, there could be barriers to seeking professional counseling. For example, parents might fear that they will be judged negatively by their social environment as well as by professional health care workers [24]. One possible way to reach affected families at an early stage is by a smartphone app. In Germany, approximately 87% of people have smartphones in their households [51,52]. With >101,000 health- and fitness-related offers in app stores worldwide in 2020 [53], apps are an increasingly popular way to obtain health-related information. Research indicates positive impacts of psychoeducational apps on both parent and child outcomes. Shorey et al [54] showed that an app-based educational program increases parental self-efficacy, social support, and parenting satisfaction during the postpartum period. App-based sleep interventions were found to reduce sleeping problems in children and improve sleep patterns in children aged 6 to 12 months [55] and from the age of 2 years [56]. Furthermore, parents who used features such as tracking feeding behavior reported higher perception of control and self-efficacy [57]. These results indicate that digitized psychoeducational interventions could be effective tools to support families with children with early crying, sleeping, and feeding problems. However, most apps provided in app stores are neither tested for effectiveness nor based on scientifically sound content [58]. Furthermore, most apps such as specialized sleeping or feeding apps target only one of the symptoms [55,56], although the symptoms frequently occur simultaneously in a complex manner [2,4,14]. To our knowledge, no apps are available to date that specifically target crying, sleeping, and feeding problems as common symptom patterns in early childhood. In addition, many apps (such as mere tracking apps) are limited to a few functions, instead of combining psychoeducational and interactive tools [59], and rarely address both child symptoms and parental outcomes, for example, parenting stress. Thus, we developed a new psychoeducational app targeting both parental and child outcomes as a low-threshold early support offer for affected families. In addition to scientifically based information on crying, sleeping, and feeding problems, the app included interactive tools such as a symptom diary function, relaxation strategies for parents, an emergency plan, and contact information for professional counseling centers.

Study Aim

In this study, we aimed to evaluate the effectiveness of the app in a clinical sample by including an intervention group (IG) and a waitlist control group (WCG). We hypothesized that, compared with the WCG, the IG using the app would have a significant reduction in the primary outcome, that is, parenting stress. Furthermore, we aimed to explore the app's effects on secondary outcomes, including parental knowledge about crying, sleeping, and feeding problems; parental self-efficacy; perceived social support; and child crying, sleeping, and feeding problems.

Methods

Study Design

The app was evaluated in a monocentric, prospective, and randomized controlled intervention study using a pretest-posttest (posttest [t2]) design and a WCG from 2019 to 2022. The methods and results of this study are presented in accordance with the CONSORT (Consolidated Standards of Reporting Trials) Statement [60], the CONSORT-EHEALTH (CONSORT of Electronic and Mobile Health Applications and Online Tele Health) checklist (Multimedia Appendix 1), and the Guidelines for Executing and Reporting Research on Internet Interventions [61].

Ethics Approval

The study protocol was approved by the Ethics Committee of the Technical University of Munich (vote number: 56/18 S). The study was registered with the German Register of Clinical Studies (DRKS; register number: DRKS00019001; Multimedia Appendix 2).

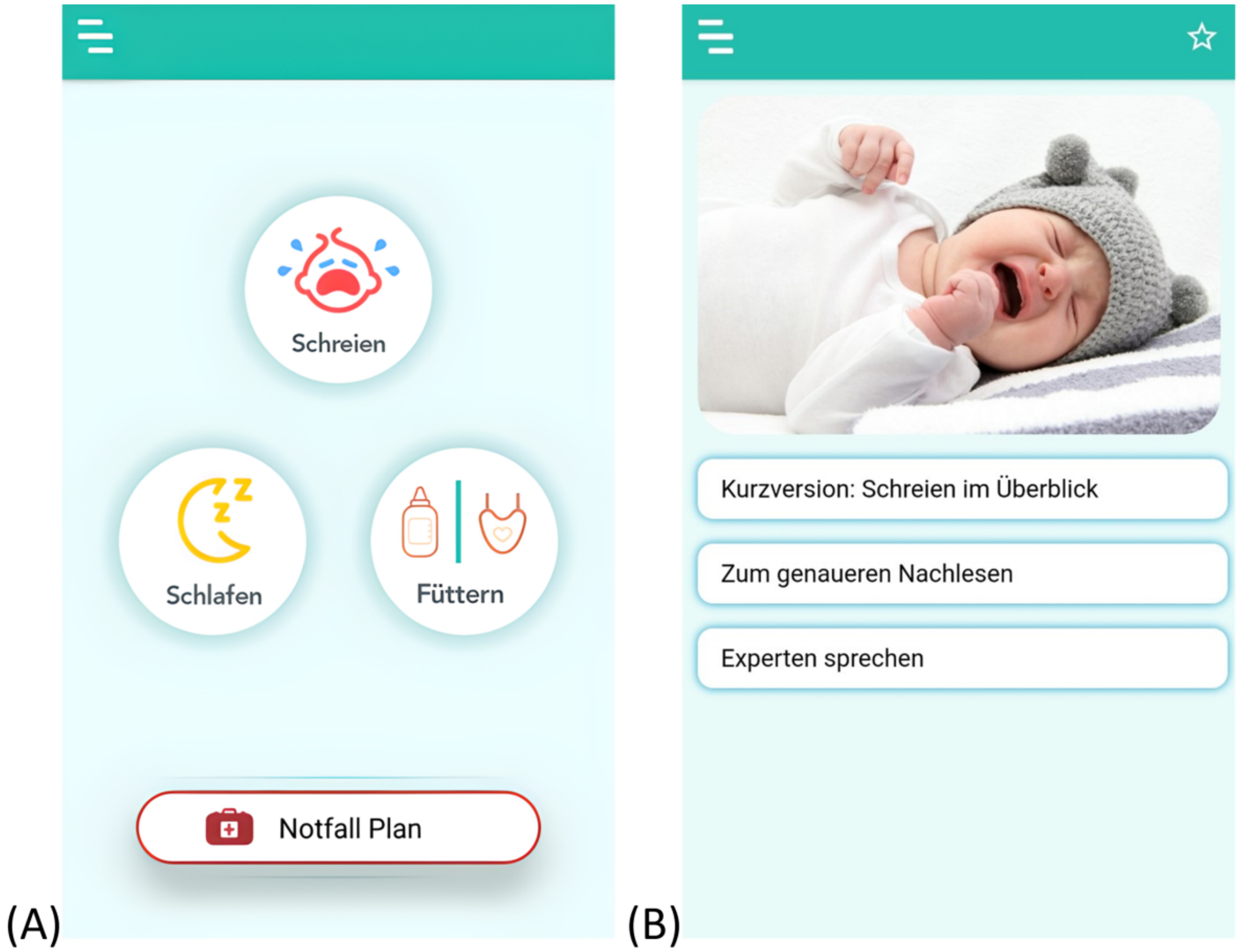
Recruitment and Procedure

The target group were German-speaking parents of children aged 0 to 24 months who contacted a cry baby outpatient clinic in Bavaria (Southern Germany) for the first consultation because of crying, sleeping, or feeding problems. To avoid confounding the effects of app use with the effects of counseling, the individual study phase ended before the first appointment at the outpatient clinic (posttest [t2]). Participants whose possible study duration was very short (<10 days from first contact with the study team until counseling appointment in the outpatient clinic) were excluded because the applied questionnaires referred to a report period of at least 1 week. During the initial phone contact with the clinic, interested parents who met the inclusion criteria were referred to the study team. After verbal consent, study information, declaration of consent, and baseline test questionnaires (t1) were sent to their home addresses. As soon as the study team received the original signed informed consent and t1 questionnaires, families were included and randomly assigned to the IG or the WCG. Randomization was conducted by an independent researcher using Research Randomizer [62]. Participants were not blinded to the study conditions, and they received access to the app by email at different time points. While the IG obtained the app for the duration of the regular waiting period until the first counseling appointment, the WCG received it only after the first counseling appointment. A few days before the counseling appointment, t2 questionnaires were sent by post to the participants, and completed forms were collected by the study team just before the counseling appointment. To avoid sequence effects, all questionnaires were administered in permuted order. The individual study duration corresponded to the average waiting time for an initial counseling appointment (mean 3 weeks). During the study, participants were contacted repeatedly via email to receive a confirmation or reminder: (1) after 1 week if they had not returned the t1 questionnaires by then; (2) after the t1 questionnaires arrived; (3) five days after the app was unlocked to see whether the installation was successful; (4) one week before the initial counseling appointment in the outpatient clinic to remind them to bring the completed t2 questionnaire; and (5) after the study ended with a brief thank you note for participation.

Intervention

The overall aim of the app was to provide psychoeducation regarding early childhood crying, sleeping, and feeding problems (for screenshot examples, see Figure 1). In line with the Medical Research Council guidelines [63], the app was developed in a working consortium of international scientists and clinical health care professionals (n=10) in the field of early crying, sleeping, and feeding problems and was piloted regarding its first impression by parents and clinical experts (n=15; M Augustin et al, unpublished data, December 2022). The first final version of the app consisted of the following main components: (1) informational texts including tips on how to deal adequately with a child's behavior were provided in a short version in simple language as well as in a more detailed version; (2) video interviews with experienced clinical experts addressing frequent questions; (3) an emergency plan in acute situations of excessive demands providing guidance for de-escalation; (4) a profile and diary function enabling parents to document their child's symptoms; (5) self-care strategies suitable for everyday use addressing parents' own needs; (6) an experience report of one family with a child with crying problems showing frequent problems of affected families; (7) a chat forum providing the opportunity to contact other affected parents; and (8) a regional register of counseling centers in Bavaria aiming to encourage affected parents to seek professional support at an early stage. Parents had the app at their free disposal and could decide how often they wanted to use it.

Figure 1. App intervention: screenshot examples. (A) Main screen and (B) menu for psychoeducational content concerning child crying.



Measures

Primary Outcome: Parenting Stress

Parenting stress was assessed at t1 and t2 with the Eltern-Belastungs-Inventar (EBI) [64], which is a German adaptation of the Parenting Stress Index [65]. The questionnaire contained 48 items covering the child domain (stress emanating from the child's behavior) and the parent domain (impairments of parental functions). In this study, the overall parent domain score was applied as the outcome variable, and the 7 parent domain subscales (attachment, isolation, parental competence, depression, health, role restriction, and spouse-related stress) were used in the secondary exploratory analysis. Responses were given on a 5-point-Likert scale (1=strongly agree to 5=strongly disagree). The internal consistency (Cronbach α) of the EBI parent domain has been shown to be excellent (.93). External validity has been examined in validation studies using different samples. Results showed, among others, moderate to high correlations of the EBI with other stress indicators, as well as related constructs of parenting stress [66].

Secondary Outcomes

Knowledge About Crying, Sleeping, and Feeding

A multiple-choice test on crying, sleeping, and feeding (self-developed and based on the contents of the app) assessed the parents' level of knowledge at t1 and t2 using 18 questions with 4 answer options, of which one or more were correct. An example item is as follows: "It is natural for children to cry (eg, due to tummy ache). However, at a certain age, the crying frequency decreases in most babies. When does this occur?—after 6 weeks—from 3 months of life—from 6 months of life—from 12 months of life." The maximum possible score was 72. The internal reliability was acceptable, with a Cronbach α of .72.

In a face validity test, participants rated all items as very suitable or mostly suitable (Table S1 in Multimedia Appendix 3).

Parental Self-efficacy

The Perceived Maternal Parenting Self-Efficacy Questionnaire (PMP-SE) [67] measured parental self-efficacy in dealing with the child at t1 and t2. A total score was computed based on 20 items with 4 subscales (caretaking procedures, evoking behaviors, reading behaviors or signaling, and situational beliefs). Items were scored on a 4-point Likert scale ranging from 1 (do not agree at all) to 4 (totally agree). Internal reliability has been proven to be excellent, with a Cronbach α of .91. The predictive and criterion validity have been previously reported [67].

Perceived Social Support

Perceived social support was assessed at t1 and t2 with the K-22 short form of the Social Support Questionnaire (F-SozU) [68], and the total score was computed based on 22 items with 3 subscales (practical support, emotional support, and social integration). All items were rated on a 5-point Likert scale (1=does not apply to 5=is exactly correct). Internal reliability of the total score has been proven to be excellent with a Cronbach α of .91 [69]. Its validity has been confirmed in numerous studies [68,69].

Child Crying, Sleeping, and Feeding Problems

The Questionnaire for Crying, Feeding, and Sleeping (CFS; German Version) [70] assessed early child problems with respect to crying, sleeping, and feeding at t1 and t2. A total score was computed based on 3 scales: "Crying, Whining, and Sleeping" (24 items), "feeding" (13 items), and "coregulation" (12 items). Cronbach α has been shown to vary from .81 to .89 for the subscales, and α is .90 for the complete questionnaire, indicating high internal reliability. The questionnaire has been validated by correlations with behavioral diaries and clinical diagnoses [70].

Sociodemographic Questionnaire (Self-developed)

Parental age, participant's relation to child, nationality, mother tongue, educational qualifications, current employment situation, partnership status, child's age and gender, siblings, and information about the child's problems were recorded at t1.

App Use

App use was measured at t2 with 1 item derived from an app evaluation questionnaire (self-developed) using 5 categories: daily (4), several times a week (3), once a week (2), <once a week (1), and app not used (0).

Statistical Analysis

Power

An a priori case number calculation was performed using G*Power software (version 3.1.9.2; Kiel University) [71]. A repeated-measures mixed ANOVA (between-group × within-subject factor) was assumed. However, to exclude between-subject effects with sufficient certainty, case number planning was completed for the between-subject factor analyses, providing lower power. The estimation was based on a type 1 error of α=.01 and a power of 1–β=.80. Findings from a randomized controlled trial (RCT) of the effectiveness of a psychoeducational app for parents in the postpartum period [54] and another RCT on the effectiveness of an information video–based intervention for early child crying [72] were used to estimate effect sizes [54]. On the basis of statistical values reported in these studies, calculations yielded medium to high effect sizes. Accordingly, in this study, the number of cases was estimated more conservatively for medium treatment effects (Cohen *f*=0.30), using the ANOVA design described previously. The calculation of the number of cases resulted in a sample size of 136 participants.

Analysis Plan

To evaluate the effectiveness of the app-based intervention in the IG compared with the WCG, analyses were based on the intention-to-treat (ITT) principle. All analyses were performed using R software (version 4.0.1; R Foundation for Statistical Computing) [73]. The code used for the analyses has been made openly available in an Open Science Framework repository [74]. Missing values were assumed to be missing at random, which is typically plausible for trials with off-treatment assessments [75,76]. Missing values were therefore imputed using groupwise multivariate imputation by chained equations algorithm (fully conditional specification) [77], with 50 iterations and 50 (m) imputation sets. Several auxiliary variables were included in the imputation model to approximate the missing data in random missingness patterns.

Main Effectiveness Analysis

We tested whether the outcomes of app-based intervention in the IG was superior to those of the WCG in terms of its effects on parenting stress and secondary outcomes from pre- to posttests. For the confirmatory primary outcome analysis, a 1-sided test was used, whereas 2-sided testing was used for exploratory secondary outcome analyses. As an additional exploratory analysis, we examined the differences between groups on the subscales of the primary outcome (EBI parental stress). The effect differences between the 2 study conditions were assessed using univariate analysis of covariance. Baseline scores were used as covariates. Child age was entered as an additional covariate for the child symptom outcome analysis (CFS). All the models were fitted to each of the multiple imputed data sets. Model estimates were then aggregated via

Rubin’s combination rules [78] using a large-sample χ^2-approximation to combine the *F*-statistics [79]. To calculate the between-group standardized mean difference (Cohen *d*), we pooled the unstandardized group coefficients of a linear model without covariate adjustment using Rubin’s rules. This estimate was then standardized using the pooled outcome SD to obtain Cohen *d* and CI.

Sensitivity Analysis

To examine the robustness of the main analysis, 2 sensitivity analyses were conducted. First, we conducted a completer analysis of individuals who had no missing data and provided data at all assessment points. Second, a per-protocol analysis was conducted, focusing on individuals in the IG who accessed the app at least once (while retaining all participants in the WCG). The applied methods exactly mirrored the ones of the main effectiveness evaluation.

Results

Participant Enrollment and Characteristics

After the first screening by the outpatient clinic, a total of 41.3% (276/669) of participants were assessed for eligibility by the study team. A total 136 individuals who met the inclusion criteria were randomly allocated to the IG (n=73, 53.7%) and WCG (n=63, 46.3%; Figure 2 shows the CONSORT participant flowchart). Participants (mean age 33.97, SD 4.03 years) were predominantly mothers (127/136, 93.3%) of German nationality (113/136, 83.1%) with higher education (105/136, 77.2% qualified for university entrance) and currently not employed or on parental leave (97/136, 71.3%). Children (70/136, 51.5% boys) aged on average 10.32 (SD 5.06) months, had no siblings (95/136, 69.9%), and experienced sleeping (61/136, 44.9%) or combined (40/136, 29.4%) problems. The demographics divided by group are presented in Table 1.

Figure 2. Participant flowchart.

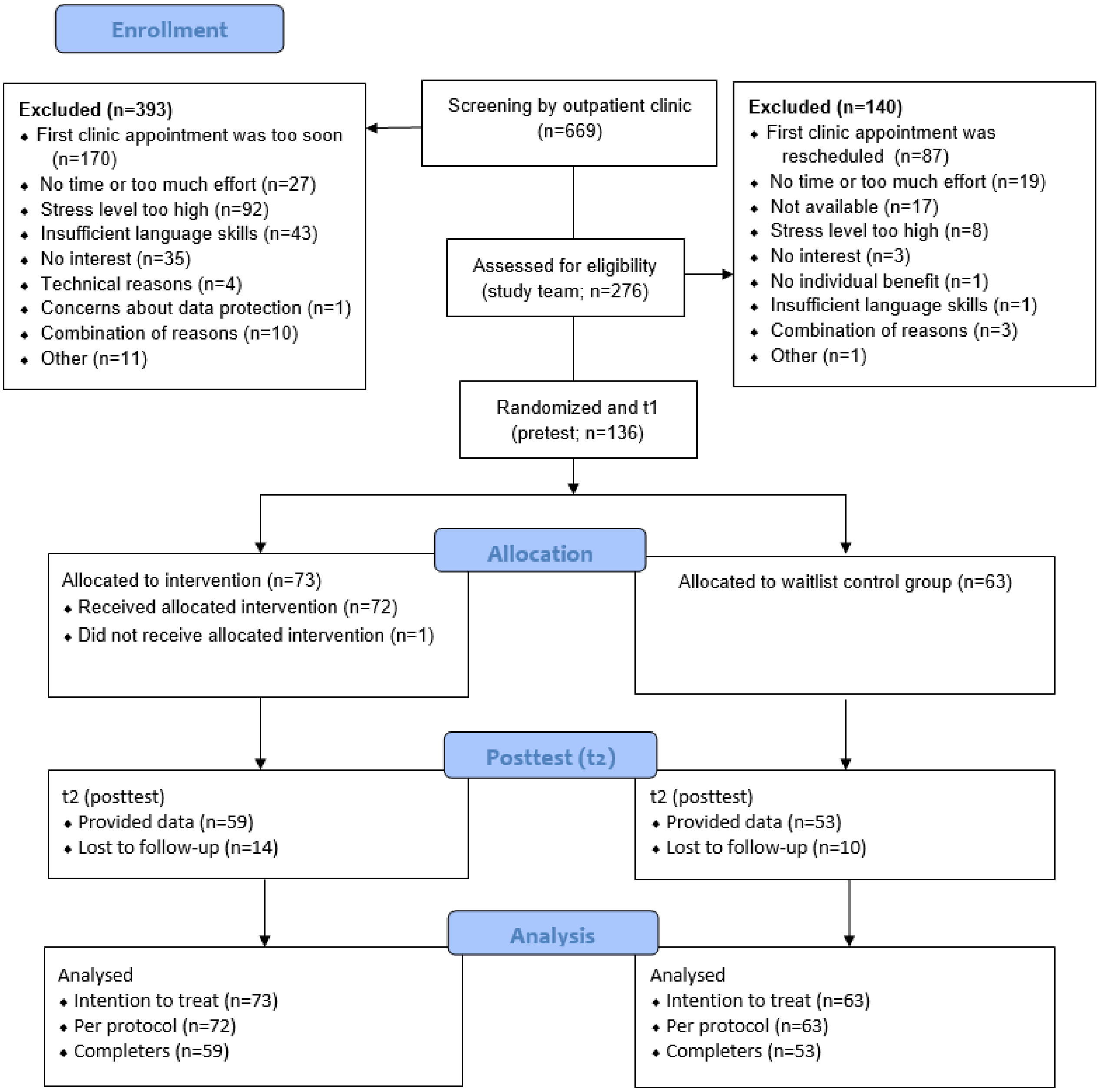


Table 1. Participant characteristics.

Characteristics	Intervention group (n=73)	Waitlist control group (n=63)
Parental age (years), mean (SD)	33.71 (3.86)	34.27 (4.22)
Participant's relation to child, n (%)		
Mother	68 (93.2)	59 (93.7)
Father	5 (6.8)	4 (6.3)
Academic qualification, n (%)		
Qualified for university entrance	57 (78.1)	48 (76.2)
Other or missing	16 (21.9)	15 (23.8)
Nationality, n (%)		
German	63 (86.3)	50 (79.4)
Other	10 (13.7)	13 (20.6)
First language, n (%)		
German	58 (79.5)	52 (82.5)
Other	15 (20.5)	11 (17.5)
Employment, n (%)		
Parental leave or currently not employed	53 (72.6)	44 (69.8)
Currently employed or apprenticeship	19 (26.0)	17 (27)
Other or missing	1 (1.4)	1 (1.6)
Single parent, n (%)	0 (0)	2 (3.2)
Child age (months), mean (SD)	10.21 (4.95)	10.44 (5.21)
Child gender, n (%)		
Girl	36 (49.3)	30 (47.6)
Boy	37 (50.7)	33 (52.4)
Siblings, n (%)		
Yes	22 (30.1)	19 (30.2)
No	51 (69.9)	44 (69.8)
Child's symptom duration (months), mean (SD)	6.91 (4.20)	7.76 (5.46)
Reason for consultation, n (%)		
Sleeping problems	34 (46.6)	27 (42.9)
Feeding problems	9 (12.3)	7 (11.1)
Crying or whining	5 (6.8)	3 (4.8)
Combined problems	21 (28.8)	19 (30.2)
Other or missing	4 (5.5)	7 (11.1)

Descriptive Statistics

Mean study duration from pre- to posttest was 23.41 (SD 10.42; range 10-49) days. In the IG, most parents used the app at least once a week (once a week: 24/73, 33%; several times a week: 21/73, 29%; daily: 7/73, 10%), whereas 27% (20/73) used the app less than once a week and 1 person did not use the app for unknown reasons. Descriptive statistics of outcome variables are displayed in Table 2.

Table 2. Mean scores in intention-to-treat sample per study group for outcome variables (pretest and posttest).

Outcome variable	Descriptive statistics of outcomes			
	Intervention group, mean (SD)		Waitlist control group, mean (SD)	
	t1 (pretest)	t2 (posttest)	t1 (pretest)	t2 (posttest)
EBI[a] parental scale	85.31 (18.97)	83.18 (19.94)	86.39 (16.65)	87.46 (16.67)
Attachment	10.05 (3.16)	10.35 (3.24)	9.86 (3.28)	10.21 (3.26)
Isolation	10.86 (4.04)	10.57 (3.86)	11.11 (3.38)	11.81 (3.51)
Parental competence	11.14 (3.94)	10.80 (3.91)	11.27 (3.76)	11.10 (3.41)
Depression	13.14 (4.00)	12.32 (4.01)	12.96 (3.06)	12.22 (3.26)
Health	13.10 (3.28)	12.23 (3.33)	13.19 (3.59)	13.02 (3.58)
Role restriction	13.15 (4.31)	13.27 (4.08)	13.48 (3.51)	14.13 (3.27)
Spouse-related stress	13.88 (3.51)	13.65 (3.90)	14.52 (3.10)	14.97 (3.11)
Knowledge test	59.39 (4.60)	62.91 (4.30)	60.66 (5.04)	61.15 (4.46)
PMP-SE[b]	65.38 (6.78)	66.96 (7.15)	65.89 (6.74)	66.62 (6.32)
F-SozU[c]	4.35 (0.58)	4.35 (0.60)	4.18 (0.68)	4.17 (0.68)
CFS[d]	2.28 (0.27)	2.25 (0.28)	2.29 (0.23)	2.22 (0.25)

[a]EBI: Eltern-Belastungs-Inventar.

[b]PMP-SE: Perceived Maternal Parenting Self-Efficacy Questionnaire.

[c]F-SozU: Social Support Questionnaire.

[d]CFS: Questionnaire for Crying, Feeding, and Sleeping.

Main Effectiveness Evaluation

Primary Outcome

Participants in the IG reported significantly lower levels of parenting stress at t2 compared with participants in the WCG ($F_{1,2683.50}$=3.38; *P*=.03; Cohen *d*=−0.23). Exploratory analysis of EBI parent domain subscales revealed a significant reduction in the social isolation subscale ($F_{1,1072.48}$=4.68; *P*=.03; Cohen *d*=−0.32), but not in the other domains (Table 3).

Table 3. Analysis of covariance (ANCOVA) results for intention-to-treat sample.

Variable	*F* test (*df*[a])	*P* value	Cohen *d*	95% CI
EBI[b] parental subscale	3.38 (1,2683.50)	.03[c]	−0.23	−0.52 to −0.06
Attachment	0.10 (1,12563.63)	.76	0.04	−0.30 to 0.38
Isolation	4.68 (1,1072.48)	.03[d]	−0.32	−0.66 to 0.02
Parental competence	0.25 (1,4046.88)	.62	−0.08	−0.42 to 0.26
Depression	0.11 (1,6522.79)	.74	0.03	−0.31 to 0.37
Health	2.24 (1,1536.52)	.13	−0.22	−0.56 to 0.12
Role restriction	2.31 (1,5263.57)	.13	−0.22	−0.57 to 0.11
Spouse-related stress	3.17 (1,1136.55)	.08	−0.36	−0.71 to −0.02
Knowledge test	19.46 (1,402.98)	<.001[d]	0.38	0.05 to 0.73
PMP-SE[e]	0.93 (1,7680.23)	.34	0.05	−0.29 to 0.39
F-SozU[f]	0.20 (1,11479.08)	.66	0.04	−0.10 to 0.17
CFS[g]	0.86 (1581.14)	.35	0.10	−0.23 to 0.45

[a]Approximated *df* based on multivariate imputation.
[b]EBI: Eltern-Belastungs-Inventar.
[c]1-tailed test based on the *t* distribution.
[d]2-tailed test (*F* test).
[e]PMP-SE: Perceived Maternal Parenting Self-Efficacy Questionnaire.
[f]F-SozU: Social Support Questionnaire.
[g]CFS: Questionnaire for Crying, Feeding, and Sleeping.

Secondary Outcomes

In the IG, a significantly higher level of knowledge about crying, sleeping, and feeding was noticeable at t2 compared with the WCG ($F_{1,402.98}$=19.46; *P*<.001; Cohen *d*=0.38).

No differences were found in parental efficacy, perceived social support, and child symptoms (Table 3).

Sensitivity Analysis

In the completer analysis, the primary and secondary outcomes remained stable. In line with the ITT analyses, in the IG, significantly lower levels of parenting stress (*P*=.01) and a significantly higher level of knowledge about crying, sleeping, and feeding (*P*<.001), but not in the other domains, were noticeable at t2 compared with the WCG. Furthermore, in the exploratory completer analysis referring to EBI parental subscales, in line with the ITT analysis, significantly lower scores on the social isolation subscale were noticeable (*P*=.004), and, in contrast to the ITT analysis, significantly lower levels on the role restriction (*P*=.04) and spouse-related stress (*P*=.01) subscales were evident in the IG at t2 (Table S2 in Multimedia Appendix 3).

In the per-protocol analysis, one person in the IG who did not use the app was excluded. In line with the ITT and completer analyses, in the IG, significantly lower levels of parenting stress (*P*=.04) and a significantly higher level of knowledge about crying, sleeping, and feeding (*P*<.001), but not in the other domains, were evident at t2 compared with the WCG. In the exploratory analysis of the EBI subscales, a significantly lower score on the spouse-related stress subscale became evident (*P*=.04), which is contrary to the ITT analysis but consistent with the completer analysis (Table S3 in Multimedia Appendix 3).

Discussion

Principal Findings

To the best of our knowledge, this is the first randomized controlled clinical study to target an app-based intervention for families having children with crying, sleeping, and feeding problems as common symptom patterns in early childhood. We found significantly lower levels of parenting stress after app use in the IG compared with the WCG. Furthermore, parents in the IG reported a higher level of knowledge about crying, sleeping, and feeding than the parents in the WCG. However, there were no significant differences between the 2 groups in terms of parental efficacy; perceived social support; and child crying, sleeping, and feeding symptoms.

Consistent with other research on families of children with crying, sleeping, and feeding problems [18,80,81], initial parenting stress levels in our sample were very high compared with normative values (EBI parental subscale scores >85 were above the 98 percentile and above the cutoff for individuals with very high stress levels). The finding that the use of our psychoeducational app contributed to the reduction in parenting stress is in line with several previous studies investigating the effects of psychoeducational interventions on stress [82,83]. Psychoeducation can bring various positive effects, such as emotional relief and promotion of coping strategies, beyond a mere increase in knowledge [84,85]. In addition, research shows that parents of children aged <5 years very frequently read web-based information on child health but tend to miss the accuracy of the content they find, are only moderately satisfied with its reliability, and often feel the need to cross-check information [86], which might be potentially time-consuming and tedious. By providing a collection of evidence-based and reliable information about their child's (problem) behavior as well as intervention strategies, the app use could potentially have spared parents' time and effort. Further research would need to be sought in this regard.

Exploratory analysis revealed a significant reduction in the parental stress (EBI) subscale of social isolation, and in the sensitivity analysis, effects on the EBI subscales of role restriction and spouse-related stress became evident. Regarding social isolation, this finding is in line with Shorey et al [54], confirming the effects of psychoeducational interventions on perceived support from the social environment. Regarding role restriction, some studies addressing different target groups indicate positive intervention effects (eg, effects of a web-based mindful parenting training for parents of toddlers [87]); however, there is a lack of comparative studies targeting role restriction in the context of child crying, sleeping, or feeding problems. Addressing spouse-related stress, findings indicate that psychoeducational interventions in the pre- and postpartum periods, including both partners, have a positive impact on partnership quality [88,89]. However, we have no information on whether in our sample both partners used the app together, so the result of reduced spouse-related stress in the context of app use still needs further investigation.

With regard to the secondary outcomes, app use led to an increase in knowledge about crying, sleeping, and feeding problems. As summarized in a systematic review by McDowall et al [43], several studies have shown increased knowledge about children's sleep after educational interventions. For instance, a pilot study by Jones et al [42] revealed that a brief brochure–based psychoeducational intervention enhanced parental knowledge about their children's sleep in a clinical sample. Although the mechanisms of psychoeducation lack systematic research [90], these findings are in line with the "Information Models" of psychoeducation, which emphasizes the importance of providing families with knowledge about psychological symptoms and their management to create awareness and to contribute to the management of the problems [85].

In this study sample, no effects of app use on parental perceived self-efficacy, social support, or children's symptoms were evident. With regard to self-efficacy, the absence of effects contrasts with the findings of Gilkerson et al [49] but is in line with the findings of a study by Missler et al [91] who also found no effects of a low-intensity psychoeducational program on parental self-efficacy. In terms of perceived social support, our findings differ from those of other studies focusing on apps for the postpartum period: Shorey et al [54] found that the use of a psychoeducational app increased perceived social support. Regarding child symptoms, studies investigating the effectiveness of psychoeducational interventions on the symptoms of excessive crying, sleeping, and feeding problems have shown inconsistent results. Although some studies indicate that educational interventions can reduce sleeping and crying problems in children [50,55], other studies cannot confirm these effects. In line with our results, Missler et al [91] found no effects of a low-intensity psychoeducational program on perceived problems of child crying, sleeping, and feeding. Hiscock et al [47] found effects only for specific subgroups, namely, very frequently fed children, showing changes in daytime sleep but not in nighttime sleep problems.

One explanation for the missing effects of app use on parental self-efficacy and perceived social support could be that, surprisingly and in contrast with other studies reporting impairment in self-efficacy and social support [24,25,92-95], the mean values of both these outcomes were in the normal range when compared with those of validation studies and normative values [67,96], indicating that these were not major issues in our clinical sample. This difference could be attributed to the characteristics of our sample, which included predominantly highly educated mothers in partnerships. Studies indicate that higher socioeconomic status is associated with higher self-efficacy and social support [97-99]. In addition, sleeping problems were predominant in our sample. Although associations between perceived social support and child crying as well as feeding problems have been investigated in various studies [24,25], studies addressing child sleeping problems are scarce. However, the effects of app use became evident in the EBI social isolation subscale, which might give us an indication that the app could have a positive effect on perceived social isolation. Further research would be useful at this point.

Another possible explanation for the absence of the effects on parental self-efficacy, perceived social support, and child problems could be that these constructs may not be as susceptible to short-term influence as parenting stress or knowledge. A meta-analysis by Amin et al [100] found that educational interventions for parents lasting at least 10 weeks produced significantly greater effects on parental self-efficacy than shorter interventions. To date, there are no comparative data on the extent to which the length of the intervention affects perceived social support. Regarding child symptom reduction, child problem behavior might be moderated by parenting stress; a regularly associated factor is poor parenting behavior, which in turn is linked to poorer child development and more behavioral problems [101,102]. Hence, an effect on child symptoms could become evident after the transfer of learned functional behavioral strategies in everyday life. A longer app use duration could have yielded different results; however, implementation was not possible in this study because the participation of the clinical sample was linked to the waiting time for an initial consultation appointment at the outpatient clinic.

Furthermore, the absence of effects could be anchored in the relevance of child problems to parental outcomes. Self-efficacy theory states that the success of strategies and actions is a major factor in increasing self-efficacy [103]. In the context of crying, sleeping, and feeding problems, this means that child symptoms would need to change because of parental strategies. Because CFS scores indicated clinically relevant problems in our sample and no changes in child symptoms were evident in this study, the effects on self-efficacy may have been absent. The lack of effect on child symptom reduction could also be related to the absence of effect on perceived social support in our study. Owing to the persistence of the child symptoms, parental networking behavior may not have changed.

Finally, we must acknowledge the possibility that an app-based psychoeducational intervention for this clientele is not sufficient to influence child symptoms, self-efficacy, or perceived social support. Parental feelings of helplessness, incompetence, and uncertainty about how to act are crucial therapeutic themes in affected families [2,21-23]. Working on these topics may require more intensive and individualized support such as parent-child psychotherapy or specialized counseling.

Our findings are of specific clinical relevance for the following reasons. First, increased knowledge may, to some extent, lead to altered and more favorable parenting behavior and thus might promote more adequate handling of the child in challenging situations [39-41,85]. Second, stress reduction provides the basis for breaking the vicious cycle of dysfunctional parent-child interactions, which is in turn an important protective factor for child mental health in the context of regulatory skills [2,40,41]. Reducing parenting stress is a prerequisite for increasing parental sensitivity, including the ability to adequately focus on a child's needs and signals [104]. Third, regarding the long-term consequences of early crying, sleeping, and feeding problems, there is evidence that parenting stress partially mediates the association between early behavioral problems and later mental health problems [102]. Thus, stress reduction might be an important factor in the prevention of long-term effects on healthy child development.

The effect sizes on parenting stress are considered small in this study. Other studies investigating the effects of psychoeducational interventions on stress have also found small effect sizes [83]. Nevertheless, our finding that an effect was evident even after a short application period is promising in terms of app use as a secondary preventive service. From a health economic perspective, it should be pointed out that an app, which is relatively cost-efficient, can reach many people in the community; thus, even small effect sizes can be highly effective at the population level. In terms of knowledge gain, the small effect size might be related to sample characteristics, and research indicates that knowledge about child sleeping problems is positively associated with parental educational level in a clinical sample [105]. In our highly educated sample, it is conceivable that the parental level of knowledge about crying, sleeping, and feeding was already high at baseline, and therefore, only a small effect on knowledge level was obtained. However, we did not compare the different educational levels in our study. Therefore, future research should be conducted in this regard. Furthermore, our sample had been dealing with the child's symptoms for an average of 7 months before seeking professional help in the outpatient clinic. Research shows that parents of children tend to seek internet-based information about child mental health daily or weekly [86]. Thus, one could assume that with such a long duration of symptoms, parents had likely already gathered information before app use, which might have resulted in a small effect on the knowledge level in our sample. Future research will have to corroborate this assumption.

Strengths and Limitations

Evaluating the app's effectiveness using a RCT design is considered the gold standard for testing the effectiveness of new interventions [106]. However, to date, few RCT intervention studies have been conducted regarding early childhood crying, sleeping, and feeding problems. To the best of our knowledge, this is the first randomized controlled clinical study to target an app-based intervention for families having children with crying, sleeping, and feeding problems as common symptom patterns in early childhood. The newly developed app stands out from other web-based offers because it contains evidence-based strategies; addresses the complexity of co-occurring symptoms of crying, sleeping, and feeding; and combines psychoeducational input with interactive elements. Data collection for effectiveness testing was based on validated measurements. To avoid sequence effects, the questionnaires were administered in permuted order.

However, our results should be interpreted against the background of some limitations of the study. First, it must be mentioned that the study was conducted in a specific clinical setting, and predominantly, mothers of German nationality who had a high level of education and were in a partnership with a sense of social support and self-efficacy participated in the study. Caution is necessary when generalizing our results to other clinical populations. However, these sample characteristics are consistent with other studies on early childhood crying, sleeping, and feeding problems, which also report predominantly academic or higher-educated families in stable partnerships [107,108].

Furthermore, regarding post hoc power, given our sample size, only medium-size effects could be detected with sufficiently high probability (β=.82 to .99 for medium effects), whereas power was too low to detect small effect sizes with sufficient probability (β=up to .79 for small effects).

Finally, because the study was implemented in a clinical setting, a variation in individual study duration is evident; the measurements were linked to clinic appointments and thus linked to individual variance and deviation because of appointment postponements. However, as pointed out in the Medical Research Council guidelines [63], ensuring very strict standardization might not be appropriate, and a specified degree of adaptation to local settings is preferred.

Conclusions

In its first practical use in an RCT, a psychoedcuational app for parents of children with crying, sleeping, and feeding problems contributes to parents' better understanding of their child's symptoms, has the potential to take the edge off parenting stress, and could therefore also serve as an effective secondary preventive measure. It can be recommended by pediatricians, gynecologists, maternity clinics, or child welfare services as an initial low-threshold information and support service when the first symptoms are developing, or it can be used to bridge waiting times for professional counseling appointments. In the future, the app will be made available free of charge as a low-threshold offer.

Acknowledgments

The authors would like to thank all participants in the study for their time and effort. They would like to thank the outpatient clinic, especially Dominika Hess, for their recruitment support; the project staff Katharina Richter, Anne-Sophie Wenzel, Sinja Schwarzwälder, Catherine Buechel, Lena Wagner, and Kristina Scherer for their contributions to the study; and Paula Kuper and Lea Schuurmans for their contributions to the analyses.

This study was funded by the initiative "Gesund.Leben.Bayern" of the Bavarian Ministry of Health (grant LP00261).

Authors' Contributions

MA, AF, MLD, DW, and VM conceptualized and designed the study. MA and AF coordinated data collection. MA, MH, DDE, AF, MLD, and LDB contributed to the analyses and interpretation of the data. MA, AF, and MH wrote the manuscript. MLD, VM, DW, LDB, AB, and MZ reviewed and revised the manuscript for important intellectual content. AF, MLD, MA, DW, LDB, AB, VM, and MZ developed the app used in the trial. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest

DDE reports to have received consultancy fees or served on the scientific advisory board from several companies such as Novartis, Sanofi, Lantern, Schön Kliniken, Minddistrict, and German health insurance companies (BARMER, Techniker Krankenkasse). DDE and MH are stakeholders of the Institute for Health Trainings Online (GET.ON), which aims to implement scientific findings related to digital health interventions in routine care. MH is an employee of the Institute for Health Trainings Online (GET.ON). AF, MLD, MA, DW, LDB, AB, VM, and MZ developed the app used in the trial.

Editorial Notice

This randomized study was only retrospectively registered, explained by authors by the fact that at the time of the study's submission, registration was not yet required by their ethics committee. On recommendation, the study was registered a few months after study initiation and well before any results were obtained. From the start of the study until the time of registration, no fundamental changes to the study design were implemented. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1712 KB - mhealth_v11i1e41804_app1.pdf]

Multimedia Appendix 2
Trial registration.
[PDF File (Adobe PDF File), 90 KB - mhealth_v11i1e41804_app2.pdf]

Multimedia Appendix 3
Knowledge test face validity and sensitivity analysis results.
[DOCX File , 25 KB - mhealth_v11i1e41804_app3.docx]

References

1. Hemmi MH, Wolke D, Schneider S. Associations between problems with crying, sleeping and/or feeding in infancy and long-term behavioural outcomes in childhood: a meta-analysis. Arch Dis Child 2011 Jul;96(7):622-629 [FREE Full text] [doi: 10.1136/adc.2010.191312] [Medline: 21508059]
2. Papoušek M. Regulationsstörungen der frühen kindheit: klinische evidenz für ein neues diagnostisches konzept. In: Papoušek M, Schieche M, Wurmser H, editors. Regulationsstörungen der frühen Kindheit: Frühe Risiken und Hilfen im Entwicklungskontext der Eltern-Kind-Beziehungen. Bern, Switzerland: Verlag Hans Huber; 2004:77-110.
3. Zero to Three. DC: 0-5: Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood. Washington, DC, USA: Zero to Three; 2016.

4. Bilgin A, Baumann N, Jaekel J, Breeman LD, Bartmann P, Bäuml JG, et al. Early crying, sleeping, and feeding problems and trajectories of attention problems from childhood to adulthood. Child Dev 2020 Jan;91(1):e77-e91. [doi: 10.1111/cdev.13155] [Medline: 30291757]
5. Lucassen PL, Assendelft WJ, van Eijk JT, Gubbels JW, Douwes AC, van Geldrop WJ. Systematic review of the occurrence of infantile colic in the community. Arch Dis Child 2001 May;84(5):398-403 [FREE Full text] [doi: 10.1136/adc.84.5.398] [Medline: 11316682]
6. Wurmser H, Laubereau B, Hermann M, Papousek M, von Kries R. Excessive infant crying: often not confined to the first 3 months of age. Early Hum Dev 2001 Aug;64(1):1-6. [doi: 10.1016/s0378-3782(01)00166-9] [Medline: 11408103]
7. Bilgin A, Wolke D. Regulatory problems in very preterm and full-term infants over the first 18 months. J Dev Behav Pediatr 2016 May;37(4):298-305. [doi: 10.1097/DBP.0000000000000297] [Medline: 27096573]
8. Wolke D, Meyer R, Ohrt B, Riegel K. Häufigkeit und Persistenz von Ein- und Durchschlafproblemen im Vorschulalter: Ergebnisse einer prospektiven Untersuchung an einer repräsentativen Stichprobe in Bayern. Prax Kinderpsychol Kinderpsychiatr 1994. [doi: 10.23668/psycharchives.10698]
9. Wurmser H. Schrei-, schlaf- und fütterstörung: prävalenz, persistenz, prädiktoren und langzeitprognose. Monatsschr Kinderheilkd 2009 Apr 25;157(6):574-579. [doi: 10.1007/s00112-008-1895-4]
10. Petzoldt J, Wittchen HU, Einsle F, Martini J. Maternal anxiety versus depressive disorders: specific relations to infants' crying, feeding and sleeping problems. Child Care Health Dev 2016 Mar;42(2):231-245. [doi: 10.1111/cch.12292] [Medline: 26490836]
11. Cook F, Giallo R, Hiscock H, Mensah F, Sanchez K, Reilly S. Infant regulation and child mental health concerns: a longitudinal study. Pediatrics 2019 Mar;143(3):e20180977. [doi: 10.1542/peds.2018-0977] [Medline: 30737245]
12. Benjasuwantep B, Chaithirayanon S, Eiamudomkan M. Feeding problems in healthy young children: prevalence, related factors and feeding practices. Pediatr Rep 2013 Jun 13;5(2):38-42 [FREE Full text] [doi: 10.4081/pr.2013.e10] [Medline: 23904965]
13. Wright CM, Parkinson KN, Shipton D, Drewett RF. How do toddler eating problems relate to their eating behavior, food preferences, and growth? Pediatrics 2007 Oct;120(4):e1069-e1075. [doi: 10.1542/peds.2006-2961] [Medline: 17908727]
14. Olsen AL, Ammitzbøll J, Olsen EM, Skovgaard AM. Problems of feeding, sleeping and excessive crying in infancy: a general population study. Arch Dis Child 2019 Nov;104(11):1034-1041. [doi: 10.1136/archdischild-2019-316851] [Medline: 31270094]
15. Wake M, Morton-Allen E, Poulakis Z, Hiscock H, Gallagher S, Oberklaid F. Prevalence, stability, and outcomes of cry-fuss and sleep problems in the first 2 years of life: prospective community-based study. Pediatrics 2006 Mar;117(3):836-842. [doi: 10.1542/peds.2005-0775] [Medline: 16510665]
16. Wolke D, Söhne B, Riegel K, Ohrt B, Osterlund K. An epidemiologic longitudinal study of sleeping problems and feeding experience of preterm and term children in southern Finland: comparison with a southern German population sample. J Pediatr 1998 Aug;133(2):224-231. [doi: 10.1016/s0022-3476(98)70224-0] [Medline: 9709710]
17. Papoušek M. Persistierendes schreien: schreiprobleme im entwicklungskontext von eltern-kind-kommunikation und -beziehung. Monatsschr Kinderheilkd 2009 Apr 29;157(6):558-566. [doi: 10.1007/s00112-008-1893-6]
18. Sidor A, Thiel-Bonney C, Kunz E, Eickhorst A, Cierpka M. Das exzessive Schreien bei 5 Monate alten Säuglingen und prä-, peri- und postnatale Belastungen ihrer Mütter in einer Hoch-Risiko-Stichprobe. Z Kinder Jugendpsychiatr Psychother 2012 Jul;40(4):239-250. [doi: 10.1024/1422-4917/a000177] [Medline: 22707121]
19. van den Berg MP, van der Ende J, Crijnen AA, Jaddoe VW, Moll HA, Mackenbach JP, et al. Paternal depressive symptoms during pregnancy are related to excessive infant crying. Pediatrics 2009 Jul;124(1):e96-103. [doi: 10.1542/peds.2008-3100] [Medline: 19564275]
20. Bilgin A, Wolke D. Development of comorbid crying, sleeping, feeding problems across infancy: neurodevelopmental vulnerability and parenting. Early Hum Dev 2017 Jun;109:37-43. [doi: 10.1016/j.earlhumdev.2017.04.002] [Medline: 28433797]
21. Leavitt LA. Mothers' sensitivity to infant signals. Pediatrics 1998 Nov;102(5 Suppl E):1247-1249. [Medline: 9794963]
22. Stifter CA, Bono MA. The effect of infant colic on maternal self-perceptions and mother-infant attachment. Child Care Health Dev 1998 Sep;24(5):339-351. [doi: 10.1046/j.1365-2214.2002.00088.x] [Medline: 9728282]
23. Vik T, Grote V, Escribano J, Socha J, Verduci E, Fritsch M, European Childhood Obesity Trial Study Group. Infantile colic, prolonged crying and maternal postnatal depression. Acta Paediatr 2009 Aug;98(8):1344-1348. [doi: 10.1111/j.1651-2227.2009.01317.x] [Medline: 19432839]
24. Garratt R, Bamber D, Powell C, Long J, Brown J, Turney N, et al. Parents' experiences of having an excessively crying baby and implications for enhancing support services. J Health Visit 2019 Mar 02;7(3):132-140. [doi: 10.12968/johv.2019.7.3.132]
25. Long T, Johnson M. Living and coping with excessive infantile crying. J Adv Nurs 2001 Apr;34(2):155-162. [doi: 10.1046/j.1365-2648.2001.01740.x] [Medline: 11430276]
26. Barr RG, Trent RB, Cross J. Age-related incidence curve of hospitalized Shaken Baby Syndrome cases: convergent evidence for crying as a trigger to shaking. Child Abuse Negl 2006 Jan;30(1):7-16. [doi: 10.1016/j.chiabu.2005.06.009] [Medline: 16406023]

27. Winsper C, Bilgin A, Wolke D. Associations between infant and toddler regulatory problems, childhood co-developing internalising and externalising trajectories, and adolescent depression, psychotic and borderline personality disorder symptoms. J Child Psychol Psychiatry 2020 Feb;61(2):182-194. [doi: 10.1111/jcpp.13125] [Medline: 31469175]
28. Talvik I, Alexander RC, Talvik T. Shaken baby syndrome and a baby's cry. Acta Paediatr 2008 Jun;97(6):782-785. [doi: 10.1111/j.1651-2227.2008.00778.x] [Medline: 18397351]
29. Barlow KM, Minns RA. Annual incidence of shaken impact syndrome in young children. Lancet 2000 Nov 04;356(9241):1571-1572. [doi: 10.1016/S0140-6736(00)03130-5] [Medline: 11075773]
30. Rautava P, Lehtonen L, Helenius H, Sillanpää M. Infantile colic: child and family three years later. Pediatrics 1995 Jul;96(1 Pt 1):43-47. [Medline: 7596720]
31. Elliott MR, Pedersen EL, Mogan J. Early infant crying: child and family follow-up at three years. Can J Nurs Res 1997;29(2):47-67. [Medline: 9355290]
32. Bilgin A, Wolke D. Infant crying problems and symptoms of sleeping problems predict attachment disorganization at 18 months. Attach Hum Dev 2020 Aug;22(4):367-391. [doi: 10.1080/14616734.2019.1618882] [Medline: 31132936]
33. Ammaniti M, Lucarelli L, Cimino S, D'Olimpio F, Chatoor I. Feeding disorders of infancy: a longitudinal study to middle childhood. Int J Eat Disord 2012 Mar;45(2):272-280. [doi: 10.1002/eat.20925] [Medline: 21495054]
34. Schmid G, Wolke D. Preschool regulatory problems and attention-deficit/hyperactivity and cognitive deficits at school age in children born at risk: different phenotypes of dysregulation? Early Hum Dev 2014 Aug;90(8):399-405. [doi: 10.1016/j.earlhumdev.2014.05.001] [Medline: 24951075]
35. Sivertsen B, Harvey AG, Reichborn-Kjennerud T, Torgersen L, Ystrom E, Hysing M. Later emotional and behavioral problems associated with sleep problems in toddlers: a longitudinal study. JAMA Pediatr 2015 Jun;169(6):575-582. [doi: 10.1001/jamapediatrics.2015.0187] [Medline: 25867179]
36. Wolke D, Schmid G, Schreier A, Meyer R. Crying and feeding problems in infancy and cognitive outcome in preschool children born at risk: a prospective population study. J Dev Behav Pediatr 2009 Jun;30(3):226-238 [FREE Full text] [doi: 10.1097/DBP.0b013e3181a85973] [Medline: 19433987]
37. Zucker N, Copeland W, Franz L, Carpenter K, Keeling L, Angold A, et al. Psychological and psychosocial impairment in preschoolers with selective eating. Pediatrics 2015 Sep;136(3):e582-e590 [FREE Full text] [doi: 10.1542/peds.2014-2386] [Medline: 26240213]
38. Bäuml JG, Baumann N, Avram M, Mulej Bratec S, Breeman L, Berndt M, et al. The default mode network mediates the impact of infant regulatory problems on adult avoidant personality traits. Biol Psychiatry Cogn Neurosci Neuroimaging 2019 Apr;4(4):333-342. [doi: 10.1016/j.bpsc.2018.11.005] [Medline: 30579926]
39. Akhnikh S, Engelberts AC, van Sleuwen BE, L'Hoir MP, Benninga MA. The excessively crying infant: etiology and treatment. Pediatr Ann 2014 Apr;43(4):e69-e75. [doi: 10.3928/00904481-20140325-07] [Medline: 24716561]
40. Keefe MR. Irritable infant syndrome: theoretical perspectives and practice implications. ANS Adv Nurs Sci 1988 Apr;10(3):70-78. [doi: 10.1097/00012272-198804000-00008] [Medline: 3132089]
41. Keefe MR, Kajrlsen KA, Lobo ML, Kotzer AM, Dudley WN. Reducing parenting stress in families with irritable infants. Nurs Res 2006;55(3):198-205. [doi: 10.1097/00006199-200605000-00006] [Medline: 16708044]
42. Jones CH, Owens JA, Pham B. Can a brief educational intervention improve parents' knowledge of healthy children's sleep? A pilot-test. Health Educ J 2012 Oct 22;72(5):601-610. [doi: 10.1177/0017896912464606]
43. McDowall PS, Galland BC, Campbell AJ, Elder DE. Parent knowledge of children's sleep: a systematic review. Sleep Med Rev 2017 Feb;31:39-47. [doi: 10.1016/j.smrv.2016.01.002] [Medline: 26899741]
44. Bäuml J, Pitschel-Walz G. Psychoedukation, quo vadis? Was will, kann und muss dieses verfahren leisten? Psychotherapeut 2012 Jun 27;57(4):289-290. [doi: 10.1007/s00278-012-0923-y]
45. Bäuml J, Pitschel-Walz G. Psychoedukative informationsvermittlung: pflicht und kür. In: Bäuml J, Pitschel-Walz G, editors. Psychoedukation Bei schizophrenen Erkrankungen: Konsensuspapier der Arbeitsgruppe "Psychoedukation bei schizophrenen Erkrankungen". Stuttgart, Germany: Schattauer Verlag; 2003:101-123.
46. Lang AV, Simhandl C. Psychoedukation – zusammen weniger allein: neue therapieansätze bei bipolaren depressionen. Psychopraxis 2009 Oct 8;12(4):29-32. [doi: 10.1007/s00739-009-0154-3]
47. Hiscock H, Cook F, Bayer J, Le HN, Mensah F, Cann W, et al. Preventing early infant sleep and crying problems and postnatal depression: a randomized trial. Pediatrics 2014 Feb;133(2):e346-e354. [doi: 10.1542/peds.2013-1886] [Medline: 24394682]
48. Long J, Powell C, Bamber D, Garratt R, Brown J, Dyson S, et al. Development of materials to support parents whose babies cry excessively: findings and health service implications. Prim Health Care Res Dev 2018 Jul;19(4):320-332 [FREE Full text] [doi: 10.1017/S1463423617000779] [Medline: 29317013]
49. Gilkerson L, Burkhardt T, Katch LE, Hans SL. Increasing parenting self-efficacy: the Fussy Baby Network intervention. Infant Ment Health J 2020 Mar;41(2):232-245. [doi: 10.1002/imhj.21836] [Medline: 32242969]
50. Mihelic M, Morawska A, Filus A. Effects of early parenting interventions on parents and infants: a meta-analytic review. J Child Fam Stud 2017 Mar 1;26(6):1507-1526 [FREE Full text] [doi: 10.1007/s10826-017-0675-y]
51. Anteil der Smartphone-Nutzer* in Deutschland in den Jahren 2012 bis 2021. Statista. 2022. URL: https://de.statista.com/statistik/daten/studie/585883/umfrage/anteil-der-smartphone-nutzer-in-deutschland/ [accessed 2022-07-01]

52. Wirtschaftsrechnungen: Laufende Wirtschaftsrechnungen Ausstattung privater Haushalte mit ausgewählten Gebrauchsgütern. Statistisches Bundesamt (Destatis). 2021. URL: https://tinyurl.com/yc28twps [accessed 2022-07-05]
53. Anzahl der bei Google Play verfügbaren mHealth-Apps weltweit vom 1. Quartal 2015 bis zum 3. Quartal 2020. Statista. 2022. URL: https://de.statista.com/statistik/daten/studie/1190033/umfrage/anzahl-der-bei-google-play-verfuegbaren-mhealth-apps/ [accessed 2022-07-05]
54. Shorey S, Lau YY, Dennis CL, Chan YS, Tam WW, Chan YH. A randomized-controlled trial to examine the effectiveness of the 'Home-but not Alone' mobile-health application educational programme on parental outcomes. J Adv Nurs 2017 Sep;73(9):2103-2117. [doi: 10.1111/jan.13293] [Medline: 28276086]
55. Leichman ES, Gould RA, Williamson AA, Walters RM, Mindell JA. Effectiveness of an mHealth intervention for infant sleep disturbances. Behav Ther 2020 Jul;51(4):548-558 [FREE Full text] [doi: 10.1016/j.beth.2019.12.011] [Medline: 32586429]
56. Hiscock H, Ng O, Crossley L, Chow J, Rausa V, Hearps S. Sleep well be well: pilot of a digital intervention to improve child behavioural sleep problems. J Paediatr Child Health 2021 Jan;57(1):33-40. [doi: 10.1111/jpc.15106] [Medline: 32770791]
57. Dienelt K, Moores CJ, Miller J, Mehta K. An investigation into the use of infant feeding tracker apps by breastfeeding mothers. Health Informatics J 2020 Sep;26(3):1672-1683 [FREE Full text] [doi: 10.1177/1460458219888402] [Medline: 31793799]
58. Simon SL, Kaar JL, Talker I, Reich J. Evidence-based behavioral strategies in smartphone apps for children's sleep: content analysis. JMIR Pediatr Parent 2022 Mar 03;5(1):e32129 [FREE Full text] [doi: 10.2196/32129] [Medline: 35238787]
59. Sardi L, Idri A, Redman LM, Alami H, Bezad R, Fernández-Alemán JL. Mobile health applications for postnatal care: review and analysis of functionalities and technical features. Comput Methods Programs Biomed 2020 Feb;184:105114 [FREE Full text] [doi: 10.1016/j.cmpb.2019.105114] [Medline: 31655305]
60. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
61. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting internet intervention research. Cogn Behav Ther 2011;40(2):82-97. [doi: 10.1080/16506073.2011.573807] [Medline: 25155812]
62. Urbaniak GC, Plous S. Research Randomizer. 2013. URL: http://www.randomizer.org/ [accessed 2022-08-10]
63. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
64. Tröster H. Eltern-Belastungs-Inventar (EBI). Deutsche Version des Parenting Stress Index (PSI) von R.R. Abidin. Göttingen, Germany: Hogrefe Verlag; 2011.
65. Abidin R, Flens JR, Austin WG. The parenting stress index. In: Archer RP, editor. Forensic Uses of Clinical Assessment Instruments. Mahwah, NJ, USA: Lawrence Erlbaum Associates; 2006:297-328.
66. Irlbauer-Müller V, Eichler A, Donhauser JA, Poehlmann NE, Stemmler M, Moll GH, et al. Das Eltern-Belastungs-Inventar (EBI): einsatz und nutzen in der kinder- und jugendpsychiatrie und -psychotherapie. Diagnostica 2018 Jan;64(1):37-48. [doi: 10.1026/0012-1924/a000191]
67. Barnes CR, Adamson-Macedo EN. Perceived Maternal Parenting Self-Efficacy (PMP S-E) tool: development and validation with mothers of hospitalized preterm neonates. J Adv Nurs 2007 Dec;60(5):550-560. [doi: 10.1111/j.1365-2648.2007.04445.x] [Medline: 17973719]
68. Fydrich T, Geyer M, Hessel A, Sommer G, Brähler E. Fragebogen zur Sozialen Unterstützung (F-SozU): normierung an einer repräsentativen stichprobe. Diagnostica 1999 Oct;45(4):212-216. [doi: 10.1026//0012-1924.45.4.212]
69. Fydrich T, Sommer G, Tydecks S, Brähler E. Fragebogen zur sozialen unterstützung (F-SozU): normierung der kurzform (K-14). Z Med Psychol 2009;18(1):43-48 [FREE Full text] [doi: 10.1026//0012-1924.45.4.212]
70. Gross S, Reck C, Thiel-Bonney C, Cierpka M. Empirische Grundlagen des Fragebogens zum Schreien, Füttern und Schlafen (SFS). Prax Kinderpsychol Kinderpsychiatr 2013;62(5):327-347. [doi: 10.13109/prkk.2013.62.5.327] [Medline: 23802351]
71. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
72. McRury JM, Zolotor AJ. A randomized, controlled trial of a behavioral intervention to reduce crying among infants. J Am Board Fam Med 2010;23(3):315-322 [FREE Full text] [doi: 10.3122/jabfm.2010.03.090142] [Medline: 20453177]
73. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: The R Foundation; 2022. URL: https://www.R-project.org/ [accessed 2023-02-21]
74. Harrer M, Kuper P. A mobile-based intervention for infant crying, sleeping, feeding problems: randomized controlled trial. Open Science Framework. 2022 Jun 13. URL: https://osf.io/8z9tv [accessed 2023-02-22]
75. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. BMC Med Res Methodol 2014 Nov 19;14:118 [FREE Full text] [doi: 10.1186/1471-2288-14-118] [Medline: 25407057]
76. van Buuren S. Flexible Imputation of Missing Data. 2nd edition. Boca Raton, FL, USA: Chapman and Hall/CRC; 2019.

77. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
78. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika 1999 Dec 01;86(4):948-955. [doi: 10.1093/biomet/86.4.948]
79. Grund S, Lüdtke O, Robitzsch A. Pooling ANOVA results from multiply imputed datasets: a simulation study. Methodology 2016 Jul;12(3):75-88. [doi: 10.1027/1614-2241/a000111]
80. Byars KC, Yeomans-Maldonado G, Noll JG. Parental functioning and pediatric sleep disturbance: an examination of factors associated with parenting stress in children clinically referred for evaluation of insomnia. Sleep Med 2011 Oct;12(9):898-905. [doi: 10.1016/j.sleep.2011.05.002] [Medline: 21940206]
81. Miller-Loncar C, Bigsby R, High P, Wallach M, Lester B. Infant colic and feeding difficulties. Arch Dis Child 2004 Oct;89(10):908-912 [FREE Full text] [doi: 10.1136/adc.2003.033233] [Medline: 15383432]
82. Luciano M, Del Vecchio V, Giacco D, De Rosa C, Malangone C, Fiorillo A. A 'family affair'? The impact of family psychoeducational interventions on depression. Expert Rev Neurother 2012 Jan;12(1):83-92. [doi: 10.1586/ern.11.131] [Medline: 22243046]
83. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
84. Wolke D, Gray P, Meyer R. Excessive infant crying: a controlled study of mothers helping mothers. Pediatrics 1994 Sep;94(3):322-332. [Medline: 8065857]
85. Bhattacharjee D, Rai AK, Singh NK, Kumar P, Munda SK, Das B. Psychoeducation: a measure to strengthen psychiatric treatment. Ind Psychiatr J 2011 Apr;14(1):33-39 [FREE Full text]
86. Bryan MA, Evans Y, Morishita C, Midamba N, Moreno M. Parental perceptions of the internet and social media as a source of pediatric health information. Acad Pediatr 2020;20(1):31-38. [doi: 10.1016/j.acap.2019.09.009] [Medline: 31648059]
87. Potharst ES, Boekhorst MG, Cuijlits I, van Broekhoven KE, Jacobs A, Spek V, et al. A randomized control trial evaluating an online mindful parenting training for mothers with elevated parental stress. Front Psychol 2019 Jul 17;10:1550 [FREE Full text] [doi: 10.3389/fpsyg.2019.01550] [Medline: 31379646]
88. Midmer D, Wilson L, Cummings S. A randomized, controlled trial of the influence of prenatal parenting education on postpartum anxiety and marital adjustment. Fam Med 1995 Mar;27(3):200-205. [Medline: 7774781]
89. Rowe HJ, Fisher JR. Development of a universal psycho-educational intervention to prevent common postpartum mental disorders in primiparous women: a multiple method approach. BMC Public Health 2010 Aug 18;10:499 [FREE Full text] [doi: 10.1186/1471-2458-10-499] [Medline: 20718991]
90. Mühlig S, Jacobi F. Psychoedukation. In: Hoyer J, Knappe S, editors. Klinische Psychologie & Psychotherapie. Berlin, Germany: Springer; 2020:557-573.
91. Missler M, van Straten A, Denissen J, Donker T, Beijers R. Effectiveness of a psycho-educational intervention for expecting parents to prevent postpartum parenting stress, depression and anxiety: a randomized controlled trial. BMC Pregnancy Childbirth 2020 Oct 31;20(1):658 [FREE Full text] [doi: 10.1186/s12884-020-03341-9] [Medline: 33129314]
92. Bond MJ, Prager MA, Tiggemann M, Tao B. Infant crying, maternal wellbeing and perceptions of caregiving. J Appl Health Behav 2001;3(1):3-9.
93. Feldman R, Keren M, Gross-Rozval O, Tyano S. Mother-child touch patterns in infant feeding disorders: relation to maternal, child, and environmental factors. J Am Acad Child Adolesc Psychiatry 2004 Sep;43(9):1089-1097. [doi: 10.1097/01.chi.0000132810.98922.83] [Medline: 15322412]
94. Kahya Y, Uluc S. The Perceived Maternal Parenting Self-Efficacy (PMP S-E) Tool: the adaptation study in the context of attachment styles and mood in the first-time mothers. Dusunen Adam 2021;35(4):50-63. [doi: 10.14744/dajpns.2021.00120]
95. Werner A, Mayer A, Lohaus A. Sleep-related parenting self-efficacy and parent-reported sleep in young children: a dyadic analysis of parental actor and partner effects. Sleep Health 2022 Feb;8(1):54-61. [doi: 10.1016/j.sleh.2021.11.004] [Medline: 34924342]
96. von Fydrich T, Sommer G, Brähler E. F-SozU: Fragebogen zur sozialen Unterstützung. Göttingen, Germany: Hogrefe; 2007.
97. Boardman JD, Robert SA. Neighborhood socioeconomic status and perceptions of self-efficacy. Sociol Perspect 2000 Mar;43(1):117-136. [doi: 10.2307/1389785]
98. Liu T, Zhang X, Jiang Y. Family socioeconomic status and the cognitive competence of very young children from migrant and non-migrant Chinese families: the mediating role of parenting self-efficacy and parental involvement. Early Child Res Q 2020;51:229-241. [doi: 10.1016/j.ecresq.2019.12.004]
99. Tinajero C, Martínez-López Z, Rodríguez MS, Guisande MA, Páramo MF. Gender and socioeconomic status differences in university students’ perception of social support. Eur J Psychol Educ 2014 Nov 14;30(2):227-244. [doi: 10.1007/s10212-014-0234-5]
100. Liyana Amin NA, Tam WW, Shorey S. Enhancing first-time parents' self-efficacy: a systematic review and meta-analysis of universal parent education interventions' efficacy. Int J Nurs Stud 2018 Jun;82:149-162. [doi: 10.1016/j.ijnurstu.2018.03.021] [Medline: 29656206]

101. Deater-Deckard K. Parenting stress and child adjustment: some old hypotheses and new questions. Clin Psychol (New York) 1998 Sep;5(3):314-332. [doi: 10.1111/j.1468-2850.1998.tb00152.x]
102. Smarius LJ, Strieder TG, Loomans EM, Doreleijers TA, Vrijkotte TG, Gemke RJ, et al. Excessive infant crying doubles the risk of mood and behavioral problems at age 5: evidence for mediation by maternal characteristics. Eur Child Adolesc Psychiatry 2017 Mar;26(3):293-302 [FREE Full text] [doi: 10.1007/s00787-016-0888-4] [Medline: 27422707]
103. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY, USA: Freeman; 1997.
104. Pereira J, Vickers K, Atkinson L, Gonzalez A, Wekerle C, Levitan R. Parenting stress mediates between maternal maltreatment history and maternal sensitivity in a community sample. Child Abuse Negl 2012 May;36(5):433-437. [doi: 10.1016/j.chiabu.2012.01.006] [Medline: 22633056]
105. Owens JA, Jones C. Parental knowledge of healthy sleep in young children: results of a primary care clinic survey. J Dev Behav Pediatr 2011;32(6):447-453. [doi: 10.1097/DBP.0b013e31821bd20b] [Medline: 21546852]
106. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018 Dec;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
107. Richter N, Reck C. Positive maternal interaction behavior moderates the relation between maternal anxiety and infant regulatory problems. Infant Behav Dev 2013 Dec;36(4):498-506. [doi: 10.1016/j.infbeh.2013.04.007] [Medline: 23728294]
108. Winsper C, Wolke D. Infant and toddler crying, sleeping and feeding problems and trajectories of dysregulated behavior across childhood. J Abnorm Child Psychol 2014;42(5):831-843. [doi: 10.1007/s10802-013-9813-1] [Medline: 24122374]

Abbreviations

CFS: Questionnaire for Crying, Feeding, and Sleeping
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Tele Health
DRKS: German Register of Clinical Studies
EBI: Eltern-Belastungs-Inventar
F-SozU: Social Support Questionnaire
IG: intervention group
ITT: intention-to-treat
PMP-SE: Perceived Maternal Parenting Self-Efficacy Questionnaire
RCT: randomized controlled trial
t1: baseline test
t2: posttest
WCG: waitlist control group

Edited by L Buis; submitted 10.08.22; peer-reviewed by J Brown, D Barsasella, A Notenbomer; comments to author 23.10.22; revised version received 16.12.22; accepted 13.01.23; published 10.03.23.

Please cite as:
Augustin M, Licata-Dandel M, Breeman LD, Harrer M, Bilgin A, Wolke D, Mall V, Ziegler M, Ebert DD, Friedmann A
Effects of a Mobile-Based Intervention for Parents of Children With Crying, Sleeping, and Feeding Problems: Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e41804
URL: https://mhealth.jmir.org/2023/1/e41804
doi:10.2196/41804
PMID:36897641

Original Paper

Mobile Prenatal Education and Its Impact on Reducing Adverse Pregnancy Outcomes: Retrospective Real-World Study

Jie Hao[1*], PhD; Lin Yang[1*], PhD; Yaxin Wang[2*], MBBS; Yushan Lan[1], BS; Xiaowei Xu[1], MS; Ziyang Wang[1], BM; Zanmei Li[1], MS; Liangkun Ma[2], MD; Jiao Li[1], PhD; Suhan Zhang[2], MBA; Yin Sun[2], MD

[1]Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yin Sun, MD
Department of Obstetrics and Gynecology
Peking Union Medical College Hospital
No.1 Shuaifuyuan Wangfujing, Dongcheng District
Beijing, 100730
China
Phone: 86 10 69151305
Email: goldhill@163.com

Abstract

Background: Pregnancy is a pivotal phase in a woman's life, demanding special attention to ensure maternal and fetal health. Prenatal education plays a vital role in promoting healthy pregnancies and reducing adverse outcomes for pregnant women. Mobile prenatal education programs have gained traction due to their accessibility and timeliness, especially in light of finite health care resources and the constraints imposed by the COVID-19 pandemic.

Objective: This study aims to develop and evaluate the effectiveness of a mobile-based prenatal education program in improving pregnancy outcomes.

Methods: We developed a mobile-based prenatal education curriculum in collaboration with a multidisciplinary maternal care team from Peking Union Medical College Hospital (PUMCH) in Beijing, China. Data were retrospectively collected from 1941 pregnant women who had registered for the PUMCH mobile prenatal education program and subsequently delivered at PUMCH between May 2021 and August 2022. The study compared pregnancy outcomes between the completing group, which were pregnant women who had completed at least 1 course, and the noncompleting group. We also analyzed differences among course topics within the completing group and assessed course topic popularity among pregnant women.

Results: The PUMCH mobile prenatal education curriculum consists of 436 courses across 9 topics. Out of the participants, a total of 1521 did not complete any courses, while 420 completed at least 1 course. Compared with the noncompleting group, pregnant women who completed courses exhibited a significant reduction in the risk of gestational diabetes mellitus, induced abortion, postpartum infection, fetal intrauterine distress, and neonatal malformation. Among those in the completing group, a total of 86% (361/420) started course completion during the first and second trimesters. Furthermore, completing courses related to topics of pregnancy psychology and pregnancy nutrition was associated with reduced risks of premature rupture of membranes and small for gestational age infants, respectively. Pregnancy psychology and postpartum recovery were the preferred topics among pregnant women.

Conclusions: The study demonstrates the potential of mobile-based prenatal education programs in improving pregnancy outcomes and supporting health care providers in delivering effective prenatal education. The rise of mobile prenatal education presents an opportunity to improve maternal and child health outcomes. Further research and broader implementation of such programs are warranted to continually improve maternal and child health.

(JMIR Mhealth Uhealth 2023;11:e46910) doi:10.2196/46910

KEYWORDS

adverse pregnancy outcome; mobile prenatal education; pregnancy; real-world study; retrospective study

Introduction

Pregnancy brings about a multitude of psychological and social changes in a woman's life [1]. Access to reliable information can aid pregnant women in making informed and healthier decisions [2]. Therefore, it is important to recognize pregnant women as the primary target audience for prenatal education, aiming to mitigate potential adverse outcomes [3]. To achieve desirable pregnancy outcomes, it is essential that pregnant women receive comprehensive and high-quality prenatal education.

Prenatal education has been widely acknowledged for its ability to increase prenatal examination usage [4], enhance the psychological well-being of expectant mothers [5], and contribute to improved delivery outcomes [6], among other benefits. Recent studies have highlighted the beneficial effects of prenatal education, particularly for women experiencing their first pregnancy. A retrospective analysis demonstrated that those who participated in childbirth classes were more likely to achieve successful normal vaginal deliveries, contributing positively to overall pregnancy outcomes [1]. Furthermore, the effectiveness of psychoeducational interventions has been confirmed through a randomized control trial, effectively reducing the fear of childbirth among anxious pregnant women [7]. Similarly, simulation-based childbirth education programs have demonstrated their capacity to alleviate the fear of childbirth among Chinese primiparas, and in high-income countries, childbirth training workshops have effectively reduced the incidence of unnecessary cesarean sections [8]. In another randomized controlled trial, a combination of 2 in-person and 11 telephone sessions focused on promoting healthy pregnancy behaviors led to a reduction in the weekly rate of gestational weight gain in pregnant women [9].

While traditional group prenatal education has been traditionally conducted in face-to-face settings, thereby limiting participant numbers, the evolving health care landscape has introduced new challenges. The increasing burden on health care systems has made it challenging for medical professionals to provide personalized health education to every expectant mother using conventional methods. In response to these accessibility issues and to better cater to the evolving needs of pregnant women and their partners, web-based prenatal education has gained significant traction [10]. Web-based education represents a novel and efficient teaching method, offering a more effective and timely approach compared to traditional face-to-face education [11]. During the COVID-19 pandemic and amid rising health care costs, the popularity of web-based education has soared [12,13]. A randomized clinical trial conducted among low-income postpartum women demonstrated that web-based education could reduce maternal weight gain during pregnancy [14]. Additionally, researchers provided web-based articles on physical activity during pregnancy and observed improvements in pregnant women's physical activity levels [15]. Furthermore, a randomized controlled trial highlighted the benefits of web-based prenatal education, including a reduction in concerns about labor, fear of childbirth, and fear of COVID-19 during the pandemic [16].

Mobile-based learning has also emerged as a promising approach for supporting prenatal education [17]. A smartphone-based prenatal education program aimed at parents at risk of preterm birth showcased its ability to raise awareness about preterm birth and establish a foundational knowledge base for making informed medical care choices [18]. In China, mobile-based prenatal education has gained significant traction [19], with thousands of maternal health-related apps available, most of which focus on prenatal education [20]. Recent statistics reveal that approximately half of the pregnant women in China have used maternal-related apps [21], with the majority participating in at least 1 mobile-based prenatal course during their early and mid-pregnancy [22]. Despite this, both national surveys have highlighted a shared desire among pregnant women for the development of evidence-based and well-informed mobile-based prenatal education programs, endorsed by obstetricians [21,22]. The needs of pregnant women underscore the prevailing tendency of existing mobile-based prenatal education programs to narrow their focus on specific topics, often lacking the comprehensive curriculum design offered by experienced obstetrician-led teams.

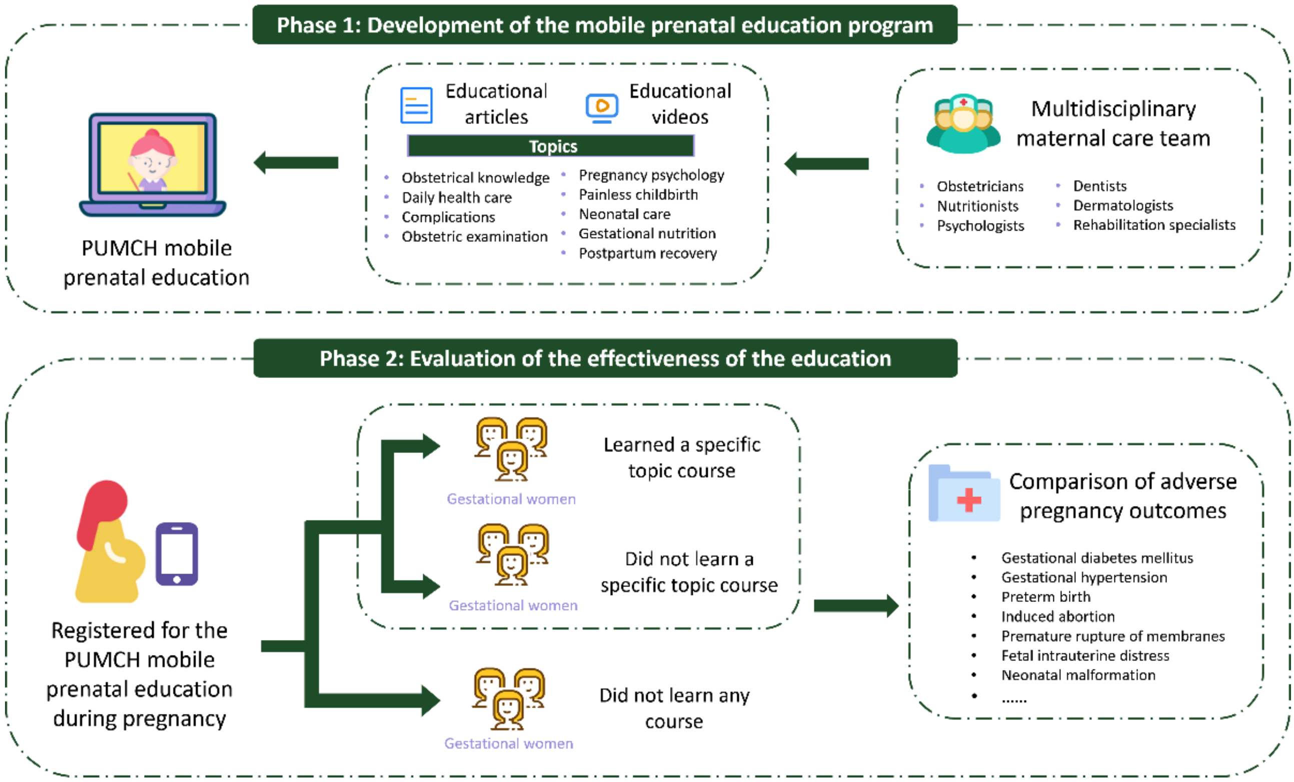
In this study, we present a comprehensive mobile-based prenatal education program tailored to the needs of pregnant women. Deployed within a hospital-authorized smartphone app, our program seeks to enhance the accessibility and cost-effectiveness of prenatal education. To evaluate the program's effectiveness, we conducted a retrospective real-world analysis of app module usage records and clinical outcomes among pregnant women. Our mobile-based prenatal education module helps to address the need for professional health care support and improve outcomes for expectant mothers.

Methods

Study Design

This study consists of 2 distinct phases. In the first phase, we developed a comprehensive curriculum for a mobile-based prenatal education program with the invaluable support and expertise of a multidisciplinary maternal care team at the Peking Union Medical College Hospital (PUMCH) in Beijing, China. All courses were designed to be easily accessible using smartphones. The second phase involved evaluating the effectiveness of our prenatal education program by analyzing the records of class attendance. The workflow of this study is shown in Figure 1.

Figure 1. Workflow of developing and evaluating the Peking Union Medical College Hospital (PUMCH) mobile prenatal education.



Development of the PUMCH Mobile Prenatal Education

The curriculum of our mobile-based prenatal education was designed and implemented by a multidisciplinary maternal care team at the PUMCH in Beijing, China. This team includes obstetricians, nutritionists, psychologists, dentists, dermatologists, and rehabilitation specialists, all possessing extensive expertise with over 20 years of traditional prenatal education. Drawing upon evidence-based insights gathered from pregnant women [21,22], the PUMCH team developed a comprehensive curriculum comprising a total of 436 courses, organized into 9 topics, which are obstetrical knowledge, gestational nutrition, daily health care, complications, obstetric examination, pregnancy psychology, painless childbirth, neonatal care, and postpartum recovery.

To optimize the learning experience, the courses were structured to span the entire duration of pregnancy. Roughly 1-2 courses were made available for each topic daily, and each course required just about 5 minutes to complete. This scheduling accommodated the busy routines of pregnant women, enabling them to learn effectively during shorter, fragmented periods of time. Furthermore, 2 external nurses conducted individual reviews of the developed courses to maintain the quality of our content. To enhance the knowledge acquisition experience, these courses were presented in a multimedia format, including articles and videos.

In addition, all courses were seamlessly integrated into the official PUMCH app, which provides a wide range of patient services, including registration, payment, and health education. Upon visiting the obstetrician and completing the registration process, pregnant women were granted authorized access to the PUMCH mobile prenatal education curriculum through this official app. This convenient platform allowed pregnant women to access comprehensive prenatal education at their convenience, free from the constraints of location and time that are often associated with traditional prenatal programs.

Recognizing that different trimesters of pregnancy require specific knowledge, we automated the course recommendations for pregnant women based on their current pregnancy trimester. Throughout the pregnancy journey, these tailored course recommendations were delivered daily, ensuring that expectant mothers received knowledge precisely suited to their specific stage of pregnancy.

Evaluation of the Effectiveness of the PUMCH Mobile Prenatal Education

To evaluate the effectiveness of the PUMCH mobile prenatal education program, we conducted multiple comparisons to assess its impacts on the pregnancy outcomes of registered pregnant women. This evaluation aimed to (1) determine whether completing courses can reduce the occurrence of adverse pregnancy outcomes in comparison to not participating in any courses, and (2) examine the impact of learning specific topics on pregnancy outcomes among pregnant women who completed the courses.

Study Population

Through the official PUMCH app, we retrospectively collected records from 1941 pregnant women who had registered for the PUMCH mobile prenatal education curriculum and subsequently delivered at the PUMCH in Beijing, China, during the period from May 2021 to August 2022. The inclusion criteria included singleton pregnancy and registration in the PUMCH mobile prenatal education program before delivery. The exclusion criteria were as follows: (1) diabetes mellitus (DM), (2) hypertension, and (3) a history of smoking [23].

Data Collection

Course-taking records, including details such as course titles, topics, and timings, were retrieved through the app from all eligible pregnant women participating in this study. Additionally, maternal characteristics and adverse pregnancy outcomes were extracted from electronic health records. Maternal characteristics included maternal age, prepregnancy BMI, parity, gravidity, history of abortion, history of gestational diabetes mellitus (GDM), history of abnormal pregnancy, family history of DM, family history of hypertension, and polycystic ovary syndrome (PCOS). Adverse pregnancy outcomes included GDM, gestational hypertension, postpartum infection, preterm birth, induced abortion, premature rupture of membranes (PROM), fetal intrauterine distress, and neonatal malformation.

Statistical Analyses

The data were analyzed with Python (version 3.8.8; Python Software Foundation). As for the maternal characteristics and adverse pregnancy outcomes, continuous variables were reported as mean (SD) and categorical variables as n (%). To determine the effectiveness of the PUMCH mobile prenatal education, the 2-tailed *t* test was performed for continuous variables and the chi-square test for categorical variables, and $P<.05$ was considered statistically significant. In addition, multiple logistic regressions, adjusted by maternal characteristics, were performed.

Ethical Considerations

As we used anonymized and deidentified data and did not constitute human research, the need for written informed consent was waived by the Ethics Review Board of the PUMCH due to the retrospective nature of the study. This study was approved by the Ethics Review Board of the PUMCH (I-22PJ122). All procedures were performed in accordance with the Declaration of Helsinki.

Results

The PUMCH Mobile Prenatal Education

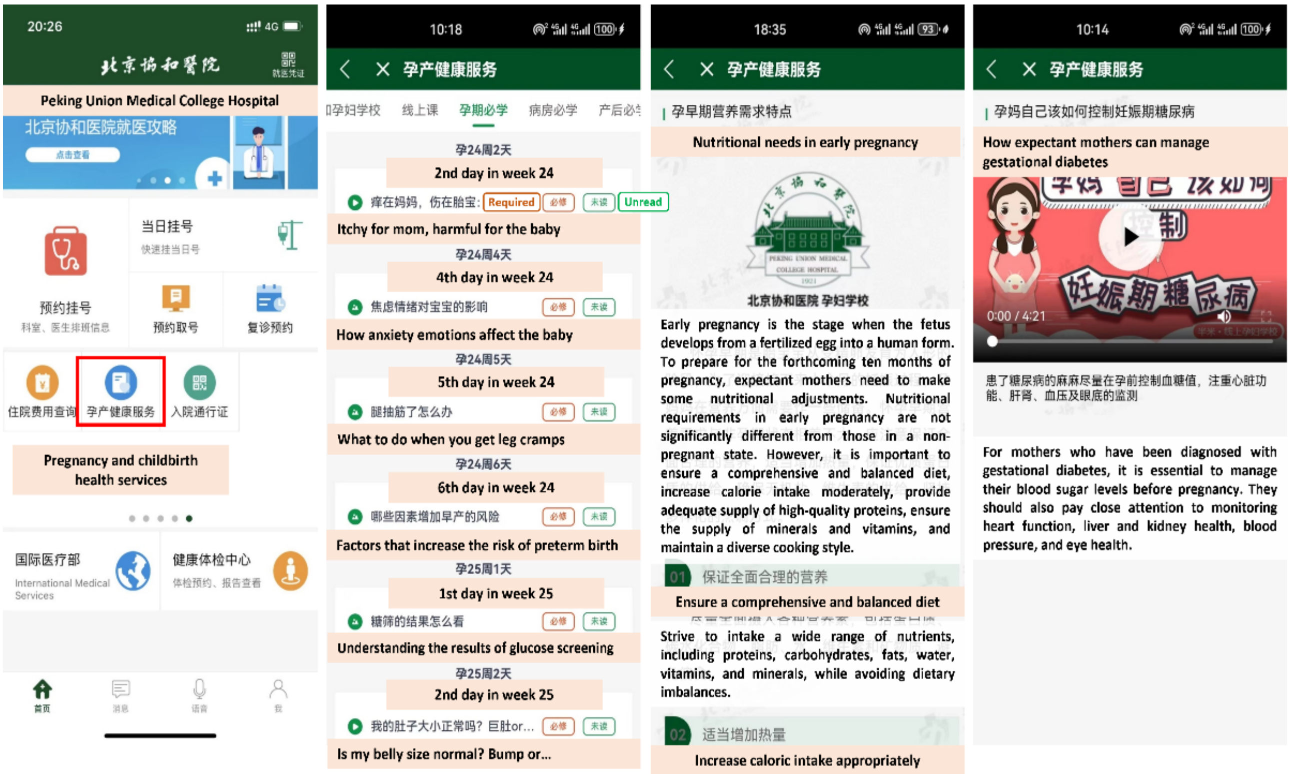
In this study, we implemented a total of 436 mobile-based prenatal courses with valuable support from the PUMCH multidisciplinary maternal care team. These courses included 234 educational videos and 202 articles. An illustrative example of course titles and formats for 3 topics is presented in Table 1.

Table 1. Example of course titles and formats for 3 topics in the Peking Union Medical College Hospital (PUMCH) mobile prenatal education program.

Topic and title	Format
Obstetrical knowledge	
How to distinguish between true and false labor contractions?	Video
How to determine if a vaginal delivery is possible?	Video
Is vaginal bleeding in late pregnancy always a sign of labor onset?	Article
Don't be misled again! Common pregnancy myths.	Article
Gestational nutrition	
Is eating pork liver effective for treating anemia during pregnancy?	Video
When encountering gestational diabetes, how should expectant mothers adjust their diet?	Video
Dietary adjustments in early pregnancy	Article
How to supplement iron?	Article
Daily health care	
Pregnant and have a cold? Be cautious about avoiding medication.	Video
Four ways to relieve abdominal and lower back pain during pregnancy.	Video
Precautions for exercise during pregnancy.	Article
How to perform chest and upper arm stretches during pregnancy?	Article

These courses were accessible to pregnant women through the official PUMCH app. In addition to receiving daily course recommendations tailored to their trimesters, pregnant women had the flexibility to explore other courses of interest, allowing them to acquire prenatal knowledge according to their individual needs. Figure 2 provides snapshots of the PUMCH mobile prenatal education interface. The first snapshot reveals the easy accessibility of our program on the main page of the official app, seamlessly integrated with other medical services. The second snapshot highlights the prenatal course recommended to pregnant women, taking into account their specific pregnancy stage in terms of the number of days. As illustrated in the third and fourth snapshots, all courses were available in either article or video format.

Figure 2. Snapshots of the Peking Union Medical College Hospital (PUMCH) mobile prenatal education program. From left to right: the main page of the official PUMCH app, recommended prenatal courses organized by the number of days, a course in article format, and a course in video format.



Participants Characteristics

A total of 1941 pregnant women were included in the study, with 1521 of them not completing any courses, while 420 completed at least 1 course. Table 2 presents the maternal characteristics of participants in these 2 groups. The average age of participants in both groups is approximately 32 years. In the group that did not complete the courses, the proportion of participants of advanced maternal age (35 years of age or older) is 25.05% (381/1521), which is about 3% higher than the group that completed courses. There is a statistically significant difference in prepregnancy BMI values (kg/m^2) between the 2 groups (P=.04), with participants in the completing group having higher BMI values than those in the not completing group. Additionally, parity and gravidity also exhibit significant differences between the 2 groups with very small P values (P<.001). In the completing group, participants have a higher proportion of nullipara, primigravida, or both.

Table 2. Maternal characteristics of participants between 2 groups of not completing and completing courses.

Characteristics	Noncompleting (n=1521)	Completing (n=420)	*P* value
Maternal age (years), mean (SD)	32.01 (3.83)	32.06 (3.69)	.80
Maternal age group (years), n (%)			.18
<35	1140 (74.95)	328 (78.10	
≥35	381 (25.05)	92 (21.90)	
Prepregnancy BMI (kg/m^2), mean (SD)	23.72 (3.41)	24.12 (3.39)	*.04* [a]
BMI group (kg/m^2), n (%)			.17
<18.5	39 (2.56)	7 (1.66)	
18.5-24	816 (53.65)	206 (49.05)	
24-28	498 (32.74)	159 (37.86)	
≥28	168 (11.05)	48 (11.43)	
Parity, n (%)			*<.001*
Nullipara	1179 (77.51)	375 (89.29)	
Multipara	342 (22.49)	45 (10.71)	
Gravidity, n (%)			*<.001*
Primigravid	876 (57.59)	278 (66.19)	
Multigravid	645 (42.41)	142 (33.81)	
History of abortion, n (%)			.79
No	1069 (70.28)	298 (70.95)	
Yes	452 (29.72)	122 (29.05)	
History of GDM[b], n (%)			.53
No	1514 (99.54)	419 (99.76)	
Yes	7 (0.46)	1 (0.24)	
History of abnormal pregnancy, n (%)			.22
No	1445 (95)	405 (96.43)	
Yes	76 (5)	15 (3.57)	
PCOS[c], n (%)			.24
No	1461 (96.06)	398 (94.76)	
Yes	60 (3.94)	22 (5.24)	
Family history of DM[d], n (%)			.53
No	1203 (79.09)	338 (80.48)	
Yes	318 (20.91)	82 (19.52)	
Family history of hypertension, n (%)			.06
No	1012 (66.54)	300 (71.43)	
Yes	509 (33.46)	120 (28.57)	

[a]Italics indicate significant values.
[b]GDM: gestational diabetes mellitus.
[c]PCOS: polycystic ovary syndrome.
[d]DM: diabetes mellitus.

Statistics on Completing the PUMCH Mobile Prenatal Courses

In the PUMCH mobile prenatal education program, we have a total of 436 courses available, with 234 presented in video format and 202 in article format. These courses are categorized into 1 topic, and the topic with the highest number of courses is obstetrical knowledge, accounting for approximately 24.5% (107/436) of the total course count. Among the 420 pregnant women who completed at least 1 course, we observed that 43.8% (184/420) of them began completing courses during their first trimester, and a significant 86% (361/420) started course completion during their first and second trimesters. Table 3 shows that the topics of obstetrical knowledge and daily health care were the most popular, with completion rates over 80% (351/420 and 344/420). Furthermore, when considering the average number of participants per course, it was obvious that the topics of pregnancy psychology and postpartum recovery had the most attention. Detailed participant counts for each topic are provided in Table 3.

Table 3. Statistics on the Peking Union Medical College Hospital (PUMCH) mobile prenatal education: topics, course count (%), participant count (%), and average participants per course.

Topics	Courses, n (%)	Participants who completed courses (n=420), n (%)	Average number of participants per course
Obstetrical knowledge	107 (24.5)	351 (83.6)	3.28
Daily health care	73 (16.7)	344 (81.9)	4.71
Complications	71 (16.3)	244 (58.1)	3.43
Gestational nutrition	44 (10.1)	295 (70.2)	6.70
Pregnancy psychology	34 (7.8)	278 (66.2)	8.17
Obstetric examination	33 (7.6)	234 (55.7)	7.09
Neonatal care	29 (6.7)	150 (35.7)	5.17
Painless childbirth	25 (5.7)	171 (40.7)	6.84
Postpartum recovery	20 (4.6)	143 (34)	7.15

Comparison of Adverse Pregnancy Outcomes Between the Noncompleting and Completing Groups

Table 4 displays the comparisons of adverse pregnancy outcomes between the 2 groups. There are statistically significant differences in adverse pregnancy outcomes for GDM, postpartum infection, induced abortion, fetal intrauterine distress, and neonatal malformation between the 2 groups. For instance, the odds ratio (OR) value for the noncompleting group compared with the completing group in GDM is 0.3043, which means that the odds of GDM in the completing group are approximately 30% lower than the odds in the noncompleting group. This significant OR value indicates that participants in the completing group have a lower risk of GDM compared with noncompleting group. Except for neonatal malformation, for which there were no cases in the completing group, multiple logistic regressions were conducted for each significant adverse pregnancy outcome, adjusted by maternal age group, BMI group, parity, and gravidity. GDM, postpartum infection, and fetal intrauterine distress remained statistically significant. Detailed results are provided in Table S1 in Multimedia Appendix 1. However, there were no significant differences observed between the 2 groups for gestational hypertension, preterm birth, PROM, macrosomia, and small for gestational age (SGA).

Table 4. Adverse pregnancy outcomes of participants between the noncompleting and completing groups.

Adverse pregnancy outcomes	Noncompleting (n=1521), n (%)	Completing (n=420), n (%)	*P* value	OR[a]
GDM[b]	285 (18.74)	98 (23.33)	*.04* [c]	0.3043
Gestational hypertension	30 (1.97)	8 (1.90)	.93	0.0194
Postpartum infection	115 (7.56)	52 (12.38)	*.002*	0.1413
Induced abortion	419 (27.55)	143 (34.05)	*.009*	0.5162
Preterm birth	98 (6.44)	23 (5.47)	.47	0.0579
PROM[d]	391 (25.71)	126 (30)	.08	0.4286
Fetal intrauterine distress	219 (14.40)	83 (19.76)	*.007*	0.2463
Neonatal malformation	17 (1.11)	0 (0)	*.03*	0.0000
SGA[e]	12 (0.79)	4 (0.95)	.74	0.0096
Macrosomia	67 (4.41)	15 (3.57)	.45	0.0370

[a]OR: odds ratio.
[b]GDM: gestational diabetes mellitus.
[c]Italics indicate significant values.
[d]PROM: premature rupture of membranes.
[e]SGA: small for gestational age.

Comparison of Adverse Pregnancy Outcomes Among Course Topics in the Completing Group

We examined whether completing specific course topics reduced adverse pregnancy outcomes among participants in the completing group. For each topic, participants were categorized into 2 groups: those who completed the topic and those who did not complete it. Table 5 presents the significant differences in adverse pregnancy outcomes between the "completing the topic" group and the "not completing the topic" group. Participants who completed the pregnancy psychology topic course experienced a reduced risk of PROM compared with the "not completing the topic" group (*P*=.03 and OR=0.5027). Additional results for pairs of topics and adverse pregnancy outcomes are available in Table S2 in Multimedia Appendix 1.

Table 5. Effectiveness of adverse pregnancy outcomes among participants in the completing group for different topics.

Adverse pregnancy outcomes	Topics	Not completing the topic, n/N (%)	Completing the topic, n/N (%)	*P* value	OR[a]
PROM[b]	Pregnancy psychology	33/142 (25.71)	93/278 (30)	*.03* [c]	0.5027
SGA[d]	Gestational nutrition	3/125 (0.79)	1/295 (0.95)	*.05*	0.0034

[a]OR: odds ratio.
[b]PROM: premature rupture of membranes.
[c]Italics indicate significant values.
[d]SGA: small for gestational age.

Discussion

Principal Findings

In this study, we have collaboratively designed a mobile-based prenatal education program with the PUMCH multidisciplinary maternal care team. This mobile-based curriculum offers 436 courses across 9 topics and is accessible through the official PUMCH app, providing pregnant women with valuable guidance throughout their entire pregnancy journey. A mixed methods study has indicated that web-based prenatal education can positively influence lifestyle choices and the ease of accessing health information during pregnancy [21]. The COVID-19 pandemic has heightened the importance of easily accessible health care resources, especially for pregnant women at increased risk of severe illness if infected [24]. The World Health Organization (WHO) advises pregnant women to avoid crowded and poorly ventilated indoor spaces to reduce the risk of COVID-19 transmission. Therefore, there is an urgent need for cost-effective and easily accessible mobile-based prenatal education programs to support expectant mothers, especially in these challenging times.

To evaluate the effectiveness of our mobile-based program, we compared adverse pregnancy outcomes between the group that completed courses and the group that registered but did not complete any courses. Our data, derived from real-world app records, revealed an imbalance between the 2 groups, with approximately 21.6% (420/1941) of registered pregnant women completing at least 1 course. Significantly, pregnant women who completed courses demonstrated a reduced risk of adverse pregnancy outcomes such as GDM, postpartum infection, induced abortion, fetal intrauterine distress, and neonatal malformation. For instance, GDM, a common pregnancy complication, is associated with the risk of hyperglycemia and adverse short- and long-term health outcomes for both mothers and infants [25]. Our curriculum addresses GDM directly, offering courses on improving physical activity and nutrition, potentially lowering the risk of GDM. Taking those courses shows the potential to lower the risk of GDM by delivering knowledge about improving physical activity, nutrition intake, etc [26,27].

Furthermore, we analyzed the behavior of pregnant women who completed specific topic courses in the completing group. While the impact of completing the specific topic courses on reducing adverse pregnancy outcomes was generally modest, we identified 2 significant findings. First, completing the pregnancy psychology course was associated with a statistically significant reduction in the risk of PROM, and second, completing the pregnancy nutrition course may prevent the occurrence of SGA infants.

Nutrition during pregnancy plays an important role in lifelong health [28], especially in terms of brain development and behavior [29]. As many adverse birth outcomes originate during pregnancy [30], awareness of nutritional balance and healthy dietary habits can reduce the risk of outcomes like SGA. SGA, typically defined as birthweight below the tenth percentile [31], is a known risk factor for stillbirth [32]. Previous research has demonstrated that nutrition interventions and adopting healthy dietary patterns can significantly lower the risk of SGA [33]. Moreover, maternal creatine supplementation during pregnancy has shown the potential to reduce the risk of neonatal asphyxia, as evidenced in small animal studies [34].

We observed that 86% (361/420) of participants started course completion during their first and second trimesters, in line with survey results on Chinese web-based prenatal education [22]. Pregnancy psychology and postpartum recovery are the most popular topics. These findings can inform the development of additional educational courses on both topics and further engage pregnant women.

Our mobile-based prenatal education program serves as a valuable source of information, enhancing pregnant women's knowledge regarding gestational nutrition, recommended practices, and anticipated challenges. Given that pregnant women increasingly seek information on the web [35], our program not only meets the rising demand for accessible mobile prenatal education but also bridges a significant gap in this field by offering a cost-effective and efficient source of information [22,35]. By promoting this curriculum more widely, we aim to encourage healthier lifestyles and enhance pregnancy outcomes.

Limitations

This study has several limitations. First, the effectiveness of our mobile-based prenatal education program was assessed among pregnant women from a single center, specifically PUMCH in Beijing, and its generalizability to other settings needs to be confirmed. Second, the analysis was based on 1 year of data, and a larger sample size with a longer duration would provide more robust estimates of the effectiveness of our mobile-based program. Third, as this is a retrospective study with outcomes and information collected post pregnancy, potential confounders and biases may have influenced the results. To establish the effectiveness of mobile-based prenatal education more conclusively, future research should consider conducting randomized controlled trials or prospective cohort studies.

Conclusion

This study outlined the development of a mobile-based prenatal education program and assessed its effectiveness by analyzing adverse pregnancy outcomes among pregnant women. The findings indicate that the mobile-based prenatal education curriculum has the potential to reduce adverse outcomes in pregnant women, offering valuable support for health care providers in delivering effective prenatal education services. The growing prominence of mobile prenatal education presents an opportunity to enhance pregnancy outcomes and the health of both mothers and children.

Acknowledgments

The authors would like to thank Nana Liu, Xuanjin Yang, and Hang Lin for their support with follow-up data collection and data management.

This study was supported by the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS; grant 2021-I2M-1-056, 2020-I2M-2-009, and 2021-I2M-1-023), the Key Research and Development (R&D) Program of Hunan Province (grant 2021SK2024), the National Key R&D Program of China (grant 2022YFC2703304), Recommendations for weight gain in women with gestational diabetes mellitus (grant 20191901), and the Clinical and Translational Medicine Research Fund of the Central Public Welfare Research Institutes of the CAMS (grant 2019XK320007).

Authors' Contributions

JH, LY, JL, and YS contributed to the concept and design of the study. YW, LM, SZ, and YS developed the methodology and collected the data. YL and ZW analyzed the data. JH, LY, YL, XX, and ZL interpreted the results. YL drafted the initial manuscript. JH and LY revised the manuscript. JL and YS supervised the work. All authors contributed to the review of the manuscript and provided approval for the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Result of all comparisons on adverse pregnancy outcomes.
[DOCX File , 23 KB - mhealth_v11i1e46910_app1.docx]

References

1. Gluck O, Pinchas-Cohen T, Hiaev Z, Rubinstein H, Bar J, Kovo M. The impact of childbirth education classes on delivery outcome. Int J Gynaecol Obstet 2020;148(3):300-304. [doi: 10.1002/ijgo.13016] [Medline: 31912479]
2. Vogels-Broeke M, Daemers D, Budé L, de Vries R, Nieuwenhuijze M. Sources of information used by women during pregnancy and the perceived quality. BMC Pregnancy Childbirth 2022;22(1):109 [FREE Full text] [doi: 10.1186/s12884-022-04422-7] [Medline: 35135487]
3. Gourounti K, Sarantaki A, Dafnou ME, Hadjigeorgiou E, Lykeridou A, Middleton N. A qualitative study of assessing learning needs and digital health literacy in pregnancy: Baby Buddy forward Greek findings. Eur J Midwifery 2022;6:55 [FREE Full text] [doi: 10.18332/ejm/150770] [Medline: 36119404]
4. Shi Y, Wang D, Yuan Y, Jiang Y, Zeng Q, Chang C. The effect of prenatal education curriculum on mother's prenatal examination utilization, delivery mode and recovery status: a cross-sectional survey in China. Environ Health Prev Med 2015;20(6):397-403 [FREE Full text] [doi: 10.1007/s12199-015-0480-4] [Medline: 26201848]
5. Shen Q, Huang CR, Rong L, Ju S, Redding SR, Ouyang YQ, et al. Effects of needs-based education for prenatal anxiety in advanced multiparas: a randomized controlled trial. BMC Pregnancy Childbirth 2022;22(1):301 [FREE Full text] [doi: 10.1186/s12884-022-04620-3] [Medline: 35395734]
6. Vanderlaan J, Sadler C, Kjerulff K. Association of delivery outcomes with the number of childbirth education sessions. J Perinat Neonatal Nurs 2021;35(3):228-236 [FREE Full text] [doi: 10.1097/JPN.0000000000000579] [Medline: 34330134]
7. Fenwick J, Toohill J, Gamble J, Creedy DK, Buist A, Turkstra E, et al. Effects of a midwife psycho-education intervention to reduce childbirth fear on women's birth outcomes and postpartum psychological wellbeing. BMC Pregnancy Childbirth 2015;15:284 [FREE Full text] [doi: 10.1186/s12884-015-0721-y] [Medline: 26518597]
8. Chen I, Opiyo N, Tavender E, Mortazhejri S, Rader T, Petkovic J, et al. Non-clinical interventions for reducing unnecessary caesarean section. Cochrane Database Syst Rev 2018;9(9):CD005528 [FREE Full text] [doi: 10.1002/14651858.CD005528.pub3] [Medline: 30264405]
9. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai AL, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. Lancet Diabetes Endocrinol 2020;8(6):490-500 [FREE Full text] [doi: 10.1016/S2213-8587(20)30107-8] [Medline: 32445736]
10. Roch G, Da Silva RB, de Montigny F, Witteman HO, Pierce T, Semenic S, et al. Impacts of online and group perinatal education: a mixed methods study protocol for the optimization of perinatal health services. BMC Health Serv Res 2018;18(1):382 [FREE Full text] [doi: 10.1186/s12913-018-3204-9] [Medline: 29843691]
11. Longhini J, Rossettini G, Palese A. Massive open online courses for nurses' and healthcare professionals' continuous education: a scoping review. Int Nurs Rev 2021;68(1):108-121. [doi: 10.1111/inr.12649] [Medline: 33855697]
12. Wijesooriya NR, Mishra V, Brand PLP, Rubin BK. COVID-19 and telehealth, education, and research adaptations. Paediatr Respir Rev 2020;35:38-42 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.009] [Medline: 32653468]
13. Costa SA, Kavouras I, Cohen N, Huang TTK. Moving education online during the COVID-19 pandemic: thinking back and looking ahead. Front Public Health 2021;9:751685 [FREE Full text] [doi: 10.3389/fpubh.2021.751685] [Medline: 34760865]
14. Phelan S, Hagobian T, Brannen A, Hatley KE, Schaffner A, Muñoz-Christian K, et al. Effect of an internet-based program on weight loss for low-income postpartum women: a randomized clinical trial. JAMA 2017;317(23):2381-2391 [FREE Full text] [doi: 10.1001/jama.2017.7119] [Medline: 28632867]
15. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women—the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]
16. Uludağ E, Serçekuş P, Vardar O, Özkan S, Alataş SE. Effects of online antenatal education on worries about labour, fear of childbirth, preparedness for labour and fear of covid-19 during the covid-19 pandemic: a single-blind randomised controlled study. Midwifery 2022;115:103484 [FREE Full text] [doi: 10.1016/j.midw.2022.103484] [Medline: 36155390]
17. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
18. Kim UO, Barnekow K, Ahamed SI, Dreier S, Jones C, Taylor M, et al. Smartphone-based prenatal education for parents with preterm birth risk factors. Patient Educ Couns 2019;102(4):701-708 [FREE Full text] [doi: 10.1016/j.pec.2018.10.024] [Medline: 30396713]
19. Chen XW, Jiang LY, Chen Y, Guo LF, Zhu XH. Analysis of online antenatal education class use via a mobile terminal app during the COVID-19 pandemic. BMC Pregnancy Childbirth 2022;22(1):412 [FREE Full text] [doi: 10.1186/s12884-022-04745-5] [Medline: 35578202]

20. Zhang P, Dong L, Chen H, Chai Y, Liu J. The rise and need for mobile apps for maternal and child health care in China: survey based on app markets. JMIR Mhealth Uhealth 2018;6(6):e140 [FREE Full text] [doi: 10.2196/mhealth.9302] [Medline: 29884601]
21. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the use of smartphone apps for health information among pregnant Chinese women: mixed methods study. JMIR Mhealth Uhealth 2019;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
22. Huang X, Sun W, Wang R, Wu H, Yu S, Fang X, et al. Present situation and the future development of web-based prenatal education in China: cross-sectional web-based survey. J Med Internet Res 2022;24(6):e28637 [FREE Full text] [doi: 10.2196/28637] [Medline: 35771613]
23. Rogers JM. Smoking and pregnancy: epigenetics and developmental origins of the metabolic syndrome. Birth Defects Res 2019;111(17):1259-1269 [FREE Full text] [doi: 10.1002/bdr2.1550] [Medline: 31313499]
24. Coronavirus disease (COVID-19): pregnancy, childbirth and the postnatal period. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-pregnancy-and-childbirth [accessed 2023-12-12]
25. Johns EC, Denison FC, Norman JE, Reynolds RM. Gestational diabetes mellitus: mechanisms, treatment, and complications. Trends Endocrinol Metab 2018;29(11):743-754. [doi: 10.1016/j.tem.2018.09.004] [Medline: 30297319]
26. Kc K, Shakya S, Zhang H. Gestational diabetes mellitus and macrosomia: a literature review. Ann Nutr Metab 2015;66(Suppl 2):14-20 [FREE Full text] [doi: 10.1159/000371628] [Medline: 26045324]
27. Davenport MH, Ruchat SM, Poitras VJ, Garcia AJ, Gray CE, Barrowman N, et al. Prenatal exercise for the prevention of gestational diabetes mellitus and hypertensive disorders of pregnancy: a systematic review and meta-analysis. Br J Sports Med 2018;52(21):1367-1375. [doi: 10.1136/bjsports-2018-099355] [Medline: 30337463]
28. Huang LT. Maternal and early-life nutrition and health. Int J Environ Res Public Health 2020;17(21):7982 [FREE Full text] [doi: 10.3390/ijerph17217982] [Medline: 33143058]
29. Moreno-Fernandez J, Ochoa JJ, Lopez-Frias M, Diaz-Castro J. Impact of early nutrition, physical activity and sleep on the fetal programming of disease in the pregnancy: a narrative review. Nutrients 2020;12(12):3900 [FREE Full text] [doi: 10.3390/nu12123900] [Medline: 33419354]
30. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
31. Blankenship SA, Brown KE, Simon LE, Stout MJ, Tuuli MG. Antenatal corticosteroids in preterm small-for-gestational age infants: a systematic review and meta-analysis. Am J Obstet Gynecol MFM 2020;2(4):100215 [FREE Full text] [doi: 10.1016/j.ajogmf.2020.100215] [Medline: 33345924]
32. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. BMJ 2013;346:f108 [FREE Full text] [doi: 10.1136/bmj.f108] [Medline: 23349424]
33. Crovetto F, Crispi F, Casas R, Martín-Asuero A, Borràs R, Vieta E, et al. Effects of mediterranean diet or mindfulness-based stress reduction on prevention of small-for-gestational age birth weights in newborns born to at-risk pregnant individuals: the IMPACT BCN randomized clinical trial. JAMA 2021;326(21):2150-2160 [FREE Full text] [doi: 10.1001/jama.2021.20178] [Medline: 34874420]
34. LaRosa DA, Ellery SJ, Parkington HC, Snow RJ, Walker DW, Dickinson H. Maternal creatine supplementation during pregnancy prevents long-term changes in diaphragm muscle structure and function after birth asphyxia. PLoS One 2016;11(3):e0149840 [FREE Full text] [doi: 10.1371/journal.pone.0149840] [Medline: 26930669]
35. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]

Abbreviations

DM: diabetes mellitus
GDM: gestational diabetes mellitus
PCOS: polycystic ovary syndrome
PROM: premature rupture of membranes
PUMCH: Peking Union Medical College Hospital
SGA: small for gestational age
WHO: World Health Organization

Edited by L Buis; submitted 02.03.23; peer-reviewed by H Wu; comments to author 18.05.23; revised version received 03.11.23; accepted 29.11.23; published 20.12.23.

Please cite as:
Hao J, Yang L, Wang Y, Lan Y, Xu X, Wang Z, Li Z, Ma L, Li J, Zhang S, Sun Y
Mobile Prenatal Education and Its Impact on Reducing Adverse Pregnancy Outcomes: Retrospective Real-World Study
JMIR Mhealth Uhealth 2023;11:e46910
URL: https://mhealth.jmir.org/2023/1/e46910
doi:10.2196/46910
PMID:38117555

Feasibility of a Smartphone-Based Hearing Aid App for Mild-to-Moderate Hearing Loss: Prospective Multicenter Randomized Controlled Trial

Sungmin Jo[1,2], BA; Moo Kyun Park[1,3], MD, PhD; Jae-Hyun Seo[4], MD, PhD; Ki-eun Lee[1], BA, MD; Jae Sang Han[4], MD, PhD; Ji Hyung Lim[4], MD; Jun Ho Lee[1,3], MD, PhD; Seung-Ha Oh[1,3], MD, PhD

1

2

3

4

Corresponding Author:
Moo Kyun Park, MD, PhD

Abstract

Background: Hearing loss is a growing health concern worldwide. Hearing aids (HAs) are the treatment of choice for hearing rehabilitation in most cases of mild-to-moderate hearing loss. However, many patients with hearing loss do not use HAs due to their high cost, stigma, and inaccessibility. Since smartphones are widely used, many apps that mimic the amplification function of HAs have been introduced. Smartphone-based HA apps (SHAAs) are affordable and easy to access. However, the audiological benefit of SHAAs has not been determined.

Objective: We compared the audiological performance between an SHAA and a conventional HA in a prospective, multicenter randomized controlled trial.

Methods: Patients with mild-to-moderate hearing loss were prospectively enrolled from 2 tertiary hospitals and randomly assigned to either an SHAA (Petralex; IT4YOU Corp LLC) or a conventional HA (Siya 1 miniRITE; Oticon A/S). For the cross-over study design, participants used the alternate device and repeated the same 2-month trial. Audiological measurements were obtained using hearing tests, real-ear measurements, and the hearing-in-noise test (HINT). Subjective satisfaction was evaluated using the Abbreviated Profile of Hearing Aid Benefit (APHAB) and International Outcome Inventory for Hearing Aids (IOI-HA).

Results: Overall, 63 participants were screened and 38 completed the study. In sound-field audiometry testing, the SHAA showed a 20- to 60-dB gain in the low-to-high frequencies of the hearing threshold level. The HA provided adequate gain in the middle-to-high frequencies (55, 65, and 75 dB in real-ear measurements), which is the sound level for most speaking volumes. However, the SHAA could not improve word recognition at 50 dB. The HA showed better audiological performance than the SHAA in both quiet and noisy conditions in the HINT. The IOI-HA scores were significantly improved by both the HA and SHAA versus unaided conditions. Among the SHAA users, 37% (14/38), 42% (16/38), 24% (9/38), and 32% (12/38) showed improvement in APHAB scores for ease of communication, reverberation, background noise, and aversiveness of sounds, respectively. There were no differences in adverse events between the 2 study groups.

Conclusions: The HA showed better performance than the SHAA in word recognition and the HINT. However, the SHAA was significantly better than unaided hearing in terms of amplification. The SHAA may be a useful hearing assistance device for patients with mild-to-moderate hearing loss when listening to soft sounds in quiet conditions. The SHAA demonstrated poorer performance than the HA in the mid- to high-frequency sounds that are important for word recognition, sound quality, and hearing in noisy conditions. Further development of the signal technology of SHAAs is needed to improve the sound quality of mid- to high-frequency sounds and overcome noisy environments.

Trial Registration: ClinicalTrials.gov NCT05644106; http://clinicaltrials.gov/ct2/show/NCT05644106

(JMIR Mhealth Uhealth 2023;11:e46911) doi:10.2196/46911

KEYWORDS

hearing aid apps; hearing aids; hearing loss; prospective; smartphone; hearing rehabilitation

Introduction

Hearing loss is a common and growing global health issue. The World Health Organization (WHO) reported that approximately 2.5 billion people will experience hearing loss and 700 million will need hearing rehabilitation by 2050 [1]. The prevalence of hearing loss increases with age [2]. Further, 9 in 10 people with hearing loss are adults and >70% are older adults. Among those aged >60 years, one-fourth have hearing loss [3]. Patients with hearing loss have difficulty communicating and understanding speech [4]. Patients with hearing loss often perform poorly in school and have a higher unemployment rate [5,6]. Hearing loss decreases quality of life and cognitive function, as well as increasing anxiety and depression. Hearing loss is the largest population-attributable fraction for dementia [7]. The WHO suggests that the annual global cost for unaddressed hearing loss is approximately US $980 billion [8].

Hearing rehabilitation improves audiological performance, daily activity functioning, and quality of life. Hearing aids (HAs) are the most common and important method of hearing rehabilitation [9,10]. Conventional HAs have shown significantly improved amplification in patients with mild-to-moderate hearing loss [11]. HAs benefit physical, social, emotional, and mental well-being [12]. However, due to the global high costs, stigma, difficult accessibility, and inconvenience, only 17% of those who need HAs for hearing loss have them [13]. People who have internalized a strong sense of stigma against hearing loss can exhibit denial of the condition, leading them to delay or reject hearing habilitation interventions [14]. Other barriers include low levels of social support from family and friends, which is an important factor in empowering patients to receive hearing rehabilitation [15]. Furthermore, the HA utilization rate is primarily affected by HA prices and insurance reimbursement policies [16]. In South Korea, the proportion of people receiving HA subsidies increased after reimbursement for HAs increased [16]. HAs are used by only 6% of the population in India and 9% in Peru [17], whereas in other low-income countries, HA use is almost nonexistent. Recently, the US government approved over-the-counter HAs to improve access and provide HAs at a lower cost [18]. Hearing loss is the dementia cause with the largest population-attributable fraction, and HA use is the most important factor protecting against cognitive decline [1].

Smartphones can be useful devices for amplification, and many smartphone-based HA apps (SHAAs) have been developed to improve the audibility of sounds for both individuals with and without hearing impairment [19]. They use signal-processing algorithms similar to those used in conventional HAs. Accessibility and low cost are the main advantages of SHAAs [20]. They are either free or require a subscription after a free trial. The subscription fee is very low compared to HA costs and even to other personal sound amplification devices [21,22]. For example, the subscription fee of the Petralex HA app is currently US $12.43/mo or US $59.50/y, but the retail price of the Siya 1 miniRITE HA is US $2863/piece. Smartphone users can easily locate and download these apps from their internet store. SHAAs allow users to access hearing amplification through earbuds or headphones. The basic features of SHAAs include adjustable amplification, an equalizer, and sometimes a noise reduction feature. Other options include self-audiometry via earbuds and feedback control. However, if a patient wants to purchase an HA, receive follow-up to control sound quality, and benefit from various advanced functions such as noise reduction, visiting a specialist is necessary. A growing number of individuals with hearing impairment who are not ready to invest in conventional HAs have shown an interest in SHAAs due to their accessibility, ease of use, and low cost [19]. SHAA use may also improve a person's attitude toward amplification and conventional HAs [23].

Although some SHAAs provide the recommended level of processing delay, conventional HAs have shown better performance than SHAAs in most aspects of amplification. Our previous study on the electroacoustic performance of various SHAAs [19] demonstrated that some SHAAs provided satisfactory amplification [24]. In addition, other studies have demonstrated the feasibility of SHAAs for patients with hearing loss [24]. In this study, we investigated the listening performance of an SHAA using objective assessments and subjective satisfaction in patients with mild-to-moderate hearing loss.

This study compared the audiological performance of an SHAA to that of a conventional HA. The study settings included unaided and aided conditions in quiet and noisy situations. The objective benefit of hearing amplification was evaluated in sound-field tests and real-ear measurements (REMs). We used the sound-field test to evaluate the hearing threshold and word recognition scores (WRSs) at 50 dB. REMs were studied using 55-, 65-, and 75-dB stimuli in the 250-8000 Hz range. The subjective benefits were evaluated using the Abbreviated Profile of Hearing Aid Benefit (APHAB) and International Outcome Inventory for Hearing Aids (IOI-HA).

Methods

Study Design

A prospective randomized controlled trial was conducted at 2 tertiary hospitals in South Korea (the Department of Otolaryngology, Seoul National University Hospital and the Department of Otolaryngology, The Catholic University of Korea, Seoul St. Mary's Hospital) from August 2020 to November 2022. The inclusion criteria were (1) acquired symmetrical, sensorineural, and mild-to-moderate hearing loss; (2) current use of a smartphone; and (3) no previous experience with hearing amplification. Mild and moderate hearing loss ranged from the 26- to 40-db hearing level and from the 41- to 55-dB hearing level, respectively (according to the American Speech-Language-Hearing Association) [25], on pure-tone averages of 0.5, 1, 2, and 4 kHz. Patients with communication problems, central nervous system diseases, lesions of the auditory nerve, or external and middle ear diseases or anomalies were excluded.

To identify sensorineural hearing loss, otoscopy, pure-tone audiometry (PTA) by air and bone conduction, and tympanometry were performed. All participants visited 4 times and all tests were conducted in a sound-treated room. At the first visit, demographic assessment, otoscopy, tympanogram, tinnitogram, PTA, and speech audiometry were conducted. At the second visit, the participants who passed the screening test were enrolled; otoscopy, sound-field unaided and aided PTA, and WRS testing were performed; and our questionnaire was completed. At this time, the participants were randomly given their first hearing amplification to use for 2 months. At the 2-month visit, the same tests were conducted except for unaided PTA and WRS testing. The hearing amplification method was then switched to be used for another 2 months. At the final visit, participants completed the same tests as in the previous visit. Hearing impairment (>20-dB change in 1 frequency, >10-dB change in 2 consecutive frequencies, or >5-dB change in 3 consecutive frequencies) and tinnitus (>1 mo for at least 8 h/d) were considered adverse events in this study.

Ethical Considerations

This study was registered in the clinical trials registry in South Korea (Clinical Research Information Service; KCT0005458) [26] and at ClinicalTrials.gov (NCT05644106). This study was approved by the Institutional Review Board of Seoul National University Hospital (No. 2003-028-1109), Seoul, South Korea. Informed consent was obtained from all participants.

Hearing Amplification Systems (HA and SHAA)

A pilot study that conducted behavioral evaluations of 3 SHAAs provided the evidence used to select an SHAA for this study [24]. We chose the Petralex SHAA (IT4YOU Corp LLC) because it was available for download on both Android and iOS phones and users showed a greater improvement in WRSs compared to the other apps [24]. Petralex provides various functions, including a hearing test, speech recognition, acoustic amplification, and dynamic compression. It can amplify sound up to 30 dB, and the gain can be adjusted based on the hearing test results. In addition, a noise reduction function reduces background noise and focuses on speech recognition.

For the conventional HA, we used the Siya 1 miniRITE (Oticon A/S), coupled with two 85-dB receivers and single closed ear tips. The Siya 1 miniRITE is an essential-level HA, first introduced in 2018 with 48 channels and 10 adjustable bands. The Siya 1 miniRITE has basic modern digital HA technology including advanced feedback management, noise reduction, and multiband directionality. During the study, special functions of the Siya 1 miniRITE were activated, including instantaneous noise management, binaural bandwidth processing, and a feedback shield. However, noise reduction was not activated. All participants used each method, alternating at 2 months. Before the study, all measurement materials were calibrated.

Sound-Field Audiometry and WRS Tests

Sound-field audiometry (SFA) and WRS tests were performed using a calibrated Interacoustics AC40 (Interacoustics A/S). To determine the difference between the SHAA and the conventional HA, SFA was obtained at 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz for both hearing amplification systems. Warble tone was used to avoid standing waves. The WRS was evaluated using the Korean Standard Monosyllabic Word List for Adults [27], which comprised 25 monosyllabic words per condition and was presented at the most comfortable sound level. Lastly, the mean and SD of the WRS were calculated under unaided and aided conditions to measure the improvement. The WRS test in an aided condition was presented at a hearing level of 50 dB.

REM Test

This test measures actual HA gain by comparing the difference between the HA target value and the sound pressure measured using a microphone probe tip in the ear canal. We used Affinity 2.0 software (version 2.6.0; Serif Ltd) to conduct the REMs. To validate the HA performance, the REMs were performed using speech stimuli divided into a soft level (55-dB sound pressure level [SPL]), a medium level (65-dB SPL), and a loud level (75-dB SPL) based on the National Acoustic Laboratory–NL2 prescription. For the SHAA, the gain control and parameters were modified manually to match the targets.

Hearing-in-Noise Test

The hearing-in-noise test (HINT) was developed to measure binaural speech recognition ability under quiet and noisy situations [28]. We used the Korean version of the HINT, which is composed of 12 lists with 20 sentences per list [29]. Tests were performed under 4 conditions: quiet, front noise, left noise, and right noise. The participants were positioned 1 m in front of the speaker, and the speech or noise was first presented at a 65-dB SPL (0-dB signal-to-noise ratio [SNR]) through the speaker. If the participant responded correctly, we adjusted the level of the subsequent sentence using the transformed up-down methods recommended by Levitt [30] in 1971. When the listener achieved 50% of correct answers, the SNR was fixed at the corresponding level. These measurements were conducted using both hearing amplification methods.

Questionnaires

For the measurement of subjective satisfaction, the APHAB and IOI-HA were used. The APHAB consists of 4 items: ease of communication, reverberation, background noise, and aversiveness of sounds. Each item has 6 questions, and higher scores indicate greater disability. This survey can be used to evaluate the performance of an HA in unaided and aided conditions. The IOI-HA is a self-estimation tool designed to check the efficiency of HA fitting [31]. It contains 7 items including use of the HA, benefit of the HA, residual activity limitations, satisfaction, residual participation restrictions, impact on others, and quality of life.

Statistical Analysis

A sample size of 26 would achieve 90% power to detect noninferiority with a 1-sided significance level of .025, and a cross-over design would have a margin of noninferiority of −10%, a true mean difference of 0, and an SD of the paired differences of 15, based on the test results from a previous study for understanding speech in noise [32]. Therefore, we aimed to enroll 33 participants with the anticipation of a 20% dropout rate. Although the noninferiority margin for otolaryngological interventions has not been established, a noninferiority margin of 10% is commonly used for drugs intended for treatment. Continuous variables were expressed as mean (SD), and all statistical analyses were performed using SPSS (version 25.0; IBM Corp). The 2-tailed *t* test was used to compare continuous variables, and one-way ANOVA with the Scheffé post hoc test was used to compare the questionnaire scores and HINT performances of 3 groups, including in unaided conditions. One-way ANOVA *P* values of <.05 were considered to indicate statistical significance.

Results

Participants

A total of 63 participants were screened; 8 participants were excluded due to withdrawal of consent and greater than mild-to-moderate hearing loss. During the 2-month follow-up, 10 participants in group A dropped out for personal reasons and 7 participants in group B dropped out because they withdrew consent. The final analysis was performed on 38 participants (Figure 1). REMs and the HINT were performed in 25 (66%) out of 38 participants. The mean age of the study participants was 66 (range 41-84) years, and 42% (16/38) were male. Most participants were without tinnitus (22/38, 58%), and the degree of hearing loss was mild (23/38, 61%) to moderate (15/38, 39%). The characteristics of the participants are shown in Table 1. Although SHAAs may have potential adverse effects, such as tinnitus and discomfort, because patients can increase the SHAA volume to a risky sound level [33], there were no serious complications (hearing impairment or tinnitus) in either group during this study.

Figure 1. Flow diagram of the study design. HA: hearing aid; SHAA: smartphone-based hearing aid app.

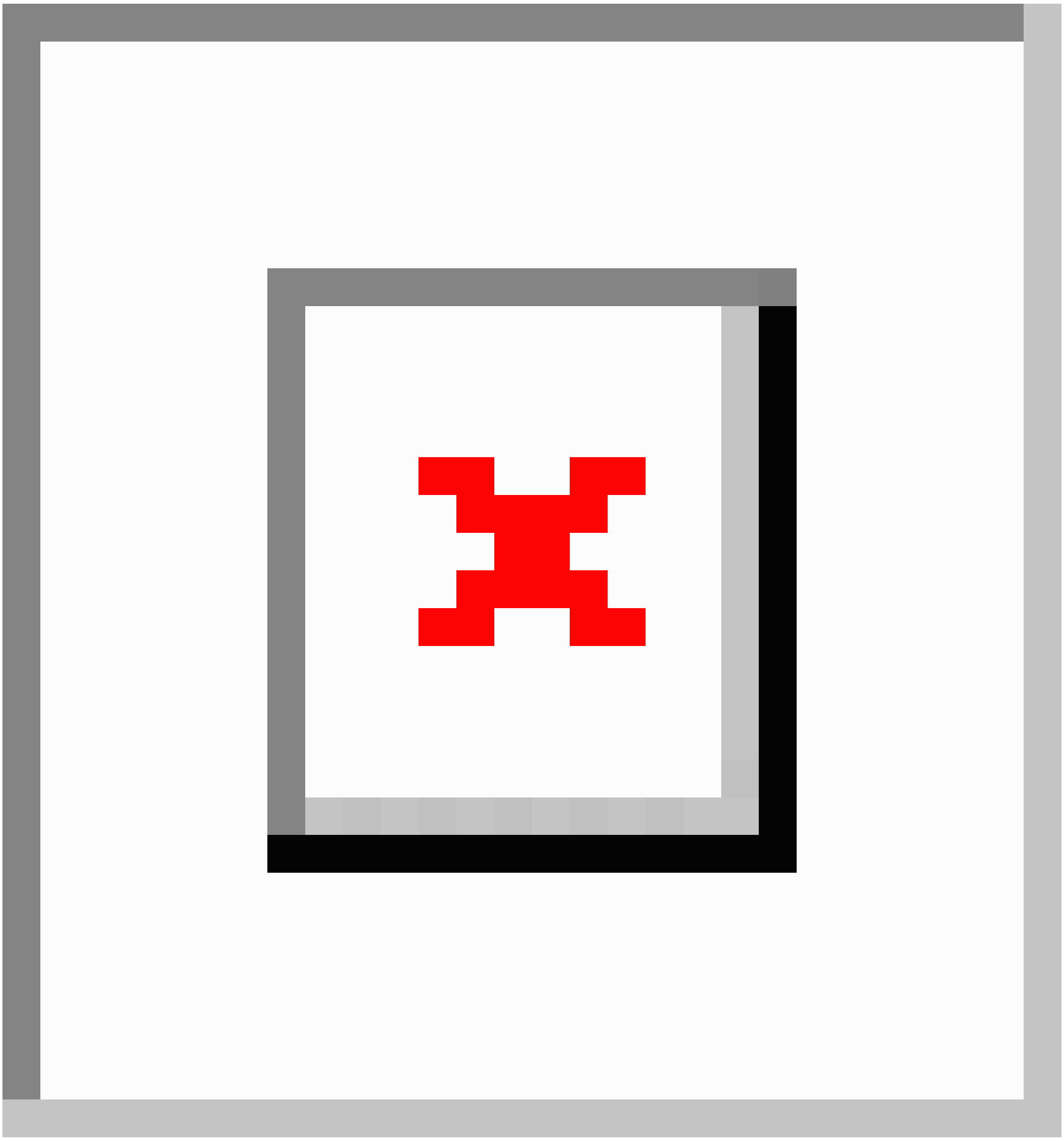


Table . Characteristics of the participants.

Characteristics		Values (n=38)
Age (years), mean (SD)		66 (12)
Sex, n (%)		
	Male	16 (42)
	Female	22 (58)
Tinnitus, n (%)		
	Yes	16 (42)
	No	22 (58)
Hearing loss (dB HL[a])		
	4-frequency average[b], mean (SD)	39 (7)
	26-40, n (%)	23 (61)
	41-55, n (%)	15 (39)
Word recognition score, mean (SD)		66 (20)

[a]dB HL: decibel of hearing level.

[b]4 frequency averages: 500, 1000, 2000, and 4000 Hz.

SFA and WRS Results

The SHAA showed a slightly lower gain at all frequencies than the conventional HA, with statistical significance at 1000 Hz (*P*=.001), 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), 4000 Hz (*P*<.001), 6000 Hz (*P*=.001), and 8000 Hz (*P*=.04; Figure 2). The mean improvement in WRS was −2.6 (SD 18.3) and 16.0 (SD 12.8) for the SHAA and HA groups, respectively (*P*<.001). Most participants (24/38, 63%) showed no improvement with the SHAA compared to unaided conditions (Figure 3).

Figure 2. Sound-field (warble-tone) audiometry thresholds in the HA and SHAA groups. Error bars indicate the SD. db HL: decibel of hearing level; HA: hearing aid; SHAA: smartphone-based hearing aid app. **P*<.05.

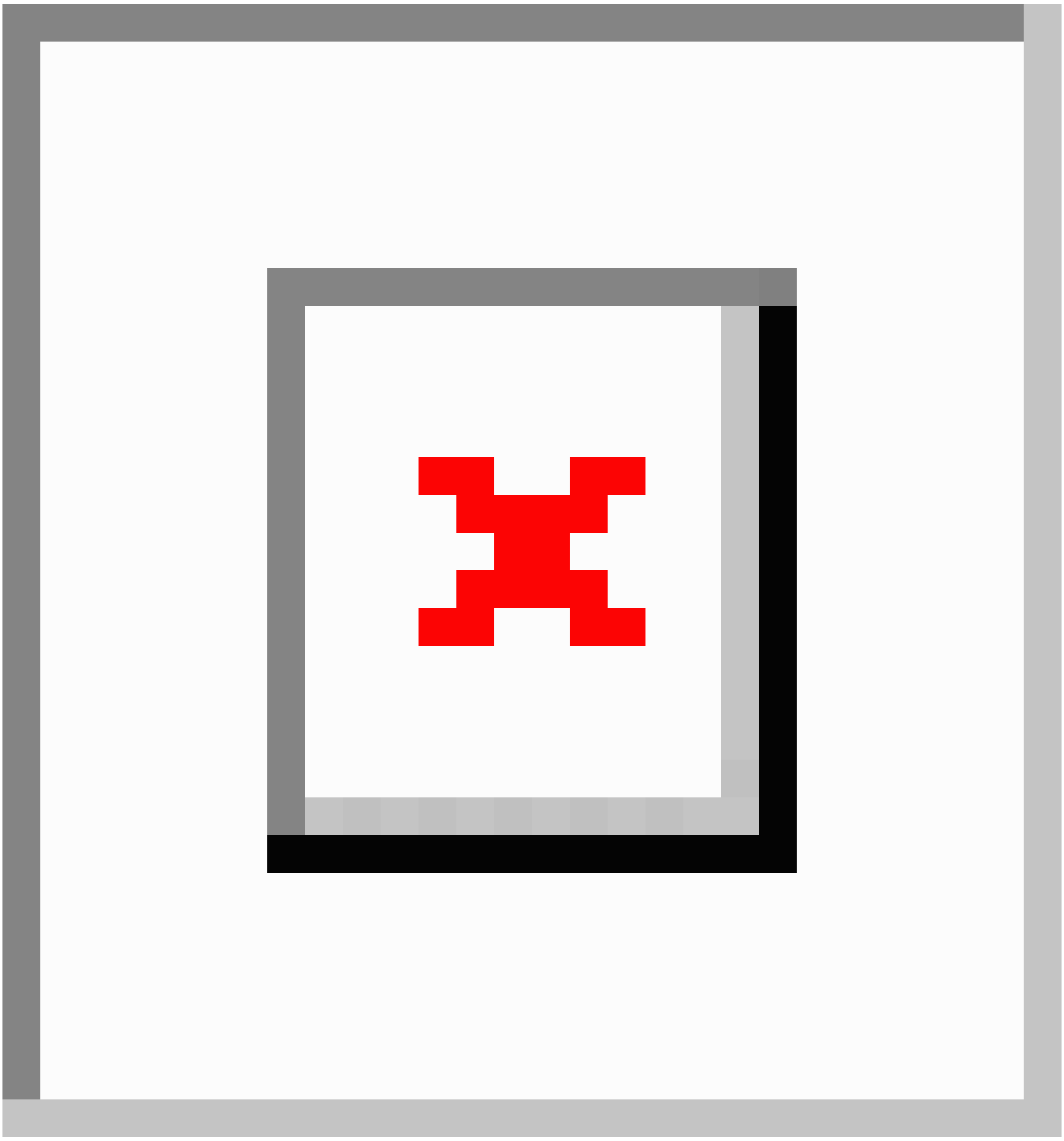
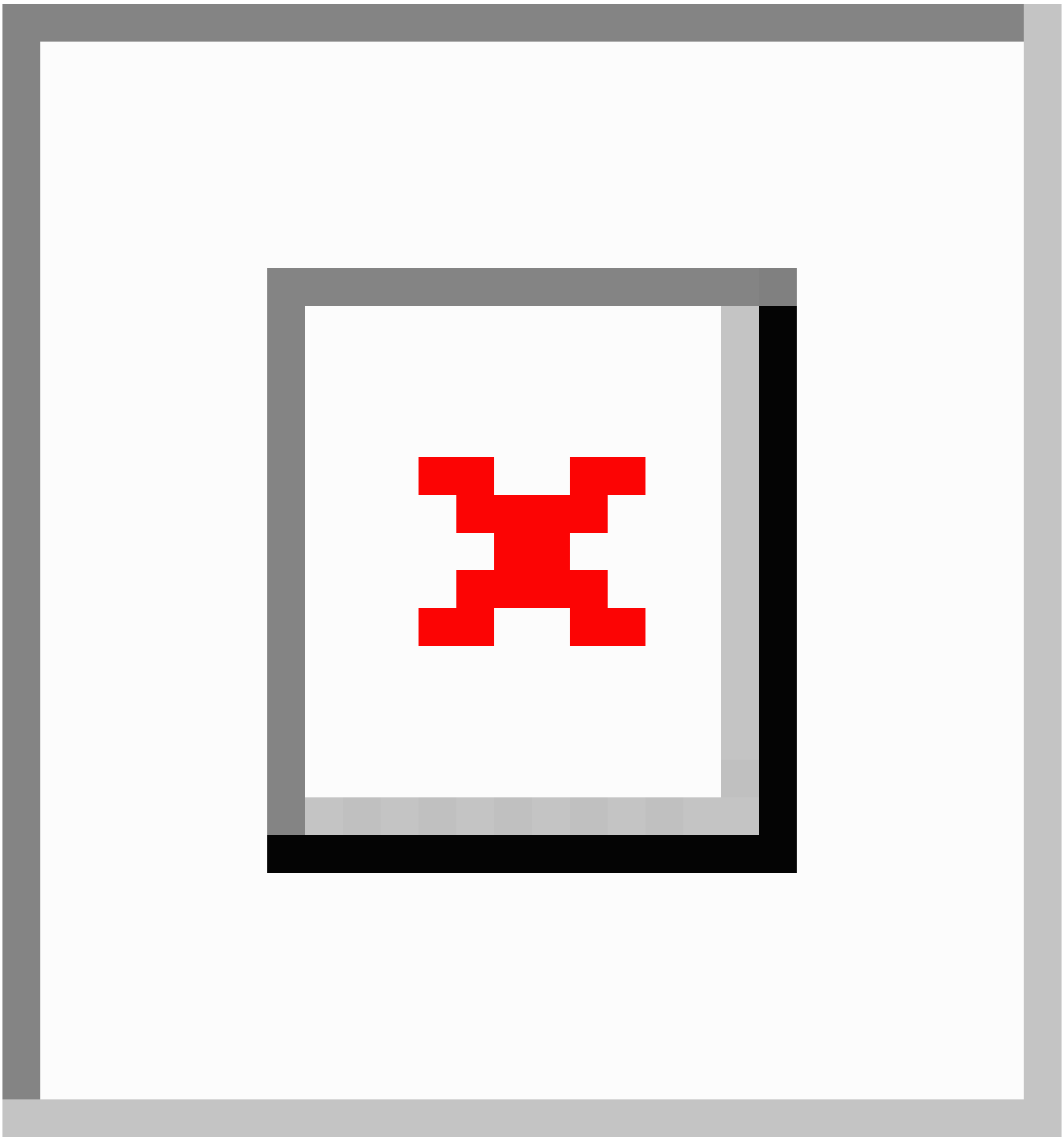


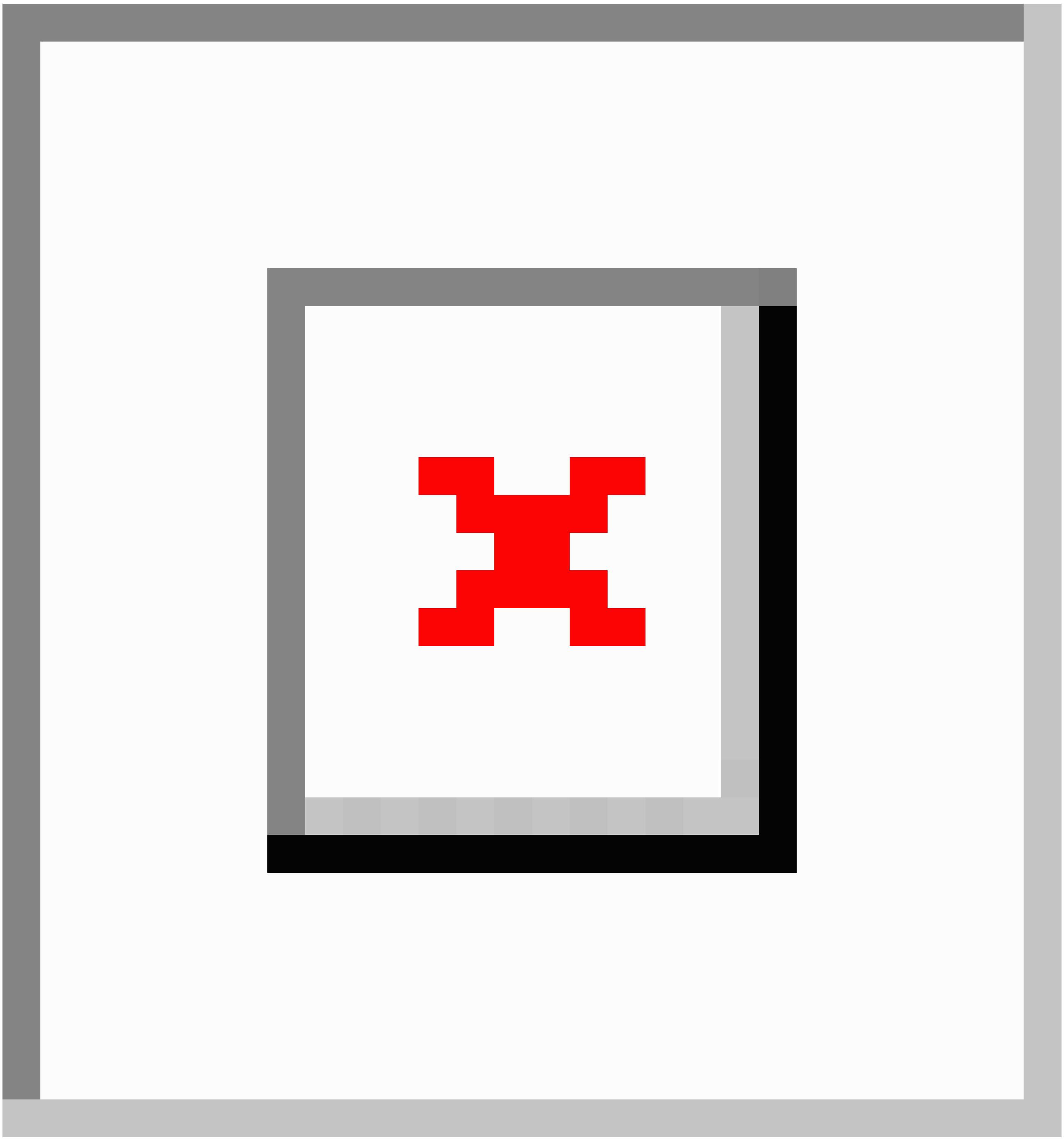
Figure 3. Improvements in WRS, comparing unaided and aided conditions at 50-dB hearing level stimuli. Error bars indicate the SD. SHAA: smartphone-based hearing aid app; WRS: word recognition score.



REM Results

In real-ear aided responses at 55 dB, the HA showed a larger gain at 1000, 2000, 3000, 4000, and 6000 Hz than the SHAA. The SHAA showed a larger gain at low frequencies (250 and 500 Hz) than the HA. Statistical significance was found at 500 Hz (*P*=.01), 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), and 4000 Hz (*P*=.001) in the right ear and at 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), and 4000 Hz (*P*<.001) in the left ear. At 65 dB, the HA revealed greater gains at 1000, 2000, 3000, 4000, and 6000 Hz than the SHAA. Statistical significance was found at 500 Hz (*P*=.02), 1000 Hz (*P*=.01), 2000 Hz (*P*<.001), 3000 Hz (*P*=.001), and 4000 Hz (*P*=.002) in the right ear and at 1000 Hz (*P*<.001), 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), and 4000 Hz (*P*=.002) in the left ear. At 75 dB, the HAs showed a larger gain at 1000, 2000, 3000, and 4000 Hz than the SHAA. As with the results at 65 db, statistical significance was found at 500 Hz (*P*=.03), 1000 Hz (*P*=.007), 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), and 4000 Hz (*P*=.004) in the right ear and at 1000 Hz (*P*<.001), 2000 Hz (*P*<.001), 3000 Hz (*P*<.001), and 4000 Hz (*P*=.001) in the left ear. At all speech levels, the SHAA showed greater gain at low frequencies (250 and 500 Hz) than the HA (Figure 4).

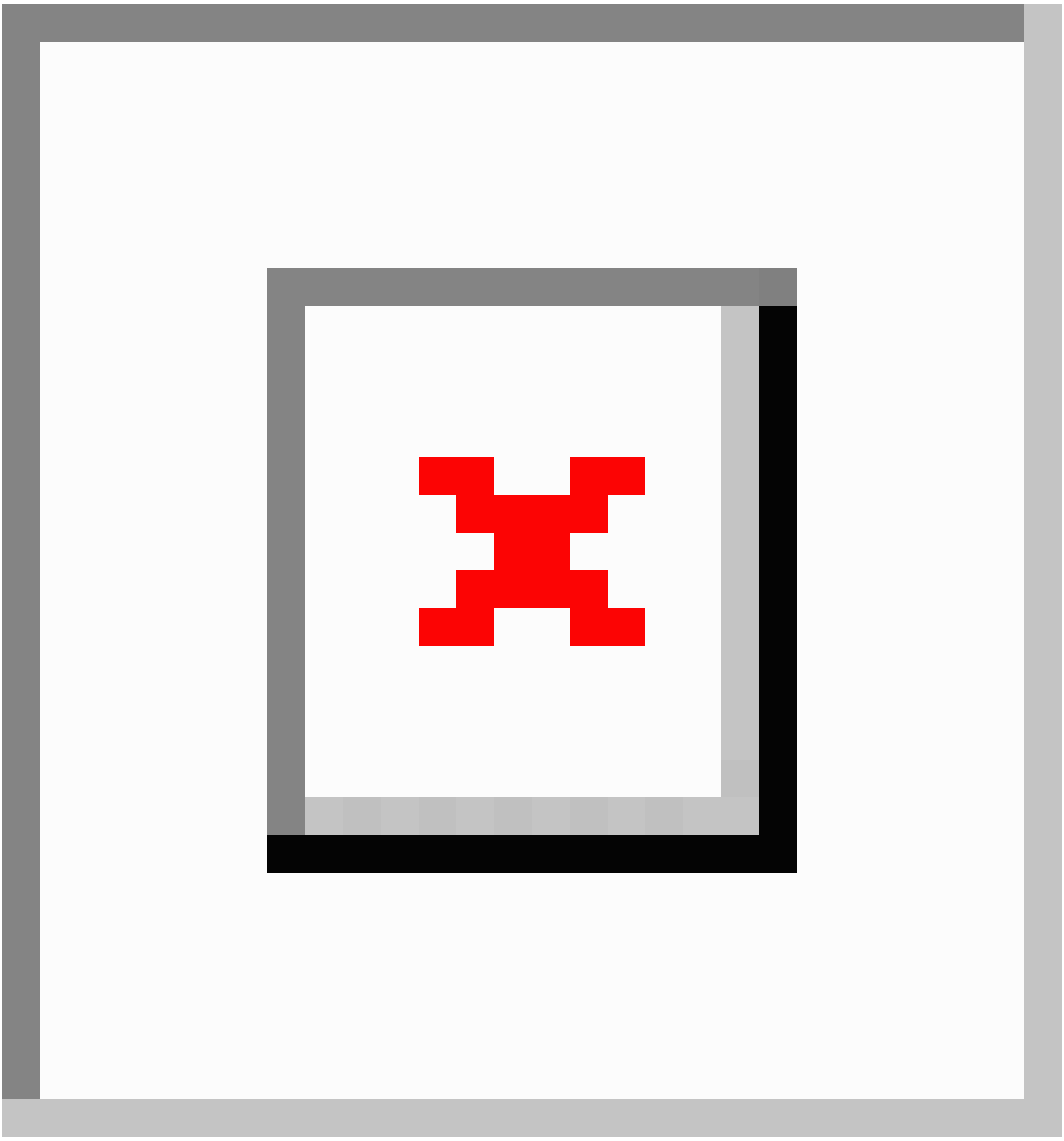
Figure 4. Real-ear aided responses (REARs) of the hearing aid and SHAA groups measured at the 55-, 65-, and 75-dB sound pressure levels (SPLs): average REAR in the right ear at the (A) 55-, (B) 65-, and (C) 75-dB SPLs and average REAR in the left ear at the (D) 55-, (E), 65-, and (F) 75-dB SPLs. Error bars indicate the SD. SHAA: smartphone-based hearing aid app. *$P<.05$.



HINT Results

The mean SNRs (HINT performance) in the unaided, HA, and SHAA groups were 51.3 (SD 5.6), 47 (SD 3.4), and 56 (SD 6.4) dB, respectively, in a quiet situation ($F_{2,72}$=17.164; P<.001). The Scheffé post hoc analysis showed a statistically significant difference between the unaided and SHAA groups (P=.006), the unaided and HA groups (P=.046), and the SHAA and HA groups (P<.001). The mean SNR (performance) was −0.1 (SD 2.1), –0.3 (SD 2.1), and 2.4 (SD 2.8) dB in a front noise situation and −1.7 (SD 3.4), –1.6 (SD 2.7), and 1.4 (SD 3.6) dB in a right noise situation ($F_{2,72}$=7.110; P=.002) for the unaided, HA, and SHAA groups, respectively. In the Scheffé post hoc analysis, there were statistically significant differences between the HA and SHAA groups (P=.009) and between the unaided and SHAA groups (P=.006; Figure 5). The mean SNR (performance) was −3.6 (SD 3.0), –4.5 (SD 3.6), and −0.8 (SD 4.1) dB in a left noise situation for the unaided, HA, and SHAA groups, respectively, showing statistical significance ($F_{2,72}$=7.008; P=.002)

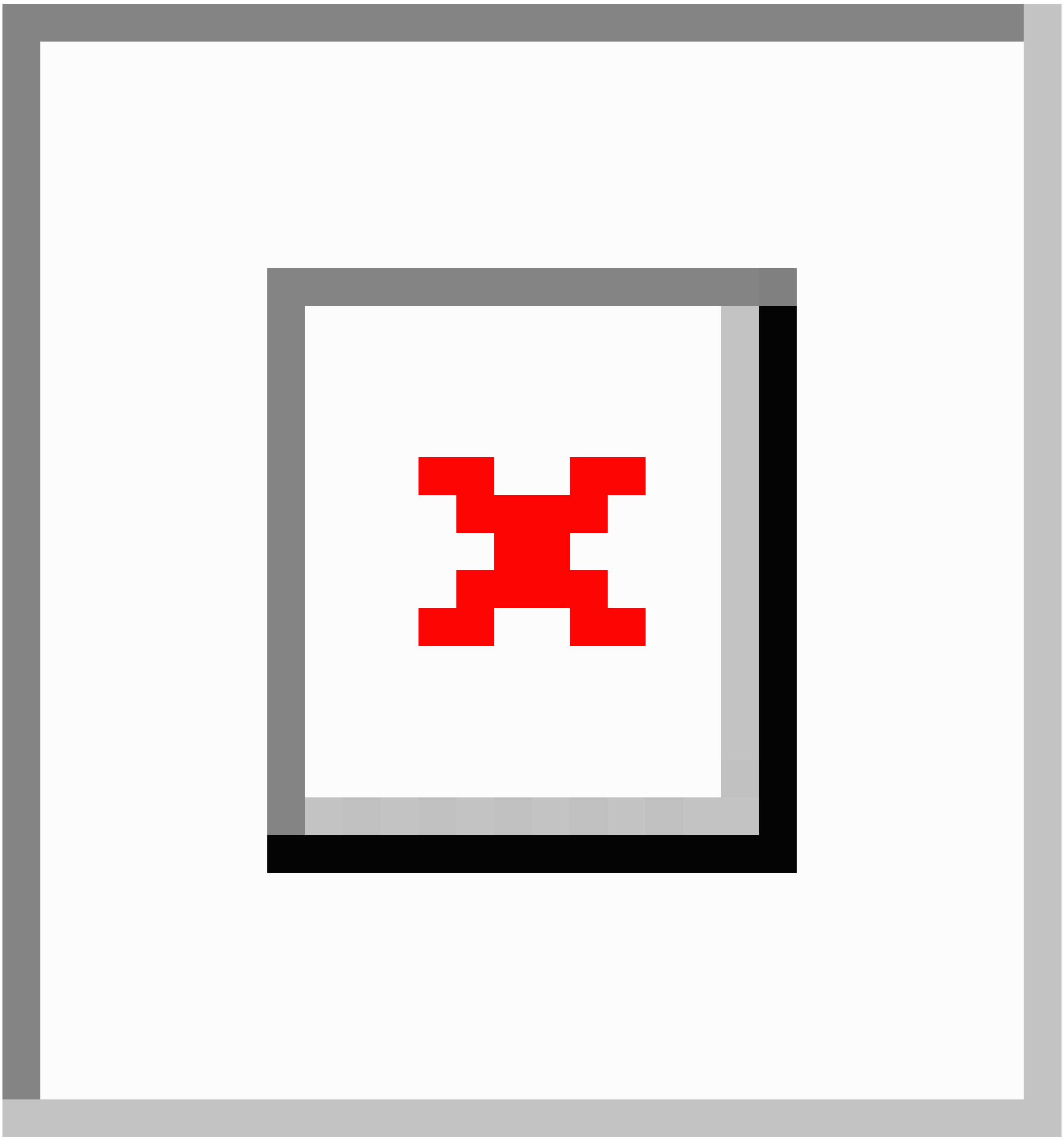
Figure 5. Hearing-in-noise test (HINT) in quiet, front noise, left noise, and right noise conditions for the unaided, hearing aid, and SHAA groups. Error bars indicate the SD. RTS: reception threshold for speech; SHAA: smartphone-based hearing aid app; SNR: signal-to-noise ratio.



Questionnaire Results

The mean score of the IOI-HA was 7 (SD 0), 23.4 (SD 4.0), and 16.3 (SD 4.9) in the unaided, HA, and SHAA groups, respectively ($F_{2,111}$=187.469; *P*<.001). The Scheffé post hoc analysis showed statistically significant differences between all groups (*P*<.001). The HA group scored the highest, followed by the SHAA and unaided groups (Figure 6).

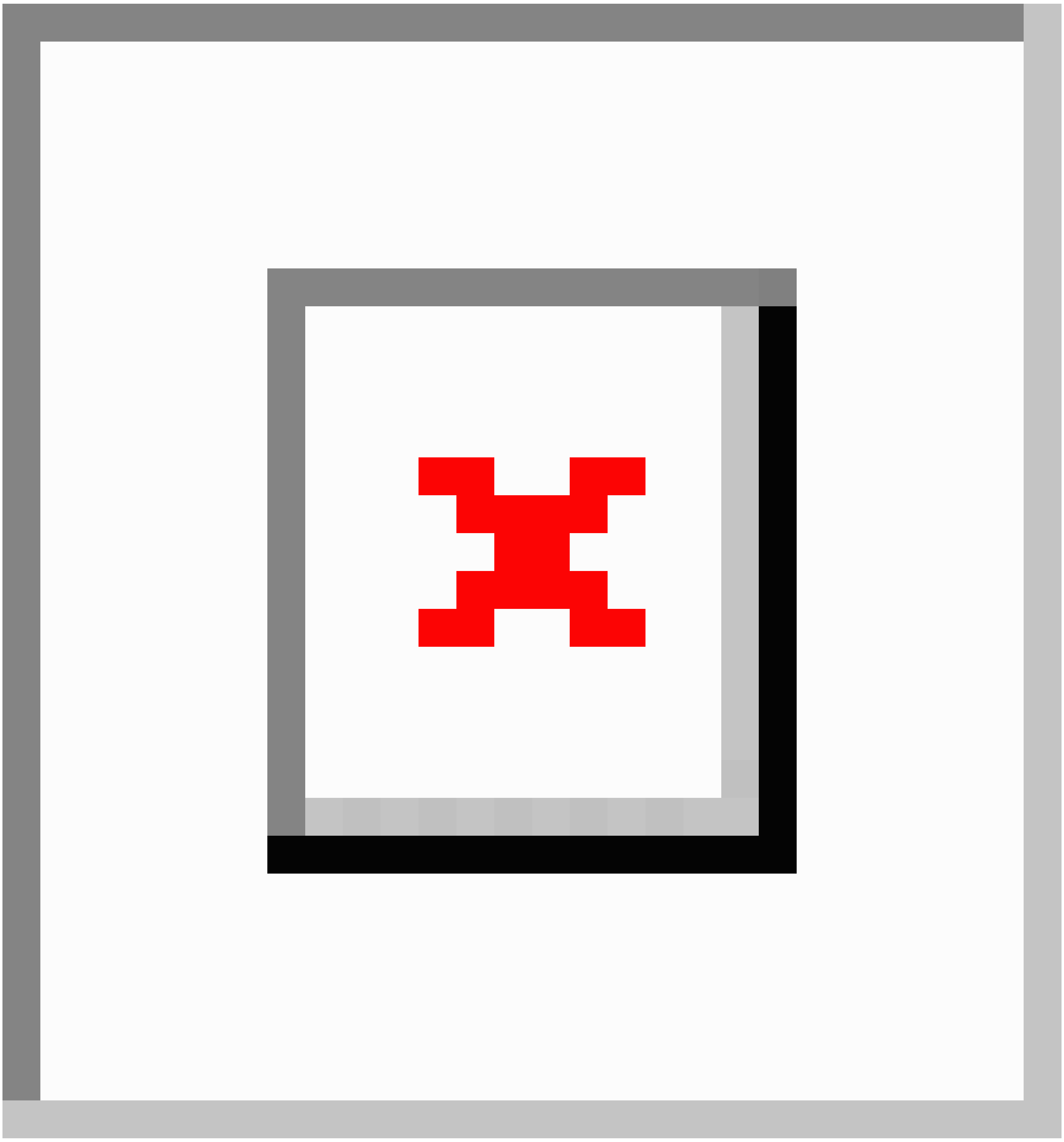
Figure 6. International Outcome Inventory for Hearing Aids (IOI-HA) items for the unaided, hearing aid, and SHAA groups. Error bars represent the SD. SHAA: smartphone-based hearing aid app. *P<.05.



In the ease of communication subscale item of the APHAB, the mean scores were 45.6 (SD 20.0), 51.2 (SD 20.9), and 33.2 (SD 18.0) in the unaided, SHAA, and HA groups, respectively ($F_{2,111}$=8.009; P=.001). The Scheffé post hoc analysis showed statistically significant differences between the unaided and HA groups (P=.03) and between the HA and SHAA groups (P=.001). The HA group had the lowest scores, followed by the unaided and SHAA groups. In the reverberation subscale item of the APHAB, the mean scores were 48.0 (SD 17.6), 53.8 (SD 13.2), and 43.0 (SD 14.8) in the unaided, SHAA, and HA groups, respectively ($F_{2,111}$=5.525; P=.006). The Scheffé post hoc analysis indicated that the mean score of the HA group was significantly lower than that of the SHAA group (P=.01). In the background noise subscale item of the APHAB, the mean scores were 43.1 (SD 17.7), 53.7 (SD 14.7), and 41.3 (SD 17.4) in the unaided, SHAA, and HA groups, respectively ($F_{2,111}$=6.010; P=.003). After the Scheffé post hoc analysis, significant differences were observed between the unaided and SHAA groups (P=.03) and between the HA and SHAA groups (P=.007). The SHAA group had the highest scores, followed by the unaided and HA groups. In the aversiveness of sound subscale of the APHAB, the mean scores were 42.8 (SD 17.4), 52.0 (SD 19.3), and 62.0 (SD 20.0) in the unaided, SHAA, and HA groups, respectively ($F_{2,111}$=9.509; P<.001). The Scheffé post hoc analysis indicated that the mean score of the unaided group was significantly lower than that of the HA group

(P<.001). In particular, the SHAA group's score was lower than that of the HA group, but without statistical significance (P=.08). Lastly, the mean global scores were 45.6 (SD 16.4), 52.9 (SD 14.1), and 39.1 (SD 14.7) in the unaided, SHAA, and HA groups, respectively ($F_{2,111}$=7.689; P=.001). The Scheffé post hoc analysis showed a significant difference between the HA and SHAA groups (P<.001; Figure 7).

Figure 7. Mean score of the Abbreviated Profile of Hearing Aid Benefit (APHAB) items for the unaided, hearing aid, and smartphone-based hearing aid app (SHAA) groups. Error bars indicate the SD. AV: aversiveness of sounds; BN: background noise; EC: ease of communication; RV: reverberation. *P<.05.



A comparison of individuals in the SHAA group between unaided and aided conditions found that 37% (14/38), 42% (16/38), 24% (9/38), and 32% (12/38) of users showed improvements in the ease of communication, reverberation, background noise, and aversiveness of sound APHAB subscales, respectively. In the same subscales of the APHAB, the mean postintervention scores were 43.1 (SD 24.0), 50.2 (SD 10.8), 45.8 (SD 8.0), and 35.4 (SD 16.1), respectively (Table 2).

Table . Mean scores of the Abbreviated Profile of Hearing Aid Benefit (APHAB) items for the unaided and aided conditions in the SHAA group, with improvements in ease of communication, reverberation, background noise, and aversiveness of sound.

Subscale	Participants (n=38), n (%)	Pretest score, mean (SD)	Posttest score, mean (SD)
Ease of communication	14 (37)	57.1 (20.4)	43.1 (24.0)
Reverberation	16 (42)	60.6 (9.6)	45.8 (8.0)
Background noise	9 (24)	57.6 (8.2)	50.2 (10.8)
Aversiveness of sound	12 (32)	53.3 (9.4)	35.4 (16.1)

Discussion

Principal Findings

This study used a noninferiority study design to compare the audiological performance between an SHAA and a conventional HA. The SHAA did not show noninferiority to the HA in the WRS test, which was the primary measurement used in the original study. However, the SHAA showed feasibility in some users, especially in unaided conditions and for the amplification of soft sounds. In this study, 63 participants were screened and 38 finished the study. The objective audiological performance was evaluated using the warble-tone hearing test, the WRS test, REMs, and the HINT, and subjective satisfaction was evaluated using the APHAB and IOI-HA questionnaires. The aided hearing thresholds for the SHAA were 20-60 dB, from low to high frequencies. Although the SHAA group showed a lower gain at the 1000- to 8000-Hz levels than the HA group, it showed gains similar to those of the HA group at low frequencies. The REMs results at 75 dB, as well as the common speaking levels of 55 and 65 dB, indicated that the HA group showed greater gains in the middle-to-high frequencies than the SHAA group. One-third (14/38, 37%) of the SHAA group showed slight improvements in WRS. However, most SHAA users (24/38, 63%) showed no difference in their WRS compared to the unaided group. The HA group showed significantly better audiological performance than the unaided and SHAA groups in all HINT situations. The HA group showed the highest scores in the IOI-HA, but the SHAA users also showed significantly greater improvement than the unaided group. A comparison of individuals in the SHAA group between unaided and aided conditions found that 37% (14/38), 42% (16/38), 24% (9/38), and 32% (12/38) of users showed improvements in the ease of communication, reverberation, background, and aversiveness of sound subscales of the APHAB, respectively.

Comparisons With Previous Research

de Sousa et al [34] investigated the electroacoustic and self-reported performance of SHAAs in the Google Play and Apple App stores. The Petralex app in the Android Samsung S6 showed a 3.9-dB improvement in the SNR at full noise reduction. The SNR became worse with no noise suppression. Of 5 participants, 4 agreed that conversations were easy to follow while using the app. However, all 5 participants preferred the HA to the SHAA. The study suggested that Petralex showed the best performance in noise reduction. de Sousa et al [34] also evaluated the subjective listening experience of participants using the Petralex app. However, they used only the electroacoustic analysis to evaluate noise reduction and investigated the subjective listening experience of only 5 participants.

Amlani et al [23] reported that SHAAs improved the perceived benefits and reduced the perceived barriers in first-time users of amplification and those who had quit using HAs. They conducted a survey before and after SHAA use. However, their study did not evaluate hearing performance with hearing tests, and there was no control group.

Reed et al [32] compared 5 personal sound amplification products (PSAPs) to conventional HAs for understanding speech in noise. Three of the PSAPs improved speech understanding for individuals with hearing loss, similar to the results obtained with HAs, whereas 1 product demonstrated little improvement and 1 product degraded speech understanding. This study suggested that the alternative devices to HAs, which are more affordable and easier to access, can improve speech understanding. Several studies have compared PSAPs to conventional HAs with varying results. However, only a few studies of SHAAs have been conducted.

Lin et al [35] reported that smartphone-bundled earphones can be used as PSAPs for mild-to-moderate hearing loss. They used AirPods Pro (Apple Inc) and compared it to 5 other PSAPs. In quiet conditions, there was no difference in speech perception between the AirPods Pro and the HAs, but they differed in noisy conditions. That study did not evaluate the performance of SHAAs and did not include a subjective satisfaction evaluation. Most studies, including this study, have suggested that HAs lead to the best performance in speech perception in noisy conditions.

Our previous study compared the electroacoustic characteristics of SHAAs to those of HAs and found that only a few performed reliably [19]. In another study, we evaluated the behavioral performance of a selection of currently available SHAAs in patients with mild hearing loss. However, it was an exploratory pilot study with only 7 participants [24].

Martinez-Beneyto et al [36] studied whether audiological smartphone apps can improve audiological performance in groups without hearing loss or groups with various grades of hearing loss. Most participants showed improvement in PTA and word recognition. They used the Apple iPhone 6S iOS (version 10.1.1) and the Sony MDR-EX15LP in-ear headphones. In our study, we used Android devices, and the audiological performance of the SHAA showed a significant difference according to the smartphone device and bundled earphone. Furthermore, that study was conducted among patients with mild-to-severe hearing loss. Although this study included only patients with mild-to-moderate hearing loss, we think that the findings could be generalized to those with more severe hearing loss. A comprehensive comparison with currently used methods is summarized in Table 3.

Table . Comparison with currently used methods.

Study	Purpose	Type of amplification	Test measure	Key findings
de Sousa et al [34]	Investigate electroacoustic and self-reported performance through various apps and smartphone manufacturers	• 4 SHAAs[a] (Petralex, Super Ear, Earshot, and Hearing Aid Master) on Android devices • 4 SHAAs (Petralex, Fennex, Mobile Ears, and Super Hearing Aid) on iPhone devices	Objective sound quality (latency and SNR[b]) and subjective listening experience	• Petralex and Fennex on iPhone 6 showed the shortest latency and highest SNR improvement • All SHAAs showed longer latency and lower SNR improvement on Android devices • Participants preferred using HAs[c] over SHAAs
Amlani et al [23]	Determine the feasibility of improving attitude during 4 weeks with an SHAA	• 1 SHAA (Ear Machine) with iPod Touch	N/A[d]	• The SHAA can modify the perception toward amplification and reduce the perceived barriers in first-time users of amplification
Reed et al [32]	Compare PSAPs[e] with a conventional HA for mild-to-moderate hearing loss	• 5 PSAPs (Sound World Solutions CS50+, Soundhawk, Etymotic BEAN, Tweak Focus, and MSA 30X Sound Amplifier) and 1 HA (Oticon Nera2)	AZBio sentence-in-noise task	• 3 PSAPs improved speech understanding, similar to the results with the HA • 1 PSAP showed worse speech understanding than the unaided condition
Lin et al [35]	Examine electroacoustic properties of AirPods and compare hearing performance for mild-to-moderate hearing loss	• 5 PSAPs, 2 HAs (Oticon Opn 1 and Benafon MD 1), AirPods Pro, and AirPods 2	Mandarin HINT[f]	• AirPod Pro met 4 PSAP standards • No difference in speech perception between the AirPod Pro and HA in quiet conditions
Koo et al [24]	Evaluate the behavioral performance of 3 SHAAs	• 3 SHAAs (Ear Machine, Sound Amplifier, and Petralex)	REM[g], warble-tone audiometry, WRS[h], and the HINT (quiet and noise-front conditions)	• HAs showed greater gain than SHAAs at 2 and 3 kHz in sound-field audiometry • 6% showed improvement with Petralex in WRS • No improvement was found with SHAAs in the HINT • Some SHAAs were beneficial for patients with mild-to-moderate hearing loss
Martinez-Beneyto et al [36]	Assess whether SHAAs can improve audiological performance in patients without hearing loss or those with varying grades of hearing loss	• 1 SHAA (Petralex) with iPhone 6S	PTA[i], WRS in quiet and noisy conditions, and a questionnaire	• SHAA can improve word recognition and PTA • 61% answered "good" or "excellent" for the app sound quality

[a]SHAA: smartphone-based hearing aid app.
[b]SNR: signal-to-noise ratio.
[c]HA: hearing aid.
[d]N/A: not applicable.

[e]PSAP: personal sound amplification product.
[f]HINT: hearing-in-noise test.
[g]REM: real-ear measurement.
[h]WRS: word recognition score.
[i]PTA: pure-tone audiometry.

Strengths and Limitations

To the best of our knowledge, this is the first prospective, multicenter randomized controlled trial comparing an SHAA and a conventional HA. Hearing performance was evaluated for comprehensive objective functional gain and for subjective satisfaction.

In this study, we enrolled Android users who used their own smartphones. Although the performance varies depending on which app is used on different smartphone models, de Sousa et al [34] reported that SHAAs on iPhone 6 and iOS with wired earphones showed better SNR improvement and sound quality than SHAAs on Samsung S7. However, newly introduced devices and later versions of Android may show better amplification performance. Furthermore, Nguyen et al [19] found that most apps in iOS provided better electroacoustic performance than the corresponding Android versions. Another limitation of our study is a high dropout rate because of the COVID-19 pandemic, which hindered many patients from visiting. Finally, audiological performance was evaluated only after a 2-month trial, and 2 months may not be enough time to fully evaluate and clarify the long-term performance of HAs and SHAAs. Longer-term research is needed in the future.

Conclusion

In our study, the SHAA demonstrated a significant benefit when compared to an unaided situation. Our results indicate that the SHAA could be a useful assistive device for patients with mild-to-moderate hearing loss in quiet conditions. In addition, people with poor access to hearing amplification devices can more easily download an SHAA at a more reasonable price than a conventional HA. Although, the SHAA performed poorly when compared with the HA at a conversational sound level and in noisy conditions, the future signal technology of SHAAs should improve and perform better in noisy environments.

Acknowledgments

This research was supported by a grant from the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant HC19C0128).

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 18290 KB - mhealth_v11i1e46911_app1.pdf]

References

1. World report on hearing. World Health Organization. 2021. URL: www.who.int/publications/i/item/9789240020481 [accessed 2023-09-13]
2. Howarth A, Shone GR. Ageing and the auditory system. Postgrad Med J 2006 Mar;82(965):166-171. [doi: 10.1136/pgmj.2005.039388] [Medline: 16517797]
3. GBD 2019 Hearing Loss Collaborators. Hearing loss prevalence and years lived with disability, 1990-2019: findings from the Global Burden of Disease Study 2019. Lancet 2021 Mar 13;397(10278):S0140-6736(21)00516-X. [doi: 10.1016/S0140-6736(21)00516-X] [Medline: 33714390]
4. Kotby MN, Tawfik S, Aziz A, Taha H. Public health impact of hearing impairment and disability. Folia Phoniatr Logop 2008;60(2):58-63. [doi: 10.1159/000114646] [Medline: 18235192]
5. Most T. The effects of degree and type of hearing loss on children's performance in class. Deaf Educ Int 2004 Oct;6(3):154-166. [doi: 10.1179/146431504790560528]
6. Tucci D, Merson MH, Wilson BS. A summary of the literature on global hearing impairment: current status and priorities for action. Otol Neurotol 2010 Jan;31(1):31-41. [doi: 10.1097/mao.0b013e3181c0eaec] [Medline: 20050266]
7. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 8;396(10248):413-446. [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
8. Deafeness and hearing loss. World Health Organization. 2023 Feb 27. URL: www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2023-02-10]
9. Ferguson M, Henshaw H. How does auditory training work? Joined-up thinking and listening. Semin Hear 2015 Nov;36(4):237-249. [doi: 10.1055/s-0035-1564456] [Medline: 27587911]

10. Mulrow CD, Aguilar C, Endicott JE, Tuley MR, Velez R, Charlip WS, et al. Quality-of-life changes and hearing impairment. A randomized trial. Ann Intern Med 1990 Aug 1;113(3):188-194. [doi: 10.7326/0003-4819-113-3-188] [Medline: 2197909]
11. Ferguson MA, Kitterick PT, Chong LY, Edmondson-Jones M, Barker F, Hoare DJ. Hearing aids for mild to moderate hearing loss in adults. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD012023. [doi: 10.1002/14651858.CD012023.pub2] [Medline: 28944461]
12. Chisolm TH, Johnson CE, Danhauer JL, Portz LJP, Abrams HB, Lesner S, et al. A systematic review of health-related quality of life and hearing aids: final report of the American Academy of Audiology Task Force On the Health-Related Quality of Life Benefits of Amplification in Adults. J Am Acad Audiol 2007 Feb;18(2):151-183. [doi: 10.3766/jaaa.18.2.7] [Medline: 17402301]
13. Orji A, Kamenov K, Dirac M, Davis A, Chadha S, Vos T. Global and regional needs, unmet needs and access to hearing aids. Int J Audiol 2020 Mar;59(3):166-172. [doi: 10.1080/14992027.2020.1721577] [Medline: 32011190]
14. Wallhagen MI. The stigma of hearing loss. Gerontologist 2010 Feb;50(1):66-75. [doi: 10.1093/geront/gnp107] [Medline: 19592638]
15. Meyer C, Hickson L. What factors influence help-seeking for hearing impairment and hearing aid adoption in older adults? Int J Audiol 2012 Feb;51(2):66-74. [doi: 10.3109/14992027.2011.611178] [Medline: 22256793]
16. Byun H, Kim EM, Kim I, Lee SH, Chung JH. The trend in adoption of hearing aids following changes in provision policy in South Korea. Sci Rep 2022 Aug 4;12(1):13389. [doi: 10.1038/s41598-022-17459-7] [Medline: 35927560]
17. Bernabe-Ortiz A, Diez-Canseco F, Vásquez A, Miranda JJ. Disability, caregiver's dependency and patterns of access to rehabilitation care: results from a national representative study in Peru. Disabil Rehabil 2016;38(6):582-588. [doi: 10.3109/09638288.2015.1051246] [Medline: 26017542]
18. FDA FINALIZES historic rule enabling access to over-the-counter hearing AIDS for millions of Americans. US Food and Drug Administration. 2022 Aug 16. URL: www.fda.gov/news-events/press-announcements/fda-finalizes-historic-rule-enabling-access-over-counter-hearing-aids-millions-americans [accessed 2023-09-1]
19. Nguyen W, Koo M, Lee JH, Oh SH, Park MK. Electroacoustic evaluation of smartphone-based hearing aid applications. Clin Exp Otorhinolaryngol 2022 May;15(2):135-143. [doi: 10.21053/ceo.2021.01004] [Medline: 35124945]
20. Amlani AM, Taylor B, Levy C, Robbins R. Utility of smartphone-based hearing aid applications as a substitute to traditional hearing AIDS. Hearing Review 2013 Sep;20(13):16-18 [FREE Full text]
21. Han JS, Park YH, Song JJ, Moon IJ, Lee W, Kim Y, et al. Knowledge and expectations of hearing aid apps among smartphone users and hearing professionals: cross-sectional survey. JMIR Mhealth Uhealth 2022 Jan 7;10(1):e27809. [doi: 10.2196/27809] [Medline: 34994699]
22. Han JS, Kim Y, Park MK, Song JJ, Moon IJ, Lee W, et al. Perception and expectations of personal sound amplification products in Korea: a hospital-based, multi-center, cross-sectional survey. PLoS One 2022 May 26;17(5):e0269123. [doi: 10.1371/journal.pone.0269123] [Medline: 35617328]
23. Amlani AM, Smaldino J, Hayes D, Taylor B, Gessling E. Feasibility of using a smartphone-based hearing aid application to improve attitudes toward amplification and hearing impairment. Am J Audiol 2019 Mar 15;28(1):125-136. [doi: 10.1044/2018_AJA-17-0068] [Medline: 30938556]
24. Koo M, Nguyen W, Lee JH, Oh SH, Park MK. Behavioral evaluation of 3 smartphone-based hearing aid apps for patients with mild hearing loss: an exploratory pilot study. J Int Adv Otol 2022 Sep;18(5):399-404. [doi: 10.5152/iao.2022.21469] [Medline: 36063096]
25. Hearing loss in adults. American Speech-Language-Hearing Association. URL: www.asha.org/practice-portal/clinical-topics/hearing-loss/#collapse_9 [accessed 2023-09-1]
26. Clinical effectiveness of personal sound amplification products and smartphone-based hearing aid applications in adults with mild-to-moderate sensorineural hearing loss. Clinical Research Information Service. URL: cris.nih.go.kr/cris/search/detailSearch.do/17736 [accessed 2023-09-1]
27. Kim JS, Lim D, Hong HN, Shin HW, Lee KD, Hong BN, et al. Development of Korean standard monosyllabic word lists for adults (KS-MWL-A). Audiol Speech Res 2008 Dec;4(2):126-140. [doi: 10.21848/audiol.2008.4.2.126]
28. Nilsson M, Soli SD, Sullivan JA. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am 1994 Feb;95(2):1085-1099. [doi: 10.1121/1.408469] [Medline: 8132902]
29. Moon SK, Kim SH, Mun HA, Jung HK, Lee JH, Choung YH, et al. The Korean hearing in noise test. Int J Audiol 2008 Jun;47(6):375-376. [doi: 10.1080/14992020701882457] [Medline: 18569115]
30. Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am 1971 Feb;49(2):Suppl 2:467+. [Medline: 5541744]
31. Cox RM, Alexander GC. The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. Int J Audiol 2002 Jan;41(1):30-35. [doi: 10.3109/14992020209101309] [Medline: 12467367]
32. Reed NS, Betz J, Kendig N, Korczak M, Lin FR. Personal sound amplification products vs a conventional hearing aid for speech understanding in noise. JAMA 2017 Jul 4;318(1):89-90. [doi: 10.1001/jama.2017.6905] [Medline: 28672306]
33. Manchaiah V, Taylor B, Dockens AL, Tran NR, Lane K, Castle M, et al. Applications of direct-to-consumer hearing devices for adults with hearing loss: a review. Clin Interv Aging 2017 May 18;12:859-871. [doi: 10.2147/CIA.S135390] [Medline: 28553093]

34. de Sousa KC, Moore DR, Motlagh-Zadeh L, Myburgh HC, Swanepoel DW. Do smartphone hearing aid apps work? Hear J 2019;72(11):34. [doi: 10.1097/01.HJ.0000612588.11985.6a]
35. Lin HY, Lai HS, Huang CY, Chen CH, Wu SL, Chu YC, et al. Smartphone-bundled earphones as personal sound amplification products in adults with sensorineural hearing loss. iScience 2022 Nov 15;25(12):105436. [doi: 10.1016/j.isci.2022.105436] [Medline: 36590464]
36. Martinez-Beneyto P, Franchella S, Rodriguez FA, Navarro-Velasquez R, Martinez-Beneito MA, Martini A, et al. Are smartphone applications (app) useful to improve hearing? Acta Otorhinolaryngol Ital 2020 Aug;40(4):304-310. [doi: 10.14639/0392-100X-N0318] [Medline: 32519992]

Abbreviations

APHAB: Abbreviated Profile of Hearing Aid Benefit
HA: hearing aid
HINT: hearing-in-noise test
IOI-HA: International Outcome Inventory for Hearing Aids
PSAP: personal sound amplification product
PTA: pure-tone audiometry
REM: real-ear measurement
SFA: sound-field audiometry
SHAA: smartphone-based hearing aid app
SNR: signal-to-noise ratio
SPL: sound pressure level
WHO: World Health Organization
WRS: word recognition score

Edited by L Buis; submitted 09.03.23; peer-reviewed by Y Park, YC Chu; revised version received 02.07.23; accepted 10.08.23; published 05.10.23.

Please cite as:
Jo S, Park MK, Seo JH, Lee KE, Han JS, Lim JH, Lee JH, Oh SH
Feasibility of a Smartphone-Based Hearing Aid App for Mild-to-Moderate Hearing Loss: Prospective Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e46911
URL: https://mhealth.jmir.org/2023/1/e46911
doi:10.2196/46911

Original Paper

Clinical Study of a Wearable Remote Rehabilitation Training System for Patients With Stroke: Randomized Controlled Pilot Trial

Liquan Guo[1,2*], MS; Jiping Wang[1,2], MS; Qunqiang Wu[3], BS; Xinming Li[4], PhD; Bochao Zhang[1,2], BS; Linfu Zhou[5*], PhD; Daxi Xiong[1,2*], PhD

[1]School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Suzhou, China

[2]Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China

[3]Department of Rehabilitation Medicine, Tangdu Hospital Airforce Medicine University, Xi‘an, China

[4]Department of Rehabilitation Medicine, Xi’an Gaoxin Hospital, Xi'an, China

[5]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital, Nanjing Medical University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Daxi Xiong, PhD
School of Biomedical Engineering (Suzhou)
Division of Life Sciences and Medicine
University of Science and Technology of China
No.88, Keling Road
Suzhou, 215163
China
Phone: 86 18662576055
Email: xiongdx@sibet.ac.cn

Abstract

Background: In contrast to the large and increasing number of patients with stroke, clinical rehabilitation resources cannot meet their rehabilitation needs. Especially for those discharged, ways to carry out effective rehabilitation training without the supervision of physicians and receive guidance from physicians remain urgent problems to be solved in clinical rehabilitation and have become a research hot spot at home and abroad. At present, there are many studies on home rehabilitation training based on wearable devices, Kinect, among others, but these have disadvantages (eg, complex systems, high price, and unsatisfactory rehabilitation effects).

Objective: This study aims to design a remote intelligent rehabilitation training system based on wearable devices and human-computer interaction training tasks, and to evaluate the effectiveness and safety of the remote rehabilitation training system for nonphysician-supervised motor rehabilitation training of patients with stroke through a clinical trial study.

Methods: A total of 120 inpatients with stroke having limb motor dysfunction were enrolled via a randomized, parallel-controlled method in the rehabilitation institutions, and a 3-week clinical trial was conducted in the rehabilitation hall with 60 patients in the experimental group and 60 in the control group. The patients in the experimental group used the remote rehabilitation training system for rehabilitation training and routine clinical physical therapy (PT) training and received routine drug treatment every day. The patients in the control group received routine clinical occupational therapy (OT) training and routine clinical PT training and routine drug treatment every day. At the beginning of the training (baseline) and after 3 weeks, the Fugl-Meyer Motor Function Rating scale was scored by rehabilitation physicians, and the results were compared and analyzed.

Results: Statistics were performed using SAS software (version 9.4). The total mean Fugl-Meyer score improved by 11.98 (SD 8.46; 95% CI 9.69-14.27) in the control group and 17.56 (SD 11.65; 95% CI 14.37-20.74) in the experimental group, and the difference between the 2 groups was statistically significant (P=.005). Among them, the mean Fugl-Meyer upper extremity score improved by 7.45 (SD 7.24; 95% CI 5.50-9.41) in the control group and 11.28 (SD 8.59; 95% CI 8.93-13.62) in the experimental group, and the difference between the 2 groups was statistically significant (P=.01). The mean Fugl-Meyer lower extremity score improved by 4.53 (SD 4.42; 95% CI 3.33-5.72) in the control group and 6.28 (SD 5.28; 95% CI 4.84-7.72) in the experimental group, and there was no significant difference between the 2 groups (P=.06). The test results showed that the experimental group was better than the control group, and that the patients’ motor ability was improved.

Conclusions: The remote rehabilitation training system designed based on wearable devices and human-computer interaction training tasks can replace routine clinical OT training. In the future, through medical device registration certification, the system will be used without the participation of physicians or therapists, such as in rehabilitation training halls, and in remote environments, such as communities and homes.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200061310; https://tinyurl.com/34ka2725

(JMIR Mhealth Uhealth 2023;11:e40416) doi:10.2196/40416

KEYWORDS

remote rehabilitation; wearable devices; human-computer interaction; rehabilitation training; stroke

Introduction

Stroke is a disease of cerebral blood circulation disorder and brain tissue function and structural damage caused by cerebral vascular obstruction or rupture. It is the third leading cause of death and the second leading cause of disability worldwide. The high disability rate increases economic burden and mental pressure on both society and families [1]. According to the "Report on Stroke Prevention and Treatment in China 2021" [2], in 2020, the standardized prevalence of stroke among people aged over 40 years in China was 2.61%, the incidence rate was 505.23/100,000, and there were about 17.8 million patients with stroke. In addition, 3.4 million new patients are diagnosed with stroke each year in China, compared with approximately 13.7 million annually worldwide.

According to statistics, about 70%-85% of patients with first-time stroke have limb motor dysfunction, which seriously affects the quality of life and brings a heavy burden for the family and society. Timely and effective rehabilitation training can help them restore certain motor functions [3]. However, compared with the large and increasing number of patients with stroke, the resources of rehabilitation medical care are very limited. Therefore, rehabilitation training with the participation of nonrehabilitation physicians or therapists, especially remote and home-based rehabilitation, has received increasing attention.

According to some studies [4,5], the effect of rehabilitation training in a remote environment is comparable to or even better than that in a hospital environment. Relevant studies show [6-8] that sustained and effective remote rehabilitation can activate the neuroplasticity of patients with stroke and greatly improve the rehabilitation effect. Remote rehabilitation training can save medical resources, promote the motor function of patients, and improve the rehabilitation level after discharge in view of poor compliance of discharged patients [9-11]. Indeed, patient adherence and acceptability of rehabilitative practices need to be actively enhanced, overcoming pitfalls due to motor (eg, endurance), nonmotor (eg, fatigue, pain, dysautonomic symptoms, and motivational), and cognitive deficits [12].

Active and effective rehabilitation with nonphysician involvement, such as remote and home rehabilitation, and uploading training data and results to physicians for analysis and guidance are effective solutions to the problems of lack of clinical rehabilitation resources and poor adherence of discharged patients and are hot spots of international research; however, they still face many challenges.

Currently, to carry out effective rehabilitation training with nonphysician involvement, 2 main technical support solutions are proposed for training data acquisition and human-computer interaction control around application scenarios such as patient limb movements [13-15], activity detection [16-18], and motion recognition [19-22].

The first is vision-based solution, such as using a depth camera or Kinect. Placidi et al [23] designed a simple motion analysis system based on the use of a depth camera and a 3D real-time model of the human body. Their experimental results showed no significant differences in more than 95% of the data. However, the experiment could not achieve the rehabilitation training goals for fine motor movements and was not suitable for patients with severe disabilities. Webster and Celik [24] summarized the application of Kinect in geriatric care and stroke rehabilitation, based on which it was pointed out that the current application should be simulated toward the real situation, there was the need to capture obscuring movements, and in addition, the Kinect application is vulnerable to the spatial environment.

The second is a wearable sensor-based solution that integrates inertial sensors such as accelerometers to assess functional activities related to patient mobility in terms of type, intensity, time, and quality of the activity. Rau et al [25] developed a triaxial accelerometer-based remote assessment system for acquiring kinematic data on upper extremity anterior extension movements in patients with stroke. Spearman analysis showed a strong correlation between this remote assessment system and standard kinematics. However, the experiment was only for upper extremity movements and required an expert on hand for guidance. Yang et al [14] proposed a stroke rehabilitation system combining inertial measurement sensors with physiological sensors, with an average recognition accuracy of 96.20% for hand gesture movements. However, this study only focused on identifying patient-specific movements and the training and validation data were from the same patients.

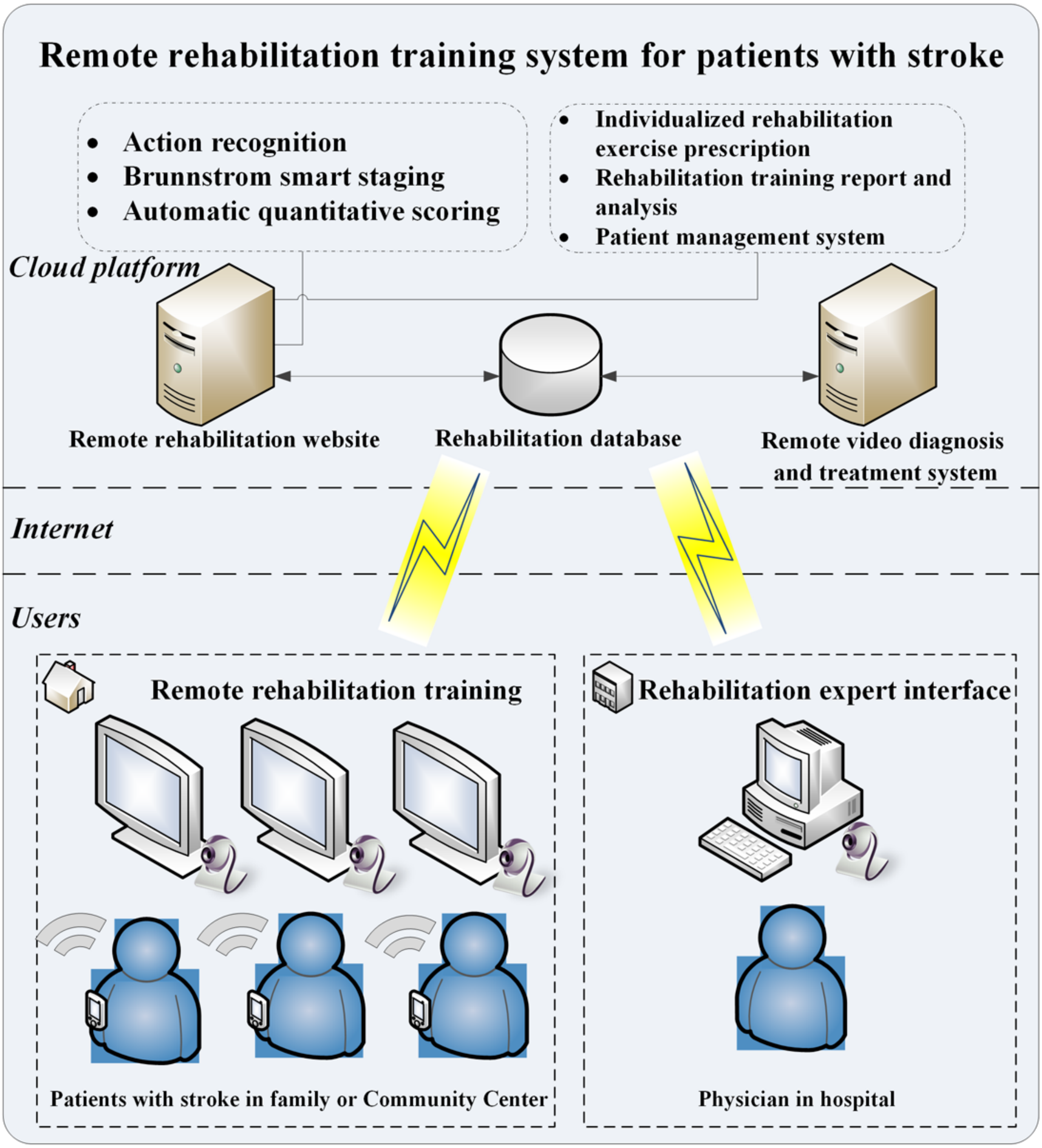
In addition, most studies have been conducted in clinical settings under physician supervision, and there is a lack of studies on rehabilitation training without physician supervision and validation in standardized clinical trials [26,27]. Therefore, to conduct effective rehabilitation training without physician supervision for use in remote and home settings, wearable devices based on the inertial measurement unit (IMU) and flex sensor are designed to be worn on the affected limb. Patients undergo interactive rehabilitation training based on standard training videos combined with human-computer interaction games. The feasibility, efficacy, and safety of the system are evaluated by conducting a 120-case parallel-controlled, 2-center clinical trial.

Methods

Overview of the System Framework

In this study, we designed wearable devices such as rehabilitation training gloves and upper and lower limb rehabilitation training modules. Further, a remote rehabilitation system integrating training equipment hardware, man-machine communication training games, rehabilitation training software, a remote rehabilitation management platform, and a mobile app were developed. The overall architecture of the system is shown in Figure 1. The system consists of 3 parts, the patients with stroke side, the physician side, and the cloud server. Through the remote server, the rehabilitation physician in the hospital can view, analyze, and guide the patients' rehabilitation training remotely in the training hall, community, and home, and prescribe new rehabilitation exercises for the patients. Through this system, patients can perform rehabilitation training without physician involvement.

Figure 1. Remote rehabilitation training system for patients with stroke.



Training Equipment Hardware

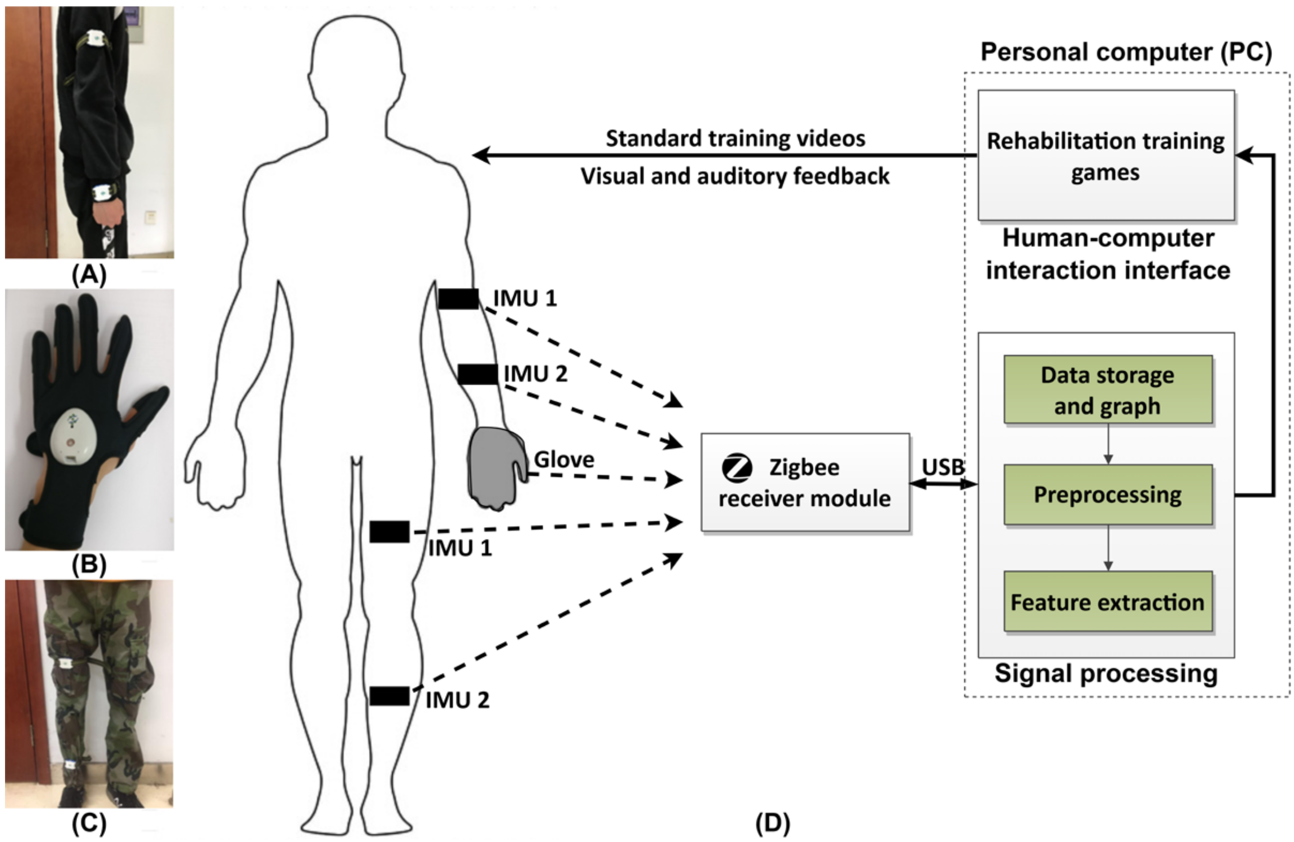
The wearable remote rehabilitation training equipment mainly includes the IMU modules for upper and lower limb rehabilitation training, rehabilitation training gloves for hand rehabilitation training, and Zigbee wireless receiver, among others. There are 2 IMU modules containing 9-axis motion sensors, including a 3-axis accelerometer, a 3-axis angular velocity meter, and a 3-axis magnetometer. The 2 IMU sensors are fixed to the upper and lower arms, respectively, by straps during upper limb training, and to the thigh and calf, respectively, during lower limb training. The rehabilitation glove contains 1 IMU and 5 flex sensors inside to monitor the movement of the wrist and individual fingers. The gloves are designed for left and right hands, with large, medium, and small sizes available to suit different patients.

The sensor is fixed on the affected side by the patients or their family according to the instructions and the wearing process is not complicated. In addition, the sensor works in headless mode, and its current position is automatically defined as the initial position through coordinate transformation at the beginning of each movement, so the deviation of the placement position does not affect the rehabilitation training.

Each wearable device has a 400-mAh battery and a power consumption of 20 mAh, with a full charge meeting the rehabilitation training for about 20 hours. Patients can train 2 times a day for half an hour each time, so the wearable devices can be used continuously for 10 days with a full charge. Each sensor was previously networked through the ZigBee2007 wireless communication protocol, which is very convenient for expansion and synchronous data collection. The sampling rate of each wearable device is 30 times per second, which is sufficient for rehabilitation training exercise data collection and analysis.

The rehabilitation training process based on the wearable device and interactive game is shown in Figure 2. Panels A-C are the schematic diagrams of the wearable device worn on the affected upper limb, hand, and lower limb, respectively, while D is the schematic diagram of the human-computer interaction rehabilitation training process. Patients wear the wearable device according to the instruction manual and undergo rehabilitation training according to the standard training video on the software. The sensor collects the rehabilitation training data, receives them through the Zigbee wireless receiver, and transmits them to the computer through USB. The rehabilitation training software on the PC side collects, stores, and displays data; improves signal quality through preprocessing such as sliding filtering; and then extracts patient motion features. The system controls rehabilitation training games through motion features, conducts human-computer interaction training for patients, and provides video and auditory feedback to patients to improve their enthusiasm for rehabilitation training.

Figure 2. Rehabilitation training process based on wearable devices and interactive games. (A) Instructions for wearing upper limbs; (B) Instructions for wearing gloves; (C) Instructions for wearing lower limbs; (D) Human-computer interaction rehabilitation training process. IMU: inertial measurement unit.



Under the advice of the rehabilitation physician and according to the characteristics of rehabilitation training movements, the motion features extracted by the system mainly include motion amplitude, direction, dynamic energy, motion smoothness, and motion force size, as shown in Table 1. The training difficulty is set according to the Fugl-Meyer score of the patient at the time of enrollment, and the system automatically adjusts the difficulty of the next training according to the previous training, with different features selected for different rehabilitation training movements and difficulty levels. For example, the simple mode of the Bobath handshake training uses AMP (amplitude) as the training game control parameter, while the hard mode collects all 5 features (amplitude, mean value, root-mean-square, JERK, strength) for weighted calculation results as the training game control parameter, with the weighting coefficients of 0.5, 0.2, 0.1, 0.1, and 0.1, respectively.

Table 1. Extracted motion features and physical meaning.

Number	Feature	Definition	Physical meaning
1	AMP[a]	AMP=max(x) – max(x)	Describes the magnitude of the movement
2	MEAN[b]		Describes the direction of the movement
3	RMS[c]		Describes motion dynamic energy
4	JERK		Describes motion smoothness
5	Strength	Value	Describes the magnitude of the exercise effort

[a]AMP: amplitude.

[b]MEAN: mean value.

[c]RMS: root-mean-square.

Rehabilitation Training Software

This is a rehabilitation equipment system that integrates human-computer interaction. It uses software games to simulate daily life scenarios and guide patients in rehabilitation operation training. The computer software collects the strength, speed, distance, and other movement features of the rehabilitation training to control the game tasks, and gives feedback to the patients in a visual or auditory form, so as to guide the patients to continuously adjust their movements. Virtual games can provide clear training goals and tasks. The process of patients completing game tasks is the process of rehabilitation training. The higher the similarity between the patient's rehabilitation training data and the standard data in terms of characteristics, the higher the patient's score on the game task. Therefore, the training mode of human-computer interaction can greatly mobilize the enthusiasm of patients for rehabilitation training. In the absence of visual and auditory feedback, the patient is not fully aware of the abnormal movement pattern of the affected limb, often has a trunk or proximal limb compensation, and is more prone to fatigue [28].

During the trial, the patient opens the Rehabilitation Training and Assessment client software (Figure 3), wears the wearable devices (2 IMU modules and rehabilitation training gloves) based on the physician's prescription for rehabilitation training, and trains according to the standard rehabilitation video. The software can choose different games and can set different game difficulties according to the patient's recovery status. Patients are rehabilitated by a threshold to determine whether the training action is effective or not. The thresholds are specifically 80% for difficult, 60% for moderate, and 40% for easy. By completing a valid action, the game will increase the corresponding score, and if the action is invalid, the score will remain the same. Patients will try to follow the movements of the standard training video to get a higher score during training. During the training process, the patient and the game perform human-computer interaction and receive feedback, and can simultaneously see the effect of each rehabilitation training action.

Figure 3. Rehabilitation training and assessment of client software: (A) rehabilitation prescription; (B) virtual game; (C) patient training; (D) exercise time; (E) action guidance; and (F) equipment operating status.

For example, in the apple picking game, each effective rehabilitation training exercise is defined as picking an apple. After the training is completed, the training score is given according to the parameters and features of the patient's rehabilitation training, and the training data and results are automatically uploaded to the remote server so that the physician increases or decreases the length and intensity of the relevant movements according to the patient's rehabilitation score. The new exercise prescription is automatically updated in the patient's rehabilitation software. The clinical Fugl-Meyer score at enrollment and the game score during training were used as the basis for updating the exercise prescription. Adjustments were made once a day, and no adjustment was required for score changes of 5 points or less.

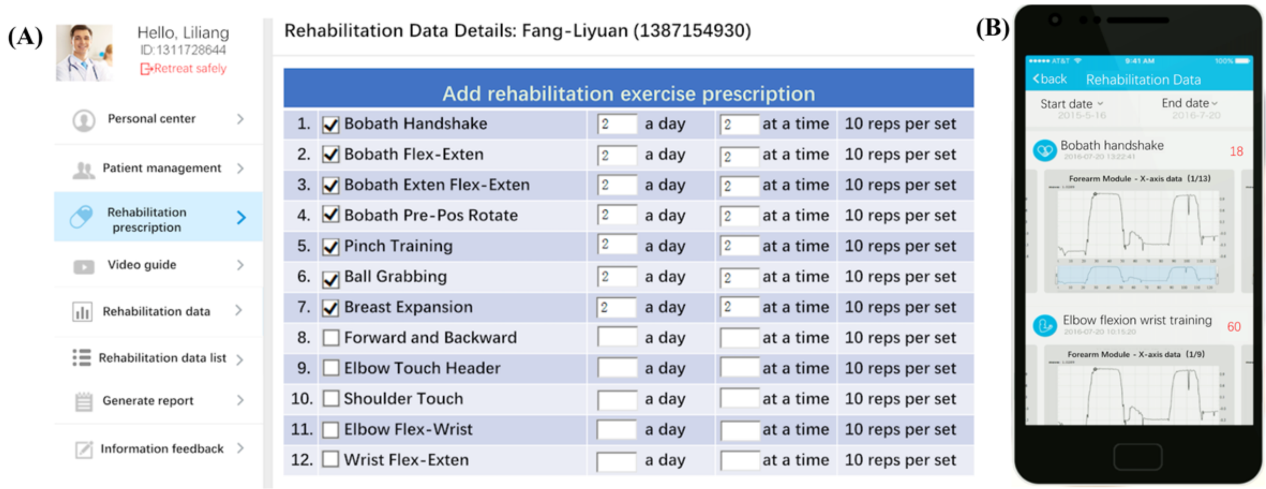
The software, games, and cloud platform included in the rehabilitation training system have been inspected by the National Medical Device Inspection Center and tested in accordance with the testing standard *GB/t 25000.1-2010 Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE)* [29], which meets the relevant requirements of software and network security and can be used for clinical research.

Remote Rehabilitation Cloud Management Platform

To help physicians view and analyze the rehabilitation training situation of remote patients more timely and effectively, data are visualized through the remote rehabilitation cloud management platform, thus allowing them to manage the rehabilitation training of their patients. The remote rehabilitation management platform consists of a front-end interactive interface and a back-end data analysis system. The front-end interactive interface comprises multiple pages for users to query and edit related information. The back-end data analysis system mainly realizes the functions of analyzing multisource sensor data and generating analysis reports and pushes the evaluation results to the corresponding rehabilitation physicians.

The remote rehabilitation cloud management platform (Figure 4) includes the web terminal and the mobile app, both of which have the functions of patient information management, updating rehabilitation prescriptions, remote video guidance, viewing rehabilitation data, generating analysis reports, and constructing patient rehabilitation files to assist physicians in better managing remote rehabilitation and guiding them in rehabilitation training. Patients or their families can also log-in to the management platform to consult and communicate with rehabilitation physicians.

Figure 4. (A) Remote rehabilitation cloud management platform; (B) mobile app. Reps: repetitions.



Clinical Trial: Parallel Controlled 2-Center Study

Participants and Setting

For this pilot study, we chose patients in Tangdu Hospital and Xi'an Gaoxin Hospital with limb motor dysfunction caused by stroke 15-180 days after the onset (recovery period) and requiring rehabilitation training.

Patient inclusion criteria were as follows: (1) stroke diagnosed by computed tomography or magnetic resonance imaging within 90 days; (2) age between 30 and 75 years, male or female; (3) stable rehabilitation patients with limb motor dysfunction (with hemiplegic motor function evaluated according to the Brunnstrom upper or lower extremity grading stages II-VI) caused by stroke 15-180 days after its onset (recovery period); (4) cognition is clear and can follow the research protocol; (5) the patient can understand the study's purpose, as well as showing sufficient compliance with the study protocol and signed the informed consent.

The following patients were excluded: (1) significant impairment of cognition and consciousness so that the Fugl-Meyer test could not be completed, (2) other significant limb lesions, such as fractures, severe arthritis, or amputation; (3) formation of limb joint contractures; (4) patients with disability, as specified by the International Classification of Functioning, Disability, and Health; (5) patients with a combination of severe primary diseases involving the cardiovascular, liver, kidney, and hematopoietic systems and mentally ill patients, as well as other circumstances that the investigator considers inappropriate to participate in this trial.

Experimental Design

This clinical trial is planned to be carried out in 2 clinical trial institutions at the same time and is divided into an experimental group and a control group. Patients in the experimental group received exercise training guided by the remote rehabilitation training system, routine clinical physical therapy (PT) training, and routine drug treatment. By contrast, patients in the control group received routine clinical occupational therapy (OT) training, routine clinical PT training, and routine drug treatment (see Multimedia Appendix 1 for the CONSORT [Consolidated Standards of Reporting Trials] checklist).

According to the inclusion criteria, this study selects all patients who conform to the entire trial process, that is, those who conform to the trial protocol, have good compliance, and can complete the corresponding tasks for analysis. The patients are randomly allocated to the experimental and control group in a 1:1 ratio, and the main efficacy index (Fugl-Meyer score of patients) was used as the basis for case estimation, with the sample size calculated according to the following formula [30]:

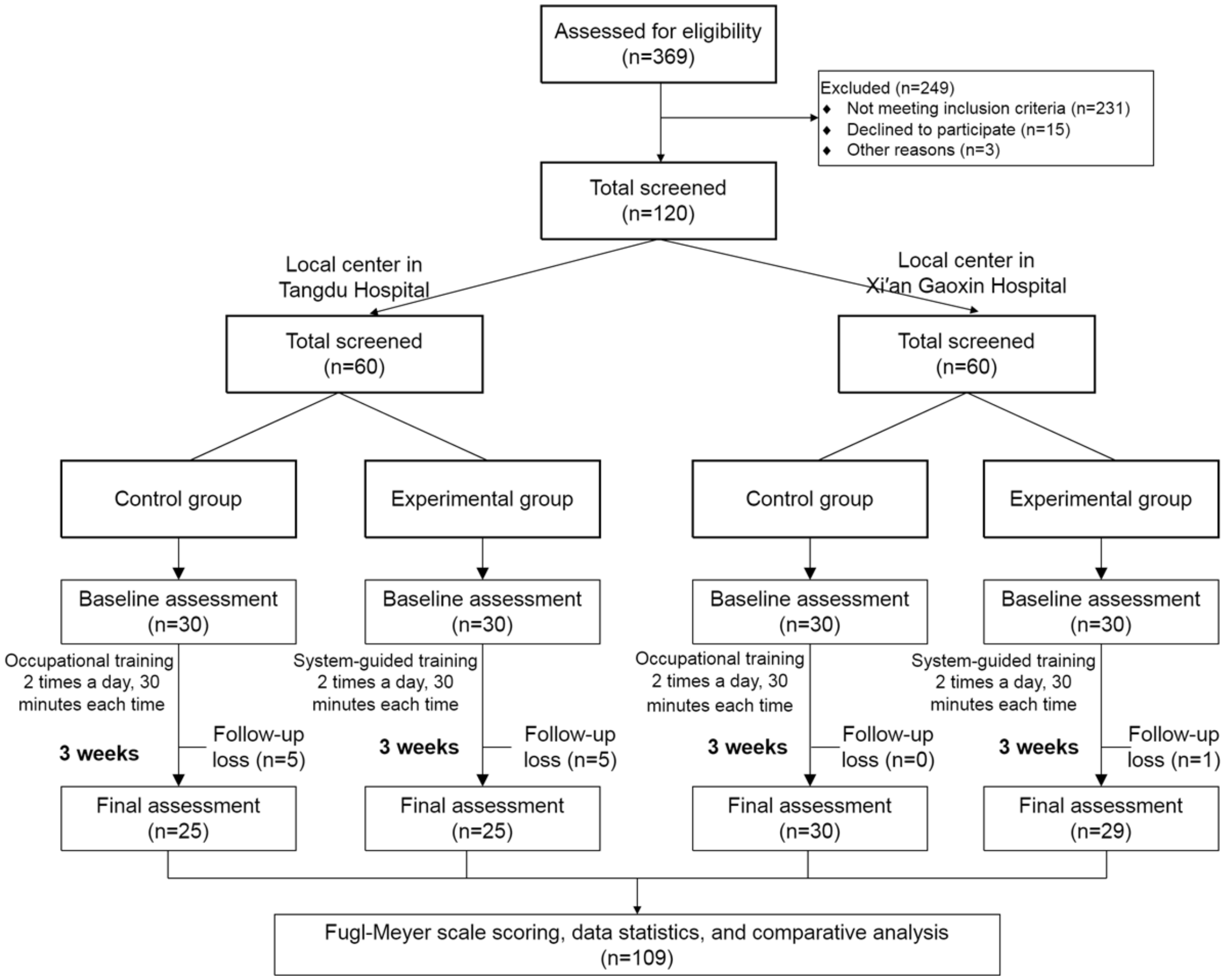
According to class II medical devices recognized by the industry, when the probability of type I error α is set to 1-sided .025, $z_{\alpha/2}$=1.96, and the probability of type II error β is set to .2, that is, the power (1–β=80%) is 80%, z_β=0.84. According to the aforesaid formula for the number of classic cases, this study predicted that the Fugl-Meyer change value of the experimental group is µ1=11.0, the mean change of the control group is µ2=10.0, and the mean SD ()=5.5. The noninferiority margin was 40% of the mean SD, and if δ=2.2, the number of cases was calculated as n=47. Assuming a 20% dropout rate, the number of patients in each group should be at least 59. This clinical trial determined 60 cases in the experimental group and 60 in the control group, thus there were a total of 120 cases.

Considering whether the patients were exposed to PT/OT training and reducing the associated effects, this clinical trial protocol used a randomized grouping approach in which all patients who met the inclusion criteria were randomly assigned to either the test or control group according to randomization rules. The randomization method and steps were as follows: (1) Patients were randomized according to the stratified block randomization method. First, the random seed was set, then the block length was determined and stratified according to the center. SAS version 9.4 (SAS Institute) was used to generate a random grouping table of 120 patients receiving the trial (experimental group or control group). Each center was assigned consecutive random numbers that connect with each other. The patients were randomly assigned to either the experimental group or the control group according to the order in which the cases were enrolled and the randomization table.

As the training methods were different for the 2 groups, an open randomized trial was used. After randomizing patients to 1 of the 2 groups, neither the investigator nor the patients knew the grouping when baseline scoring was performed. At the beginning of the training, the randomized envelope for rehabilitation training corresponding to each randomization number was opened to know the corresponding training method. Thus, the investigator and patients only became aware of the grouping and scoring results after starting training. The statistical analysts, however, before unblinding, were not aware of the patients' grouping.

The flowchart of this clinical trial is shown in Figure 5. In the 2 rehabilitation medical centers, 369 patients were assessed for eligibility. Among these 249 were excluded, of which 231 did not meet the inclusion criteria, 15 were unwilling to participate, and 3 had other reasons. Finally, 120 patients were randomly assigned, including 60 in the experimental group and 60 in the control group. Tangdu Hospital enrolled 60 cases, including 30 in the experimental group and 30 in the control group, while Xi'an Gaoxin Hospital enrolled 60 cases, including 30 in the experimental group and 30 in the control group. However, in the actual process of the trial, according to the principle of patients' voluntariness, 11 dropped out, and finally, a total of 109 patients completed the entire trial process.

Figure 5. Clinical trial flowchart.



According to related studies [28], active exercise training is more conducive to functional improvement and cortical function remodeling than passive training. According to the general rehabilitation guidelines and operating norms at home and abroad [31-33], in combination with the current commonly used clinical rehabilitation training movements and training methods, and under the advice and recommendation of many rehabilitation experts and physicians, 16 typical rehabilitation exercises were designed. The designed rehabilitation movements are used for the coordinated movement training of upper extremity, hand, and lower extremity. The training actions of the remote rehabilitation training system were as follows:

Upper extremity movements: (1) Bobath handshake training, (2) Bobath flexion and extension, (3) Bobath external anterior flexion and extension, (4) Bobath pre- and postrotation, (5) breast expansion exercise, (6) shoulder joint internal and external rotation, (7) shoulder touch training, and (8) elbow joint flexion and touch. Hand movements: (1) flexion-pressure rotation forward and backward, (2) wrist flexion and extension, (3) elbow flexion and wrist compression training, (4) finger-to-finger training, and (5) ball gripping training. Lower extremity movements: (1) squat training, (2) knee flexion and extension, and (3) knee internal and external rotation.

The experimental group adopted the training method of the remote rehabilitation training system and routine clinical PT training, whereas the control group used routine clinical rehabilitation training methods for limb motor dysfunction (ie, routine clinical OT training and routine clinical PT training).

For the inpatients in the experimental and control groups, the specific diagnosis and treatment methods were based on the condition and the test content, and the corresponding training and rehabilitation exercise methods that could be completed independently were selected in the rehabilitation hall. In addition, they performed system-guided training or routine clinical OT training 2 times a day (each session lasted 30 minutes) and conventional PT training 2 times a day (each session lasted 30 minutes). They trained no less than 10 times a week, for a total of no less than 30 times, for a total of 3 weeks.

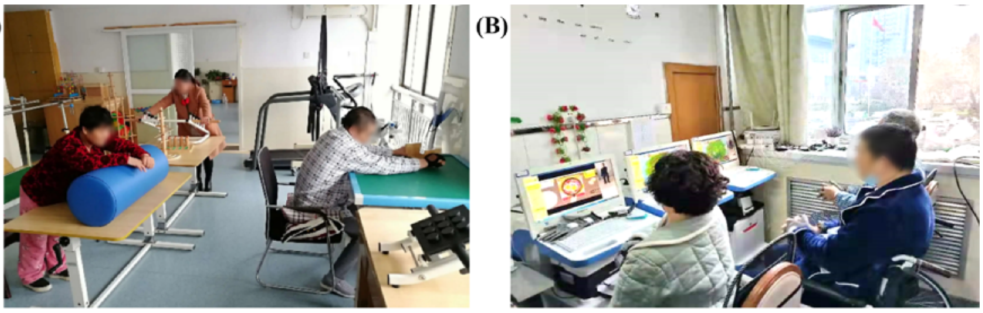
In this study, the simplified Fugl-Meyer Motor Function Assessment scale was used for evaluation. The scale has good reliability and validity, Cronbach reliability coefficient >.80, and intraclass correlation coefficient >0.70 [34].

The scale consists of 50 items, including 33 for the upper extremities and 17 for the lower extremities, with each item rated on a scale of 0 (unable to complete the specified movement), 1 (able to partially complete), or 2 (can fully complete). The total score is 100 points. The higher the score, the better the motor function of the patient. At baseline and after 3 weeks of training, the patients were assessed by the rehabilitation physician according to the Fugl-Meyer Assessment (FMA) scale and the related results were recorded, respectively.

To study the rehabilitation training under the real nonphysician involvement scenario, the rehabilitation physician or therapist was next to the patient during the whole rehabilitation training process, only to ensure the patient's safety. In addition, the rehabilitation data from the experimental group were uploaded to the rehabilitation website so that the physician could view the training data and update the exercise prescription as necessary from the office. The actual training of patients with stroke is shown in Figure 6. Figure 6A shows the control group receiving conventional OT training and Figure 6B shows the experimental group wearing the wearable device and following the video and human-computer interaction game for autonomous rehabilitation training.

Figure 6. Practical application of the 2 training methods: (A) patients in the control group using conventional occupational therapy training; (B) patients in the experimental group using a remote rehabilitation training system.

(A) (B)



Statistical Analysis

In this study, the rehabilitation status of patients with limb motor dysfunction (based on the change in the Fugl-Meyer Motor Function Rating scale score) after 3 weeks of clinical observation was used as the primary endpoint. The Fugl-Meyer score was used as the evaluation index to evaluate the clinical effectiveness of the remote rehabilitation training and evaluation system, and the safety of the system was judged by the number of adverse events and the relationship with the test system.

Descriptive statistics were used in this study to characterize demographic parameters and other baseline characteristic values. In this pilot study, a total of 109 patients ultimately completed the full trial, and statistical analyses and discussions of the data were conducted for these patients.

For descriptive statistics, demographic data, and other baseline characteristic values, parametric analysis was performed using targeted statistical methods, and the *P* value of inferential statistics was listed as the descriptive result. For the change in Fugl-Meyer score from baseline to 3 weeks of treatment, the difference between the 2 groups and its bilateral 95% CI were calculated.

SAS version 9.4 was used for analysis in this study. All statistical tests were 2-sided, and a *P* value ≤.05 was considered statistically significant.

Ethics Approval

This study was approved by the Ethics Committees of Tangdu Hospital (approval number 201912-08) and Xi'an Gaoxin Hospital (2020 ethics review number 001). All patients participating in this study have signed the informed consent form.

Results

Baseline Data Analysis

The statistical results of demographic parameters and other baseline characteristic values are presented in Table 2. Different parameters of the experimental and control groups were statistically analyzed by different statistical methods. There was no significant difference in age (P=.81), BMI (P=.39), systolic blood pressure (P=.25), and diastolic blood pressure (P=.41) between the 2 groups by (1-sided) *t* test (P>.05).

Table 2. Analysis of demographic parameters of patients with stroke (n=60).

Characteristics	Control group			Experimental group			
	Values[a]	95% CI (lower-upper)	Range (median)	Values[a]	95% CI (lower-upper)	Range (median)	*P* value
Age (year)	55.82 (9.68)	53.32-58.32	34.00-73.00 (55.00)	56.25 (9.83)	53.71-58.79	33.00-73.00 (56.00)	.81
Gender (male)	43 (71.67)			43 (71.67)			>.99
Course of disease (days)	61.20 (46.67)	49.15-73.25	15.00-169.00 (41.00)	46.22 (36.63)	36.75-55.68	15.00-165.00 (33.00)	.15
Systolic blood pressure (mmHg)	127.47 (11.56)	124.48-130.45	96.00-151.00 (127.00)	130.13 (13.73)	126.59-133.68	92.00-166.00 (130.00)	.25
Diastolic blood pressure (mmHg)	81.05 (9.00)	78.73-83.37	60.00-105.00 (80.00)	82.45 (9.51)	79.99-84.91	57.00-107.00 (80.50)	.41
BMI (kg/m^2)	24.68 (4.08)	23.59-25.76	18.34-44.92 (23.70)	24.66 (2.59)	23.99-25.33	18.34-44.92 (24.57)	.39
Stroke type (cerebral infarction)	37 (61.67)	—[b]	—	32 (53.33)	—	—	.36
Hypertension	44 (73.33)	—	—	43 (71.67)	—	—	.84
Hyperlipidemia	7 (11.67)	—	—	4 (6.67)	—	—	.53
Arteriosclerotic coronary disease/myocardial infarction	10 (16.67)	—	—	12 (20.00)	—	—	.64

[a]Data are mean (SD) or n (%).

[b]Not applicable.

Although the mean course of stroke in the 2 groups was 61.20 and 46.22, respectively, in the Wilcoxon test for these 2 nonnormally distributed data, *P*=.15 (*P*>.05), indicating that there was no statistically significant difference between the 2 groups. The reason for the difference in the means of the 2 groups was that there were 2 cases in the control group with stroke duration days close to 180 days, which increased the mean, but did not affect the overall experimental results.

The chi-square test showed that there was no significant difference between the 2 groups in gender (*P*>.99), stroke type (*P*=.36), hypertension (*P*=.84), and arteriosclerotic coronary disease/myocardial infarction (*P*=.64). Fisher test showed that there was no significant difference in hyperlipidemia between the 2 groups (*P*=.53). These statistical results showed that in terms of various parameters, there was no statistical difference between the control group and the experimental group.

Results of the Clinical Trial

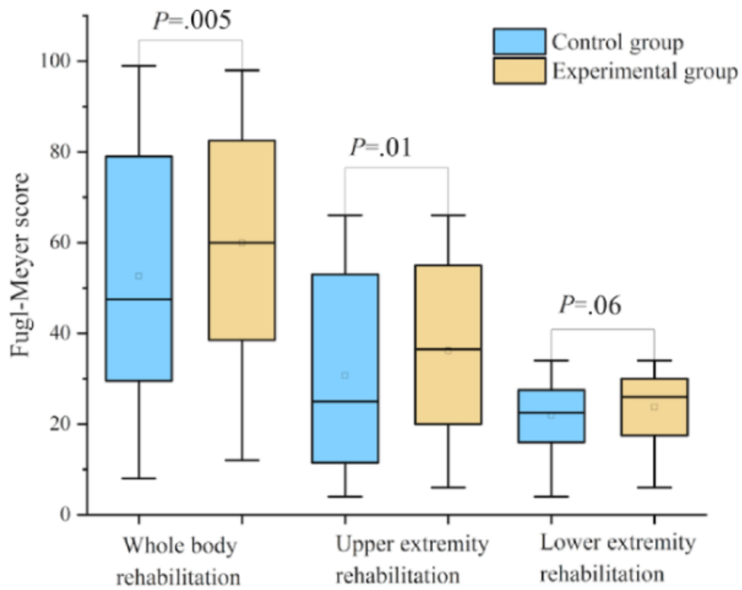
At baseline and 21 days, patients in the experimental group and patients in the control group were evaluated for motor function according to the FMA scale by experienced clinical rehabilitation physicians. The results of the experimental and control groups were statistically analyzed using the *t* test, and the relevant results are presented in Table 3. A total of 55 patients (92%) in the control group completed all trials, whereas a total of 54 patients (90%) in the experimental group completed all trials. Physician Fugl-Meyer mean total score changes in the control group were 11.98 (SD 8.46; 95% CI 9.69-14.27), whereas those in the experimental group were 17.56 (SD 11.65; 95% CI 14.37-20.74; *P*=.005). Physician Fugl-Meyer mean upper extremity score changes in the control group were 7.45 (SD 7.24; 95% CI 5.50-9.41), whereas those in the experimental group were 11.28 (SD 8.59; 95% CI 8.93-13.62; *P*=.01). Physician Fugl-Meyer mean lower extremity score changes in the control group were 4.53 (SD 4.42; 95% CI 3.33-5.72), whereas those in the experimental group were 6.28 (SD 5.28; 95% CI 4.84-7.72; *P*=.06).

Table 3. Statistical analysis of physician scores according to the Fugl-Meyer scale.

Analysis	Control group			Experimental group			*P* value
	Mean (SD)	95% CI (lower-upper)	Range (median)	Mean (SD)	95% CI (lower-upper)	Range (median)	
Total score							
0 days	41.11 (27.49)	33.68-48.54	8.00 to 96.00 (35.00)	43.28 (25.09)	33.68-48.54	13.00 to 97.00 (35.50)	.67
21 days	53.09 (28.40)	45.41-60.77	8.00 to 99.00 (48.00)	60.83 (23.80)	54.34-67.33	20.00 to 98.00 (60.00)	.13
21 to 0 days	11.98 (8.46)	9.69-14.27	0.00 to 39.00 (10.00)	17.56 (11.65)	14.37-20.74	1.00 to 49.00 (14.00)	.005
Upper extremity							
0 days	23.67 (19.93)	18.28-29.06	4.00 to 65.00 (16.00)	25.48 (18.69)	20.38-30.58	4.00 to 64.00 (19.50)	.63
21 days	31.13 (21.67)	25.27-36.99	4.00 to 66.00 (25.00)	36.76 (18.65)	31.67-41.85	7.00 to 66.00 (36.50)	.15
21 to 0 days	7.45 (7.24)	5.50-9.41	0.00 to 36.00 (5.00)	11.28 (8.59)	8.93-13.62	0.00 to 32.00 (9.50)	.01
Lower extremity							
0 days	17.44 (9.36)	14.91-19.97	4.00 to 34.00 (17.00)	17.80 (8.03)	15.61-19.99	5.00 to 33.00 (16.50)	.83
21 days	21.96 (8.36)	19.70-24.22	4.00 to 34.00 (23.00)	24.07 (7.04)	22.15-26.00	10.00 to 34.00 (26.00)	.16
21 to 0 days	4.53 (4.42)	3.33-5.72	–7.00 to 16.00 (4.00)	6.28 (5.28)	4.84-7.72	–3.00 to 22.00 (5.00)	.06

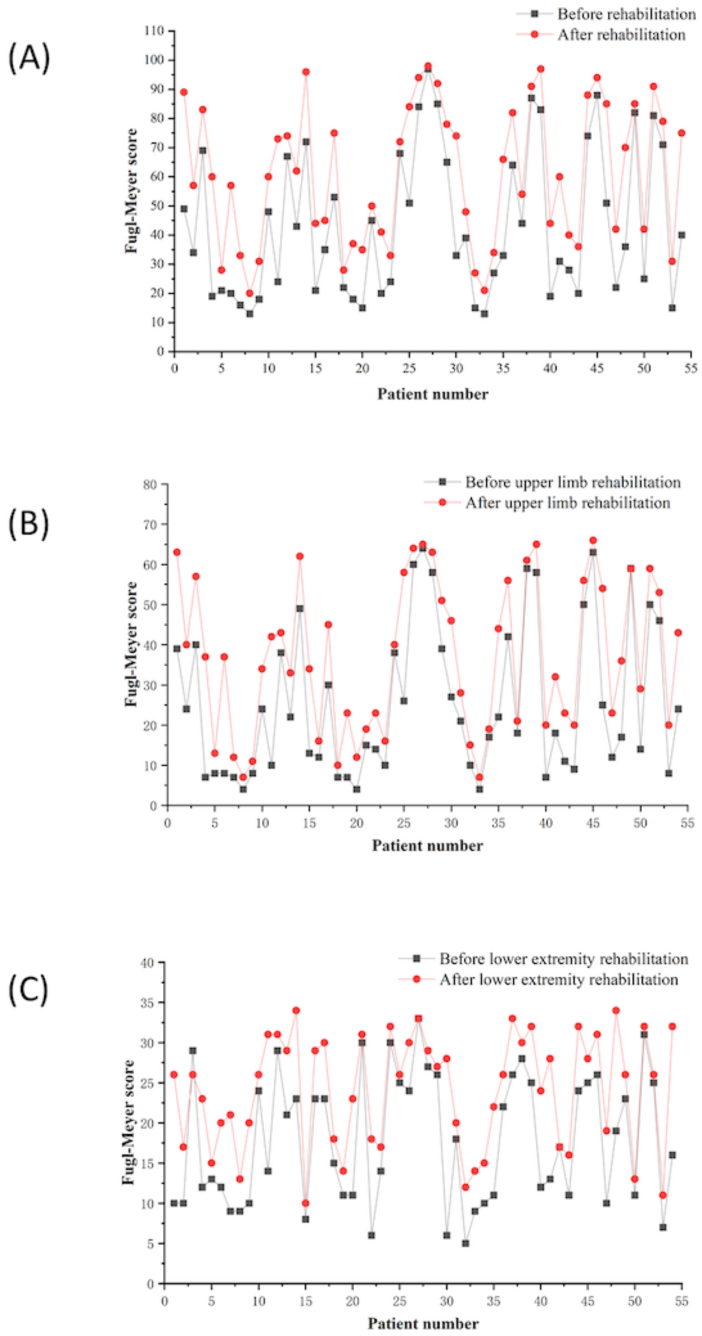
Figure 7 shows the change distribution of Fugl-Meyer scores in the control group and the experimental group after 21 days of rehabilitation training, including the total score, upper limb score, and lower extremity score. The test results showed that the experimental group was better than the control group in the improvement of the total score, upper limb score, and lower extremity score. In the general evaluation and upper limb rehabilitation training, there were significant differences in the changes between the 2 groups (*P*=.005 and .01, respectively), and there was no significant difference in the changes in the lower extremity score between the 2 groups (*P*=.06). The reason may be that, on the one hand, there are only 3 lower extremity rehabilitation exercises, and on the other hand, because the patients also undergo exercise rehabilitation training for the lower extremities in daily walking and other activities, there is no significant difference in the lower extremity rehabilitation effects between the 2 groups (*P*=.06).

Figure 7. Fugl-Meyer Scale score results of the control group and the experimental group.



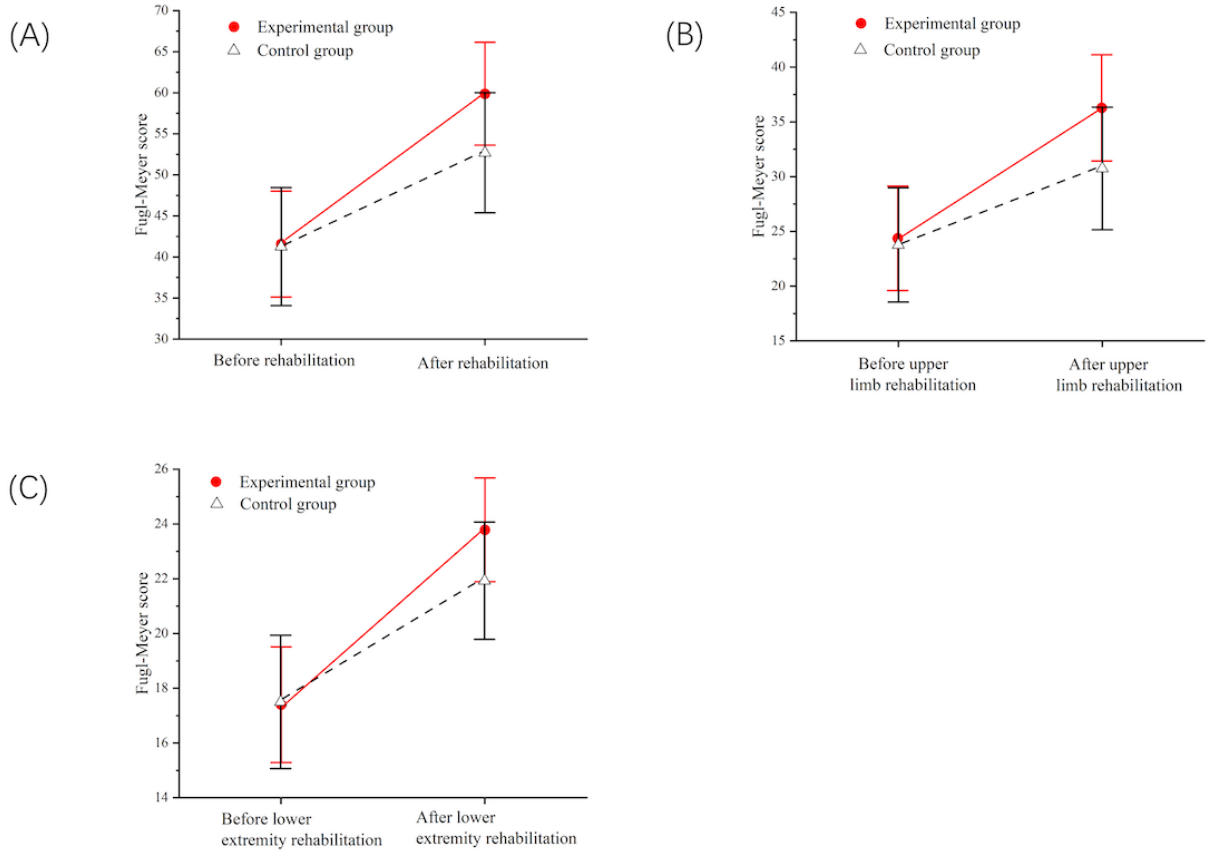
To analyze the effect of a single patient's use of a remote rehabilitation training system on motor function recovery, this study compared the results of the total score, upper limb score, and lower limb score of all patients in the experimental group before and after rehabilitation, as shown in Figure 8. Combined with data in Table 3, the average score of patients before rehabilitation was 43.28, and the average score after rehabilitation training was 60.83, with an average increase of 17.56. The average score of upper limbs before rehabilitation was 25.48, and the average score of upper limbs after rehabilitation training was 36.76, with an average increase of 11.28. The average score of the front lower extremity was 17.80, and the average score of the lower extremity after rehabilitation training was 24.07, an average increase of 6.28. These results show that for all patients using the remote rehabilitation training system, after 21 days of rehabilitation training, the FMA total score, the upper limb score, and the lower extremity score have improved significantly, that is, the patient's exercise ability has been effectively recovered.

Figure 8. (A) Comparison of total scores of the Fugl-Meyer Scale before and after rehabilitation of patients in the experimental group; (B) comparison of Fugl-Meyer Scale upper limb scores of patients in the experimental group before and after rehabilitation; (C) comparison of Fugl-Meyer Scale lower extremity scores before and after rehabilitation in the experimental group.



To compare the effects of using the remote rehabilitation training system and receiving conventional OT training on the recovery of patients' exercise ability, this study compared the control group and the experimental group before and after rehabilitation, as shown in Figure 9. Compared with receiving conventional OT training, the patient's exercise ability improved significantly through the remote rehabilitation training system, and the difference was significant (*P*=.005; Table 3). Among them, the Fugl-Meyer score change value of the upper limb was greater than that of the lower extremity, and all patients using the remote rehabilitation training system had a better rehabilitation effect on the upper limb remote rehabilitation.

Figure 9. (A) Changes in the total score of different rehabilitation methods in the control group and the experimental group; (B) changes in upper limb scores in different rehabilitation methods in the control group and experimental group; (C) changes in lower extremity scores in different rehabilitation methods in the control group and experimental group.



Finally, adverse events in this trial were analyzed. Adverse events are unfavorable medical events that occur during a clinical trial, whether related to a device or not. During the entire clinical trial, 28 adverse events were reported in the control group, with an incidence rate of 46.67%, and 22 adverse events in the experimental group, with an incidence rate of 36.67%; however, there was no significant difference between the 2 groups (*P*=.27). The adverse events that occurred were judged by the investigator to be irrelevant to this test system.

Discussion

Principal Findings

In this study, aiming at the rehabilitation training of patients with limb movement dysfunction such as stroke, a number of wireless wearable devices were developed based on IMU inertial device and bending sensor. Using the Zigbee wireless networking technology, the movement data from patients' rehabilitation training can be collected at the same time. Through data fusion and signal processing, real-time rehabilitation training exercise monitoring and exercise ability analysis are realized. Using rehabilitation training games based on daily life scenes, human-computer interaction rehabilitation training is realized. Further, the patient's rehabilitation training data and results are recorded and uploaded to the remote server platform, so that the remote-end rehabilitation physician can view, analyze, and guide the patient to undergo effective rehabilitation training in a timely manner and improve the patient's enthusiasm and compliance for rehabilitation training.

In addition, the game scenes correspond to the rehabilitation training actions, and the actions and games are matched according to the parts of the patient's body that need rehabilitation. At the same time, the patient can modify the default game and the system will automatically save the record of the game selected by the patient and use it for subsequent rehabilitation training. With the improvement of the patient's exercise ability, the difficulty of training will increase, and the rehabilitation exercise prescription will become more diversified. Presenting continually challenging new tasks helps patients stay motivated and interested in rehabilitation therapy. The virtual training scene based on daily life can reduce the danger caused by the wrong operation of patients with stroke in the real environment.

Based on the clinical trial of 109/120 (90.8%) patients with stroke, those in the experimental and control groups were scored according to the FMA scale at baseline and 21 days, respectively, and the scores of the 2 groups were compared and statistically analyzed. The results showed that the experimental group outperformed the control group in terms of changes relative to baseline in Fugl-Meyer total scores, upper extremity scores, and lower extremity scores, and that the patients' upper and lower extremity motor abilities were better restored and improved, with significant improvement in upper extremity and total scores and some improvement in lower extremity scores. Other studies also found that lower extremity training improved motor function [35,36]. This clinical trial shows that the remote rehabilitation training system is used for the rehabilitation training of limb motor function of patients with stroke, and that the effect is better than that of routine clinical OT training.

In addition, in terms of safety, no adverse events related to this system occurred during the entire trial. Therefore, the designed remote rehabilitation training system based on wearable devices and human-computer interaction is used for rehabilitation training of patients with stroke and other limb motor dysfunction, which has good efficacy and good safety.

Comparison With Prior Work

According to the literature [37], long-term and specific rehabilitation training can maximize the recovery of patients' health and confidence. However, patients are less willing to participate in rehabilitation programs for daily repetitive and passive training [38,39]. By contrast, active training of patients is more effective than passive training and can enhance patient outcomes [40,41].

Chae et al [42] proposed a smartwatch and machine learning–based remote rehabilitation system for home training of patients' upper limbs. However, the number of patients was small, and the system is only for upper limb training; besides, the actual accuracy of home motion detection was not evaluated. Held et al [43] proposed a method of gait rehabilitation for patients with stroke, combining mobile augmented reality technology and sensor technology to adjust and train patients to walk. However, the device requires set up and calibration, making it more difficult for patients to use. The use of robotic technology for the rehabilitation of patients with stroke has been greatly developed. Ren et al [13] developed a wearable ankle joint rehabilitation robot to perform active and passive training on patients, but only for patients with acute stroke requiring ankle rehabilitation. Zhang et al [44] designed a desktop rehabilitation robot to train and evaluate the motor function of the upper limbs of patients.

Most experimental systems are complicated to use, expensive, inconvenient for patients to perform home training, and have few training movements, and therefore, they cannot undergo comprehensive training for the whole body. In addition, most of the aforesaid studies were performed under the supervision of physicians on-site, and cannot be applied in remote environments such as in home.

In similar clinical trials of the efficacy of home remote rehabilitation, Cramer et al [45] conducted a comparison trial with clinical rehabilitation modalities for patients with stroke having upper extremity motor deficits and showed that activity-based training significantly improved arm motor function, but the trial was only for upper extremity and lacked further analysis for patients requiring lower extremity rehabilitation. In a trial comparing lower extremity rehabilitation, Kang et al [46] compared patients' activities of daily living abilities through treadmill training, and reported that the Nordic treadmill training was an effective aid. However, the trial was performed for patients with mild issues under the supervision of the therapist, and there was no random allocation method, so caution should be exercised when interpreting the findings. In terms of human-computer interaction, Lee's study [35] found that mobile phone–based virtual reality applied to patients' stationary bicycle training improved lower extremity motor function recovery, but the movement of both legs was easily dominated by the healthy side of the body and lacked targeted training for the affected side of the body.

Therefore, the wearable remote rehabilitation training system for patients with stroke designed in this study can effectively overcome the aforesaid technology problems. Besides, the system was further designed and optimized based on the previous versions. Consequently, patients can receive effective training and guidance at home or in the community. In addition, the effectiveness and safety of the designed stroke active rehabilitation training system were verified by analyzing the results of the finalized clinical trial of 109 patients with stroke.

Limitations and Prospects

During the clinical experiment, almost all patients in the experimental group and rehabilitation physicians expressed strong interest in the designed rehabilitation training system owing to wearable devices and human-computer interaction training games.

However, according to the recommendations of rehabilitation physicians and patients, the system still has some limitations and needs further improvement for its better application in remote and home environments. In future work, the following improvements will be made.

First, according to the patients' suggestion, the size/resolution of the standard training video on the software interface needs to be increased, with the action details and precautions also displayed, to facilitate the patient to standardize the rehabilitation training according to the standard video. Second, the software needs to have built-in instructions and videos on how to wear the wearable device so that patients who are unfamiliar with the system can adapt more quickly and actively participate in rehabilitation training. Third, we need to add more rehabilitation training actions and more human-computer interaction sports games in daily life scenarios to meet the needs for more refined and diversified rehabilitation training.

In terms of the experimental design, the following limitations apply:

1. The pilot was set up in a hospital rehabilitation hall rather than in a remote and decentralized home setting to more fully assess the effectiveness of patient rehabilitation training and the overall management of the rehabilitation process.

2. Considering that hospital patients usually recover in the hospital for about 3 weeks, the trial was shorter than other studies [47,48]. After the trial, only 3 months of telephone follow-up was conducted for patients, and no abnormalities related to the trial were found. The clinical follow-up results are not taken into account in this study, which is one of the limitations of the design scheme of this study.

3. This trial only studied the patients' performance in the FMA scale. The follow-up research will include the Activities of Daily Living scale, the Wolf Motor Function Test, the patients' psychological status and satisfaction level, the impact of stroke publicity and education, among others, to further explore the rehabilitation effect of the remote rehabilitation training system.

Conclusions

This study found that the use of the remote intelligent rehabilitation training system designed based on wearable devices and human-computer interaction training tasks has a significant effect on the rehabilitation of motor function of patients with stroke, which can replace routine clinical OT training and improve the motivation, compliance, and rehabilitation effect of the training. In the future, improvements to the system will be made based on physician and patient recommendations, and through the medical device registration certification, it will be used without the participation of physicians or therapists, such as in rehabilitation training halls, and in remote environments, such as communities and homes.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (2018YFC1313602), and the clinical trial was approved by the Ethics Committee of Tangdu Hospital and the Ethics Committee of Xi'an Gaoxin Hospital. All patients participating in the clinical trial research have signed the informed consent. We thank them for their contributions to this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 584 KB - mhealth_v11i1e40416_app1.pdf]

References

1. Lin B, Zhang Z, Mei Y, Wang C, Xu H, Liu L, et al. Cumulative risk of stroke recurrence over the last 10 years: a systematic review and meta-analysis. Neurol Sci 2021 Jan;42(1):61-71. [doi: 10.1007/s10072-020-04797-5] [Medline: 33040195]
2. Wang LD, Chang JL, Jiao YH. Report on Stroke Prevention and Treatment in China 2021. Beijing, China: China Population Publishing House; 2022.
3. Dobkin BH. Strategies for stroke rehabilitation. The Lancet Neurology 2004 Sep;3(9):528-536. [doi: 10.1016/s1474-4422(04)00851-8]
4. Langan J, Delave K, Phillips L, Pangilinan P, Brown S. Home-based telerehabilitation shows improved upper limb function in adults with chronic stroke: a pilot study. J Rehabil Med 2013 Feb;45(2):217-220 [FREE Full text] [doi: 10.2340/16501977-1115] [Medline: 23319181]
5. Peng Q, Park HS, Zhang LQ. A low-cost portable tele-rehabilitation system for the treatment and assessment of the elbow deformity of stroke patients. 2005 Presented at: ICORR 2005: 9th International Conference on Rehabilitation Robotics, 2005; June 28 - July 1, 2005; Chicago, IL p. 149-151. [doi: 10.1109/icorr.2005.1501072]
6. Hsieh Y, Chang K, Hung J, Wu C, Fu M, Chen C. Effects of Home-Based Versus Clinic-Based Rehabilitation Combining Mirror Therapy and Task-Specific Training for Patients With Stroke: A Randomized Crossover Trial. Arch Phys Med Rehabil 2018 Dec;99(12):2399-2407. [doi: 10.1016/j.apmr.2018.03.017] [Medline: 29702070]
7. Hiengkaew V, Vongsirinavarat M. Home-Based Physical Therapy for Individuals With Stroke in Thailand. Home Health Care Management & Practice 2016 Jul 09;28(4):209-215. [doi: 10.1177/1084822316630278]
8. Wilkins KB, Owen M, Ingo C, Carmona C, Dewald JPA, Yao J. Neural Plasticity in Moderate to Severe Chronic Stroke Following a Device-Assisted Task-Specific Arm/Hand Intervention. Front Neurol 2017 Jun 14;8:284 [FREE Full text] [doi: 10.3389/fneur.2017.00284] [Medline: 28659863]
9. Muñoz Esquivel K. Trends in Telemedicine and E-Health 2018 Aug 27;1(1):1-19 [FREE Full text] [doi: 10.31031/tteh.2018.01.000503]
10. Jurkiewicz MT, Marzolini S, Oh P. Adherence to a home-based exercise program for individuals after stroke. Top Stroke Rehabil 2011 Jan 05;18(3):277-284. [doi: 10.1310/tsr1803-277] [Medline: 21642065]
11. Roderick P, Low J, Day R, Peasgood T, Mullee M, Turnbull J, et al. Stroke rehabilitation after hospital discharge: a randomized trial comparing domiciliary and day-hospital care. Age Ageing 2001 Jul;30(4):303-310. [doi: 10.1093/ageing/30.4.303] [Medline: 11509308]
12. Calvaresi D, Marinoni M, Dragoni AF, Hilfiker R, Schumacher M. Real-time multi-agent systems for telerehabilitation scenarios. Artif Intell Med 2019 May;96:217-231. [doi: 10.1016/j.artmed.2019.02.001] [Medline: 30827696]

13. Ren Y, Wu Y, Yang C, Xu T, Harvey RL, Zhang L. Developing a Wearable Ankle Rehabilitation Robotic Device for in-Bed Acute Stroke Rehabilitation. IEEE Trans Neural Syst Rehabil Eng 2017 Jun;25(6):589-596. [doi: 10.1109/TNSRE.2016.2584003] [Medline: 27337720]
14. Yang G, Deng J, Pang G, Zhang H, Li J, Deng B, et al. An IoT-Enabled Stroke Rehabilitation System Based on Smart Wearable Armband and Machine Learning. IEEE J Transl Eng Health Med 2018;6:2100510 [FREE Full text] [doi: 10.1109/JTEHM.2018.2822681] [Medline: 29805919]
15. Kim SH, Huizenga DE, Handzic I, Ditwiler RE, Lazinski M, Ramakrishnan T, et al. Relearning functional and symmetric walking after stroke using a wearable device: a feasibility study. J Neuroeng Rehabil 2019 Aug 28;16(1):106 [FREE Full text] [doi: 10.1186/s12984-019-0569-x] [Medline: 31455358]
16. Yang C, Hsu Y. A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors (Basel) 2010 Aug 20;10(8):7772-7788 [FREE Full text] [doi: 10.3390/s100807772] [Medline: 22163626]
17. Gebruers N, Vanroy C, Truijen S, Engelborghs S, De Deyn PP. Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures. Arch Phys Med Rehabil 2010 Feb;91(2):288-297. [doi: 10.1016/j.apmr.2009.10.025] [Medline: 20159136]
18. Zhang S, McCullagh P, Zhang J, Yu T. A smartphone based real-time daily activity monitoring system. Cluster Comput 2014 Jan 1;17(3):711-721. [doi: 10.1007/s10586-013-0335-y]
19. Dobkin BH. A Rehabilitation-Internet-of-Things in the Home to Augment Motor Skills and Exercise Training. Neurorehabil Neural Repair 2017 Mar 24;31(3):217-227 [FREE Full text] [doi: 10.1177/1545968316680490] [Medline: 27885161]
20. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May 5;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
21. Ordóñez FJ, Roggen D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors (Basel) 2016 Jan 18;16(1):115 [FREE Full text] [doi: 10.3390/s16010115] [Medline: 26797612]
22. Chen Y, Xue Y. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. 2015 Presented at: 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015; October 9-12, 2015; Hong Kong City, Hong Kong p. 1488-1492. [doi: 10.1109/smc.2015.263]
23. Placidi G, Avola D, Ferrari M, Iacoviello D, Petracca A, Quaresima V, et al. A low-cost real time virtual system for postural stability assessment at home. Comput Methods Programs Biomed 2014 Nov;117(2):322-333. [doi: 10.1016/j.cmpb.2014.06.020] [Medline: 25070756]
24. Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. J Neuroeng Rehabil 2014 Jul 03;11:108 [FREE Full text] [doi: 10.1186/1743-0003-11-108] [Medline: 24996956]
25. Rau C, Chen Y, Lai J, Chen S, Kuo T, Jaw F, et al. Low-cost tele-assessment system for home-based evaluation of reaching ability following stroke. Telemed J E Health 2013 Dec;19(12):973-978 [FREE Full text] [doi: 10.1089/tmj.2012.0300] [Medline: 24138613]
26. Sarfo FS, Ulasavets U, Opare-Sem OK, Ovbiagele B. Tele-Rehabilitation after Stroke: An Updated Systematic Review of the Literature. J Stroke Cerebrovasc Dis 2018 Sep;27(9):2306-2318 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2018.05.013] [Medline: 29880211]
27. Cramer SC, Dodakian L, Le V, McKenzie A, See J, Augsburger R, et al. A Feasibility Study of Expanded Home-Based Telerehabilitation After Stroke. Front Neurol 2020 Feb 3;11:611453 [FREE Full text] [doi: 10.3389/fneur.2020.611453] [Medline: 33613417]
28. Huang J, Gong WJ. Changes of lower extremity scores in different rehabilitation methods in control group and experimental group M. Beijing, China: People's Health Publishing House; 2019:283.
29. National Information Technology Standardization Technical Committee, Administration of Quality Supervision, Inspection and Quarantine of People's Republic of China, Standardization Administration of China. Software engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Guide to SQuaRE[S]. 2010. URL: https://www.chinesestandard.net/PDF/English.aspx/GBT25000.1-2010 [accessed 2022-12-29]
30. Sun ZQ. Medical Statistics. Beijing, China: People's Health Publishing House; 2014:167.
31. Zhang WW, Speare S, Churilov L, Thuy M, Donnan G, Bernhardt J. Stroke rehabilitation in China: a systematic review and meta-analysis. Int J Stroke 2014 Jun;9(4):494-502. [doi: 10.1111/ijs.12029] [Medline: 23621107]
32. Li JA, Huang XL. Rehabilitation Medicine. Beijing, China: People's Health Publishing House; 2020:80.
33. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, American Heart Association Stroke Council, Council on Cardiovascular Stroke Nursing, Council on Clinical Cardiology, Council on Quality of Care Outcomes Research. Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2016 Jun;47(6):e98-e169. [doi: 10.1161/STR.0000000000000098] [Medline: 27145936]
34. Lin J, Hsueh I, Sheu C, Hsieh C. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. Clin Rehabil 2004 Jun 01;18(4):391-397. [doi: 10.1191/0269215504cr737oa] [Medline: 15180122]

35. Lee K. Speed-Interactive Pedaling Training Using Smartphone Virtual Reality Application for Stroke Patients: Single-Blinded, Randomized Clinical Trial. Brain Sci 2019 Oct 27;9(11):295 [FREE Full text] [doi: 10.3390/brainsci9110295] [Medline: 31717888]
36. Lv X, Chen H. Effect of Virtual Reality Combined with Intelligent Exercise Rehabilitation Machine on the Nursing Recovery of Lower Limb Motor Function of Patients with Hypertensive Stroke. J Healthc Eng 2022;2022:2106836. [doi: 10.1155/2022/2106836] [Medline: 35281535]
37. Wade D. Rehabilitation - a new approach. Part four: a new paradigm, and its implications. Clin Rehabil 2016 Feb 29;30(2):109-118. [doi: 10.1177/0269215515601177] [Medline: 26715679]
38. Dipietro L, Krebs HI, Fasoli SE, Volpe BT, Stein J, Bever C, et al. Changing motor synergies in chronic stroke. J Neurophysiol 2007 Aug;98(2):757-768 [FREE Full text] [doi: 10.1152/jn.01295.2006] [Medline: 17553941]
39. van Duijn M, Lötters F, Burdorf A. Influence of modified work on return to work for employees on sick leave due to musculoskeletal complaints. J Rehabil Med 2005 May;37(3):172-179 [FREE Full text] [doi: 10.1080/16501970410023434] [Medline: 16040475]
40. Van Peppen RP, Kwakkel G, Wood-Dauphinee S, Hendriks HJ, Van der Wees PJ, Dekker J. The impact of physical therapy on functional outcomes after stroke: what's the evidence? Clin Rehabil 2004 Dec 01;18(8):833-862. [doi: 10.1191/0269215504cr843oa] [Medline: 15609840]
41. Hölmich P, Uhrskou P, Ulnits L, Kanstrup I, Nielsen MB, Bjerg AM, et al. Effectiveness of active physical training as treatment for long-standing adductor-related groin pain in athletes: randomised trial. The Lancet 1999 Feb;353(9151):439-443. [doi: 10.1016/s0140-6736(98)03340-6]
42. Chae SH, Kim Y, Lee K, Park H. Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e17216 [FREE Full text] [doi: 10.2196/17216] [Medline: 32480361]
43. Held JPO, Yu K, Pyles C, Veerbeek JM, Bork F, Heining S, et al. Augmented Reality-Based Rehabilitation of Gait Impairments: Case Report. JMIR Mhealth Uhealth 2020 May 26;8(5):e17804 [FREE Full text] [doi: 10.2196/17804] [Medline: 32452815]
44. Zhang M, Chen J, Ling Z, Zhang B, Yan Y, Xiong D, et al. Quantitative Evaluation System of Upper Limb Motor Function of Stroke Patients Based on Desktop Rehabilitation Robot. Sensors (Basel) 2022 Feb 03;22(3):1170 [FREE Full text] [doi: 10.3390/s22031170] [Medline: 35161913]
45. Cramer SC, Dodakian L, Le V, See J, Augsburger R, McKenzie A, National Institutes of Health StrokeNet Telerehab Investigators. Efficacy of Home-Based Telerehabilitation vs In-Clinic Therapy for Adults After Stroke: A Randomized Clinical Trial. JAMA Neurol 2019 Sep 01;76(9):1079-1087 [FREE Full text] [doi: 10.1001/jamaneurol.2019.1604] [Medline: 31233135]
46. Kang T, Lee J, Cynn H. Six-Week Nordic Treadmill Training Compared with Treadmill Training on Balance, Gait, and Activities of Daily Living for Stroke Patients: A Randomized Controlled Trial. J Stroke Cerebrovasc Dis 2016 Apr;25(4):848-856. [doi: 10.1016/j.jstrokecerebrovasdis.2015.11.037] [Medline: 26796052]
47. Torriani-Pasin C, Demers M, Polese JC, Bishop L, Wade E, Hempel S, et al. mHealth technologies used to capture walking and arm use behavior in adult stroke survivors: a scoping review beyond measurement properties. Disabil Rehabil 2022 Oct;44(20):6094-6106. [doi: 10.1080/09638288.2021.1953623] [Medline: 34297652]
48. Torriani-Pasin C, Demers M, Polese J, Bishop L, Wade E, Hempel S, et al. Measurement Properties of mHealth Technologies to Capture Functional Movement Behaviors in Stroke Survivors: a Scoping Review. Archives of Physical Medicine and Rehabilitation 2021 Oct;102(10):e115. [doi: 10.1016/j.apmr.2021.07.462]

Abbreviations

AMP: amplitude
FMA: Fugl-Meyer Assessment
IMU: inertial measurement unit
MEAN: mean value
OT: occupational therapy
PT: physical therapy
RMS: root-mean-square
SQuaRE: Software Engineering-Software Product Quality Requirements and Evaluation

Edited by L Buis; submitted 20.06.22; peer-reviewed by KS Sahu, B Dobkin; comments to author 25.07.22; revised version received 19.10.22; accepted 09.12.22; published 23.02.23.

Please cite as:
Guo L, Wang J, Wu Q, Li X, Zhang B, Zhou L, Xiong D
Clinical Study of a Wearable Remote Rehabilitation Training System for Patients With Stroke: Randomized Controlled Pilot Trial
JMIR Mhealth Uhealth 2023;11:e40416
URL: https://mhealth.jmir.org/2023/1/e40416
doi:10.2196/40416
PMID:36821348

Original Paper

Engagement and Utilization of a Complete Remote Digital Care Program for Musculoskeletal Pain Management in Urban and Rural Areas Across the United States: Longitudinal Cohort Study

Justin Scheer[1], MD; Anabela C Areias[2], PhD; Maria Molinos[2], PhD; Dora Janela[2], BSc; Robert Moulder[3], PhD; Jorge Lains[4,5], MD; Virgílio Bento[2], PhD; Vijay Yanamadala[2,6,7], MD; Fernando Dias Correia[2,8], MD, PhD; Fabíola Costa[2], PhD

[1]Department of Neurological Surgery, University of California, San Francisco, CA, United States

[2]Sword Health Inc, Draper, UT, United States

[3]Institute for Cognitive Science, University of Colorado Boulder, Boulder, CO, United States

[4]Rovisco Pais Medical and Rehabilitation Centre, Coimbra, Portugal

[5]Faculty of Medicine, Coimbra University, Coimbra, Portugal

[6]Department of Neurosurgery, Hartford Healthcare Medical Group, Westport, CT, United States

[7]Department of Surgery, Frank H Netter School of Medicine, Quinnipiac University, Hamden, CT, United States

[8]Neurology Department, Centro Hospitalar e Universitário do Porto, Porto, Portugal

Corresponding Author:
Fabíola Costa, PhD
Sword Health Inc
13937 Sprague Lane
Draper, UT, 84020
United States
Phone: 1 1 385 308 8034
Fax: 1 1 801 206 3433
Email: f.costa@swordhealth.com

Abstract

Background: Musculoskeletal (MSK) conditions are the number one cause of disability worldwide. Digital care programs (DCPs) for MSK pain management have arisen as alternative care delivery models to circumvent challenges in accessibility of conventional therapy. Despite the potential of DCPs to reduce inequities in accessing care, the outcomes of such interventions in rural and urban populations have yet to be studied.

Objective: The aim of this study was to assess the impact of urban or rural residency on engagement and clinical outcomes after a multimodal DCP for MSK pain.

Methods: This study consists of an ad hoc analysis of a decentralized single-arm investigation into engagement and clinical-related outcomes after a multimodal DCP in patients with MSK conditions. Patients were coded according to their zip codes to a specific rural-urban commuting area code and grouped into rural and urban cohorts. Changes in their engagement and clinical outcomes from baseline to program end were assessed. Latent growth curve analysis was performed to estimate change trajectories adjusting for the following covariates: age, gender, BMI, employment status, and pain acuity. Outcomes included engagement, self-reported pain, and the results of the Generalized Anxiety Disorder 7-item, Patient Health Questionnaire 9-item, and Work Productivity and Activity Impairment scales. A minimum clinically important difference (MCID) of 30% was considered for pain.

Results: Patients with urban and rural residency across the United States participated in the program (n=9992). A 73.8% (7378/9992) completion rate was observed. Both groups reported high satisfaction scores and similar engagement with exercise sessions, with rural residents showing higher engagement with educational content ($P<.001$) and higher program completion rates ($P=.02$). All groups showed a significant improvement in all clinical outcomes, including pain, mental health, and work productivity, without statistically significant intergroup differences. The percentage of patients meeting the MCID was similar in both groups (urban: 67.1%, rural: 68.3%; $P=.30$).

Conclusions: This study advocates for the utility of a DCP in improving access to MSK care in urban and rural areas alike, showcasing its potential to promote health equity. High engagement, satisfaction, and completion rates were noted in both groups, as well as significant improvements in clinical outcomes.

Trial Registration: ClinicalTrials.gov NCT04092946; https://clinicaltrials.gov/ct2/show/NCT04092946

(JMIR Mhealth Uhealth 2023;11:e44316) doi:10.2196/44316

KEYWORDS

physical therapy; physiotherapy; remote care; telerehabilitation; digital therapy; eHealth; telehealth; telemedicine; musculoskeletal; musculoskeletal conditions; urban; rural; pain; health inequity; digital care; pain management; clinical outcome; health equity; engagement

Introduction

Musculoskeletal (MSK) conditions are highly prevalent worldwide, resulting in significant disability and suffering [1], and were associated with up to US $380.9 billion of total medical expenditures in 2016 in the United States alone [2]. Exercise-based physical therapy is the mainstay treatment for such conditions [1,3-6]. Recently, telerehabilitation and digital physical therapy have emerged as alternative care delivery systems for a wide range of MSK conditions [7-10]. These alternative care delivery systems have shown to be effective and feasible compared to traditional physical therapy [11-18] while increasing access and affordability to patients and easing the burdens of conventional programs [19]. Accessibility is increased by reducing travel limitations and time barriers while eliminating any geographic restrictions. Additionally, digital therapy can increase compliance by allowing patients to undergo treatment at their convenience and at their own pace, increasing patient empowerment and self-management [8,10,20].

Despite the many benefits of telehealth, inequities in health other than underlying health status still exist based on age, geography, respective availability of health care facilities, and socioeconomic factors [21-27]. In fact, compared to those in urban areas, patients from rural areas tend to be older, are more likely to be obese, have higher rates of disability, have more chronic health conditions, and have higher fall rates [24,28]. Rural areas are known to have higher proportions of uninsured and underinsured individuals and higher costs of health care services when compared to urban areas [29]. Overall, 65% of rural US counties are designated as health professional shortage areas [30], and rural areas have lower patient-to-primary care physician ratios [31]. Patients in rural areas of the United States have fewer opportunities for in-person physical activity programs due to limited access to indoor facilities and limited transportation when compared to urban patients [32]. These inequities are further compounded by lower educational levels, higher rates of poverty, and lower rates of internet access in rural areas [24,33]. Also, individuals with limited or no digital literacy or with limited access to digital technology may not have the means to pursue and maintain a telehealth intervention [23]. It is therefore crucial to identify strategies for improved access and quality of physical therapy in these historically disinvested areas.

To our knowledge, no study has been conducted on the impact of urban or rural location on engagement and clinical outcomes following a telerehabilitation program for MSK conditions. We have previously reported a multimodal digital care program (DCP) combining exercise-based physical therapy with psychoeducational components, which provided a comprehensive approach to pain management. This program encourages patients to develop strategies and self-management skills to manage their pain and has been validated in several acute and chronic MSK conditions [15-17,34-36]. Additionally, the impact of race and ethnicity [37], as well as baseline mental health [38] and fear-avoidance beliefs [39], on final clinical outcomes have also been explored. The purpose of this study was to assess the impact of geographical location on engagement and clinical outcomes after a multimodal DCP, with the hypothesis being that patients from both rural and urban areas would have similar engagement and significant improvement in outcomes after program completion.

Methods

Study Design

This study is an ad hoc analysis of a decentralized, single-arm investigation into clinical and engagement-related outcomes following a multimodal DCP in patients with musculoskeletal (MSK) pain conditions. The DCP was administered at the patients' homes and delivered between March 1, 2021, and March 10, 2022.

Ethics Approval

This study is part of a trial that was prospectively registered on ClinicalTrials.gov (NCT04092946) on September 17, 2019, and approved by the New England Institutional Review Board (120190313) on June 18, 2020.

Population

The study population included adults (≥18 years of age) who were beneficiaries of employer health plans from 50 US states and the District of Columbia. Employees and their dependents who reported either acute or chronic MSK pain in the spine, upper limbs, or lower limbs were eligible and were invited to apply to the DCP of Sword Health (located in Draper, Utah) through a dedicated website. Throughout enrollment, participants were asked to provide demographic data, including zip codes and baseline clinical information (eg, initial pain levels). Participants were informed about the study and invited to provide consent. The exclusion criteria were as follows: (1) a health condition (eg, cardiac or respiratory) not allowing a participant to engage in at least 20 minutes of light to moderate exercise, (2) being under treatment for active cancer, and (3) rapid loss of strength or numbness in the arms or legs or change in bowel or urinary function in the previous 2 weeks.

Intervention

The DCP has been described previously [15-17,34-36]. In brief, this multimodal program consisted of 4-, 8-, or 12-week telerehabilitation interventions comprising exercise, education, and cognitive behavioral therapy (CBT). This program digitally interfaced between the patient and an assigned physical therapist (PT), who monitored the patient for the study duration. Participants who lacked internet access at home were given a Wi-Fi hotspot. A US Food and Drug Administration–listed class II medical device that consisted of inertial motion trackers, a mobile app in a dedicated tablet, and a cloud-based portal was made available to all patients. Briefly, the personalized exercises were displayed on the tablet, with trackers allowing real-time video and audio biofeedback on performance. At session end, the data related to the exercise sessions, such as compliance, presence or absence of movement errors, and level of pain and fatigue during the exercise, were registered and stored in a cloud-based portal. This portal enabled remote and asynchronous monitoring by the assigned PT, who revised the prescribed exercises if needed. Patients were recommended a frequency of 3 exercise sessions per week. The education and CBT components of the program were developed by a multidisciplinary team following current clinical guidelines and state-of-the-art research [40-44]. The education component delved into topics focused on anatomy, physiology, symptoms, evidence-based treatments, fear avoidance, and active coping skills (including managing feelings of anxiety and depression). The CBT program was based on third-generation techniques—mindfulness, acceptance, and commitment therapy; empathy-focused therapy; fear-avoidance behavior; and constructive coping. The education and CBT materials were delivered to the patients through written articles, audio content, and interactive modules. Bidirectional communication with the assigned PT was ensured through built-in secure chat within a smartphone app and through video calls. Participants who did not perform any exercise session for 28 consecutive days were considered dropouts.

Demographic Data

Demographic data included age, BMI, patient gender, educational level, and employment status. The gender category included "man," "woman," "nonbinary," "other," and "prefer not to specify." The employment status categories were defined as the following: full-time employed, part-time employed, or not employed. The educational levels were (1) high school or less (including technical or vocational training), (2) some college, including a bachelor's degree, community college, or an associate degree, (3) some graduate school, including a master's or doctoral degree, and (4) "not available" or "prefer not to answer."

Patients were coded according to their zip codes to a specific rural-urban commuting area (RUCA) code [45]. RUCA codes characterize all census areas regarding their rural and urban status and relationships. This classification system uses the standard Bureau of Census urbanized area and urban cluster definitions in combination with work-commuting information [46,47]. Rural areas have been defined as having an urban core of 50,000 people or less [24]. Therefore, using primary RUCA codes, we defined urban areas by scores from 1 to 3, and rural areas by aggregating codes 4 to 10 (Multimedia Appendix 1, Table S1 [47]).

Outcomes

Outcomes were collected at baseline and 4, 8, and 12 weeks, and mean changes were calculated between baseline and program end. Engagement and clinical outcomes are described in Table 1.

Table 1. Engagement and clinical outcomes in this study.

Outcome	Description
Engagement	Measured through the following: • Completion of the program (considered as the retention rate) • Number of completed exercise sessions over the 12-week digital care program • Weekly session frequency • Time spent performing exercise sessions • Articles read • Interactions with the physical therapist • Satisfaction, assessed through the question "On a scale from 0 to 10, how likely is it that you would recommend this intervention to a friend or neighbor?"
Numerical Pain Rating Scale [48,49]	Assessed through the question "Please rate your average pain over the last 7 days, from 0 (no pain at all) to 10 (worst pain imaginable)"; the number of patients reaching the minimum clinically important difference of 30% between baseline and treatment end was also assessed
Generalized Anxiety Disorder 7-item scale (range 0-21) [50]	Used to assess anxiety; higher scores are associated with worse outcomes
Patient Health Questionnaire 9-item scale (range 0-27) [51]	Used to assess depression; higher scores are associated with worse outcomes
WPAI[a] for general health questionnaire (version 2.0) [52]	Evaluated in employed participants to assess overall work impairment (WPAI overall: total presenteeism and absenteeism from work), presenteeism (WPAI work), absenteeism (WPAI time), and non–work-related activity impairment (WPAI activity); higher scores represent higher impairment.

[a]WPAI: Work Productivity and Activity Impairment.

Statistical Analysis

Analyses of baseline characteristics (demographics and clinical data), as well as engagement metrics, were performed using a 2-tailed, 2-sample *t* test, a 2-way analysis of variance (ANOVA) with Bonferroni post hoc test, a chi-square test, or a 2-proportion *z* test. Patients who completed the program were defined as "completers" and those that did not were defined as "noncompleters."

Latent growth curve analysis (LGCA) was used to estimate trajectories of each outcome over time [35,53]. The LGCA has been recognized as one of the most powerful methods to analyze longitudinal data, since it provides a measure of fitness and addresses missing data through full information maximum likelihood (FIML) [54-57]. FIML uses all available data at each time point from all participants to calculate maximum likelihood estimates, outperforming multiple imputation by chained equations or listwise deletion [58,59]. In addition, the LGCA uses a structural equation model to define trajectories through intercept, slope, and curvature for each variable, allowing analysis of the recovery pace and leveling of the effect for each outcome. In order to account for unbalanced group sizes, a multiple-group LGCA was conducted. This allows for creating separate models for rural and urban groups while simultaneously performing intergroup comparisons (eg, mean change). A conditional analysis was conducted to assess the influence of age, gender, BMI, employment status, and education level and was fitted as a random effect. Additionally, analysis of subpopulations was performed by focusing on participants who met the following criteria at baseline: Generalized Anxiety Disorder 7-item (GAD-7) and Patient Health Questionnaire 9-item (PHQ-9) scores equal or greater than 5 points [50,51] and a Work Productivity and Activity Impairment (WPAI; comprising overall, work, time, and activity) score greater than 0 points. A robust sandwich estimator was used in all models for standard errors.

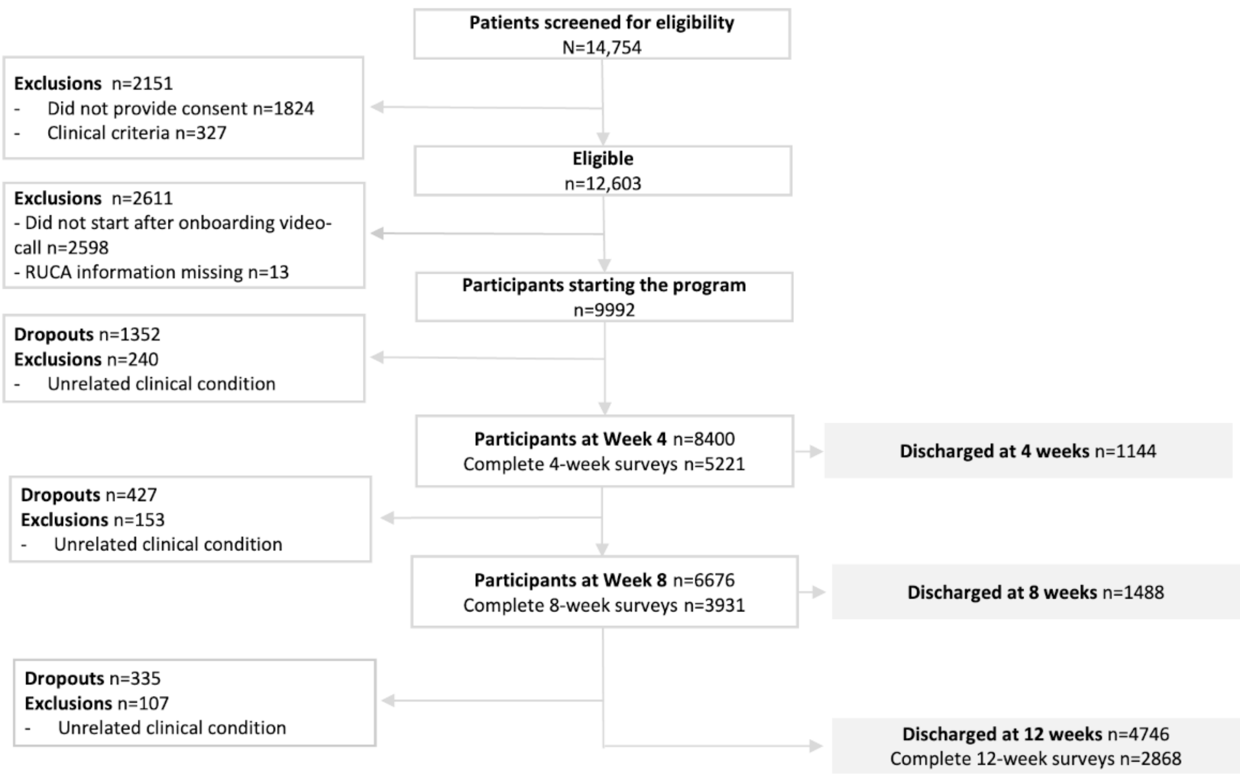
All statistical analyses were conducted using commercially available software (SPSS version 22; IBM Corp), and the level of significance was set at *P*<.05 for all tests. The LGCA was coded using R (version 4.2.2; R Foundation for Statistical Computing).

Results

Participant Inclusion

A total of 14,754 participants were screened for eligibility (Figure 1). Of these, 2151 were excluded, for a total of 12,603 (85.4%) eligible patients, of whom 2611 were excluded due to unavailable RUCA data or not starting the program, resulting in a total of 9992 patients at program start. A total of 7378 of 9992 (73.8%) patients completed the program. The study flow diagram is presented in Figure 1.

Figure 1. Study flow chart showing the number of participants who were excluded, included, and dropped out. RUCA: rural-urban commuting area.



Baseline Characteristics

Patients' baseline demographics grouped by urban and rural areas are presented in Table 2, while baseline characteristics stratified by completers and noncompleters can be found in Multimedia Appendix 1, Table S2.

Table 2. Baseline characteristics for urban and rural groups following an intention-to-treat analysis. Filtered cases correspond to participants who reported relevant impairment at baseline (>0 or ≥5 points). Statistically significant *P* values are italicized.

	Total (n=9992)	Urban (n=8809)	Rural (n=1183)	*P* value
Age (years), mean (SD)	48.55 (12.45)	48.11 (12.37)	51.85 (12.57)	*<.001*
Age categories (years), n (%)				*<.001*
<25	127 (1.3)	114 (1.3)	13 (1.1)	
25-40	2753 (27.6)	2520 (28.6)	233 (19.7)	
40-60	5279 (52.8)	4649 (52.8)	630 (53.3)	
>60	1833 (18.3)	1526 (17.3)	307 (26)	
BMI (kg/m^2), mean (SD)	29.18 (6.74)	28.96 (6.60)	30.83 (7.48)	*<.001*
BMI categories (kg/m^2), n (%)				*<.001*
Underweight (<18.5)	90 (0.9)	84 (1)	6 (0.5)	
Normal (18.5-25)	2798 (28)	2548 (28.9)	250 (21.1)	
Overweight (25-30)	3373 (33.8)	3011 (34.2)	362 (30.6)	
Obese (30-40)	2957 (29.6)	2525 (28.7)	432 (36.5)	
Obese grade III (>40)	743 (7.4)	614 (7)	129 (10.9)	
Gender, n (%)				.12
Woman	5502 (55.1)	4818 (54.7)	684 (57.8)	
Man	4457 (44.6)	3963 (45)	494 (41.8)	
Nonbinary	24 (0.2)	19 (0.2)	5 (0.4)	
Other	3 (0)	3 (0)	0 (0)	
Prefer not to specify	6 (0.1)	6 (0.1)	0 (0)	
Employment status, n (%)				*<.001*
Employed full-time	6271 (62.8)	5616 (63.8)	655 (55.4)	
Employed part-time	2348 (23.5)	2076 (23.6)	272 (23)	
Not employed	1067 (10.7)	853 (9.7)	214 (18.1)	
Education level, n (%)				*.001*
High school or less	866 (8.7)	700 (7.9)	166 (14)	
Some college, including bachelor's or associate degree	4543 (45.5)	4031 (45.8)	512 (43.3)	
Some graduate school, including master's or doctoral degree	2082 (20.8)	1876 (21.3)	206 (17.4)	
Not available or prefer not to answer	2501 (25)	2202 (25)	299 (25.3)	
Acuity, n (%)[a]				*<.001*
Acute	2147 (21.5)	1952 (22.2)	195 (16.5)	
Chronic	7845 (78.5)	6857 (77.8)	988 (83.5)	
Anatomical pain region, n (%)				*.003*
Ankle	422 (4.2)	380 (4.3)	42 (3.6)	
Elbow	286 (2.9)	259 (2.9)	27 (2.3)	
Hip	900 (9)	786 (8.9)	114 (9.6)	
Knee	1438 (14.4)	1292 (14.7)	146 (12.3)	
Low back	3976 (39.8)	3441 (39.1)	535 (45.2)	
Neck	936 (9.4)	834 (9.5)	102 (8.6)	
Shoulder	1632 (16.3)	1461 (16.6)	171 (14.5)	

	Total (n=9992)	Urban (n=8809)	Rural (n=1183)	*P* value
Wrist or hand	402 (4)	356 (4)	46 (3.9)	
Clinical outcomes (score)				
Pain, mean (SD)	4.83 (1.99)	4.83 (1.99)	4.85 (1.98)	.72
GAD-7[b] ≥5, n (%)	2751 (27.5)	2430 (27.6)	321 (27.1)	.74
GAD-7 ≥5, mean (SD)	8.89 (4.08)	8.89 (4.07)	8.89 (4.19)	.99
GAD-7, mean (SD)	3.03 (4.35)	3.04 (4.35)	2.98 (4.37)	.69
PHQ-9[c] ≥5, n (%)	2071 (20.7)	1790 (20.3)	281 (23.8)	*.006*
PHQ-9 ≥5, mean (SD)	9.21 (4.30)	9.20 (4.29)	9.31 (4.41)	.69
PHQ-9, mean (SD)	2.37 (4.15)	2.33 (4.11)	2.7 (4.41)	*.004*
WPAI[d] overall >0, mean (SD)	29.89 (20.11)	29.9 (20.11)	29.80 (20.16)	.91
WPAI overall, mean (SD)	17.32 (21.26)	17.23 (21.25)	17.9 (21.39)	.34
WPAI work >0, mean (SD)	28.63 (18.80)	28.62 (18.77)	28.71 (19.05)	.91
WPAI work, mean (SD)	16.27 (20.05)	16.16 (20.0)	17.04 (20.35)	.21
WPAI time >0, mean (SD)	18.08 (18.07)	18.45 (18.45)	15.38 (14.94)	.11
WPAI time, mean (SD)	1.91 (8.07)	1.94 (8.22)	1.65 (6.82)	.31
WPAI activity >0, mean (SD)	37.37 (22.86)	37.33 (22.85)	37.70 (22.91)	.65
WPAI activity, mean (SD)	29.04 (25.46)	28.92 (25.45)	29.93 (25.48)	.20
Medications, n (%)	2364 (23.7)	2062 (23.5)	302 (25.6)	.11

[a]A total of 1.1% (114/9992) of patients were postsurgical.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PHQ-9: Patient Health Questionnaire 9-item scale.

[d]WPAI: Work Productivity and Activity Impairment scale.

Patients from rural areas were significantly older than patients from urban areas (51.85, SD 12.57 years vs 48.11, SD 12.37 years, respectively; *P*<.001), had higher BMI (30.83, SD 7.48 kg/m^2 vs 28.96, SD 6.60 kg/m^2, respectively), had a higher percentage of unemployed individuals (214/1183, 18.1% vs 853/8809, 9.7%, respectively; *P*<.001), and had a higher percentage of individuals with a lower educational level (166/1183, 14% vs 700/8809, 7.9%, respectively; *P*<.001). No statistically significant differences in gender distribution were found between groups.

Patients in rural areas also presented a higher prevalence of chronic pain and low back pain conditions than patients from urban areas (535/1183, 45.2% vs 3441/8809, 39.1%, respectively). In opposition, a higher prevalence of knee-related conditions was observed in patients from urban areas than from rural areas (1292/8809, 14.7% vs 146/1183, 12.3%, respectively; *P*<.001).

Overall, similar clinical metrics were observed between patients from rural and urban areas at baseline. The statistically significant differences found were in baseline depression (rural PHQ-9 score 2.70, SD 4.41 vs urban PHQ-9 score 2.33, SD 4.22; *P*=.004) and percentage of patients with at least mild or moderate depression at baseline (281/1183, 23.8% vs 1790/8809, 20.3%, respectively; *P*=.006), where more patients from rural areas were depressed.

When comparing completers with noncompleters, the latter were younger (46.23, SD 12.61 years vs 49.38, SD 12.29 years, respectively; *P*<.001) and reported higher BMI (30.22, SD 7.36 kg/m^2 vs 28.81, SD 6.47 kg/m^2, respectively; *P*<.001). A larger proportion of noncompleters were employed full-time (1732/2614, 66.3% vs 4539/7378, 61.5%, respectively; *P*<.001) and reported a lower educational level. Noncompleters also reported higher levels of impairment in productivity (WPAI overall, *P*=.004 and WPAI work, *P*=.001) and non–work-related activities (WPAI activity, *P*=.01) at baseline.

Engagement

Individuals from rural areas were more likely to complete the program than patients from urban areas (906/1183, 76.6% vs 6472/8809, 73.5%, respectively; *P*=.02). However, independently of dropout rates, both groups had a similar pattern of engagement. Engagement data stratified by patients from urban and rural areas is presented in Table 3. The 2 groups had similar time dedicated to exercise (*P*=.48), number of sessions (*P*=.77), sessions per week (*P*=.11), and interactions with the PT (*P*=.14). Average satisfaction scores were similarly high in both groups (rural score 8.6, SD 1.7 and urban score 8.6, SD 1.8; *P*=.95). The single significant difference in engagement between groups was the number of educational articles consulted, with patients from rural areas reading more articles than patients from urban areas (*P*<.001; Table 3).

Table 3. Engagement data across the groups. Statistically significant *P* values are italicized.

Engagement outcomes	Urban, mean (SD)	Rural, mean (SD)	*P* value
Sessions, n	33.88 (32.23)	34.18 (30.94)	.77
Sessions per week, n	2.76 (1.14)	2.78 (1.12)	.11
Training time, minutes	472.02 (485.56)	482.54 (485.96)	.48
Articles read, n	2.72 (5.27)	3.44 (6.18)	*<.001*
Interactions with physical therapist, n	11.79 (12.54)	12.37 (13.90)	.14
Average satisfaction score	8.6 (1.7)	8.6 (1.8)	.95

Clinical Outcomes

Mean changes in clinical outcomes for both urban and rural groups following an intent-to-treat analysis are presented in Table 4, while the corresponding model estimates and model fitness are presented in Multimedia Appendix 1, Tables S3 and S4, respectively [60,61]. Change trajectories of each outcome are depicted in Figure 2. The impact of the covariates in clinical outcomes is presented in Multimedia Appendix 1, Table S5. The same analysis following a per-protocol approach is presented in Multimedia Appendix 1, Tables S6-S8. Similar results were observed from both intention-to-treat and per-protocol approaches. Since intention-to-treat analysis offers an overview change of the entire cohort, the following section is focused on these results.

Table 4. Mean changes between baseline and program end and mean differences between groups for the studied clinical outcomes following an intent-to-treat analysis. Statistically significant *P* values are italicized.

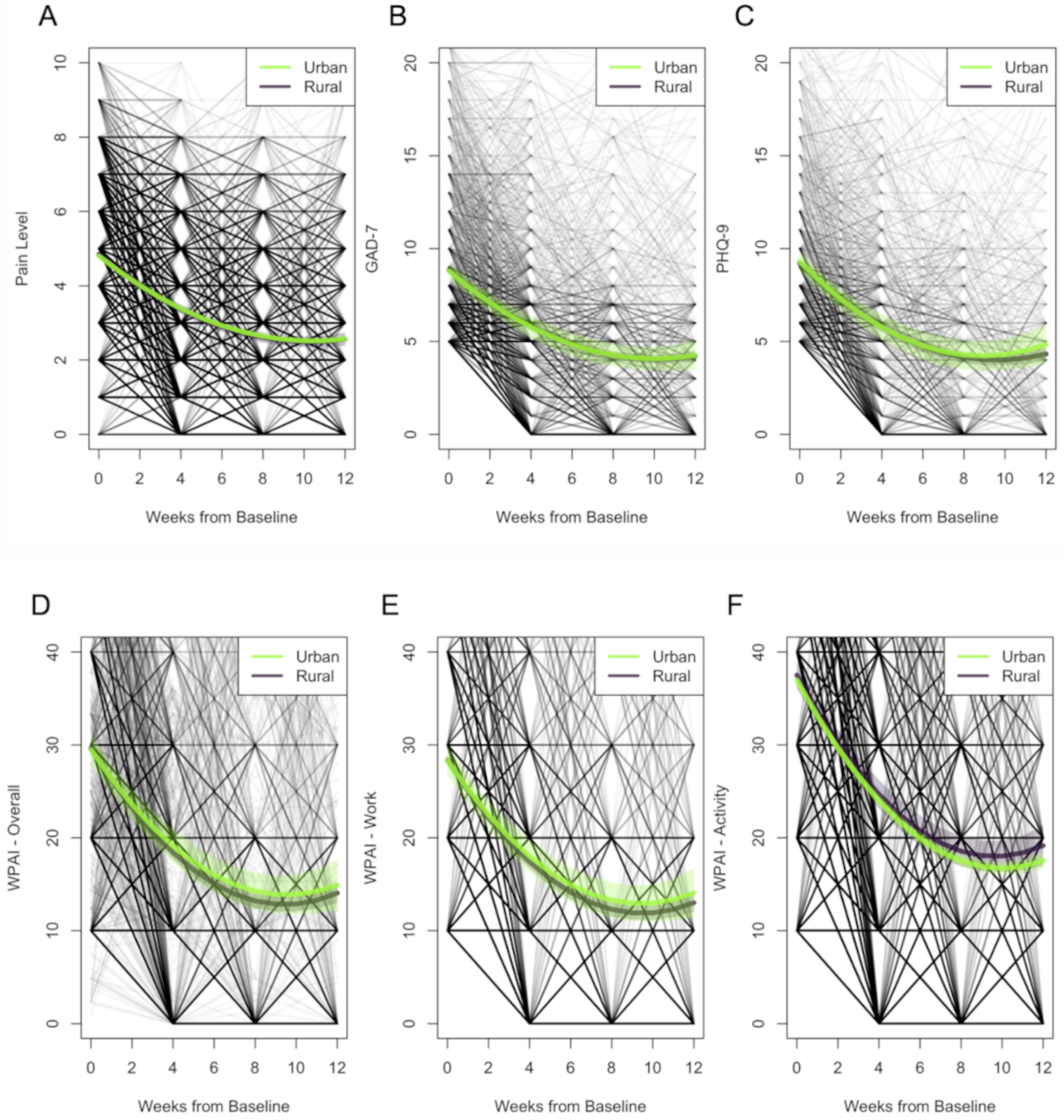
Scores	Urban		Rural		Mean difference	
	Mean change (95% CI)	*P* value	Mean change (95% CI)	*P* value	Difference (95% CI)	*P* value
Pain	2.2 (2.2 to 2.3)	*<.001*	2.3 (2.1 to 2.5)	*<.001*	–0.1 (–0.3 to 0.2)	.62
GAD-7[a]	1.26 (1.16 to 1.37)	*<.001*	1.16 (0.86 to 1.47)	*<.001*	0.1 (–0.22 to 0.43)	.53
GAD-7 ≥5	4.5 (3.7 to 5.4)	*<.001*	4.6 (4.3 to 4.9)	*<.001*	–0.1 (–1.0 to 0.8)	.85
PHQ-9[b]	0.93 (0.82 to 1.03)	*<.001*	1.14 (0.84 to 1.45)	*<.001*	–0.22 (–0.54 to 0.1)	.19
PHQ-9 ≥5	4.5 (3.39 to 5.53)	*<.001*	4.9 (4.5 to 5.2)	*<.001*	–0.41 (–1.5 to 0.7)	.48
WPAI[c] overall	7.37 (6.65 to 8.09)	*<.001*	7.19 (5.28 to 9.11)	*<.001*	0.18 (–1.87 to 2.22)	.87
WPAI overall >0	14.6 (11.7 to 17.4)	*<.001*	15.6 (14.5 to 16.7)	*<.001*	–1.0 (–4.1 to 2.0)	.50
WPAI work	13.73 (12.99 to 14.47)	*<.001*	13.59 (11.67 to 15.51)	*<.001*	0.15 (–1.91 to 2.2)	.89
WPAI work >0	14.3 (11.6 to 17.0)	*<.001*	15.4 (14.3 to 16.4)	*<.001*	–1.1 (–4 to 1.8)	.46
WPAI time missed	7.14 (6.47 to 7.81)	*<.001*	6.82 (5 to 8.63)	*<.001*	0.32 (–1.62 to 2.25)	.75
WPAI time missed >0	11.4 (6.9 to 16.0)	*<.001*	11.8 (9.9 to 13.7)	*<.001*	–0.35 (–5.2 to 4.5)	.89
WPAI activity	0.66 (0.37 to 0.96)	*<.001*	0.42 (–0.34 to 1.19)	.28	0.24 (–0.58 to 1.06)	.57
WPAI activity >0	19.4 (18.6 to 20.3)	*<.001*	18.4 (16.1 to 20.7)	*<.001*	1.06 (–1.4 to 3.5)	.40

[a]GAD-7: Generalized Anxiety Disorder 7-item scale.

[b]PHQ-9: Patient Health Questionnaire 9-item scale.

[c]WPAI: Work Productivity and Activity Impairment scale.

Figure 2. Longitudinal changes across time using intent-to-treat analysis. A: pain level; B and C: mental health (GAD-7-and PHQ-9 scores, respectively) for cases with at least mild or moderate anxiety or depression at baseline; D-F: work productivity (WPAI overall, WPAI work, and WPAI activity scores, respectively) for cases reporting impairment at baseline. The shaded areas are the 95% CI. GAD-7: Generalized Anxiety Disorder 7-item scale; PHQ-9: Patient Health Questionnaire 9-item scale; WPAI: Work Productivity and Activity Impairment scale.



Pain

Pain levels decreased similarly in both groups from baseline to program end (*P*=.62), corresponding to a significant mean change of 2.2 points (95% CI 2.2-2.3) in the urban cohort and 2.3 points (95% CI 2.1-2.5) in the rural cohort (Table 4 and Figure 2A). The rate of patients meeting pain MCID at program end was not statistically different between urban and rural cohorts (67.1% versus 68.3%, *P*=.30).

Mental Health

Among those who reported at least mild or moderate anxiety (GAD-7 score ≥5) at baseline, we observed a significant decrease in anxiety in both groups (urban score 4.5, 95% CI 3.7-5.4; *P*<.001 and rural score 4.6, 95% CI 4.3-4.9; *P*<.001); this decrease was similar in the groups (*P*=.85). Likewise, among those reporting at least mild or moderate depression at baseline (ie, PHQ-9 score ≥5), we observed a significant decrease in depression scores in both groups (urban score 4.5, 95% CI 3.39-5.53; *P*<.001 and rural score 4.9, 95% CI 4.5-5.2; *P*<.001); this decrease was also similar in the groups (*P*=.48). In patients from urban areas, chronic MSK conditions were associated with steeper initial recovery of depression, followed by a stronger leveling effect. In contrast, in patients from rural areas, chronic pain was associated with slower improvement that was more sustained over time (Figure 2C, difference in slope between groups –0.07, *P*=.004; difference in curve between groups 1.06, *P*<.001; Multimedia Appendix 1, Table S5).

Productivity

Productivity improvements were observed in both groups with no differences between them (Figure 2D-F). For WPAI overall, mean changes of 14.6 points (95% CI 11.7-17.4) and 15.6 points (95% CI 14.5-16.7) were reported for urban and rural groups, respectively (*P*=.50). Presenteeism, measured through WPAI work, improved by 14.3 points (95% CI 11.6-17.0) in the urban group and by 15.4 points (95% CI 14.3-16.4) in the rural group (*P*=.46). Absenteeism, measured through WPAI time missed, was reduced by 11.4 points (95% CI 6.9-16.0) and 11.8 points (95% CI 9.9-13.7) in the urban and rural groups, respectively. Similarly, impairment in non–work-related activities (ie, WPAI activity) had improvements of 19.4 points (95% CI 18.6-20.3) and 18.4 points (95% CI 16.1-20.7) in the urban and rural groups, respectively.

In urban areas, individuals with higher BMI reported a greater leveling effect on absenteeism improvement than those from rural areas (the difference in curve between groups was 0.14, *P*=.04; Multimedia Appendix 1, Table S5). In patients from rural areas, the presence of chronic pain was associated with a faster recovery in absenteeism in comparison with that in patients from urban areas (the difference in slope between groups was 0.16, *P*=.03; Multimedia Appendix 1, Table S5).

Discussion

Principal Findings

The multimodal DCP herein reported was able to reach all US states in both urban and rural locations and had a completion rate of 73.8% (7378/9992), which is similar to previous studies reporting the use of digital interventions for MSK pain management [37,62]. The percentage of participants in urban areas was significantly higher than in rural locations, which was expected given that only approximately 19.6% of the US population is located in rural areas according to the US Census Bureau [63].

Health inequities between urban and rural populations are prevalent in the United States. [64]. Rural populations have been reported to have demographics associated with a poorer prognosis for MSK pain [65-69]. In accord with this, in this study, patients from rural areas were older [67], had higher BMI [69], had lower educational levels [65], and had a higher prevalence of depression [67]. Some of these factors have previously been associated with lower chances of receiving physical therapy [33]. Lower education (and consequently lower digital literacy) have been associated with lower adherence, for example [21-23].

Studies have shown that patients in rural areas of the United States may face additional difficulties in recovery due to fewer opportunities for in-person physical activity programs as a consequence of limited access to indoor facilities, limited transportation, and a lower overall health status when compared to urban patients [24,32]. Additionally, rural residents are less likely to report having home broadband than those living in urban or suburban areas [70], which seriously impacts their access to digital health care tools and electronic communication with health providers [71].

In this study, engagement was similar between both rural and urban areas (eg, the number of sessions and interactions with a PT), and completion rates were higher in the rural cohort. The reasons behind these observations may be multifactorial, but one can speculate that the lack of access to alternative health care resources, as well as the provision of a Wi-Fi hotspot to those without internet, might have prompted patients from rural areas to not only engage with the exercise sessions but also to achieve higher completion rates [70,71]. Also, despite lower educational levels, patients from rural areas engaged more with curated health educational articles advocating for telerehabilitation programs as enablers of health literacy.

Despite the worse clinical outcomes reported at baseline by those in rural communities, in line with what has been described before [32,66-69], similar improvements in pain, mental health, and productivity impairment were observed in both groups, again reinforcing the notion that higher MSK pain burden in rural areas may be associated with lack of access to care. Pain improvements were above a 2-point change independently of the studied group, with 67.1% to 68.3% of participants meeting the MCID for pain [48,49]. The percentages of patients meeting the MCID were within the ranges previously reported for digital interventions (49%-75.6%) [72-74] and in-person physical therapy [75].

The prevalence of depression and anxiety has been reported to be higher in residents of rural areas compared to urban areas [76], with those from rural areas facing a shortage of mental health services [77]. Since mental health and MSK pain are tightly associated [38,78], the scarcity of psychological support can seriously impact the recovery rates of rural populations. This study confirms a higher depression burden in patients from rural areas but found similar improvement in mental health scores in both rural and urban patients, reinforcing the notion that lack of access to mental health resources may be the main driver for the higher burden of disease in rural areas. Additionally, MSK pain has been reported to be a main driver for loss of work productivity [79,80]. The factors weighing on absenteeism recovery were BMI and the presence of chronic pain, both previously reported to negatively impact MSK pain recovery [81,82]. Nevertheless, we did not observe significant differences in improvement in any productivity domains between urban and rural groups; all of these domains showed significant improvement following the DCP.

Despite the wide reach of telerehabilitation, many areas across the United States are still facing unmet needs. The results observed herein support the need for further research and investment in digital rehabilitation to mitigate inequities in health care access and care delivery optimization.

Strengths and Limitations

There are many strengths to this study, namely the novelty of investigating the urban-rural dichotomy within a digital therapy program in a large sample size from a real-world context, including patients from 50 US states and the District of Columbia, which allows for a diverse population and thus better generalizability. Another strength is the DCP itself, which uses a multimodal approach that includes exercises with real-time biofeedback, mental support, regular communication with the PT, and a digital format. All these components favor accessibility and maximize engagement and clinical outcomes, allowing us to study different aspects of the problem, from pain to mental health to productivity.

The classification of rural and urban areas is a challenging topic considering the multitude of factors that can highly influence the obtained readings. Despite the application of a recognized classification system [45-47], we cannot rule out the existence of other confounding factors with contributions that were not taken into account during this exploratory analysis, including desirability bias. Other limitations include the lack of control groups (to account for nonspecific treatment effects) and the lack of long-term outcome and objective outcome measures (ie, through activity trackers). Nevertheless, this exploratory study may lay a foundation for future work in this field, identifying areas in need of improvement for future telerehabilitation programs. Further prospective controlled studies are warranted to better characterize the effect of rural and urban inequities on digital therapy outcomes.

Conclusion

This study provides important insights regarding the impact of a multimodal digital program for MSK pain management in rural and urban settings. The DCP was able to reach all areas across the United States with high completion rates in both settings. Despite the inherent health inequities between patients from rural and urban areas, similarly high satisfaction and engagement, alongside significant improvements in pain, mental health, and productivity, were observed in both groups. This showcases the potential of the DCP to mitigate inequities by improving the accessibility of MSK care independently of geographic location.

Acknowledgments

The authors acknowledge the team of physical therapists responsible for managing the participants. The authors also acknowledge the contributions of João Tiago Silva and Guilherme Freches in data validation (both employees of Sword Health). Critical revision of the manuscript for important intellectual content was done by all authors. All authors were involved with the final approval of the manuscript. The study sponsor, Sword Health, was involved in the study design, data collection, and interpretation and writing of the manuscript.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors made a significant contribution to the work. FC, FDC, and JL were responsible for the study concept and design. MM acquired the data. RM performed the statistical analysis. JS, FC, ACA, DJ, MM, and FC interpreted the data. JS was responsible for drafting the work. VB was responsible for funding.

Conflicts of Interest

Sword Health owns the Digital Care Program and as such this program is only available to those covered by a commercial agreement with Sword Health. FC, DJ, AA, MM, FDC, VB, and VY are employees of Sword Health, the sponsor of this study. DJ, FC, FDC, VY, and VB also hold equity in Sword Health. RM is an independent scientific consultant responsible for the statistical analysis. JS and JL are independent scientific and clinical consultants who received adviser honorariums from Sword Health.

Multimedia Appendix 1

Information regarding (1) Rural-urban commuting area (RUCA) codes, (2) baseline characteristics of completers and non-completers, (3) Latent growth curve analysis (LGCA) model following intent-to-treat analysis with corresponding model fitness and conditional analysis, and (4) 12-week mean changes for the per-protocol analysis and corresponding LGCA model and model fitness.

[DOCX File , 78 KB - mhealth_v11i1e44316_app1.docx]

References

1. El-Tallawy SN, Nalamasu R, Salem GI, LeQuang JAK, Pergolizzi JV, Christo PJ. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. Pain Ther 2021 Jun;10(1):181-209 [FREE Full text] [doi: 10.1007/s40122-021-00235-2] [Medline: 33575952]
2. Dieleman JL, Cao J, Chapin A, Chen C, Li Z, Liu A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020 Mar 03;323(9):863-884 [FREE Full text] [doi: 10.1001/jama.2020.0734] [Medline: 32125402]
3. Blyth FM, Noguchi N. Chronic musculoskeletal pain and its impact on older people. Best Pract Res Clin Rheumatol 2017 Apr;31(2):160-168. [doi: 10.1016/j.berh.2017.10.004] [Medline: 29224694]

4. George SZ, Fritz JM, Silfies SP, Schneider MJ, Beneciuk JM, Lentz TA, et al. Interventions for the management of acute and chronic low back pain: revision 2021. J Orthop Sports Phys Ther 2021 Nov;51(11):CPG1-CPG60. [doi: 10.2519/jospt.2021.0304] [Medline: 34719942]
5. Gianola S, Bargeri S, Del Castillo G, Corbetta D, Turolla A, Andreano A, et al. Effectiveness of treatments for acute and subacute mechanical non-specific low back pain: a systematic review with network meta-analysis. Br J Sports Med 2022 Jan;56(1):41-50 [FREE Full text] [doi: 10.1136/bjsports-2020-103596] [Medline: 33849907]
6. Stochkendahl MJ, Kjaer P, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, Andersen, et al. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. Eur Spine J 2018 Jan;27(1):60-75. [doi: 10.1007/s00586-017-5099-2] [Medline: 28429142]
7. Grona SL, Bath B, Busch A, Rotter T, Trask C, Harrison E. Use of videoconferencing for physical therapy in people with musculoskeletal conditions: A systematic review. J Telemed Telecare 2018 Jun;24(5):341-355. [doi: 10.1177/1357633X17700781] [Medline: 28403669]
8. Lee AC, Davenport TE, Randall K. Telehealth physical therapy in musculoskeletal practice. J Orthop Sports Phys Ther 2018 Oct;48(10):736-739. [doi: 10.2519/jospt.2018.0613] [Medline: 30270782]
9. Levy CE, Silverman E, Jia H, Geiss M, Omura D. Effects of physical therapy delivery via home video telerehabilitation on functional and health-related quality of life outcomes. J Rehabil Res Dev 2015;52(3):361-370 [FREE Full text] [doi: 10.1682/JRRD.2014.10.0239] [Medline: 26230650]
10. Seron P, Oliveros M, Gutierrez-Arias R, Fuentes-Aspe R, Torres-Castro RC, Merino-Osorio C, et al. effectiveness of telerehabilitation in physical therapy: a rapid overview. Phys Ther 2021 Jun 01;101(6):pzab053 [FREE Full text] [doi: 10.1093/ptj/pzab053] [Medline: 33561280]
11. Horsley S, Schock G, Grona SL, Montieth K, Mowat B, Stasiuk K, et al. Use of real-time videoconferencing to deliver physical therapy services: A scoping review of published and emerging evidence. J Telemed Telecare 2020 Dec;26(10):581-589. [doi: 10.1177/1357633X19854647] [Medline: 31213166]
12. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-Week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
13. Nelson M, Bourke M, Crossley K, Russell T. Telerehabilitation is non-inferior to usual care following total hip replacement - a randomized controlled non-inferiority trial. Physiotherapy 2020 Jun;107:19-27. [doi: 10.1016/j.physio.2019.06.006] [Medline: 32026820]
14. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120. [doi: 10.2106/JBJS.I.01375] [Medline: 21248209]
15. Correia FD, Molinos M, Luís S, Carvalho D, Carvalho C, Costa P, et al. Digitally assisted versus conventional home-based rehabilitation after arthroscopic rotator cuff repair: a randomized controlled trial. Am J Phys Med Rehabil 2022 Mar 01;101(3):237-249 [FREE Full text] [doi: 10.1097/PHM.0000000000001780] [Medline: 33935152]
16. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]
17. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299 [FREE Full text] [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
18. Fatoye F, Gebrye T, Fatoye C, Mbada CE, Olaoye MI, Odole AC, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375 [FREE Full text] [doi: 10.2196/15375] [Medline: 32357128]
19. Fiani B, Siddiqi I, Lee SC, Dhillon L. Telerehabilitation: development, application, and need for increased usage in the COVID-19 era for patients with spinal pathology. Cureus 2020 Sep 21;12(9):e10563 [FREE Full text] [doi: 10.7759/cureus.10563] [Medline: 33101809]
20. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. Musculoskelet Sci Pract 2020 Aug;48:102193 [FREE Full text] [doi: 10.1016/j.msksp.2020.102193] [Medline: 32560876]
21. Dixit N, Van Sebille Y, Crawford GB, Ginex PK, Ortega PF, Chan RJ. Disparities in telehealth use: How should the supportive care community respond? Support Care Cancer 2022 Feb;30(2):1007-1010 [FREE Full text] [doi: 10.1007/s00520-021-06629-4] [Medline: 34668075]
22. Lattimore CM, Kane WJ, Fleming MA, Martin AN, Mehaffey JH, Smolkin ME, et al. Disparities in telemedicine utilization among surgical patients during COVID-19. PLoS One 2021;16(10):e0258452 [FREE Full text] [doi: 10.1371/journal.pone.0258452] [Medline: 34624059]
23. Qian L, Sy LS, Hong V, Glenn SC, Ryan DS, Morrissette K, et al. Disparities in outpatient and telehealth visits during the COVID-19 pandemic in a large integrated health care organization: retrospective. J Med Internet Res 2021 Sep 01;23(9):e29959 [FREE Full text] [doi: 10.2196/29959] [Medline: 34351865]

24. Felter CE, Zalewski K, Jermann R, Palmer PL, Baier AE, Falvey JR. Rural health: the dirt road less traveled. Phys Ther 2022 Aug 04:pzac112. [doi: 10.1093/ptj/pzac112] [Medline: 35925820]
25. Huber GM, Bitzer G, Corazzi C, Fitzsimmons S, Melissa H, Shelley J, et al. Access to physical therapy in a medically underserved, urban community. J Health Care Poor Underserved 2019;30(2):768-788. [doi: 10.1353/hpu.2019.0055] [Medline: 31130550]
26. Liu X, Seidel JE, McDonald T, Patel AB, Waters N, Bertazzon S, et al. Rural-urban disparities in realized spatial access to general practitioners, orthopedic surgeons, and physiotherapists among people with osteoarthritis in Alberta, Canada. Int J Environ Res Public Health 2022 Jun 23;19(13):7706 [FREE Full text] [doi: 10.3390/ijerph19137706] [Medline: 35805363]
27. Mercier MR, Galivanche AR, Wiggins AJ, Kahan JB, McLaughlin W, Radford ZJ, et al. Patient demographic and socioeconomic factors associated with physical therapy utilization after uncomplicated meniscectomy. J Am Acad Orthop Surg Glob Res Rev 2022 Jul 01;6(7):e22.00135 [FREE Full text] [doi: 10.5435/JAAOSGlobal-D-22-00135] [Medline: 35816646]
28. Castle ME, Tak CR. Self-reported vs RUCA rural-urban classification among North Carolina pharmacists. Pharm Pract (Granada) 2021;19(3):2406 [FREE Full text] [doi: 10.18549/PharmPract.2021.3.2406] [Medline: 34522240]
29. Surbhi S, Tolley EA, Cossman RE, Dashputre AA, Bailey JE. Refining a traditional urban-rural classification approach to better assess heterogeneity of treatment effects in patient-centered outcomes research. MethodsX 2021;8:101299 [FREE Full text] [doi: 10.1016/j.mex.2021.101299] [Medline: 34434819]
30. Probst JC, Moore CG, Glover SH, Samuels ME. Person and place: the compounding effects of race/ethnicity and rurality on health. Am J Public Health 2004 Oct;94(10):1695-1703. [doi: 10.2105/ajph.94.10.1695] [Medline: 15451735]
31. Hing E, Hsiao C. State variability in supply of office-based primary care providers: United States, 2012. NCHS Data Brief 2014 May(151):1-8 [FREE Full text] [Medline: 24813076]
32. Pelletier C, Cornish K, Amyot T, Pousette A, Fox G, Snadden D, et al. Physical activity promotion in rural health care settings: A rapid realist review. Prev Med Rep 2022 Oct;29:101905 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101905] [Medline: 35879935]
33. Carter SK, Rizzo JA. Use of outpatient physical therapy services by people with musculoskeletal conditions. Phys Ther 2007 May;87(5):497-512. [doi: 10.2522/ptj.20050218] [Medline: 17374630]
34. Correia FD, Molinos M, Neves C, Janela D, Carvalho D, Luis S, et al. Digital rehabilitation for acute ankle sprains: prospective longitudinal cohort study. JMIR Rehabil Assist Technol 2021 Sep 30;8(3):e31247 [FREE Full text] [doi: 10.2196/31247] [Medline: 34499038]
35. Costa F, Janela D, Molinos M, Lains J, Francisco GE, Bento V, et al. Telerehabilitation of acute musculoskeletal multi-disorders: prospective, single-arm, interventional study. BMC Musculoskelet Disord 2022 Jan 04;23(1):29 [FREE Full text] [doi: 10.1186/s12891-021-04891-5] [Medline: 34983488]
36. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Francisco GE, et al. Asynchronous and tailored digital rehabilitation of chronic shoulder pain: a prospective longitudinal cohort study. J Pain Res 2022;15:53-66 [FREE Full text] [doi: 10.2147/JPR.S343308] [Medline: 35035234]
37. Scheer J, Costa F, Molinos M, Areias A, Janela D, Moulder RG, et al. Racial and ethnic differences in outcomes of a 12-week digital rehabilitation program for musculoskeletal pain: prospective longitudinal cohort study. J Med Internet Res 2022 Oct 31;24(10):e41306 [FREE Full text] [doi: 10.2196/41306] [Medline: 36189963]
38. Costa F, Janela D, Molinos M, Moulder R, Bento V, Lains J, et al. Impacts of digital care programs for musculoskeletal conditions on depression and work productivity: longitudinal cohort study. J Med Internet Res 2022 Jul 25;24(7):e38942 [FREE Full text] [doi: 10.2196/38942] [Medline: 35714099]
39. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Scheer JK, et al. Fear avoidance beliefs in upper-extremity musculoskeletal pain conditions: secondary analysis of a prospective clinical study on digital care programs. Pain Med 2022 Oct 06:pnac149. [doi: 10.1093/pm/pnac149] [Medline: 36200858]
40. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020 Jan;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
41. Williams ACDC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11(11):CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
42. Low Back Pain and Sciatica in Over 16s: Assessment and Management. National Institute for Health and Care Excellence. 2016. URL: https://www.nice.org.uk/guidance/ng59 [accessed 2023-02-10]
43. Joypaul S, Kelly F, McMillan SS, King MA. Multi-disciplinary interventions for chronic pain involving education: A systematic review. PLoS One 2019;14(10):e0223306 [FREE Full text] [doi: 10.1371/journal.pone.0223306] [Medline: 31577827]
44. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians, Denberg TD, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline

from the American College of Physicians. Ann Intern Med 2017 Apr 04;166(7):514-530 [FREE Full text] [doi: 10.7326/M16-2367] [Medline: 28192789]
45. Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. Am J Public Health 2005 Jul;95(7):1149-1155. [doi: 10.2105/AJPH.2004.042432] [Medline: 15983270]
46. RUCA Data - Using RUCA Data. RUCA Rural Health Research Center. URL: https://depts.washington.edu/uwruca/ruca-uses.php [accessed 2022-11-01]
47. Economic Research Service, US Department of Agriculture. Rural-Urban Commuting Area Codes. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx [accessed 2022-11-01]
48. Gatchel RJ, Mayer TG, Choi Y, Chou R. Validation of a consensus-based minimal clinically important difference (MCID) threshold using an objective functional external anchor. Spine J 2013 Aug;13(8):889-893. [doi: 10.1016/j.spinee.2013.02.015] [Medline: 23523434]
49. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008 Jan 01;33(1):90-94. [doi: 10.1097/BRS.0b013e31815e3a10] [Medline: 18165753]
50. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
51. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
52. Ospina MB, Dennett L, Waye A, Jacobs P, Thompson AH. A systematic review of measurement properties of instruments assessing presenteeism. Am J Manag Care 2015 Feb 01;21(2):e171-e185 [FREE Full text] [Medline: 25880491]
53. Burant CJ. Latent growth curve models: tracking changes over time. Int J Aging Hum Dev 2016 Apr;82(4):336-350. [doi: 10.1177/0091415016641692] [Medline: 27076490]
54. Mbada CE, Olaoye MI, Dada OO, Ayanniyi O, Johnson OE, Odole AC, et al. Comparative efficacy of clinic-based and telerehabilitation application of Mckenzie therapy in chronic low-back pain. Int J Telerehabil 2019;11(1):41-58 [FREE Full text] [doi: 10.5195/ijt.2019.6260] [Medline: 31341546]
55. Enders C, Bandalos D. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Struct Equ Modeling 2001 Jul 1;8(3):430-457. [doi: 10.1207/s15328007sem0803_5]
56. Newman DA. Longitudinal modeling with randomly and systematically missing data: a simulation of ad hoc, maximum likelihood, and multiple imputation techniques. Organ Res Methods 2016 Jun 29;6(3):328-362. [doi: 10.1177/1094428103254673]
57. Olinsky A, Chen S, Harlow L. The comparative efficacy of imputation methods for missing data in structural equation modeling. Eur J Oper Res 2003 Nov;151(1):53-79. [doi: 10.1016/S0377-2217(02)00578-7]
58. Pfaffel A, Schober B, Spiel C. A comparison of three approaches to correct for direct and indirect range restrictions: a simulation study. Pract Assess Res Eval 2016;21(6):1-15 [FREE Full text] [doi: 10.7275/x4ep-fv42]
59. Xiao J, Bulut O. Evaluating the performances of missing data handling methods in ability estimation from sparse data. Educ Psychol Meas 2020 Oct;80(5):932-954 [FREE Full text] [doi: 10.1177/0013164420911136] [Medline: 32855565]
60. Iacobucci D. Structural equations modeling: Fit Indices, sample size, and advanced topics. J Consum Psychol 2009 Oct 03;20(1):90-98. [doi: 10.1016/j.jcps.2009.09.003]
61. Brown TA. Confirmatory factor analysis for applied research. New York, NY, US: The Guilford Press; 2006.
62. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
63. Urban Areas Facts. US Census Bureau. URL: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/ua-facts.html [accessed 2022-11-01]
64. Caldwell JT, Ford CL, Wallace SP, Wang MC, Takahashi LM. Intersection of living in a rural versus urban area and race/ethnicity in explaining access to health care in the United States. Am J Public Health 2016 Aug;106(8):1463-1469. [doi: 10.2105/AJPH.2016.303212] [Medline: 27310341]
65. Rethorn ZD, Garcia AN, Cook CE, Gottfried ON. Quantifying the collective influence of social determinants of health using conditional and cluster modeling. PLoS One 2020;15(11):e0241868 [FREE Full text] [doi: 10.1371/journal.pone.0241868] [Medline: 33152044]
66. Tseli E, Boersma K, Stålnacke BM, Enthoven P, Gerdle B, Äng BO, et al. Prognostic factors for physical functioning after multidisciplinary rehabilitation in patients with chronic musculoskeletal pain: a systematic review and meta-analysis. Clin J Pain 2019 Feb;35(2):148-173 [FREE Full text] [doi: 10.1097/AJP.0000000000000669] [Medline: 30371517]
67. Mallen CD, Peat G, Thomas E, Dunn KM, Croft PR. Prognostic factors for musculoskeletal pain in primary care: a systematic review. Br J Gen Pract 2007 Aug;57(541):655-661 [FREE Full text] [Medline: 17688762]
68. Green D, Lewis M, Mansell G, Artus M, Dziedzic K, Hay E, et al. Clinical course and prognostic factors across different musculoskeletal pain sites: A secondary analysis of individual patient data from randomised clinical trials. Eur J Pain 2018 Jul;22(6):1057-1070. [doi: 10.1002/ejp.1190] [Medline: 29356210]

69. Seaman DR. Body mass index and musculoskeletal pain: is there a connection? Chiropr Man Therap 2013 May 20;21(1):15 [FREE Full text] [doi: 10.1186/2045-709X-21-15] [Medline: 23687943]
70. Vogels E. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021. URL: https://www.pewresearch.org/fact-tank/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/ [accessed 2023-02-08]
71. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. J Rural Health 2018 Feb;34 Suppl 1(Suppl 1):s30-s38 [FREE Full text] [doi: 10.1111/jrh.12228] [Medline: 28075508]
72. Kent P, Laird R, Haines T. The effect of changing movement and posture using motion-sensor biofeedback, versus guidelines-based care, on the clinical outcomes of people with sub-acute or chronic low back pain-a multicentre, cluster-randomised, placebo-controlled, pilot trial. BMC Musculoskelet Disord 2015 May 29;16:131 [FREE Full text] [doi: 10.1186/s12891-015-0591-5] [Medline: 26022102]
73. Wang G, Yang M, Hong M, Krauss J, Bailey JF. Clinical outcomes one year after a digital musculoskeletal (MSK) program: an observational, longitudinal study with nonparticipant comparison group. BMC Musculoskelet Disord 2022 Mar 11;23(1):237 [FREE Full text] [doi: 10.1186/s12891-022-05188-x] [Medline: 35277161]
74. Dahlberg LE, Dell'Isola A, Lohmander LS, Nero H. Improving osteoarthritis care by digital means - Effects of a digital self-management program after 24- or 48-weeks of treatment. PLoS One 2020;15(3):e0229783 [FREE Full text] [doi: 10.1371/journal.pone.0229783] [Medline: 32130276]
75. Trulsson Schouenborg A, Rivano Fischer M, Bondesson E, Jöud A. Physiotherapist-led rehabilitation for patients with chronic musculoskeletal pain: interventions and promising long-term outcomes. BMC Musculoskelet Disord 2021 Oct 28;22(1):910 [FREE Full text] [doi: 10.1186/s12891-021-04780-x] [Medline: 34711194]
76. Probst JC, Laditka SB, Moore CG, Harun N, Powell MP, Baxley EG. Rural-urban differences in depression prevalence: implications for family medicine. Fam Med 2006 Oct;38(9):653-660 [FREE Full text] [Medline: 17009190]
77. Jameson JP, Blank MB. Diagnosis and treatment of depression and anxiety in rural and nonrural primary care: national survey results. Psychiatr Serv 2010 Jun;61(6):624-627. [doi: 10.1176/ps.2010.61.6.624] [Medline: 20513688]
78. Daré LO, Bruand P, Gérard D, Marin B, Lameyre V, Boumédiène F, et al. Co-morbidities of mental disorders and chronic physical diseases in developing and emerging countries: a meta-analysis. BMC Public Health 2019 Mar 13;19(1):304 [FREE Full text] [doi: 10.1186/s12889-019-6623-6] [Medline: 30866883]
79. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Jun 09;391(10137):2356-2367 [FREE Full text] [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
80. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2021 Dec 19;396(10267):2006-2017 [FREE Full text] [doi: 10.1016/S0140-6736(20)32340-0] [Medline: 33275908]
81. Moreira-Silva I, Santos R, Abreu S, Mota J. Associations between body mass index and musculoskeletal pain and related symptoms in different body regions among workers. SAGE Open 2013 Jun 05;3(2):215824401349195 [FREE Full text] [doi: 10.1177/2158244013491952]
82. LoMartire R, Dahlström Ö, Björk M, Vixner L, Frumento P, Constan L, et al. Predictors of sickness absence in a clinical population with chronic pain. J Pain 2021 Oct;22(10):1180-1194 [FREE Full text] [doi: 10.1016/j.jpain.2021.03.145] [Medline: 33819574]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
DCP: digital care program
FIML: full information maximum likelihood
GAD-7: Generalized Anxiety Disorder 7-item
LGCA: latent growth curve analysis
MCID: minimum clinically important difference
MSK: musculoskeletal
PHQ-9: Patient Health Questionnaire 9-item
PT: physical therapist
RUCA: rural-urban commuting area
WPAI: Work Productivity and Activity Impairment

Edited by G Eysenbach, L Buis; submitted 17.11.22; peer-reviewed by R Jamison; comments to author 05.01.23; revised version received 12.01.23; accepted 02.02.23; published 16.03.23.

Please cite as:
Scheer J, Areias AC, Molinos M, Janela D, Moulder R, Lains J, Bento V, Yanamadala V, Dias Correia F, Costa F
Engagement and Utilization of a Complete Remote Digital Care Program for Musculoskeletal Pain Management in Urban and Rural Areas Across the United States: Longitudinal Cohort Study
JMIR Mhealth Uhealth 2023;11:e44316
URL: https://mhealth.jmir.org/2023/1/e44316
doi:10.2196/44316
PMID:36735933

Original Paper

Testing the Effect of a Smartphone App on Hospital Admissions and Sedentary Behavior in Cardiac Rehabilitation Participants: ToDo-CR Randomized Controlled Trial

Kacie Patterson[1], PhD; Rachel Davey[1], PhD; Richard Keegan[2], PhD; Theo Niyonsenga[1], PhD; Itismita Mohanty[1], PhD; Sarah Bowen[3], BPT; Elizabeth Regan[4], BSN; Michelle Lander[5], BSN; Sander van Berlo[6], MSc; Nicole Freene[1,7], PhD

[1]Health Research Institute, Faculty of Health, University of Canberra, Bruce, ACT, Australia
[2]Research Institute for Sports and Exercise (UCRISE), Faculty of Health, University of Canberra, Bruce, ACT, Australia
[3]National Capital Private Hospital, Garran, ACT, Australia
[4]Calvary Public Hospital Bruce, Bruce, ACT, Australia
[5]The Canberra Hospital, Garran, ACT, Australia
[6]Onmi BV, Torenallee, Eindhoven, Netherlands
[7]Physiotherapy, Faculty of Health, University of Canberra, Bruce, ACT, Australia

Corresponding Author:
Kacie Patterson, PhD
Health Research Institute
Faculty of Health
University of Canberra
University Drive
Bruce, ACT, 2617
Australia
Phone: 61 2 6201 5550
Email: Kacie.Patterson@canberra.edu.au

Abstract

Background: People with coronary heart disease are at an increased risk of morbidity and mortality even if they attend cardiac rehabilitation. High sedentary behavior levels potentially contribute to this morbidity. Smartphone apps may be feasible to facilitate sedentary behavior reductions and lead to reduced health care use.

Objective: We aimed to test the effect of a sedentary behavior change smartphone app (Vire app and ToDo-CR program) as an adjunct to cardiac rehabilitation on hospital admissions and emergency department (ED) presentations over 12 months.

Methods: A multicenter, randomized controlled trial was conducted with 120 participants recruited from 3 cardiac rehabilitation programs. Participants were randomized 1:1 to cardiac rehabilitation plus the fully automated 6-month Vire app and ToDo-CR program (intervention) or usual care (control). The primary outcome was nonelective hospital admissions and ED presentations over 12 months. Secondary outcomes including accelerometer-measured sedentary behavior, BMI, waist circumference, and quality of life were recorded at baseline and 6 and 12 months. Logistic regression models were used to analyze the primary outcome, and linear mixed-effects models were used to analyze secondary outcomes. Data on intervention and hospital admission costs were collected, and the incremental cost-effectiveness ratios (ICERs) were calculated.

Results: Participants were, on average, aged 62 (SD 10) years, and the majority were male (93/120, 77.5%). The intervention group were more likely to experience all-cause (odds ratio [OR] 1.54, 95% CI 0.58-4.10; P=.39) and cardiac-related (OR 3.26, 95% CI 0.84-12.55; P=.09) hospital admissions and ED presentations (OR 2.07, 95% CI 0.89-4.77; P=.09) than the control group. Despite this, cardiac-related hospital admission costs were lower in the intervention group over 12 months (Aus $252.40 vs Aus $859.38; P=.24; a currency exchange rate of Aus $1=US $0.69 is applicable). There were no significant between-group differences in sedentary behavior minutes per day over 12 months, although the intervention group completed 22 minutes less than the control group (95% CI −22.80 to 66.69; P=.33; Cohen d=0.21). The intervention group had a lower BMI (β=1.62; P=.05), waist circumference (β=5.81; P=.01), waist-to-hip ratio (β=.03, P=.03), and quality of life (β=3.30; P=.05) than the control group. The intervention was more effective but more costly in reducing sedentary behavior (ICER Aus $351.77) and anxiety (ICER Aus

$10,987.71) at 12 months. The intervention was also more effective yet costly in increasing quality of life (ICER Aus $93,395.50) at 12 months.

Conclusions: The Vire app and ToDo-CR program was not an outcome-effective or cost-effective solution to reduce all-cause hospital admissions or ED presentations in cardiac rehabilitation compared with usual care. Smartphone apps that target sedentary behavior alone may not be an effective solution for cardiac rehabilitation participants to reduce hospital admissions and sedentary behavior.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12619001223123; https://australianclinicaltrials.gov.au/anzctr/trial/ACTRN12619001223123

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-040479

(JMIR Mhealth Uhealth 2023;11:e48229) doi:10.2196/48229

KEYWORDS

mobile health; secondary prevention; cardiovascular disease; sedentary behavior; hospital admissions; cost-effectiveness; mobile phone

Introduction

Overview

Cardiovascular disease, including coronary heart disease (CHD), is the leading cause of death in Australia and globally [1-3]. Exercise-based cardiac rehabilitation for people with CHD is associated with significant risk reductions in cardiovascular mortality, hospitalizations, and repeat cardiac events [4]. Even so, 1 in 3 cardiac events are repeat events, and repeat cardiac events increase the risk of premature mortality [5].

Cardiac Rehabilitation and Sedentary Behavior

Cardiac rehabilitation aims to reduce morbidity through positive lifestyle changes, including increasing physical activity and decreasing sedentary behavior [6,7]. Despite this, the sedentary behavior levels of cardiac rehabilitation participants remain high before, during, and after the program [8-16]. Sedentary behavior is associated with an increased risk of morbidity and all-cause mortality [17-21]. Accelerometry-measured sedentary times of ≥9 waking hours per day place healthy individuals at a significantly higher risk of death [18]. Television viewing times (a self-reported marker of sedentary behavior) in adults with diagnosed CHD or stroke of ≥4 hours per day is associated with a 52% increased risk of all-cause mortality compared with those watching television for ≤2 hours [21]. Breaking up sedentary time more frequently is also associated with decreased systolic blood pressure in cardiac rehabilitation participants [8]. Cardiac rehabilitation participants are likely to benefit by engaging in interventions that reduce sedentary behavior, and further options to support participants may be needed.

Smartphone Apps to Reduce Hospital Admissions and Sedentary Behavior

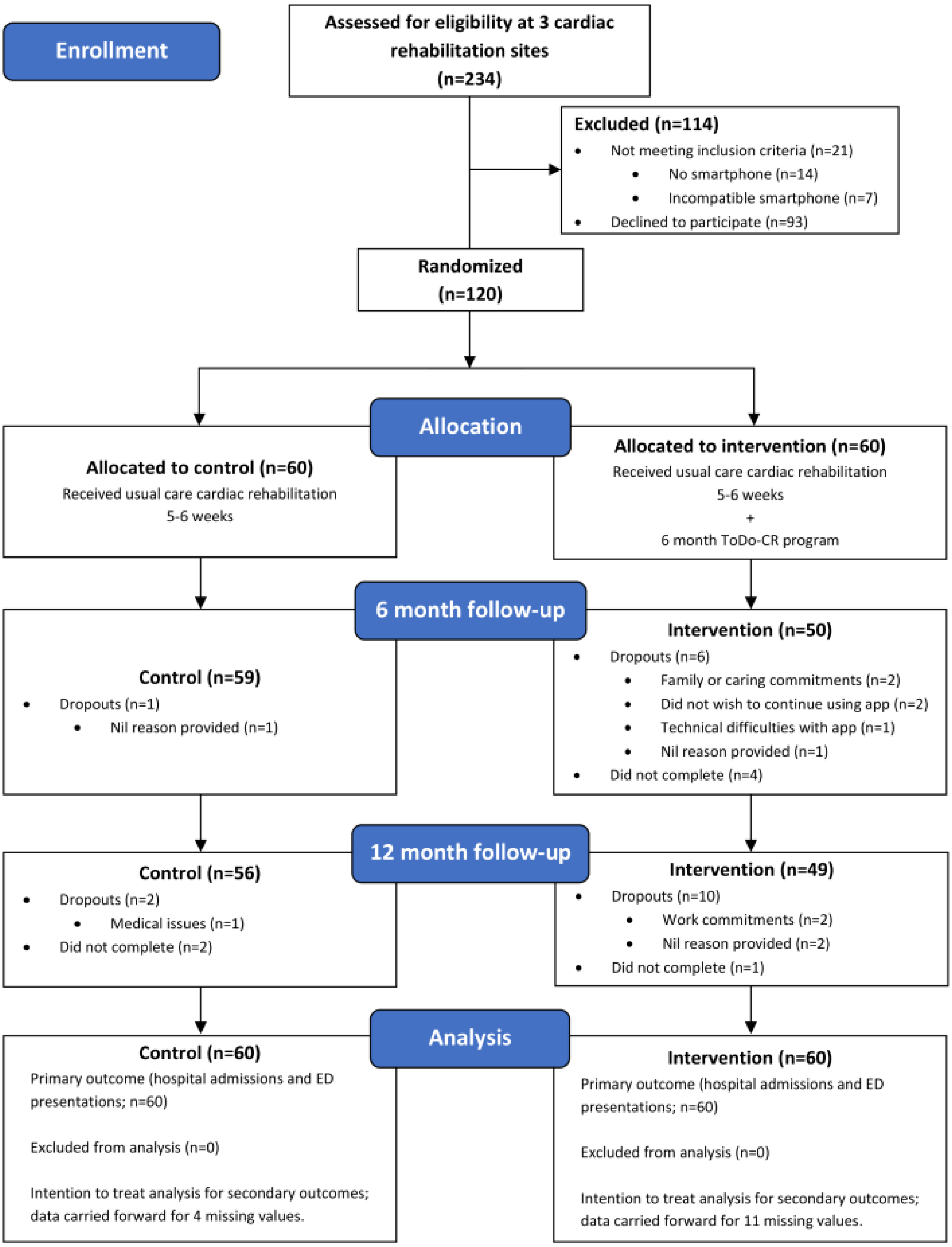
Evidence suggests that cardiac rehabilitation participants are interested in support via the internet and mobile phones [22,23], and they have the potential to reduce hospital readmissions [24-26]. Widmer et al [25] reported that cardiac rehabilitation participants using a digital health intervention had significant (P=.04) reductions in weight and blood pressure and a 28% reduction in rehospitalizations and emergency department (ED) presentations compared with those who received traditional cardiac rehabilitation only. We hypothesized that better secondary prevention management of risk factors through cardiac rehabilitation plus the digital health intervention can lead to reduced hospital admissions. Similarly, smartphone apps that reduce sedentary behavior may be a feasible option to reduce hospital admissions. Few studies have targeted sedentary behavior change through smartphone apps in people with CHD [27-29]. These studies are generally small (≤50 participants), short in duration (≤3 mo), and aimed at examining the feasibility of such interventions [30]. Effect sizes have varied in these studies, with one reporting no change in self-reported sitting time [27], and in the feasibility trial preceding this study, a medium reduction in accelerometer-measured sedentary behaviors was reported [28]. One of the largest studies to date targeting sedentary behavior involving a smartphone app, pocket-worn activity tracker, and participants with CHD reported no significant between-group reduction in device-measured sedentary time [29]. However, cardiac rehabilitation participants using the SIT LESS protocol had reduced odds of sitting >9.5 hours per day [29]. Despite varying results, these studies concluded that smartphone apps may be feasible in reducing sedentary behavior in people with CHD and that larger-scale randomized controlled trials are warranted to determine their effectiveness. Therefore, we aimed to test the effectiveness of a sedentary behavior change smartphone app (Vire app and ToDo-CR program) as an adjunct to cardiac rehabilitation for hospital admissions and ED presentations over 12 months. As secondary aims, we examined the effectiveness of the Vire app and ToDo-CR program in decreasing accelerometer-measured sedentary behavior and cost-effectiveness.

Methods

Design

Using an assessor-blind, parallel randomized controlled trial design, participants were recruited from 3 phase-2 hospital-based cardiac rehabilitation programs in Canberra, Australia, between January 2020 and December 2021. The study duration for each participant was 12 months (Figure 1). Following the baseline assessment, participants were randomly assigned 1:1 to either usual care cardiac rehabilitation or the intervention: cardiac rehabilitation plus the 6-month behavior change smartphone app commencing at the start of cardiac rehabilitation (Vire app and ToDo-CR program) and followed up at 6 and 12 months at the University of Canberra by a blinded research assistant. The study protocol has been published elsewhere [31]. Participants were aged ≥18 years, had stable CHD, owned a smartphone, had no serious medical or functional impairments, and had adequate English language and cognitive skills [31].

Figure 1. Flow of participants through the trial. ED: emergency department.



Ethical Considerations

All participants provided written informed consent. Ethics approval was obtained from the Australian Capital Territory Health (2019.ETH.00162), Calvary Public Hospital Bruce (20-2019), and the University of Canberra (HREC-2325) Human Research Ethics Committees. Reporting was informed by the CONSORT (Consolidated Standards of Reporting Trials)–EHEALTH [32] and CONSERVE statements [33] and guidelines for studies affected by the COVID-19 pandemic [34].

Intervention

The Vire app and ToDo-CR behavior change program was informed by the *Do Something Different* approach focusing on breaking existing sedentary behavior habits and becoming behaviorally flexible [31,35,36]. The Vire app integrated data from a Fitbit Inspire wearable activity tracker provided to the participants and smartphone GPS data through machine learning to create a comprehensive digital profile of the participants' current behavior. Using the data, the Vire app sends short, personalized behavior change messages known as *Do's* in the form of push notifications 2 to 3 times per week at random over 6 months. The *Do's* targeted sedentary behavior, suggesting microbehavioral alternatives designed to disrupt usual habits and encourage small lifestyle changes. Participants were encouraged to download the app, and technological support was provided throughout the trial.

Outcome Measures

Nonelective Hospital Admissions and ED Presentations

The primary outcome was the total number of all-cause hospital admissions (nonelective admission to an acute hospital) and ED presentations within the 12 months after commencing cardiac rehabilitation. Data on cardiac-related hospital admissions and time frames to admission were also collected. Participants self-reported hospital admissions, which were verified by a comprehensive hospital patient records audit for all participants, regardless of self-report admission.

Sedentary Behavior and Other Secondary Outcomes

Sedentary behavior and physical activity were measured using a triaxial commercial accelerometer (ActiGraph wGT3X-BT) worn by participants on the right hip for 7 consecutive days during waking hours. ActiLife software (ActiLife V.6.13.4) was used to download raw data (30 Hz), which were converted to 15-second epochs and counts per minute (cpm) [9,16]. All data were screened and excluded if there was <10 hours per day wear time and if there were less than 4 days of valid data [9,16,37,38]. The following vector magnitude (VM) cut-off points were used: sedentary behavior, <150 cpm; light-intensity physical activity, 150 to 2689 cpm; and moderate to vigorous intensity physical activity, ≥2690 cpm [9,16,37,39,40].

Additional secondary outcomes included BMI (kg/m^2); waist circumference; waist-to-hip ratio; blood pressure; exercise capacity (6-min walk test distance [41]); health-related quality of life (Assessment of Quality of Life [AQoL]-6D, a score of 100 reflects best health [42,43]); anxiety and depression (Hospital Anxiety and Depression Scale, a score of 0 reflects best outcomes [44]); and the stage of behavior change for physical activity (modified University of Rhode Island Change Assessment Scale-E2, scores of ≤8 indicate precontemplation, 9-11 indicate contemplation, and 12-14 indicate action and maintenance [45]).

Smartphone App Usability and Engagement

For those in the intervention group, the usability and acceptance of the Vire app and ToDo-CR program were assessed using the Unified Theory of Acceptance and Use of Technology questionnaire [46]. Engagement with the Vire app was assessed by viewing app logs showing the completion of *Do's*. The total number of completed *Do's* across the study was recorded.

Economic Evaluation

The costs of implementing and delivering the intervention were recorded prospectively, including payment for the Vire app and maintenance of the server, the purchase of Fitbit Inspire wearable activity trackers, and phone call and email support related to the app from a cardiac rehabilitation clinician.

A hospital patient records audit was completed for all participants to obtain the associated Australian Refined Diagnosis Related Group (AR-DRG) classification code for any nonelective admissions and the Urgent Related Group classification codes for any ED presentation. The Independent Health and Aged Care Pricing Authority Australian version of the National Weighted Activity Unit calculators were used to obtain hospital cost information [47,48]. Costs are reported in Aus $ throughout. A currency exchange rate of Aus $1=US $0.69 is applicable

Sample Size Calculation

Sample size was based on detecting a significant difference in hospital admissions and ED presentations between usual care and a digital health intervention. A similar study [25] noted a 28% difference in rehospitalization and ED presentations between usual care cardiac rehabilitation (a standard rate of 44%) and cardiac rehabilitation plus a digital health intervention (a standard rate of 16%). Using a 2-sided significance of P<.05 and power of 88% (calculated using G*Power 3.1.9.4), a minimum of 108 participants were needed. Accounting for a 25% dropout rate, 72 participants per group (144 total sample = 108 / [1 – 0.25]) were calculated.

Changes in Response to the COVID-19 Pandemic

To comply with public health recommendations, there were variations in the methods used in the protocol [31]. The trial was extended by 12 months owing to the closure of the cardiac rehabilitation programs. The required sample size was also reduced from 144 to a minimum of 108 (removing the 30% dropout buffer) owing to the impact of the COVID-19 pandemic [31]. In addition, all participants experienced varying restrictions and closures of nonessential health care services. Due to COVID-19 pandemic restrictions, some follow-up assessments were completed via telehealth (Zoom video call). Participants used their own blood pressure monitors, scales, and tape measures under the instruction of a blinded assessor. The method of assessment was documented throughout the study. All changes were approved by the overseeing ethics review committees.

Analysis

Data were analyzed according to group assignment using intention-to-treat analyses. Missing data were handled by bringing the last value forward (carryover approach). An on-protocol analysis was also performed. All descriptive statistics were reported using means and SDs, medians, and IQRs or proportions, as appropriate. Normality was assessed using the Kolmogorov-Smirnov test for samples ≥50.

The primary analysis was the comparison of rates of nonelective hospital admissions and ED presentations. Binary logistic regression for "yes" versus "no" admissions (odds ratios) and negative binomial with log link regression for the rate or number of admissions (incidence rate ratio) were completed. Survival analyses (Cox regression) were completed to consider the time frame to admission (hazard ratios). Adjustments were made for sociodemographic variables (eg, age and sex) and other covariates (eg, diabetes and other chronic diseases).

Linear mixed-effects models for repeated measures were used to analyze all other secondary outcomes. The maximum likelihood method was used for parameter estimation. Time and between-group comparisons were explored as fixed effects while adjusting for demographic characteristics (eg, age and sex) and other covariates (eg, education, employment, and accelerometer counts/d) [49-51]. Participants were treated as random effects (ie, random intercept models), and the intraclass correlation coefficient was reported. The model with best fit was informed by the Akaike Information Criteria [52]. All estimated effects (β) are reported with their associated 95% CIs. The correlation between outcomes and engagement with the Vire app and ToDo-CR program was also explored.

Sensitivity analyses were completed for outcomes self-administered by participants under telehealth conditions (eg, waist circumference and blood pressure) and for participants wearing their accelerometer belt during COVID-19 pandemic lockdowns and restrictions. Subanalyses were completed for participants excluded versus participants consented, dropouts versus those who completed the study, and those with prior experience with a physical activity tracker versus those with no experience. Data were analyzed using SPSS (version 27; IBM Corp).

Cost-Effectiveness Analysis

Effectiveness was measured in terms of the secondary outcomes. The incremental cost-effectiveness ratios (ICERs) were calculated using the following formula:

The mean difference in health care resource use costs included the costs of implementing the intervention plus the indirect costs associated with health care use (eg, hospital admissions and ED presentations). The difference in effects was calculated as the change from baseline to 6 months and from baseline to 12 months for each secondary outcome per individual participant. The mean difference of these individual differences was then calculated using 2-tailed independent samples *t* tests with 95% CIs. From these analyses, the additional cost per unit of health benefit gained (by using the Vire app and ToDo-CR program compared with usual care alone) was determined.

Results

Overview

A total of 120 participants were recruited for this trial (Figure 1). Participant characteristics are reported in Table 1. The majority were male (93/120, 77.5%), tertiary educated (95/120, 79.2%), and employed (62/120, 51.7%). Approximately half (55/120, 45.8%) of the participants had prior experience using physical activity tracking apps or wearable trackers. The participants who were assessed for eligibility and excluded (n=114) were significantly older (67 vs 63 y; *P*=.003), from public hospitals (*P*=.001), and less likely to have had a percutaneous coronary intervention and more likely to have had a myocardial infarction (*P*=.01; Table S1 in Multimedia Appendix 1). The main reason for exclusion was declining to participate, with the primary reason for declining being "not interested in smartphone apps" (28/93, 30%). The only reasons for not meeting the inclusion criteria were not having a smartphone (14/21, 67%) and having an incompatible smartphone (7/21, 33%; Table S2 in Multimedia Appendix 1). Retention rates were high in both groups (intervention 49/60, 82% vs control 56/60, 93%); however, there was a significant difference between the 2 groups' retention rates (*P*=.02) but no difference in characteristics (Table S3 in Multimedia Appendix 1).

Table 1. Characteristics of participants at baseline.

Characteristics	Control (n=60)	Intervention (n=60)	Total (N=120)
Age (years), mean (SD)	64.10 (9.99)	61.12 (10.06)	62.61 (10.10)
Sex (male), n (%)	48 (80)	45 (75)	93 (77.5)
Country of birth (Australia), n (%)	39 (65)	40 (66.7)	79 (65.8)
Education, n (%)			
Secondary	10 (16.7)	15 (25)	25 (20.8)
Tertiary	50 (83.3)	45 (75)	95 (79.2)
Employment, n (%)			
Full time	23 (38.3)	25 (41.7)	48 (40)
Part time	4 (6.7)	10 (16.7)	14 (11.7)
Voluntary work	4 (6.7)	1 (1.7)	5 (4.2)
Not in the labor force	29 (48.3)	24 (40)	53 (44.2)
Relationship status (partner), n (%)	55 (91.7)	48 (80)	103 (85.8)
Previous experience with physical activity trackers, n (%)			
Smartphone app	8 (13.3)	13 (21.7)	21 (17.5)
Smartwatch	11 (18.3)	12 (20)	23 (19.2)
Both	6 (10)	5 (8.3)	11 (9.2)
Neither	35 (58.3)	30 (50)	65 (54.2)
Diagnosis, n (%)			
Stable coronary heart disease	1 (1.7)	2 (3.3)	3 (2.5)
CABG[a]	12 (20)	12 (20)	24 (20)
PCI[b]	32 (53.3)	27 (45)	59 (49.2)
Myocardial infarction	1 (1.7)	2 (3.3)	3 (2.5)
Myocardial infarction + PCI	14 (23.3)	17 (28.3)	31 (25.8)
Type 2 diabetes (yes), n (%)	12 (20)	13 (22)	25 (20.8)
Other chronic disease (yes), n (%)	10 (16.7)	21 (35)	31 (25.8)
Blood pressure medication (yes), n (%)	50 (83.3)	46 (76.7)	97 (80.8)
Cholesterol medication (yes), n (%)	57 (95)	58 (96.7)	115 (95.8)
Other cardiac medications (yes), n (%)	58 (96.7)	55 (91.7)	113 (94.2)
Current smoker (yes), n (%)	1 (1.7)	2 (3.3)	3 (2.5)
Cardiac rehabilitation system, n (%)			
Public	37 (61.7)	38 (63.3)	75 (62.5)
Private	23 (38.3)	22 (36.7)	45 (37.5)
Cardiac rehabilitation model, n (%)			
Face-to-face	51 (85)	53 (88.3)	104 (86.7)
Hybrid	2 (3.3)	2 (3.3)	4 (3.3)
Telehealth	7 (11.7)	5 (8.3)	12 (10)
Cardiac rehabilitation sessions attended (%), mean (SD)	86 (24)	83 (28)	84 (26)
Previous attendance to cardiac rehabilitation (yes), n (%)	7 (11.7)	9 (15)	16 (13.3)

[a]CABG: coronary artery bypass graft.

[b]PCI: percutaneous coronary intervention.

Nonelective Hospital Admissions and ED Presentations

Nonelective hospital admissions and ED presentations for all participants are shown in Table 2. The most frequent cause of admission was chest pain (AR-DRG code F74B). The results of the logistic regression models on the likelihood that participants have a nonelective hospital admission or ED presentation are reported in Table 3. After adjustment for age, sex, and the presence of diabetes and other chronic diseases, those in the intervention group were 1.54 times more likely to have an admission, 3.26 times more likely to have a cardiac-related admission, and 2.07 times more likely to have an ED presentation than those in the control group.

Group allocation was not a significant predictor of the incidence (number) of all-cause or cardiac-related admissions or ED presentations (Table 4). The admission rate among participants in the intervention group was 1.69 times higher than the rate in the control group for all-cause admission, 1.56 times higher for cardiac-related admissions, and 1.90 times higher for ED presentations following adjustment for age, sex, and the diagnosis of diabetes or other chronic diseases (Table 4).

Results of the Cox regression analysis to determine if there were differences in time to admission or ED presentation between the intervention and control groups are presented in Table S4 in Multimedia Appendix 1. The intervention group participants had 1.52 times the probability of experiencing an all-cause admission in 12 months compared with the control group. They also had 3.14 times the probability of experiencing a cardiac-related admission and 1.84 times the probability of experiencing an ED presentation when adjusted for age, sex, and diagnosis of diabetes and other chronic diseases (Table S4 in Multimedia Appendix 1).

Table 2. Nonelective hospital admissions and emergency department presentations within 12 months of commencing cardiac rehabilitation.

	Control (n=60)	Intervention (n=60)	Difference between groups (intervention-control)	*P* value
Nonelective hospital admissions				
Proportion of participants who had at least one hospital admission (yes), n (%)	10 (17)	12 (20)	2 (3)	.64
Total number of admissions, n	13	18	5	.48
Cardiac-related admissions[a], n (%)	6 (46)	12 (67)	6 (10)	.22
Time frame from cardiac rehabilitation to admission (d), mean (SD)	131.50 (86.85)	117.25 (94.30)	−14.25 (81.30)	.83
Emergency department presentations				
Proportion who had an emergency department presentation (yes), n (%)	15 (25)	22 (37)	7 (12)	.17
Total number of emergency department presentations, n	20	36	16	.10

[a]Cardiac-related admission determined using Australian Refined Diagnosis Related Group codes (F01A to F10B, F12A to F12B, F14A to F19B, F22Z to F60B, F66A to F67B, and F69A to F76B).

Table 3. The odds of nonelective hospital admissions and emergency department presentations within 12 months using logistic regression models.

Dependent variable[a]	Model 1[b]		Model 2[c]		Model 3[d]	
	OR[e] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Nonelective all-cause hospital admissions	1.24 (0.49-3.16)	.64	1.31 (0.51-3.38)	.58	1.54 (0.58-4.10)	.39
Nonelective cardiac-related hospital admissions[f]	2.15 (0.617.58)	.23	2.24 (0.63-8.00)	.22	3.26 (0.84-12.55)	.09
Emergency department presentations	1.74 (0.79-3.81)	.17	1.79 (0.80-3.98)	.16	2.07 (0.89-4.77)	.09

[a]Reference=control.

[b]Model 1: nil adjustments.

[c]Model 2: adjusted for age and sex.

[d]Model 3: adjusted for age, sex, diabetes, and presence of other chronic diseases.

[e]OR: odds ratio.

[f]Cardiac-related admission determined using Australian Refined Diagnosis Related Group codes (F01A to F10B, F12A to F12B, F14A to F19B, F22Z to F60B, F66A to F67B, and F69A to F76B).

Table 4. The incidence of nonelective hospital admissions and emergency department presentations within 12 months using negative binomial regression with log link.

Dependent variable[a]	Model 1[b]		Model 2[c]		Model 3[d]	
	IRR[e] (95% CI)	*P* value	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value
Nonelective all-cause hospital admissions	1.39 (0.62-3.08)	.42	1.46 (0.65-3.31)	.36	1.69 (0.73-3.92)	.23
Nonelective cardiac-related hospital admissions[f]	1.83 (0.51-6.57)	.35	1.76 (0.46-6.73)	.41	1.56 (0.38-6.48)	.54
Emergency department presentations	1.80 (0.94-3.46)	.08	1.78 (0.91-3.44)	.09	1.90 (0.95-3.79)	.07

[a]Reference=control.

[b]Model 1: nil adjustments.

[c]Model 2: adjusted for age and sex.

[d]Model 3: adjusted for age, sex, diabetes, and the presence of other chronic diseases.

[e]IRR: incidence rate ratio.

[f]Cardiac-related admission determined using Australian Refined Diagnosis Related Group codes (F01A to F10B, F12A to F12B, F14A to F19B, F22Z to F60B, F66A to F67B, and F69A to F76B).

Secondary Outcomes

Sedentary behavior and physical activity measures are reported in Tables 5 and 6 and Table S5 in Multimedia Appendix 1. Both the intervention and control groups showed an increase in sedentary behavior over 12 months, spending approximately 10 hours per day in sedentary behaviors (Table S5 in Multimedia Appendix 1). There were no significant between- or within-group differences in any sedentary behavior measures, although there was a small effect size for the reduction in sedentary behavior (min/d) at 6 (Cohen *d*=0.11) and 12 months (Cohen *d*=0.21) in favor of the intervention group compared with the control group. At 6 and 12 months, the control group engaged in 15 minutes (*P*=.54) and 22 minutes (*P*=.33) more sedentary behavior than the intervention group when adjusted for age, sex, VM counts, employment, and education (Tables 5 and 6). In subanalyses, those in the intervention group who had prior experience with physical activity trackers (n=30) spent a lower percentage of the day in sedentary behavior (mean difference 6.13%, 95% CI 0.97%-11.28%; *P*=.02), had shorter sedentary bouts (mean difference 2.32 min, 95% CI 0.36-4.27; *P*=.02), and had lower overall sedentary minutes per day (mean difference 65.22 min, 95% CI −13.29 to 143.72; *P*=.10) at 6 months on the completion of the intervention (Table S6 in Multimedia Appendix 1). There were no significant differences in the control group based on prior physical activity tracker use (Table S7 in Multimedia Appendix 1).

There were no significant between-group differences at 6 or 12 months in any physical activity measures, except for wear time at 12 months (Table S5 in Multimedia Appendix 1). Nonetheless, the control group showed a significant within-group increase in VM counts at 6 months and light-intensity physical activity at 12 months. The results of the linear mixed-effects models are reported in Tables 5 and 6, and all models were nonsignificant.

All other secondary outcomes are presented in Table S8 in Multimedia Appendix 1. The control group had a higher BMI (Cohen *d*=0.30), waist circumference (Cohen *d*=0.43), and waist-to-hip ratio (Cohen *d*=0.33) at 6 and 12 months (Table S8 in Multimedia Appendix 1). Using linear mixed-effects models, BMI, waist circumference, and waist-to-hip ratio remained higher in the control group after adjustment for age, sex, employment, and education (Table S9 in Multimedia Appendix 1).

There was no significant between-group difference for systolic and diastolic blood pressure, 6-minute walking distance, anxiety, or depression (Tables S8 and S9 in Multimedia Appendix 1). There was a significant between-group difference in favor of the control group for quality of life (Cohen *d*=0.14; *P*=.03; Table S8 in Multimedia Appendix 1). In linear mixed models, the control group experienced a significantly higher overall quality of life at 6 and 12 months when adjusted for age, sex, employment, and education (Table S9 in Multimedia Appendix 1).

On the basis of the University of Rhode Island Change Assessment Scale readiness to change scores, participants were in the precontemplation stage throughout the study regarding physical activity change. There were no significant within-group or between-group differences (Table S8 in Multimedia Appendix 1) or significant differences over time (Table S9 in Multimedia Appendix 1). No significant differences were observed in sensitivity analyses for secondary outcomes between those wearing accelerometers during the COVID-19 pandemic lockdowns or completing telehealth measures.

Table 5. Difference in sedentary behavior and physical activity outcomes over 6 months between groups using linear mixed-effects models.

Dependent variable[a]	Model 1[b]			Model 2[c]			Model 3[d]		
	β (95% CI)	*P* value	ICC[e]	β (95% CI)	*P* value	ICC	β (95% CI)	*P* value	ICC
SB[f] (min/d)	18.92 (−32.16 to 70.00)	.47	0.71	18.62 (−29.13 to 66.36)	.44	0.66	14.91 (−32.85 to 62.67)	.54	0.65
Percentage of SB/d (SB/wear time)	1.21 (−2.12 to 4.53)	.47	0.75	.77 (−1.20 to 2.75)	.44	0.63	.63 (−1.34 to 2.61)	.53	0.63
Average duration of SB bouts (min)	−.50 (−1.65 to 0.66)	.40	0.62	−.90 (−1.91 to 0.12)	.08	0.55	−.93 (−1.94 to 0.08)	.07	0.54
Number of SB bouts/d	.88 (−1.08 to 2.83)	.38	0.72	.66 (−1.08 to 2.39)	.46	0.66	.50 (−1.23 to 2.23)	.57	0.65
Number of SB breaks/d	.87 (−1.08 to 2.82)	.38	0.72	.66 (−1.08 to 2.39)	.46	0.66	.50 (−1.23 to 2.23)	.57	0.65
MVPA[g] (min/d)	1.01 (−9.62 to 11.64)	.85	0.72	3.48 (−0.47 to 7.42)	.08	0.72	3.45 (−0.52 to 7.42)	.09	0.72
LPA[h] (min/d)	−11.59 (−35.96 to 12.78)	.35	0.70	−8.98 (−27.20 to 9.24)	.33	0.66	−8.44 (−26.74 to 9.86)	.36	0.66
Steps/d	−284.79 (−1323.94 to 754.36)	.59	0.69	−142.76 (−620.86 to 335.34)	.56	0.71	−146.87 (−628.19 to 334.45)	.55	0.71

[a]Reference=intervention.

[b]Model 1: nil adjustments.

[c]Model 2: adjusted for age and sex.

[d]Model 3: adjusted for age, sex, diabetes, and the presence of other chronic diseases.

[e]ICC: intraclass correlation coefficient.

[f]SB: sedentary behavior.

[g]MVPA: moderate to vigorous intensity physical activity.

[h]LPA: light-intensity physical activity.

Table 6. Difference in sedentary behavior and physical activity outcomes over 12 months between groups using linear mixed-effects models.

Dependent variable[a]	Model 1[b]			Model 2[c]			Model 3[d]		
	β (95% CI)	*P* value	ICC[e]	β (95% CI)	*P* value	ICC[e]	β (95% CI)	*P* value	ICC[e]
SB[f] (min/d)	24.52 (−24.15 to 73.20)	.32	0.66	24.50 (−20.08 to 69.09)	.28	0.62	21.94 (−22.80 to 66.69)	.33	0.62
Percentage of SB/d (SB/wear time)	1.31 (−2.01 to 4.62)	.44	0.74	.98 (−0.92 to 2.86)	.31	0.67	.85 (−1.04 to 2.74)	.38	0.67
Average duration of SB bouts (min)	−.19 (−1.42 to 1.04)	.76	0.69	−.57 (−1.68 to 0.54)	.31	0.63	−.62 (−1.72 to 0.49)	.27	0.63
Number of SB bouts/d	.90 (−0.95 to 2.75)	.33	0.65	.75 (−0.85 to 2.35)	.36	0.59	.64 (−0.95 to 2.24)	.43	0.59
Number of SB breaks/d	.89 (−0.95 to 2.74)	.34	0.65	.75 (−0.85 to 2.34)	.36	0.59	.64 (−0.96 to 2.24)	.43	0.58
MVPA[g] (min/d)	1.42 (−9.04 to 11.87)	.79	0.73	3.10 (−0.80 to 6.99)	.12	0.73	3.06 (−0.86 to 6.98)	.13	0.73
LPA[h] (min/d)	−9.48 (−33.96 to 14.99)	.44	0.73	−7.66 (−25.31 to 9.98)	.39	0.72	−7.01 (−24.66 to 10.65)	.43	0.72
Steps/d	−138.84 (−1151.26 to 873.58)	.79	0.74	−37.65 (−508.40 to 433.10)	.87	0.73	−36.24 (−510.30 to 437.83)	.88	0.73

[a]Reference=intervention.
[b]Model 1: nil adjustments.
[c]Model 2: adjusted for age and sex.
[d]Model 3: adjusted for age, sex, diabetes, and the presence of other chronic diseases.
[e]ICC: intraclass correlation coefficient.
[f]SB: sedentary behavior.
[g]MVPA: moderate to vigorous intensity physical activity.
[h]LPA: light-intensity physical activity.

Engagement and Usability of the Vire App

The median completion rate of *Dos* was 3 of 55 (IQR 0-37.75) over 6 months. There were 27% (16/60) of participants who engaged with the app for the entire 6 months (ie, completed at least 1 *Do*/mo), 33% (20/60) of participants who engaged with the app less than once per month, and 40% (24/60) of participants who did not complete any *Dos*. In logistic regression modeling, those who had used apps or wearable activity trackers before were 1.04 times (*P*=.01) more likely to complete *Dos* (χ^2_1=7.1). Those who were employed full time or part time were 0.97 times (*P*=.01) less likely to complete *Dos* (χ^2_1=6.6). There were no statistically significant correlations among the total number of *Dos* completed and age (*r*=−0.07; *P*=.60), sedentary behavior (*r*=−0.15; *P*=.25), physical activity (*r*=0.21; *P*=.11), quality of life (*r*=0.18; *P*=.17), anxiety (*r*=−0.12; *P*=.35) or depression (*r*=−0.15; *P*=.26) at 6 months (Table S10 in Multimedia Appendix 1).

The usability of the Vire app (Unified Theory of Acceptance and Use of Technology) is presented in Table S11 in Multimedia Appendix 1. Participants were relatively satisfied with the usability of the Vire app, with a median score of ≥4 in all constructs except for habit (score=3.54/7) and use (score=3.29/7).

Economic Evaluation

Overview

On average, the cost of implementing the intervention per participant was Aus $1086.55 (Table S12 in Multimedia Appendix 1). The mean cost of all-cause and cardiac-related hospital admissions was higher in the control group at 6 months and remained higher in the control group for cardiac-related admissions at 12 months. However, these costs were nonsignificant between groups (Table S13 in Multimedia Appendix 1). The health care use costs for ED presentations at 6 months were significantly higher in the intervention group (Aus $315.83 vs Aus $136.00; *P*=.04).

Cost-Effectiveness

The cost-effectiveness of the Vire app and ToDo-CR program compared with usual care is presented in Table 7 for the secondary outcomes. Although the intervention was costlier to implement and this group had higher health care use, it was also more effective at reducing the BMI, increasing light-intensity physical activity at 6 months, improving quality of life and anxiety symptoms at 6 and 12 months, and reducing sedentary behavior at 12 months. The ICERs presented in Table 6 represent the cost per unit change for each outcome. For example, the cost of a 1-minute reduction in sedentary time at 12 months was Aus $351.77.

Table 7. Health care use cost and health benefit differences at 6 and 12 months considering all-cause and cardiac-related nonelective hospital admissions per participant.

	All cause				Cardiac related[a]			
	Intervention, mean change (SD)	Control, mean change (SD)	Mean difference (95% CI)	ICER[b] (Aus $[c])	Intervention, mean change (SD)	Control, mean change (SD)	Mean difference (95% CI)	ICER (Aus $)
Baseline to 6 mo								
Health care use costs (Aus $)[d]	2041.53 (1860.55)[e]	1423.63 (6248.96)	617.90 (−1048.97 to 2284.77)	—[f]	1819.23 (1246.59)	1079.65 (6096.69)	739.58 (−851.30 to 2330.46)	—
Hospital admissions	361.00 (1287.94)	1111.48 (5948.69)	−750.48 (−2306.51 to 805.55)	—	138.70 (413.12)	767.50 (5838.20)	−628.80 (−2125.08 to 867.48)	—
ED[g]	315.83 (605.23)	136.00 (326.61)	179.83 (3.46 to 356.20)	—	—	—	—	—
Effectiveness								
SB[h] (min/d)	16.84 (109.94)	14.43 (114.34)	2.41 (−38.15 to 42.96)	256.39	16.84 (109.94)	14.43 (114.34)	2.41 (−38.15 to 42.96)	306.88
MVPA[i] (min/d)	0.37 (24.20)	4.65 (20.02)	−4.28 (−12.31 to 3.75)	−144.37	0.37 (24.20)	4.65 (20.02)	−4.28 (−12.31 to 3.75)	−172.80
LPA[j] (min/d)	11.80 (60.92)	10.84 (47.55)	0.95 (−18.80 to 20.71)	643.65	11.80 (60.92)	10.84 (47.55)	0.95 (−18.80 to 20.71)	770.40
BMI (kg/m^2)	−0.20 (1.29)	0.03 (1.35)	−0.23 (−0.71 to 0.25)	−2686.52	−0.20 (1.29)	0.03 (1.35)	−0.23 (−0.71 to 0.25)	−3215.57
Waist circumference (cm)	−1.29 (4.50)	−2.18 (6.23)	0.89 (−1.08 to 2.86)	694.27	−1.29 (4.50)	−2.18 (6.23)	0.89 (−1.08 to 2.86)	830.99
Systolic blood pressure (mm Hg)	5.52 (12.87)	3.36 (14.81)	2.16 (−3.41 to 7.74)	286.06	5.52 (12.87)	3.36 (14.81)	2.16 (−3.41 to 7.74)	342.40
6-min walk test distance (m)	55.69 (62.19)	84.42 (127.01)	−28.73 (−77.89 to 20.42)	−21.51	55.69 (62.19)	84.42 (127.01)	−28.73 (−77.89 to 20.42)	−25.74
AQoL[k]-6D Utility (0-1.0)	0.03 (0.10)	0.02 (0.07)	0.004 (−0.03 to 0.04)	61,790.00	0.03 (0.10)	0.02 (0.07)	0.004 (−0.03 to 0.04)	73,958.00
Anxiety (0-21)	−0.57 (2.13)	−0.35 (2.07)	0.38 (−0.97 to 0.54)	−2808.64	−0.57 (2.13)	−0.35 (2.07)	−0.22 (−0.97 to 0.54)	−3361.73
Depression (0-21)	−0.45 (1.67)	−0.58 (1.92)	0.13 (−0.52 to 0.78)	4753.08	−0.45 (1.67)	−0.58 (1.92)	0.13 (−0.52 to 0.78)	5689.08
Baseline to 12 mo								
Health care use costs (Aus $)[c]	3507.08 (11,730.49)	1639.17 (6359.44)	1867.92 (−1543.35 to 5279.19)	—	1932.93 (1640.97)	1171.53 (6134.12)	761.40 (−861.94 to 2384.74)	—
Hospital admissions	1826.55 (11,454.49)	1327.02 (6007.08)	499.53 (−2807.09 to 3806.16)	—	252.40 (784.37)	859.38 (5850.84)	−606.98 (−2116.14 to 092.18)	—
ED	593.98 (1046.12)	312.15 (686.72)	281.83 (−38.61 to 602.28)	—	—	—	—	—
Effectiveness								

	All cause				Cardiac related[a]			
	Intervention, mean change (SD)	Control, mean change (SD)	Mean difference (95% CI)	ICER[b] (Aus $[c])	Intervention, mean change (SD)	Control, mean change (SD)	Mean difference (95% CI)	ICER (Aus $)
SB (min/d)	19.99 (105.30)	25.30 (156.05)	−5.31 (−53.44 to 42.82)	−351.77	19.99 (105.30)	25.30 (156.05)	−5.31 (−53.44 to 42.82)	−143.39
MVPA (min/d)	−1.39 (29.49)	1.94 (19.64)	−3.33 (−12.39 to 5.72)	−9.56	−1.39 (29.49)	1.94 (19.64)	−3.33 (−12.39 to 5.72)	−228.65
LPA (min/d)	7.16 (64.21)	13.80 (52.34)	−6.64 (−27.82 to 14.54)	−281.31	7.16 (64.21)	13.80 (52.34)	−6.64 (−27.82 to 14.54)	−114.67
BMI (kg/m^2)	0.38 (1.34)	0.34 (1.44)	0.04 (−0.46 to 0.55)	46,697.75	0.38 (1.34)	0.34 (1.44)	0.04 (−0.46 to 0.55)	19,035.00
Waist circumference (cm)	0.37 (5.64)	−0.42 (5.61)	0.78 (−1.25 to 2.82)	2364.44	0.37 (5.64)	−0.42 (5.61)	0.78 (−1.25 to 2.82)	963.80
Systolic blood pressure (mm Hg)	4.59 (13.12)	7.36 (20.81)	−2.77 (−9.84 to 4.29)	−674.34	4.59 (13.12)	7.36 (20.81)	−2.77 (−9.84 to 4.29)	−274.87
6-min walk test distance (m)	96.95 (130.96)	97.37 (129.12)	−0.43 (−59.09 to 58.23)	−4344.00	96.95 (130.96)	97.37 (129.12)	−0.43 (−59.09 to 58.23)	−1770.70
AQoL-6D Utility (0-1.0)	0.04 (0.09)	0.02 (0.12)	0.02 (−0.02 to 0.06)	93,395.50	0.04 (0.09)	0.02 (0.12)	0.02 (−0.02 to 0.06)	38,070.00
Anxiety (0-21)	−0.65 (1.87)	−0.48 (2.04)	−0.17 (−0.87 to 0.54)	−10,987.71	−0.65 (1.87)	−0.48 (2.04)	−0.17 (−0.87 to 0.54)	−4478.82
Depression (0-21)	−0.35 (1.80)	−0.95 (2.03)	0.60 (−0.09 to 1.29)	3113.18	−0.35 (1.80)	−0.95 (2.03)	0.60 (−0.09 to 1.29)	1269.00

[a]Cardiac-related admission determined by Australian Refined Diagnosis Related Group codes (F01A to F10B, F12A to F12B, F14A to F19B, F22Z to F60B, F66A-F67B, and F69A to F76B).

[b]ICER: incremental cost-effectiveness ratio.

[c]A currency exchange rate of Aus $1=US $0.69 is applicable.

[d]Direct cost of intervention and indirect cost of either all-cause or cardiac-related hospital admissions and emergency department presentations per participant, Aus $.

[e]Includes cost of intervention: Aus $1086.55. ICER = (intervention cost – control cost) / (intervention effect – control effect), where effect is the health outcome.

[f]Health care use costs are accounted for in the ICER calculation for secondary outcomes.

[g]ED: emergency department.

[h]SB: sedentary behavior.

[i]MVPA: moderate to vigorous intensity physical activity.

[j]LPA: light-intensity physical activity.

[k]AQoL: Assessment of Quality of Life.

Discussion

Principal Findings

The use of the Vire app and ToDo-CR program was not effective in reducing hospital admissions and ED presentations nor did it significantly decrease sedentary behavior compared with usual care over 12 months. Participants in the intervention group were more likely to have a cardiac-related hospital admission; however, the costs of these admissions were markedly lower than those in the control group. Although the intervention was costlier to implement, it was also more effective at reducing sedentary behavior, BMI, and anxiety and increasing quality of life and light-intensity physical activity. Retention rates were high in this study; however, engagement with the Vire app and ToDo-CR program was low. Furthermore, there was no correlation between age and engagement with the Vire app and ToDo-CR program; instead, there was a correlation between prior experience with physical activity trackers and apps.

Comparison With Prior Work

Intervention participants were 50% more likely to have a hospital admission, twice as likely to have an ED presentation, and 3 times more likely to be admitted to the hospital for a cardiac-related hospital admission. This contrasts with the studies by Widmer et al [25] and Rivers et al [26], who each noted an approximate 30% decrease in the rate of hospital admissions in the groups using an app-based intervention in cardiac rehabilitation. These studies targeted a range of lifestyle risk factors (eg, physical activity, smoking, diet, and medication adherence) rather than 1 risk factor. Perhaps by targeting only sedentary behavior, the strength of the intervention was not enough to produce significant changes in the primary outcome, noting that multiple factors contribute to hospital admissions and risk factor control such as medication and comorbidity management [53].

Cardiac rehabilitation is associated with significant reductions in hospitalizations and repeat cardiac events [4]. Nonetheless, of those who are referred for cardiac rehabilitation in Australia, only 28% attend [54]. Alternate technology-based methods (such as the Vire app and ToDo-CR program) may therefore be better suited to reach those not attending traditional cardiac rehabilitation who are at a higher risk of hospitalizations compared with those who attend traditional cardiac rehabilitation. This group is missing the behavior change advice and support provided in cardiac rehabilitation and may have the most to gain by engaging. Providing the option of an app-based cardiac rehabilitation program is associated with increased overall cardiac rehabilitation participation rates [26] by removing barriers such as the need to travel. In addition, app-based cardiac rehabilitation programs have been shown to achieve outcomes comparable with traditional programs [24,55,56]. One study reported readmission data comparing no cardiac rehabilitation, traditional cardiac rehabilitation, and app-based cardiac rehabilitation [26]. Although this study was not designed to evaluate differences in readmissions, cardiac-related admissions were lower in those using the app (1/23, 4%) versus those receiving no cardiac rehabilitation (5/39, 13%) [26]. Larger-scale studies are required to investigate the effect of smartphone apps in those declining cardiac rehabilitation and the possible benefits to hospitalizations.

Those in the intervention group attended the ED more frequently than those in the control group, and the most frequent cause for admission was chest pain (AR-DRG code F74B). Presenting to the ED for chest pain is particularly relevant with advice given in the Vire app and ToDo-CR program to seek medical attention if participants experienced these symptoms. These participants may have become more proactive in self-managing their CHD because of the education provided. Public education campaigns to raise awareness of the signs and symptoms of acute chest pain have been shown to be effective in increasing the rate of ED presentations for early medical intervention [57,58]. Despite having more admissions, the intervention group costs associated with cardiac-related admissions were lower (intervention, Aus $252.40 vs control, Aus $859.30), meaning that when they were admitted, they potentially had shorter length of stays for less-severe diagnoses, requiring less-costly medical intervention and overall less impact on the health care system. Delays in presenting to the ED with acute chest pain, a potential indicator of a myocardial infarction, contribute to patient morbidity and mortality [57,59]. Smartphone apps are being developed to support patients with CHD and incorporate educational features regarding symptoms of chest pain, with a key focus being to support participants rather than just monitor them [60]. These types of features and providing education may be especially important in future studies for those not attending face-to-face cardiac rehabilitation programs.

Although the effect size for reduction in sedentary behavior was small in this study, an approximate 20-minute reduction may still be clinically meaningful. In the limited studies available that test the effect of smartphone apps on sedentary behavior [30], greater reductions have been seen compared with this study in cardiac rehabilitation participants (accelerometer-measured 100-min reduction/d at 4 mo [28]; accelerometer-measured 96-min reduction/d at 3 mo [29]); stroke (accelerometer-measured 60-min reduction/d at 6 wk [61]); and chronic stroke (self-reported 180-min reduction/d at 3 mo [62]). There is no well-established, minimal clinically important difference for sedentary behavior in people with cardiovascular disease including CHD. Studies in the general population have reported that for every 30 minutes of sedentary behavior reallocated to light-intensity or moderate to vigorous intensity physical activity, there is a 2% to 25% improvement in cardiovascular disease risk biomarkers [63], and a 1- to 2-hour reduction in television viewing time is associated with reductions in cardiovascular disease risk (eg, waist circumference, BMI, triglycerides, insulin sensitivity, and blood pressure) [64,65]. In addition, a break in sedentary behavior as short as 1 minute can reduce waist circumference and improve C-reactive protein (inflammatory marker) independent of total sedentary time [65,66]. This study showed minimal change in sedentary bouts and breaks between groups or over time. Sedentary behavior bouts and breaks are also infrequently reported [30]. Future iterations of apps such as the Vire app and ToDo-CR program may benefit from providing specific advice to reach such targets and help with creating more substantial levels of change [67].

The waist-to-hip ratio (0.03) and BMI (1.64 kg/m^2) significantly reduced in the intervention group; however, they were unlikely to be clinically significant changes. The intervention also had a statistically significant 6-cm reduction in waist circumference, which could be considered clinically significant after accounting for measurement error (≥2 cm) [68]. These results combined may indicate an overall lower central adiposity and lower cardiovascular disease risk profile, in line with previous app-based studies in CHD participants [25].

A high percentage of the intervention group did not complete a single *Do* message (24/60, 40%), which was used as the marker of engagement. There may have been a misunderstanding of the need to tick off the *Dos* after viewing them or, alternatively, the intervention did not appeal to them. Previous studies on cardiovascular disease and smartphone apps have reported similarly low engagement and adherence levels [25,62,69-71] and that levels tend to decrease with time in the intervention [30]. Similar studies have also reported technical difficulties as a key reason for low engagement [30,62,69,70,72]. Low levels of engagement, as seen in this study, could result in an engagement level that is insufficient to achieve the intended effect [73], with increasing evidence that digital health apps for chronic disease self-management require ongoing patient engagement as a key determinant of overall clinical impact [74-77].

In line with the feasibility trial preceding this study [28], not having a smartphone was a major reason for exclusion. Among those who declined to participate, the main reason provided was not being interested in smartphone apps. Our study was consistent with previous studies, with people declining to participate being significantly older [28,30] and younger cardiac rehabilitation participants being more likely to use smartphones [22,23,30,78]. Patient preference for the type of intervention may be key to their success [26,67]. In a study comparing multiple delivery methods for cardiac rehabilitation including app-based cardiac rehabilitation, approximately 25% of the group declining the app-based approach listed technology issues as the main reason for nonparticipation [26]. Perhaps key messaging moving forward is that those who want to use smartphone apps will engage regardless of age [67].

Although smartphone apps have the potential to be a cost-effective solution [79], further work is required to ensure that the health benefits offset the initial increased costs of setting up and implementing such interventions. There is 1 relevant comparison study that evaluated the cost-effectiveness of a smartphone app in people with heart failure [80]. This number highlights the relative infancy of research in this area. The CardioManager app achieved greater savings in the management of heart failure with an ICER of €9000 (US $9595.8) per patient, equating to large reductions in costs associated with hospital admissions [80]. Future research in this area would benefit from reporting the economic evaluations of smartphone apps to determine their cost-effectiveness and improve research translation and real-world implementation.

Strengths and Limitations

This study has several strengths being one of the first published studies to explore the impact of a sedentary behavior change smartphone app on hospital admissions and its cost-effectiveness. Device-measured sedentary behavior was analyzed, personalization and clinical guidelines were championed, and the study included both public and private hospitals. This is critical for improving the implementation of research apps in real-world settings.

Despite these strengths, there were several limitations. First, this trial was interrupted during the COVID-19 pandemic, impacting the recruitment and follow-up of participants and ultimately resulting in incomplete data. The number of participants who completed outcomes remotely have been reported, and sensitivity analyses were completed in line with guidelines for reporting trials affected by the pandemic [34]. No participants reported dropping out owing to the pandemic; however, there were difficulties recruiting, and hence, the target sample was not reached. Despite this limitation, the final dropout rate was lower than expected; therefore, the required sample size was maintained. Second, the majority of participants were male, tertiary educated, and working, and those excluded were significantly older, had more severe diagnoses, and were predominantly from the public health system, limiting generalizability. There were multiple assessors that may have affected the measurement error of outcomes such as waist circumference. Measuring engagement with the Vire app and ToDo-CR program was also limited. Further investigation into app engagement using back-end data is needed to better understand the relationship between app use and changes in sedentary behavior. Although half of the participants had experience using physical activity trackers, there is evidence that inexperienced users may not use all app features and therefore may not obtain the full anticipated benefits of behavior change smartphone apps [81]. In addition, although the use of ActiGraph accelerometers provides an objective measure of sedentary behavior, they are less sensitive to postural changes than devices such as ActiPal [13]. Future research would benefit from the use of inclinometers to detect changes from sitting to lying to standing and from combining this information with heart rate monitors and GPS technology to better inform the holistic picture of individual behavior surrounding sedentariness [82]. Finally, bringing the last value forward for missing data and the lack of participant blinding inherently introduced bias to the methods used.

Conclusions

It does not appear that the Vire app and ToDo-CR program targeting sedentary behavior is an outcome-effective or cost-effective solution to reduce all-cause hospital admissions or ED presentations in cardiac rehabilitation participants. Although those using the Vire app and the ToDo-CR program had more hospital admissions, these admissions were less costly. Further research is warranted to improve engagement and implementation with age appearing to be less of an impacting factor and prior experience with apps correlating more with engagement. This type of intervention may work as a better resource for those not already attending cardiac rehabilitation, who are at greater risk of hospitalizations, to influence sedentary behavior and potentially reduce costs associated with cardiac-related hospital admissions.

Acknowledgments

The authors gratefully acknowledge the participants who took the time to participate in this study. They also gratefully acknowledge Bridget Bourke and Jayden Black for their assistance with follow-up assessments; the support of Genevieve Fletcher, Heather Chadwick, and Allyce Holmes for assisting with recruitment; and the cardiac rehabilitation teams of the Canberra Hospital, Calvary Public Hospital, and National Capital Private Hospital. This work was supported by the Medical Research Future Fund (grant 1184607), and the lead author was further supported by a Digital Health Cooperative Research Centre Doctor of Philosophy top-up scholarship. These funding bodies were not involved in the design, analysis, interpretation, or writing of the manuscript.

Authors' Contributions

All authors contributed to the study design. KP drafted the manuscript and completed all the analyses. All authors contributed to the interpretation of the results and subsequent drafts and have read and approved the final manuscript.

Conflicts of Interest

The Vire app and ToDo-CR was created by a private company, Onmi. Onmi provided no funding for this study. SvB is a manager and designer for Onmi and the Vire app and the developer of the ToDo-CR behavior change program. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplementary tables for the ToDo-CR randomized controlled trial.
[DOCX File , 86 KB - mhealth_v11i1e48229_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1314 KB - mhealth_v11i1e48229_app2.pdf]

References

1. Disease expenditure in Australia. Australian Institute of Health and Welfare. 2021 Aug 25. URL: https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-australia [accessed 2023-03-25]
2. The top 10 causes of death. World Health Organization. 2020 Dec 09. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2023-03-25]
3. Timmis A, Vardas P, Townsend N, Torbica A, Katus H, De Smedt D, et al. European society of cardiology: cardiovascular disease statistics 2021: executive summary. Eur Heart J Qual Care Clin Outcomes 2022 Jun 06;8(4):377-382. [doi: 10.1093/ehjqcco/qcac014] [Medline: 35488372]
4. Dibben GO, Faulkner J, Oldridge N, Rees K, Thompson DR, Zwisler AD, et al. Exercise-based cardiac rehabilitation for coronary heart disease: a meta-analysis. Eur Heart J 2023 Feb 07;44(6):452-469 [FREE Full text] [doi: 10.1093/eurheartj/ehac747] [Medline: 36746187]
5. ACS in perspective: the importance of secondary prevention. Deloitte Access Economics. 2011. URL: https://www2.deloitte.com/au/en/pages/economics/articles/acs-perspective-importance-secondary-prevention.html [accessed 2023-03-25]
6. Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016 Jan 05;2016(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
7. Visseren F, Mach F, Smulders Y, Carballo D, Koskinas K, Bäck M. Corrigendum to: 2021 ESC guidelines on cardiovascular disease prevention in clinical practice: developed by the task force for cardiovascular disease prevention in clinical practice with representatives of the European society of cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). Eur Heart J 2022 Nov 07;43(42):4468. [doi: 10.1093/eurheartj/ehac458] [Medline: 36083202]
8. Freene N, McManus M, Mair T, Tan R, Davey R. Association of device-measured physical activity and sedentary behaviour with cardiovascular risk factors, health-related quality-of-life and exercise capacity over 12-months in cardiac rehabilitation attendees with coronary heart disease. BMC Sports Sci Med Rehabil 2022 Sep 07;14(1):169 [FREE Full text] [doi: 10.1186/s13102-022-00562-7] [Medline: 36071477]
9. Prince SA, Blanchard CM, Grace SL, Reid RD. Objectively-measured sedentary time and its association with markers of cardiometabolic health and fitness among cardiac rehabilitation graduates. Eur J Prev Cardiol 2016 May;23(8):818-825. [doi: 10.1177/2047487315617101] [Medline: 26607698]
10. Kotseva K, De Backer G, De Bacquer D, Rydén L, Hoes A, Grobbee D, EUROASPIRE Investigators*. Lifestyle and impact on cardiovascular risk factor control in coronary patients across 27 countries: results from the European Society of Cardiology ESC-EORP EUROASPIRE V registry. Eur J Prev Cardiol 2019 May 10;26(8):824-835 [FREE Full text] [doi: 10.1177/2047487318825350] [Medline: 30739508]
11. Biswas A, Oh P, Faulkner G, Alter D. A prospective study examining the influence of cardiac rehabilitation on the sedentary time of highly sedentary, physically inactive patients. Ann Phys Rehabil Med 2018 Jul;61(4):207-214 [FREE Full text] [doi: 10.1016/j.rehab.2017.06.003] [Medline: 28923367]
12. Bakker EA, van Bakel BM, Aengevaeren WR, Meindersma EP, Snoek JA, Waskowsky WM, et al. Sedentary behaviour in cardiovascular disease patients: risk group identification and the impact of cardiac rehabilitation. Int J Cardiol 2021 Mar 01;326:194-201 [FREE Full text] [doi: 10.1016/j.ijcard.2020.11.014] [Medline: 33186667]
13. Prince SA, Reed JL, Cotie LM, Harris J, Pipe AL, Reid RD. Results of the Sedentary Intervention Trial in Cardiac Rehabilitation (SIT-CR Study): a pilot randomized controlled trial. Int J Cardiol 2018 Oct 15;269:317-324. [doi: 10.1016/j.ijcard.2018.07.082] [Medline: 30072156]

14. Freene N, McManus M, Mair T, Tan R, Davey R. Objectively measured changes in physical activity and sedentary behavior in cardiac rehabilitation: a prospective cohort study. J Cardiopulm Rehabil Prev 2018 Nov;38(6):E5-E8. [doi: 10.1097/HCR.0000000000000334] [Medline: 29952812]
15. Evenson KR, Butler EN, Rosamond WD. Prevalence of physical activity and sedentary behavior among adults with cardiovascular disease in the United States. J Cardiopulm Rehabil Prev 2014 Nov;34(6):406-419 [FREE Full text] [doi: 10.1097/HCR.0000000000000064] [Medline: 25068393]
16. Ter Hoeve N, Sunamura M, van Geffen ME, Fanchamps MH, Horemans HL, Bussmann JB, et al. Changes in physical activity and sedentary behavior during cardiac rehabilitation. Arch Phys Med Rehabil 2017 Dec;98(12):2378-2384. [doi: 10.1016/j.apmr.2017.05.008] [Medline: 28623143]
17. van der Ploeg HP, Chey T, Korda RJ, Banks E, Bauman A. Sitting time and all-cause mortality risk in 222 497 Australian adults. Arch Intern Med 2012 Mar 26;172(6):494-500. [doi: 10.1001/archinternmed.2011.2174] [Medline: 22450936]
18. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
19. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
20. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, Lancet Physical Activity Series 2 Executive Committe, Lancet Sedentary Behaviour Working Group. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet 2016 Sep 24;388(10051):1302-1310. [doi: 10.1016/S0140-6736(16)30370-1] [Medline: 27475271]
21. Rogerson MC, Le Grande MR, Dunstan DW, Magliano DJ, Murphy BM, Salmon J, et al. Television viewing time and 13-year mortality in adults with cardiovascular disease: data from the Australian diabetes, obesity and lifestyle study (AusDiab). Heart Lung Circ 2016 Aug;25(8):829-836. [doi: 10.1016/j.hlc.2016.03.006] [Medline: 27262392]
22. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016 Jul 19;16(1):95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9] [Medline: 27431419]
23. Gallagher R, Roach K, Sadler L, Glinatsis H, Belshaw J, Kirkness A, et al. Mobile technology use across age groups in patients eligible for cardiac rehabilitation: survey study. JMIR Mhealth Uhealth 2017 Oct 24;5(10):e161 [FREE Full text] [doi: 10.2196/mhealth.8352] [Medline: 29066425]
24. Funahashi T, Borgo L, Joshi N. Saving lives with virtual cardiac rehabilitation. NEJM Catal 2019;5(4) [FREE Full text] [doi: 10.1056/CAT.19.0624]
25. Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]
26. Rivers JT, Smith C, Smith I, Cameron J. The impact of a mobile app on participation in cardiac rehabilitation and understanding barriers to success: comparative cohort study. JMIR Cardio 2022 Jan 17;6(1):e24174 [FREE Full text] [doi: 10.2196/24174] [Medline: 35037891]
27. Sengupta A, Beckie T, Dutta K, Dey A, Chellappan S. A mobile health intervention system for women with coronary heart disease: usability study. JMIR Form Res 2020 Jun 03;4(6):e16420 [FREE Full text] [doi: 10.2196/16420] [Medline: 32348270]
28. Freene N, van Berlo S, McManus M, Mair T, Davey R. A behavioral change smartphone app and program (ToDo-CR) to decrease sedentary behavior in cardiac rehabilitation participants: prospective feasibility cohort study. JMIR Form Res 2020 Nov 03;4(11):e17359 [FREE Full text] [doi: 10.2196/17359] [Medline: 33141091]
29. van Bakel BM, Kroesen SH, Bakker EA, van Miltenburg RV, Günal A, Scheepmaker A, et al. Effectiveness of an intervention to reduce sedentary behaviour as a personalised secondary prevention strategy for patients with coronary artery disease: main outcomes of the SIT LESS randomised clinical trial. Int J Behav Nutr Phys Act 2023 Feb 14;20(1):17 [FREE Full text] [doi: 10.1186/s12966-023-01419-z] [Medline: 36788615]
30. Patterson K, Davey R, Keegan R, Freene N. Smartphone applications for physical activity and sedentary behaviour change in people with cardiovascular disease: a systematic review and meta-analysis. PLoS One 2021 Oct 11;16(10):e0258460 [FREE Full text] [doi: 10.1371/journal.pone.0258460] [Medline: 34634096]
31. Patterson K, Davey R, Keegan R, Niyonsenga T, Mohanty I, van Berlo S, et al. A smartphone app for sedentary behaviour change in cardiac rehabilitation and the effect on hospital admissions: the ToDo-CR randomised controlled trial study protocol. BMJ Open 2020 Dec 15;10(12):e040479 [FREE Full text] [doi: 10.1136/bmjopen-2020-040479] [Medline: 33323435]
32. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

33. Orkin AM, Gill PJ, Ghersi D, Campbell L, Sugarman J, Emsley R, CONSERVE Group. Guidelines for reporting trial protocols and completed trials modified due to the COVID-19 pandemic and other extenuating circumstances: the CONSERVE 2021 statement. JAMA 2021 Jul 20;326(3):257-265. [doi: 10.1001/jama.2021.9941] [Medline: 34152382]
34. Perlis RH, Haneuse SJ, Rubenfeld GD, Fihn SD, Rivara FP. Reporting clinical studies affected by the COVID-19 pandemic: guidelines for authors. JAMA Netw Open 2021 Jan 04;4(1):e2036155 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.36155] [Medline: 33443576]
35. Habibović M, Broers E, Piera-Jimenez J, Wetzels M, Ayoola I, Denollet J, et al. Enhancing lifestyle change in cardiac patients through the do change system ("do cardiac health: advanced new generation ecosystem"): randomized controlled trial protocol. JMIR Res Protoc 2018 Feb 08;7(2):e40 [FREE Full text] [doi: 10.2196/resprot.8406] [Medline: 29422454]
36. Fletcher B, Pine KJ. Flex: Do Something Different. Hatfield, UK: University of Hertfordshire Press; 2012.
37. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
38. Freene N, Borg S, McManus M, Mair T, Tan R, Davey R, et al. Comparison of device-based physical activity and sedentary behaviour following percutaneous coronary intervention in a cohort from Sweden and Australia: a harmonised, exploratory study. BMC Sports Sci Med Rehabil 2020 May 09;12(1):17 [FREE Full text] [doi: 10.1186/s13102-020-00164-1] [Medline: 32419950]
39. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson P. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
40. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
41. Bellet RN, Adams L, Morris NR. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness--a systematic review. Physiotherapy 2012 Dec;98(4):277-286. [doi: 10.1016/j.physio.2011.11.003] [Medline: 23122432]
42. Richardson J, Atherton Day N, Peacock S, Iezzi A. Measurement of the quality of life for economic evaluation and the Assessment of Quality of Life (AQoL) mark 2 instrument. Aust Econ Rev 2004 Mar;37(1):62-88 [FREE Full text] [doi: 10.1111/j.1467-8462.2004.00308.x]
43. Richardson JR, Peacock SJ, Hawthorne G, Iezzi A, Elsworth G, Day NA. Construction of the descriptive system for the Assessment of Quality of Life AQoL-6D utility instrument. Health Qual Life Outcomes 2012 Apr 17;10(1):38 [FREE Full text] [doi: 10.1186/1477-7525-10-38] [Medline: 22507254]
44. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
45. Reed GR. The transtheoretical model and exercise behavior: a comparison of five staging methods. Open Access Master's Theses. 1993. URL: https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=2307&context=theses [accessed 2023-03-25]
46. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
47. AR-DRG version 8.0. Independent Health and Aged Care Pricing Authority. URL: https://www.ihacpa.gov.au/resources/ar-drg-version-80 [accessed 2023-01-27]
48. NWAU calculators. Independent Health and Aged Care Pricing Authority. URL: https://www.ihacpa.gov.au/health-care/pricing/nwau-calculators [accessed 2023-01-27]
49. Shaw BA, Spokane LS. Examining the association between education level and physical activity changes during early old age. J Aging Health 2008 Oct;20(7):767-787 [FREE Full text] [doi: 10.1177/0898264308321081] [Medline: 18559963]
50. Wolff-Hughes DL, Bassett DR, Fitzhugh EC. Population-referenced percentiles for waist-worn accelerometer-derived total activity counts in U.S. youth: 2003 - 2006 NHANES. PLoS One 2014 Dec 22;9(12):e115915 [FREE Full text] [doi: 10.1371/journal.pone.0115915] [Medline: 25531290]
51. Kwon S, Andersen LB, Grøntved A, Kolle E, Cardon G, Davey R, ICAD Collaborators. A closer look at the relationship among accelerometer-based physical activity metrics: ICAD pooled data. Int J Behav Nutr Phys Act 2019 Apr 29;16(1):40 [FREE Full text] [doi: 10.1186/s12966-019-0801-x] [Medline: 31036032]
52. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974 Dec;19(6):716-723 [FREE Full text] [doi: 10.1109/tac.1974.1100705]
53. Jortveit J, Halvorsen S, Kaldal A, Pripp AH, Govatsmark RE, Langørgen J. Unsatisfactory risk factor control and high rate of new cardiovascular events in patients with myocardial infarction and prior coronary artery disease. BMC Cardiovasc Disord 2019 Mar 28;19(1):71 [FREE Full text] [doi: 10.1186/s12872-019-1062-y] [Medline: 30922234]
54. Astley CM, Chew DP, Keech W, Nicholls S, Beltrame J, Horsfall M, et al. The impact of cardiac rehabilitation and secondary prevention programs on 12-month clinical outcomes: a linked data analysis. Heart Lung Circ 2020 Mar;29(3):475-482. [doi: 10.1016/j.hlc.2019.03.015] [Medline: 31072769]

55. Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
56. Yudi MB, Clark DJ, Tsang D, Jelinek M, Kalten K, Joshi S, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes [SMART-REHAB Trial]: a randomized controlled trial protocol. BMC Cardiovasc Disord 2016 Sep 05;16(1):170 [FREE Full text] [doi: 10.1186/s12872-016-0356-6] [Medline: 27596569]
57. Wright RS, Kopecky SL, Timm M, Pflaum DD, Carr C, Evers K, Wabasha Heart Attack Team. Impact of community-based education on health care evaluation in patients with acute chest pain syndromes: the Wabasha Heart Attack Team (WHAT) project. Fam Pract 2001 Oct 01;18(5):537-539. [doi: 10.1093/fampra/18.5.537] [Medline: 11604379]
58. Nehme Z, Cameron PA, Akram M, Patsamanis H, Bray JE, Meredith IT, et al. Effect of a mass media campaign on ambulance use for chest pain. Med J Aust 2017 Jan 16;206(1):30-35. [doi: 10.5694/mja16.00341] [Medline: 28076734]
59. Luepker RV, Raczynski JM, Osganian S, Goldberg RJ, Finnegan JRJ, Hedges JR, et al. Effect of a community intervention on patient delay and emergency medical service use in acute coronary heart disease: the Rapid Early Action for Coronary Treatment (REACT) trial. JAMA 2000 Jul 05;284(1):60-67. [doi: 10.1001/jama.284.1.60] [Medline: 10872014]
60. Bashi N, Varnfield M, Karunanithi M. A smartphone app for patients with acute coronary syndrome (MoTER-ACS): user-centered design approach. JMIR Form Res 2020 Dec 18;4(12):e17542 [FREE Full text] [doi: 10.2196/17542] [Medline: 33337339]
61. Paul L, Wyke S, Brewster S, Sattar N, Gill JM, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
62. Grau-Pellicer M, Lalanza JF, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020 Jul;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]
63. Buman MP, Winkler EA, Kurka JM, Hekler EB, Baldwin CM, Owen N, et al. Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, NHANES 2005-2006. Am J Epidemiol 2014 Feb 01;179(3):323-334. [doi: 10.1093/aje/kwt292] [Medline: 24318278]
64. Thorp AA, Healy GN, Owen N, Salmon J, Ball K, Shaw JE, et al. Deleterious associations of sitting time and television viewing time with cardiometabolic risk biomarkers: Australian diabetes, obesity and lifestyle (AusDiab) study 2004-2005. Diabetes Care 2010 Feb;33(2):327-334 [FREE Full text] [doi: 10.2337/dc09-0493] [Medline: 19918003]
65. Healy GN, Matthews CE, Dunstan DW, Winkler EA, Owen N. Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06. Eur Heart J 2011 Mar;32(5):590-597 [FREE Full text] [doi: 10.1093/eurheartj/ehq451] [Medline: 21224291]
66. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care 2008 Apr;31(4):661-666. [doi: 10.2337/dc07-2046] [Medline: 18252901]
67. Patterson K, Keegan R, Davey R, Freene N. Implementing a sedentary behavior change smartphone app in cardiac rehabilitation: a qualitative analysis guided by the theoretical domains framework and capability, opportunity, and motivation-behavior model. J Cardiovasc Nurs 2023 Mar 27. [doi: 10.1097/JCN.0000000000000983] [Medline: 36971651]
68. Dhaliwal SS, Welborn TA. Measurement error and ethnic comparisons of measures of abdominal obesity. Prev Med 2009 Aug;49(2-3):148-152. [doi: 10.1016/j.ypmed.2009.06.023] [Medline: 19589354]
69. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]
70. Nabutovsky I, Ashri S, Nachshon A, Tesler R, Shapiro Y, Wright E, et al. Feasibility, safety, and effectiveness of a mobile application in cardiac rehabilitation. Isr Med Assoc J 2020 Jun;22(6):357-363 [FREE Full text] [Medline: 32558441]
71. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized hypertension management using patient-generated health data integrated with electronic health records (EMPOWER-H): six-month pre-post study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
72. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
73. Gallagher R, Zhang L. Evaluating mobile health technologies: does the traditional randomized controlled trial serve our needs? Eur J Cardiovasc Nurs 2021 Aug 20;20(6):623-626. [doi: 10.1093/eurjcn/zvab053] [Medline: 34269387]
74. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
75. Hartin PJ, Nugent CD, McClean SI, Cleland I, Tschanz JT, Clark CJ, et al. The empowering role of mobile apps in behavior change interventions: the Gray matters randomized controlled trial. JMIR Mhealth Uhealth 2016 Aug 02;4(3):e93 [FREE Full text] [doi: 10.2196/mhealth.4878] [Medline: 27485822]

76. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct 22;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
77. Alvarado MM, Kum HC, Gonzalez Coronado K, Foster MJ, Ortega P, Lawley MA. Barriers to remote health interventions for type 2 diabetes: a systematic review and proposed classification scheme. J Med Internet Res 2017 Feb 13;19(2):e28 [FREE Full text] [doi: 10.2196/jmir.6382] [Medline: 28193598]
78. Santo K, Singleton A, Chow CK, Redfern J. Evaluating reach, acceptability, utility, and engagement with an app-based intervention to improve medication adherence in patients with coronary heart disease in the MedApp-CHD study: a mixed-methods evaluation. Med Sci (Basel) 2019 Jun 04;7(6):68 [FREE Full text] [doi: 10.3390/medsci7060068] [Medline: 31167489]
79. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
80. Cano Martín JA, Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Economic impact assessment from the use of a mobile app for the self-management of heart diseases by patients with heart failure in a Spanish region. J Med Syst 2014 Sep 4;38(9):96. [doi: 10.1007/s10916-014-0096-z] [Medline: 24994514]
81. Gallagher R, Parker H, Zhang L, Kirkness A, Roach K, Belshaw J, et al. Target audience and preferences related to an Australian coronary heart disease specific mobile app: a mixed methods study. Heart Lung Circ 2020 May;29(5):696-702. [doi: 10.1016/j.hlc.2019.05.178] [Medline: 31235365]
82. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. Am J Prev Med 2011 Aug;41(2):216-227 [FREE Full text] [doi: 10.1016/j.amepre.2011.05.005] [Medline: 21767730]

Abbreviations

AR-DRG: Australian Refined Diagnosis Related Group
AQoL: Assessment of Quality of Life
CHD: coronary heart disease
CONSORT: Consolidated Standards of Reporting Trials
cpm: counts per minute
ED: emergency department
ICER: incremental cost-effectiveness ratio
VM: vector magnitude

Edited by L Buis; submitted 28.04.23; peer-reviewed by L Grosman-Rimon, ST Kulnik; comments to author 20.06.23; revised version received 04.07.23; accepted 16.08.23; published 03.10.23.

Please cite as:
Patterson K, Davey R, Keegan R, Niyonsenga T, Mohanty I, Bowen S, Regan E, Lander M, van Berlo S, Freene N
Testing the Effect of a Smartphone App on Hospital Admissions and Sedentary Behavior in Cardiac Rehabilitation Participants: ToDo-CR Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e48229
URL: https://mhealth.jmir.org/2023/1/e48229
doi:10.2196/48229
PMID:37788043

Original Paper

Implementation of Remote Activity Sensing to Support a Rehabilitation Aftercare Program: Observational Mixed Methods Study With Patients and Health Care Professionals

Ziyuan Lu[1], M Med; Tabea Signer[2], MSc; Ramona Sylvester[2], MSc; Roman Gonzenbach[2], MD, PhD; Viktor von Wyl[1,3*], PhD; Christina Haag[1,3*], PhD

[1]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[2]Valens Rehabilitation Centre, Valens, Switzerland

[3]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Viktor von Wyl, PhD
Institute for Implementation Science in Health Care
University of Zurich
Universitätstrasse 84
Zurich, 8006
Switzerland
Phone: 41 446346380
Email: viktor.vonwyl@uzh.ch

Abstract

Background: Physical activity is central to maintaining the quality of life for patients with complex chronic conditions and is thus at the core of neurorehabilitation. However, maintaining activity improvements in daily life is challenging. The novel Stay With It program aims to promote physical activity after neurorehabilitation by cultivating self-monitoring skills and habits.

Objective: We examined the implementation of the Stay With It program at the Valens Rehabilitation Centre in Switzerland using the normalization process theory framework, focusing on 3 research aims. We aimed to examine the challenges and facilitators of program implementation from the perspectives of patients and health care professionals. We aimed to evaluate the potential of activity sensors to support program implementation and patient acceptance. Finally, we aimed to evaluate patients' engagement in physical activity after rehabilitation, patients' self-reported achievement of home activity goals, and factors influencing physical activity.

Methods: Patients were enrolled if they had a disease that was either chronic or at risk for chronicity and participated in the Stay With It program. Patients were assessed at baseline, the end of rehabilitation, and a 3-month follow-up. The health care professionals designated to deliver the program were surveyed before and after program implementation. We used a mixed methods approach combining standardized questionnaires, activity-sensing data (patients only), and free-text questions.

Results: This study included 23 patients and 13 health care professionals. The diverse needs of patients and organizational hurdles were major challenges to program implementation. Patients' intrinsic motivation and health care professionals' commitment to refining the program emerged as key facilitators. Both groups recognized the value of activity sensors in supporting program implementation and sustainability. Although patients appreciated the sensor's ability to monitor, motivate, and quantify activity, health care professionals saw the sensor as a motivational tool but expressed concerns about technical difficulties and potential inaccuracies. Physical activity levels after patients returned home varied considerably, both within and between individuals. The self-reported achievement of activity goals at home also varied, in part because of vague definitions. Common barriers to maintaining activity at home were declining health and fatigue often resulting from heat and pain. At the 3-month follow-up, 35% (8/23) of the patients withdrew from the study, with most citing deteriorating physical health as the reason and that monitoring and discussing their low activity would negatively affect their mental health.

Conclusions: Integrating aftercare programs like Stay With It into routine care is vital for maintaining physical activity postrehabilitation. Although activity trackers show promise in promoting motivation through monitoring, they may lead to frustration during health declines. Their acceptability may also be influenced by an individual's health status, habits, and technical skills. Our study highlights the importance of considering health care professionals' perspectives when integrating new interventions into routine care.

(JMIR Mhealth Uhealth 2023;11:e50729) doi:10.2196/50729

KEYWORDS

physical activity; activity sensor; normalization process theory; rehabilitation; chronic disease; chronic; aftercare; sensor; sensors; exercise; neurology; neuroscience; neurorehabilitation; adherence; need; needs; experience; experiences; questionnaire; questionnaires; mobile phone

Introduction

Background

Patients attending inpatient neurorehabilitation often face complex chronic health conditions, such as multiple sclerosis, or cardiovascular diseases, such as stroke [1,2]. These chronic health conditions are typically linked to increasing physical impairments, which negatively impact individuals' quality of life [3,4]. They require complex treatment and care, often involving multiple, interdisciplinary health care providers and a blend of pharmacological treatments, physical therapy, and lifestyle management [5-7].

Over the recent years, the promotion of physical activity has become state of the art in inpatient rehabilitation treatment for chronic diseases or diseases with risk for chronicity [8-14]. However, any improvements in daily activity levels and physical fitness that may be achieved in a rehabilitation setting are typically challenging to maintain in daily life [15-17]. Once back in their daily-life environment, persons with chronic diseases face multiple barriers to physical activity, such as fatigue; a less structured environment; or time restrictions associated with care responsibilities, employment, and other factors [18-23]. The Valens Rehabilitation Centre has recently developed a novel inpatient routine aftercare program, designed to prepare patients for this transition, named the Stay With It program (Swiss German: Bliib dra program). Complementing inpatient rehabilitation treatment, it combines motivational interviewing techniques with detailed action plans for promoting physical activity following rehabilitation.

The emergence of novel activity sensors designed for daily-life physical activity tracking has created novel avenues for promoting self-monitoring and self-management skills, and these sensors are well suited to support the integration of modules into patients' daily lives [19,20,24-26]. Many consumer-grade activity sensors assess a broad range of activity- and sleep-related parameters in real time, including step count, activity levels, heart rate (variability), and sleep stages. A previous umbrella meta-analysis found good evidence for the beneficial health effects of activity sensor–supported lifestyle management changes, particularly for physical activity outcomes or weight loss [27]. Conventional aftercare programs for rehabilitation, such as the Stay With It program, may, therefore, benefit from the additional systematic integration of activity sensors into their educational and aftercare monitoring procedures. Indeed, a previous study conducted in the same setting in Valens with people with multiple sclerosis suggested that many patients perceived activity sensors as devices that help maintain motivation for physical activity after discharge at home [28]. Although these novel technical innovations have potential in complementing already established routine care programs, their implementation faces multiple challenges. Potential challenges include the time required for the ongoing resolution of device-related technical issues and the commitment required to guide patients in understanding their data [26,28].

Planning the integration of a digitally supported aftercare program, such as Stay With It, into clinical routine care is a complex process, requiring the involvement of a broad range of stakeholders and the adaptation of their standard workflows. Many aspects of the planning, implementation, and evaluation of such digitally augmented programs would clearly benefit from guidance from different implementation frameworks [29]. In the context of digitally supported interventions, the normalization process theory (NPT) is a well-established framework for guiding the implementation of novel interventions in routine settings [30,31]. The NPT consists of 4 core constructs, namely "coherence," "cognitive participation," "collective action," and" "reflexive monitoring," which represent distinct steps needed in the implementation process, referred to as "normalization" [32]. The definitions of the NPT constructs and their application and relevance to both patients and health care professionals in our study are presented in Table 1.

However, there currently exists no specific guidance on the implementation of digital tools into the workflows of an established, conventional motivational program. For example, how receptive will health care providers be about the integration of activity sensors? What needs for educational support or administrative changes may arise from the add-on implementation, both for patients and health care providers?

Table 1. The 4 constructs of the normalization process theory (NPT) and how they apply to this study.

Description of NPT construct	Patients	Health care professionals
"Coherence" refers to the sense-making process at the individual and group levels [32,33].	With respect to patients, "coherence" refers to their understanding of the purpose of the program and how it, possibly in combination with an activity sensor, could support them in staying active at home.	"Coherence," as it relates to the health care professionals in this study, refers to their understanding of the purpose of the program and how an activity sensor can support implementation.
"Cognitive participation" refers to the relational work of forming and maintaining a practice community around a complex intervention or novel technology [32,33].	With respect to patients, "cognitive participation" refers to adapting daily routines to incorporate the program and related activities as well as understanding how the activity sensor can be used in and best integrated into daily life, which may involve resolving technical issues by acquiring new technical skills.	With respect to the health care professionals in this study, "cognitive participation" refers to ensuring the necessary knowledge of the program and associated procedures and collaborating as a team to facilitate implementation. Throughout program implementation, this is an ongoing process.
"Collective action" refers to the operational work required to implement new practices related to new technologies or complex interventions [32,33].	With respect to patients, "collective action" means creating an environment and circumstances that are conducive for them to benefit from the program. This may include, for example, shifting priorities or changing personal routines.	With respect to health care professionals, "collective action" means taking care of organizational tasks to facilitate program implementation and seamlessly integrate the new program and its activity sensor into their daily tasks, ensuring efficiency while minimizing their time commitment to an appropriate level.
"Reflexive monitoring" pertains to the evaluative work of assessing and understanding how a practice affects individuals themselves and their surroundings [32,33].	Reflexive monitoring, as it relates to the patients in this study, concerns how they evaluate the way in which the novel program and, if applicable, its combination with an activity sensor affect them (eg, health and daily life) and their environment (eg, family and responsibilities).	In this study, health care professionals who engaged in reflective monitoring evaluated the impact and effectiveness of the new program, especially when paired with an activity sensor, on themselves, their teams, patients, and the broader organization.

Research Aims

Overview

Building on the NPT framework, this study investigated the key implementation challenges and facilitators for the novel Stay With It aftercare program from both patient and health care professional perspectives. We also examined whether an activity sensor is appropriate to support program implementation. We had 3 overarching research aims.

Research Aim 1: Challenges and Facilitators of Program Implementation at the Valens Rehabilitation Centre

First, we explored the challenges and facilitators that both patients and health care professionals encountered in implementing the Stay With It program at the Valens Rehabilitation Centre. We also assessed whether there were any unmet needs related to program implementation. This research aim was rooted in all 4 NPT constructs.

Research Aim 2: Perceived Usefulness of an Activity Sensor to Support Program Implementation and Sustainability

Second, we examined whether both patients and health care professionals found the activity sensor useful for program implementation. We also examined patients' use of the sensor. This research aim was rooted in all 4 NPT constructs.

Research Aim 3: Maintenance of Physical Activity (Patient Adherence at Home)

Finally, we aimed to assess whether patients remained active during follow-up, patients' self-reported achievement of home activity goals, and factors that influence daily physical activity at home. This research aim was specifically rooted in the NPT construct "reflexive monitoring."

Methods

Ethical Considerations

The study was approved by the Cantonal Ethics Committee (Business Administration System for Ethics Committees [BASEC] number: 2021-02490) and preregistered on ClinicalTrials.gov (trial registration NCT05243407). All participants provided written informed consent.

Study Design

We invited both patients participating in the Stay with It program at the Valens Rehabilitation Centre and health care professionals delivering the program to participate in this study. Data were collected between February 2022 and August 2022.

With respect to the patients, the study design was an observational longitudinal cohort study. Patients were enrolled when they started with the program and followed up for 3 months after being discharged. Study participation included the optional wearing of an activity sensor throughout the study and completion of qualitative and quantitative measures. Health care professionals were assessed in February and March 2022 before program implementation and again in August 2022 after they gained experience with the program.

Study Participants

Patients

Adult inpatients (n=23) at the Valens Rehabilitation Centre who were undergoing rehabilitation and participating in the Stay With It program were eligible for the study. Those who were interested were provided with a study information sheet. The Stay With It program was designed for adults with either physical chronic health conditions or conditions at risk of becoming chronic, such as multiple sclerosis or cardiovascular disease. The program's goal was to incorporate more physical activity into participants' daily lives. Although patients were encouraged to wear an activity sensor throughout the study, this was not mandatory. We aimed to enroll approximately 20 patients in the study because this number was both feasible and would provide a variety of individual experiences.

Health Care Professionals

The participating health care professionals were occupational or physical therapists (n=13) at the Valens Rehabilitation Centre who were trained and scheduled to deliver the program.

Stay With It Program

The Stay With It program, developed by the Valens Rehabilitation Centre, was designed to help patients integrate individualized activity and action plans into their daily routines once they return home. The program is divided into 4 sequential modules: "Introduction," "Physical Strength," "Endurance," and "Application Into Daily Life." A detailed description of the 4 modules is provided in Multimedia Appendix 1 [34,35].

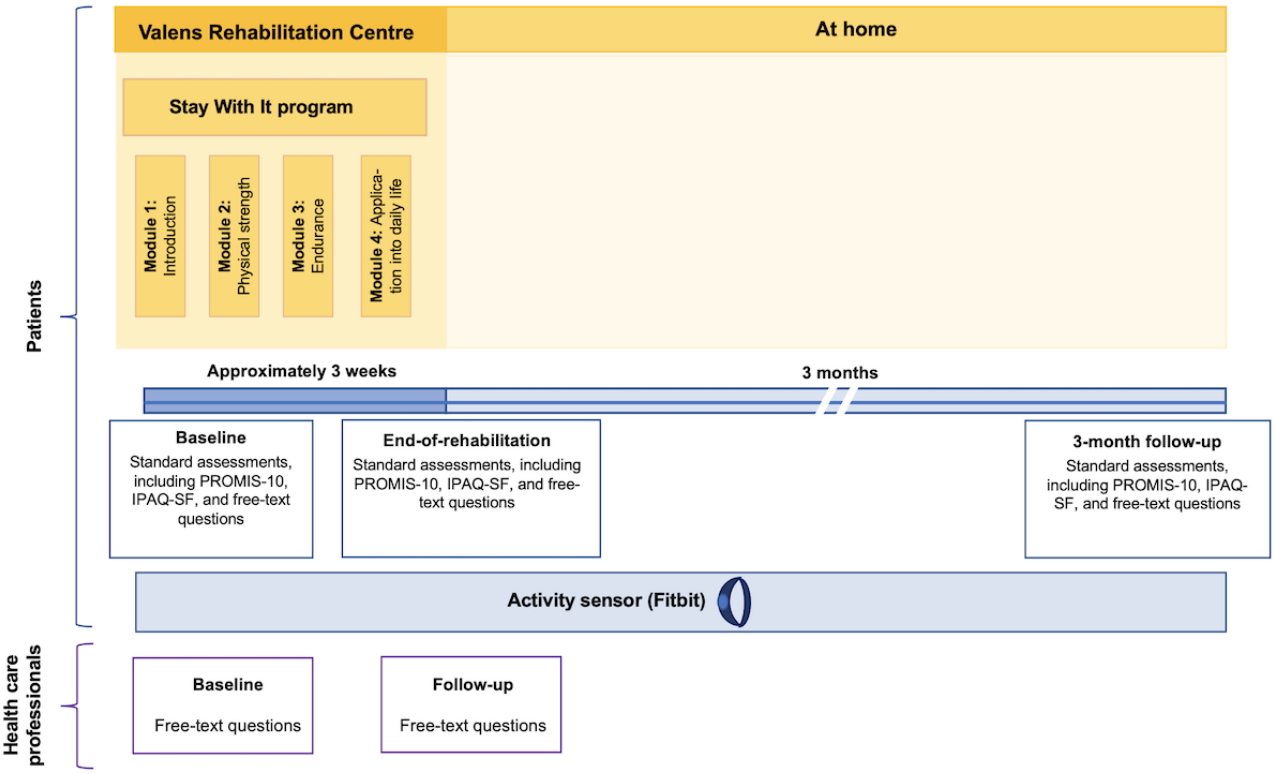
Study Procedure

Patients

For patients, the study procedure consisted of five consecutive phases (Figure 1): (1) baseline session, (2) Stay With It program participation at the Valens Rehabilitation Centre, (3) end-of-rehabilitation assessment, (4) monitoring period at home, and (5) 3-month follow-up assessment. Daily-life activity data were continuously monitored throughout the study. If patients did not provide data for 5 days, the study staff checked to see whether they were having technical problems with their device, such as syncing issues. The 5 study phases are detailed in Textbox 1.

Daily-life activity data were assessed continuously from baseline until 3-month follow-up. Study personnel regularly checked the completeness of activity sensor data and reached out to participants if they did not provide any data for 5 days to check whether they were experiencing technical problems with their device (eg, device not synchronizing) and to offer technical support.

Figure 1. Visualization of the study procedure for patients and health care professionals. Patients participated in the Stay With It program during their rehabilitation stay at the Valens Rehabilitation Centre and were encouraged, but not required, to wear an activity sensor throughout the study. Their experience with the program, quality of life, and physical activity were assessed at baseline, the end of rehabilitation, and at a 3-month follow-up. Health care professionals were asked about their experience before implementing the program and again after having acquired experience with it. PROMIS-10: Patient-Reported Outcomes Measurement Information System–Global 10; IPAQ-SF: International Physical Activity Questionnaire–Short Form.



Textbox 1. Study phases for patients.

1. Baseline session: after patients provided written informed consent, they were provided with a Fitbit Charge 4 (Fitbit LCC) activity sensor to be worn continuously on their nondominant wrist throughout the study. The study staff helped them set up the Fitbit app on their personal smartphone and connect their activity sensor to the app using a nonpersonal Fitbit study account. Patients were also given a brief introduction on how to use the app and what parameters were measured using the activity sensor. Patients then answered free-text questions about their motivation and expectations for the Stay With It program, as well as any prior experience with activity sensors. They also completed self-report questionnaires on quality of life and their daily physical activity over the previous week.
2. Stay With It program participation at the Valens Rehabilitation Centre: following the baseline session, patients took part in the Stay With It routine care program. Those with an activity sensor were encouraged to wear it continuously throughout their daily lives for monitoring to complement program participation.
3. End of rehabilitation: at the end of their rehabilitation stay, patients were phoned by the study staff and asked about their experience with the program and activity sensor and what activity goals they had set for their time at home. They were also readministered the self-report questionnaire on daily physical activity for the previous week.
4. Monitoring period at home: patients who had received an activity sensor were encouraged to continue wearing it during the 3-month monitoring period at home. The study staff routinely checked the completeness of these data.
5. Three-month follow-up: at the 3-month follow-up, the study staff contacted patients via phone. Patients were asked about their experiences and whether they felt that they had achieved their goals at home. They were also readministered questionnaires on quality of life and daily activity during the previous week.

Health Care Professionals

Health care professionals completed free-text questions before the program was implemented in February and March 2022 and again after they acquired experience with program implementation in August 2022.

Measures

The measures of this study are described in detail in the subsequent sections.

Free-Text Questions

Both patients and health care professionals responded to free-text questions focusing on the challenges and facilitators of program implementation as well as the potential benefits of supporting program implementation using an activity sensor. These questions were developed in accordance with the 4 NPT constructs. Table 1 presents the 4 NPT constructs, along with brief definitions and descriptions of how they relate to the patients and health care professionals in this study.

From these core definitions, we developed free-text questions applying the 4 NPT constructs to this study. These questions, mapped to the 3 research objectives and the NPT constructs, are presented in Multimedia Appendices 2 and 3. As most patients experienced difficulties with fine motor skills, the study personnel (ZL and CH) assisted them in documenting their responses. Patients first dictated their responses, which the staff transcribed. The responses were then read back for verification and corrected as necessary.

Patient-Only Assessments

The following data and measures were assessed only in patients.

Daily-Life Activity–Sensing Data

For patients wearing an activity sensor, we assessed daily summed step counts and active and inactive minutes (highly active, fairly active, moderately active, and sedentary time).

International Physical Activity Questionnaire–Short Form

The International Physical Activity Questionnaire–Short Form is a 7-question retrospective self-report questionnaire with good validity for assessing daily physical activity across various intensities (high and moderate activity and walking) and time spent sitting for at least 10 minutes, averaged over the past 7 days [36].

Patient-Reported Outcomes Measurement Information System–Global 10

The Patient-Reported Outcomes Measurement Information System–Global 10 is a 10-item self-report instrument that assesses physical, mental, and social health as well as pain, fatigue, and quality of life [37]. This instrument was validated in multiple studies, and scores range from 0 (severe impairment) to 20 (optimal health) [37,38].

Analytical Approach

Group comparisons for descriptive statistics were made using independent, 2-sided sample *t* tests (=.05) and were performed in R (version 4.3.1; R Core Team) using RStudio (version 2023.06.1; RStudio Inc). A detailed breakdown of how each of the free-text questions relates to the 3 research aims can be found in Multimedia Appendices 2 and 3.

Research Aim 1: Challenges and Facilitators of Program Implementation at the Valens Rehabilitation Centre

To identify the challenges and facilitators of program implementation, we used a thematic approach to analyze free-text responses from patients and health care professionals. The results are presented in a summary form, along with theme prevalence and exemplary, anonymized sample responses.

Research Aim 2: Perceived Usefulness of an Activity Sensor to Support Program Implementation and Sustainability

Consistent with the analytical approach used for research aim 1, we examined the free-text responses of patients and health care professionals using a thematic approach. Again, the results are presented in a summary form, along with theme prevalence and with exemplary, anonymized sample responses. We also visualized patients' daily activity–sensing data using individual-level plots. Visualizations were created using the R package *ggplot 2*. We also determined descriptive statistics of daily-life activity as assessed by the activity sensor and the International Physical Activity Questionnaire–Short Form self-report questionnaire.

Research Aim 3: Maintenance of Physical Activity (Patient Adherence at Home)

The analytical approach for each of the 3 subcomponents of research aim 3 is outlined in the following sections.

Maintenance of Physical Activity Through Follow-Up

Given our limited sample size, we restricted our analysis to a visual examination of daily-life physical activity for overall trends, focusing on both intraindividual and interindividual variabilities.

Self-Reported Achievement of Home Activity Goals

We manually assessed the specificity of patients' activity goal definitions using the SMART (specific, measurable, achievable, realistic, time-bound) criteria [34] and categorized them into 3 levels: high, moderate, and unspecific. Activity goals that met all the SMART criteria were considered highly specific. Activity goals that met between 1 and 4 of the SMART criteria were categorized as moderately specific. Goals that did not meet any of these criteria were considered unspecific. At 3-month follow-up, patients self-reported the degree to which they had achieved their activity goals. We classified goal attainment as "fully met," "partially met," or "not met" based on their self-reports.

Factors Influencing Daily Physical Activity at Home

We again used a thematic approach to examine free-text responses from patients at 3-month follow-up about what hindered and what facilitated physical activity in their daily lives at home.

Results

Sample Characterization

Patients

Patients' age averaged 56.26 (SD 8.58; range 43-69) years, and 61% (14/23) were female. A characterization of patients with respect to their health conditions is provided in Multimedia Appendix 4. Of the 23 patients, 8 (35%) withdrew from the study early, of whom 6 (75%) were wearing an activity sensor. Of these 6 patients, 5 (83%) dropped out due to deterioration in their physical health, which made study participation frustrating (refer to Figure 2 for the study flow). The remaining patients who dropped out (3/8, 38%; with an activity sensor: 1/8, 13%; without an activity sensor: 2/8, 25%) were unreachable by phone at 3-month follow-up and thus did not provide any reasons for withdrawal. Descriptive statistics for study completers and noncompleters are provided in Table 2.

Figure 2. Study flowchart.

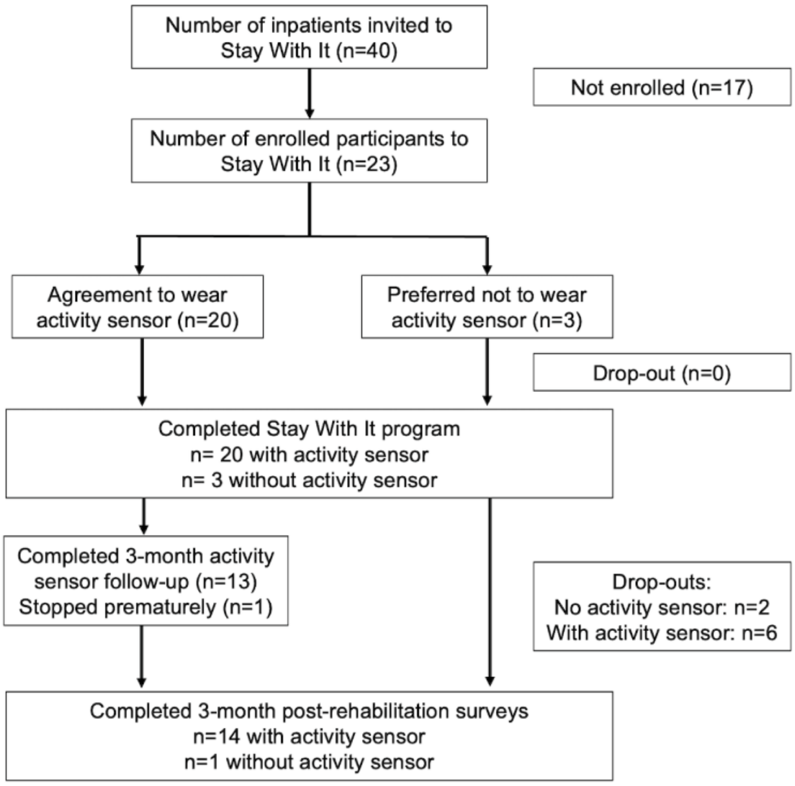


Table 2. Descriptive statistics at baseline (completers and noncompleters) and 3-month follow-up (completers).

Measure	Baseline			3-month follow-up
	Completer (15/23, 65%; with an activity sensor: 14/15, 93%; without an activity sensor: 1/15, 7%)	Noncompleter (8/23, 35%; with an activity sensor: 6/8, 75%; without an activity sensor: 2/8, 25%)	Mean difference (*P* value)	Completer (15/23, 65%; with an activity sensor: 14/15, 93%; without an activity sensor: 1/15, 7%)
PROMIS-10[a], mean (SD)				
Physical health sum score	12.2 (2.54)	12 (3.16)	.89	13.92 (3.35)
Mental health sum score	12.93 (4.03)	11.83 (4.36)	.61	10.87 (1.46)
Activity sensor, mean (SD)				
Daily steps, mean (SD)	N/A[b]	N/A	N/A	At Valens: 8072.59 (3167.73); at home: 6853.71 (4542.79)
IPAQ-SF[c](average minutes of activity intensity per day over the past 7 days), mean (SD)				
High intensity	92.14 (80.74)	46.61 (48.4)	.28	22.19 (23.33)
Moderate intensity	102.6 (90.63)	70.98 (63.17)	.29	42.34 (40.34)
Walking	51.62 (48.2)	29.2 (24.88)	.32	33.73 (48.16)
Sitting	338 (254.42)	363.75 (237)	.81	303.21 (171.95)
Attrition over the course of the study for patients who wore an activity sensor and patients who did not wear a sensor, n (%)				
Baseline: patients wearing an activity sensor (n=20)	N/A	N/A	N/A	Health deterioration: 5 (25); not available by phone: 1 (5)
Baseline: patients not wearing an activity sensor (n=3)	N/A	N/A	N/A	Not available by phone: 2 (67)

[a]PROMIS-10: Patient-Reported Outcomes Measurement Information System–Global 10.

[b]N/A: not applicable.

[c]IPAQ-SF: International Physical Activity Questionnaire—Short Form.

Health Care Professionals

We assessed a total of 13 health care professionals, all of whom were occupational or physical therapists at the Valens Rehabilitation Centre and formed the team trained to deliver the program. Of the 13 health care professionals, 3 (23%) delivered all the program modules, whereas 10 (77%) delivered only a subset of the modules for organizational reasons.

Research Aim 1: Challenges and Facilitators of Program Implementation at the Valens Rehabilitation Centre

Findings on challenges and facilitators are presented separately from the perspectives of patients and health care professionals, and the presentation is organized according to the 4 NPT constructs. The text detailing our findings is complemented by sample quotes.

Patient Perspective

Coherence

Before program participation, individuals reported a variety of challenges in attempting to maintain physical activity, which also sparked their interest in the program. On a personal level, many struggled with a lack of intrinsic motivation at home and physical limitations owing to pain, sensitivity to weather changes, and health issues (eg, falls). Outside the home, daily responsibilities; irregular working hours, especially for those who work from home; and the lack of therapeutic guidance and encouragement add to the burden ("Practicality at home is a challenge. In rehabilitation, there's a set routine that's often missing at home"). A person mentioned struggling with moving while others are watching, as this evokes feelings of shame.

Cognitive Participation and Collective Action

While at the Valens Rehabilitation Centre, patients had their rehabilitation training sessions integrated into their daily schedules to avoid conflicts with other appointments. The group program sessions (modules 2 through 4) were scheduled on Saturday mornings. On a few occasions, this conflicted with patients' desire to visit their families. During their stay, primary care providers were informed of patients' participation in the program to facilitate discussions about their program experience.

Reflexive Monitoring

Although the group setting was generally appreciated, of the 23 patients, 5 (22%) expressed a desire for a more homogeneous group composition, particularly in terms of both mobility and disease, to have more fruitful exchanges and avoid frustration. Moreover, among the 23 patients, 2 (9%) expressed a preference for more visual aids and fewer textual descriptions.

Health Care Professional Perspective

Coherence

Health care professionals had a good understanding of the program's purpose before its implementation, as evidenced by their extensive knowledge of the barriers to physical activity when patients return home. Specifically, health care professionals identified the following reasons why patients have difficulty maintaining physical activity at home: lack of structure and support (6/13, 46%), difficulty maintaining motivation without constant motivational support from the health care team (6/13, 46%), responsibilities at home regaining priority (3/13, 23%), slipping back into old habits (2/13, 15%), lack of knowledge about how to apply the basic principles of physical activity (1/13, 8%), and difficulty creating an exercise plan at home that is both realistic and effective (1/13, 8%).

Cognitive Participation

Health care professionals identified several elements of the program that they expected to work well. Specifically, they (7/13, 54%) expected patients to share strategies for staying active with their peers in group discussions ("strategy sharing"). Some (2/13, 15%) also expected the program to provide patients with concrete guidance on how to incorporate and monitor exercise in their daily lives when they return home ("personalized plan"). Of the 13 health care professionals, 2 (15%) anticipated that the program would make patients aware of their current activity levels, which might lead them to find new ways to incorporate physical activity into their routines, such as taking the stairs more often ("awareness"), and 2 (15%) anticipated that the basic knowledge of the exercise principles documented in the program materials would help patients exercise more effectively at home ("expertise").

Health care professionals also expressed some concerns about what might not work smoothly and need specific attention. The most common concern was that a long-term commitment without a structured environment or regular contact with their therapist may lead to a tendency to fall back into old patterns ("old patterns"; 4/13, 31%). Another concern was that for patients who are severely impaired, their limitations may make it difficult to find independent activities of daily living, and this might also be difficult to cope with emotionally. Patients who are less severely impaired should also not be overburdened to avoid demotivation ("demotivation"; 3/13, 23%). Finally, among the 13 health care professionals, 1 (8%) noted that it would be more realistic to plan activities of daily living that do not require extensive equipment or preparation to minimize time commitment ("feasibility").

Collective Action

Health care professionals managed the logistical aspects of implementing the program, such as room reservations, weekly scheduling, and staffing for group sessions. The organization also integrated the program into Valens' digital scheduling system to make it fit seamlessly into patients' daily schedules. In addition, the process was integrated with the internal accounting system in consultation with the IT department.

Reflexive Monitoring

After having gained experience with the program, all (13/13, 100%) health care professionals were optimistic that a considerable number of patients would be able to incorporate and maintain more physical activity in their daily lives in the long term. Some (5/13, 38%) believed that the main factors determining whether patients would stay active were their internal motivation and the concreteness of their plan for returning home, such as outpatient exercise sessions. Among the 13 health care professionals, 1 (8%) also expected that, regardless of program participation, patients who were active before rehabilitation would be more likely to maintain activity at home than those with low activity levels before rehabilitation. A total of 2 (15%) health care professionals suggested that only patients who have the intention to be more active at home should be invited to participate in the program, as large differences in motivation negatively affect the group dynamic.

Research Aim 2: Perceived Usefulness of an Activity Sensor to Support Program Implementation and Sustainability

Findings on the perceived potential of an activity sensor to support program implementation and sustainability are presented separately from the perspectives of patients and health care professionals, and the presentation is organized according to the 4 NPT constructs. The text detailing our findings is complemented by sample quotes.

Patient Perspective

Coherence

In terms of sense making, patients (20/23, 87%) who chose to use an activity sensor generally understood its potential to help them integrate program learning into their daily routines. Most (21/23, 91%) patients already used a smartphone and apps on a daily basis, and more than half (14/23, 61%) of them had previous experience with activity sensors.

Cognitive Participation

Of the 20 patients who chose to use an activity sensor, most had mostly positive expectations about using the activity sensor (n=17, 85%) and expected that it would give them control over their daily activity and remind them when they were not moving much (n=9, 45%). Among the 20 patients, 6 (30%) expected the activity sensor to motivate them to stay active, and 4 (20%) appreciated that the device could quantify movement ("The activity sensor helps me stay motivated"). Of the 23 patients recruited, 3 (13%) were open to trying the activity sensor, although they were unsure whether it would be useful for them. Some patients had concerns about using the device. Among the 23 patients, 2 (9%) expressed concerns about the technical operation of the device; 1 (4%) cited privacy issues; 1 (4%) felt that tracking their daily activity could reduce their enjoyment of the exercise; and 1 (4%) noted that devices such as watches often cause skin irritation, making direct wear on the body undesirable.

Collective Action

Patients who chose an activity sensor and remained in the study (14/23, 61%) tended to wear the activity sensor continuously, as displayed in Figures 3 and 4.

Figure 3. Individual-level time-series plots of daily step counts over the course of study participation.

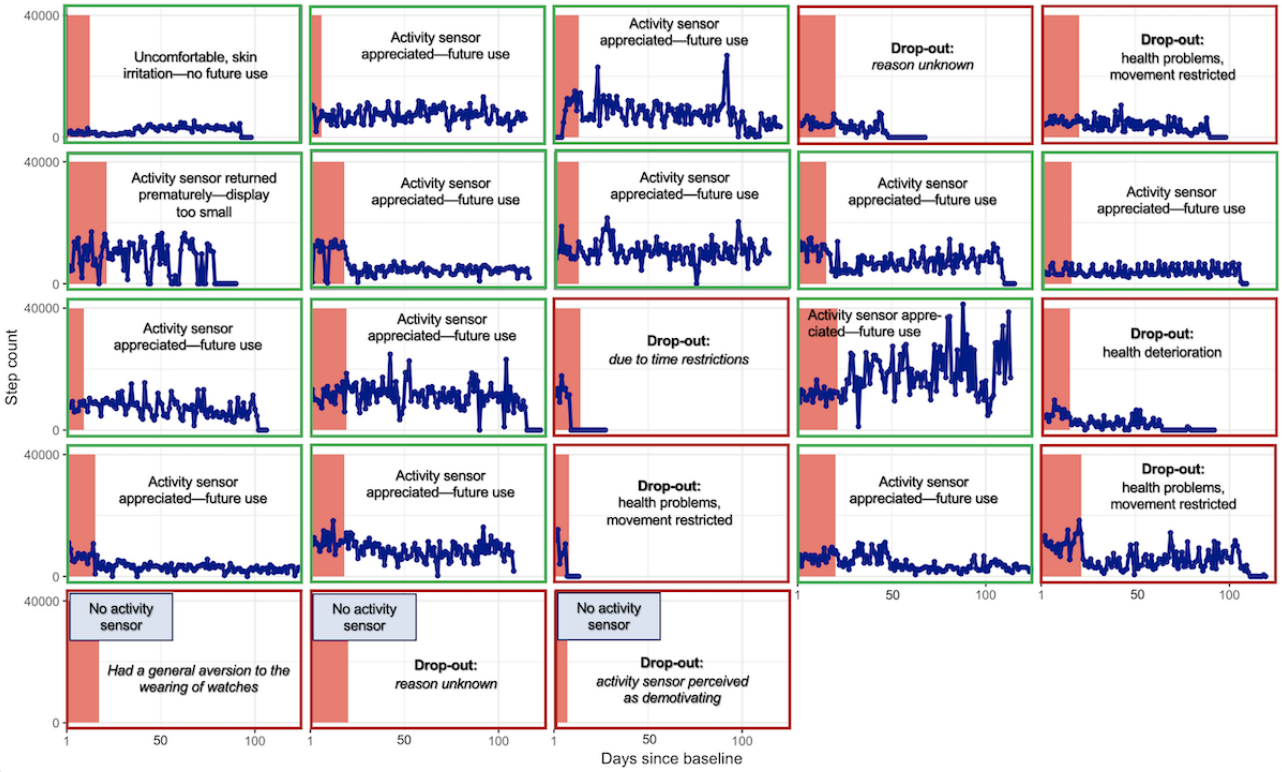
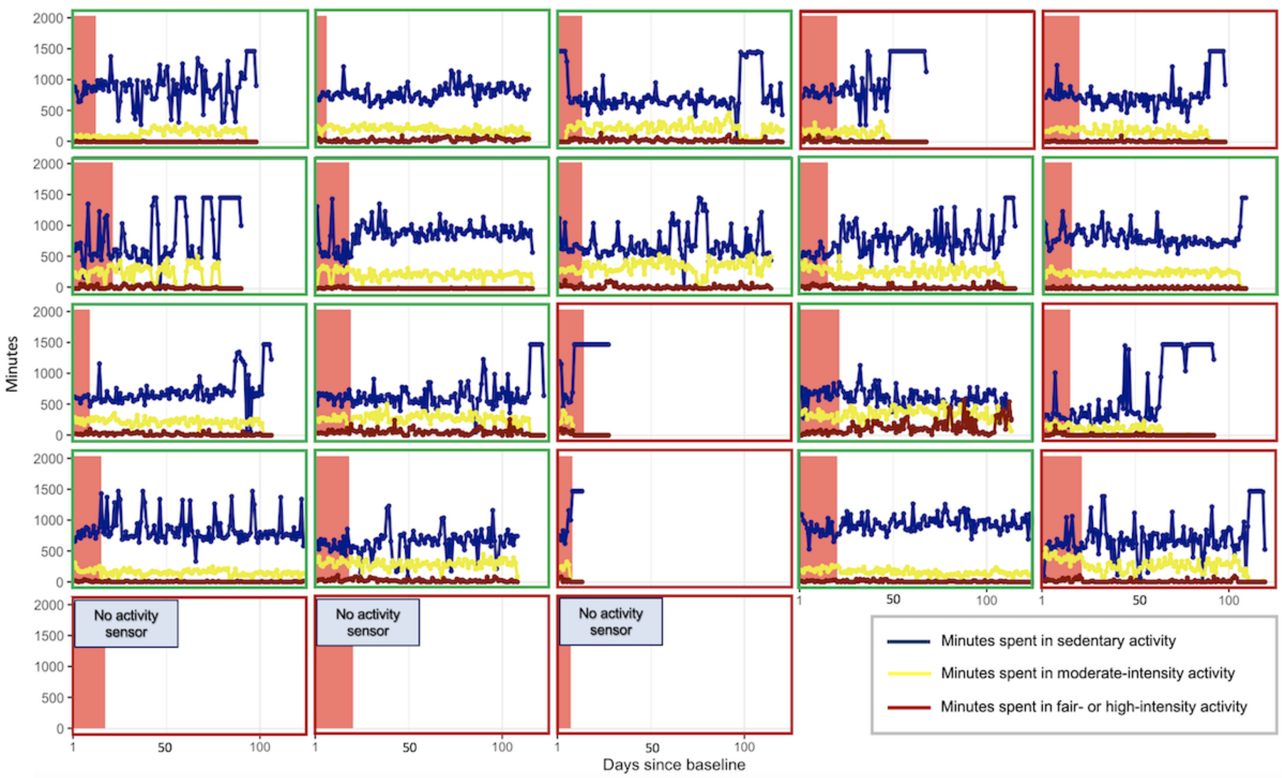


Figure 4. Individual-level time-series plots of raw minutes spent on different levels of daily activity.



Reflexive Monitoring

Of the 15 patients who were available for the 3-month follow-up, most (n=11, 73%) found the activity sensor to be an aid in staying physically active after returning home. A total of 10 (67%) patients planned to also wear an activity sensor in the future. Interestingly, patients who opted for an activity sensor and later dropped out of the study showed reduced sensor wear compliance just before dropout (refer to Figure 3). Baseline comparison between completers and noncompleters revealed no significant differences in self-reported average activity levels or quality of life (Table 2).

Health Care Professional Perspective

Coherence

Before the launch of the program, all (13/13, 100%) health care professionals were optimistic about using activity sensors to support the program. Of the 13 health care professionals, 3 (23%) viewed the sensor primarily as a motivational tool, whereas 2 (15%) believed that its main benefit would be to quantify daily exercise ("An activity sensor simplifies quantifying movement and activity"). Some concerns were raised, such as that some patients may not be technically adept enough (4/13, 31%) and the possibility of demotivation if the sensor showed unmet goals ("Failure to achieve goals can lead to frustration"). In addition, 2 (15%) of the 13 health care professionals were concerned that patients with fine motor difficulties might have difficulty wearing or removing the sensor.

Cognitive Participation and Collective Action

Before the start of the study, all (13/13, 100%) health care professionals received a sample of the activity sensor used in the study and performed a test run with it. This ensured that everyone was familiar with the specific type of activity sensor.

Reflexive Monitoring

During program implementation, health care professionals identified motivation and self-monitoring (both 7/13, 54%) as the most valuable aspects for patients. However, they also identified challenges. Among the 13 health care professionals, 6 (46%) expressed that some patients felt frustrated when comparing themselves with others or when not meeting their (at times unrealistic) goals. Moreover, 3 (23%) health care professionals reported that the sensor's measurements were occasionally inaccurate with certain physical impairments or when patients used assistive devices such as walkers or canes. A total of 3 (23%) health care professionals also observed that some patients had difficulties with the technical aspects of the sensor.

Research Aim 3: Maintenance of Physical Activity (Patient Adherence at Home)

Maintenance of Physical Activity Through Follow-Up (Collective Action)

The levels of physical activity over the 3-month follow-up period varied considerably, both within and between patients, as displayed in Figures 3 and 4.

Self-Reported Achievement of Home Activity Goals (Collective Action)

As part of the final Stay With It session, patients were guided to define 3 activity goals for their time back at home that they wished to pursue. In the event that the patients missed this session and thus did not set goals as part of the program, they were asked whether they had any self-defined goals for their time back home, and all of them did. One participant could not be reached by phone for the end-of-rehabilitation assessment, resulting in activity goals for 22 patients. The patients in our study defined, on average, 2.60 (SD 0.68; range 1-3) goals for their time back home, resulting in a total sample of 51 activity goals. On the basis of the SMART criteria, the majority of goals were of moderate specificity (20/51, 39%), whereas 31% (16/51) were highly specific, and 29% (15/51) were unspecific. At 3-month follow-up, patients self-reported their achievement of these goals, the results of which are detailed in Table 3.

Highly and moderately specific goals were reported as fully achieved in 53% (19/36) of the goals, whereas only 20% (3/15) of the nonspecific activity goals were reported as fully achieved. For 47% (7/15) of the nonspecific goals, goal achievement could not be assessed because the respective participants had dropped out before the 3-month follow-up. For highly or moderately specific goals, this was the case for 19% (7/36) of the activity goals.

Table 3. Activity goals of varying precision of the patients who completed the 3-month follow-up assessment[a].

Specificity of activity goals (N=51 goals)	Anonymized examples for activity goals[b]	Self-reported goal achievement at 3-month follow-up, n (%)
Highly specific goals (n=16, 31%)	• "10,000 steps per day" • "At least 6 hours sleep per night" • "Going outside for a walk twice a week. Instead, I do less of the housework, or I do it with less perfection" • "Climbing 6 to 10 floors a week" • "Working out 3 times a week for about 30 minutes each session"	• Fully: 8 (50) • Partially: 3 (19) • Not achieved: 1 (6) • Missing information because of participant dropouts: 4 (25)
Moderately specific goals (n=20, 39%)	• "I would like to work out less but more often (ie, 2 to 3 times a week). I plan to use resistance bands to build muscle strength." • "I want to do more stretching. I also plan to go to physical therapy once a week for instructed stretching." • "Cycling/walking as far as possible" • "Going to the gym again, that I exercise more, doing something every day—be it mowing the lawn, cleaning the windows or something else. It doesn't really matter."	• Fully: 11 (55) • Partially: 2 (10) • Not achieved: 4 (20) • Missing information because of participant dropouts: 3 (15)
Unspecific goals (n=15, 29%)	• "I want to adapt my fitness program in a way that is not overwhelming, but moderate." • "My goal is to become more fit." • "I'm aiming to incorporate some form of exercising into my daily life." • "I want to keep on going, to keep on overcoming, to keep on participating."	• Fully: 3 (20) • Partially: 2 (13) • Not achieved: 3 (20) • Missing information because of participant dropouts; 7 (47)

[a]Because of rounding, percentages may not add to 100%.

[b]Translated and edited for privacy.

Factors Influencing Daily Physical Activity at Home (Reflexive Monitoring)

In terms of reasons for not achieving the activity goals, the mentioned barriers to physical activity included the heat in summer or being sensitive to the weather in general (5/15, 33%), as this leads to reduced energy levels and fatigue. Pain and deterioration in physical health (3/15, 20%), for example, due to an illness or accident, were also common barriers. Distraction from potentially time-consuming indoor activities, such as the use of the internet, was also cited as a barrier to physical activity (2/15, 13%). The activity sensor was mentioned because it was frustrating for patients to see how little they were moving when they experienced barriers or because steps were not counted when using a rollator. In terms of facilitators for staying active, 73% (11/15) of the patients found the activity sensor helpful, 46% (7/15) mentioned regular exercise routines (eg, physical therapy) and group dance or exercise classes as helpful, 33% (5/15) had a partner who participated in their physical activity, and 13% (2/15) benefited from having a dog to help them stay active.

Discussion

Principal Findings

Using the NPT framework, this study examined the challenges and facilitators of implementing an aftercare program into routine care and whether activity sensors could support program implementation. Over a 3-month follow-up period, we examined several aspects of patients' physical activity. We found that the consideration of both patient and health care professional perspectives is critical to the successful implementation of complex interventions. In addition, there was considerable variability in patients' activity levels and goal attainment at home, influenced by multiple limiting factors.

Our findings for research aim 1 add to previous implementation science research emphasizing the need to consider both patient and health care professional perspectives when integrating complex interventions into routine care (eg, the study by Naef et al [39]). Patients in our study faced various barriers to daily physical activity but were committed to increasing activity after rehabilitation. Health care professionals saw the potential of the program and believed that success depended on individual drive and a robust postrehabilitation plan. The main challenges in implementing the program were organizational, such as integrating the program into the scheduling and accounting system, which was essential for its seamless integration into routine care. In addition, the importance of appropriate timing was highlighted by the occasional conflict between patients' personal commitments and weekend group sessions. Our findings resonate well with a 2021 review that examined exercise adherence in patients with chronic conditions and older adults [7]. The review highlighted 14 key factors, with the most relevant factors to our research being the use of technology, the initial assessment of participant characteristics, challenges and facilitators, participant education, clear expectations, and goal setting.

As for the perceived usefulness of an activity sensor in supporting the implementation and sustainability of the Stay With It program (research aim 2), most patients who used the activity sensor appreciated monitoring and "objectifying" their daily activity. Consistent with this finding, patients who opted to wear an activity tracker at baseline had high compliance rates for daily wear time. This conclusion aligns with a 2020 systematic review and meta-analysis suggesting that activity sensors may be a useful tool in promoting active lifestyles in patients [40]. However, our results also suggest that the perceived usefulness of an activity sensor may vary depending on individual situations, even within the same individual. In our study, 35% (8/23) of the patients dropped out prematurely. Our results align with a recent mobile health study of chronic low back pain with a 6-month follow-up [41]. This study compared an intervention group that received a face-to-face home visit, 12 telephone sessions, and an activity sensor with a control group that received only physical activity information and advice to stay active. The intervention group had a dropout rate of 9%, whereas the control group had a dropout rate of 42%. Although the intervention group was equipped with an activity tracker, they also benefited from ongoing personal support. This highlights the potential influence of personal engagement and a lack of support as factors that may have played a key role in preventing dropout. Notably, in our study, of the 8 patients who dropped out, 7 (88%) had previously indicated that declining physical health made monitoring daily activity burdensome and believed that discussing their minimal progress could negatively affect their mental well-being. Interestingly, patients tended to stop wearing the activity sensor just before withdrawal. This potentially transitional period suggests an opportune time for targeted interventions to increase support for participants to achieve their goals. Future research would benefit from providing tailored support to inspire individuals to reintegrate physical activity and regain motivation after a decline, for example, through proactive telephone outreach. An example of achieving high compliance using proactive telephone outreach is a recent study by Sieber et al [28]. This longitudinal activity-sensing study examined people with multiple sclerosis, consisted of a 2- to 3-week inpatient stay and then a 1-month follow-up, and reported a dropout rate of only 4% [28].

Regarding research aim 3, our study found that patients exhibited a wide range of activity levels after returning home, both on an individual basis and in comparison with each other. In addition, the self-reported achievement of home activity goals varied. In addition, successful monitoring of self-reported home activity goals was complicated by the fact that a number of these goals did not meet the SMART criteria set. Interestingly, although a subset of patients with vague goals missed the final Stay With It module, which guided patients in setting activity goals for their time at home, leading them to set activity goals for their return home autonomously, ambiguities in goal setting persisted even among patients who completed the module. This underscores that although group settings can promote the sharing of ideas and increase motivation, personalized guidance and direct collaboration with health care professionals may be more effective for some patients in defining realistic and motivating goals. In terms of barriers to physical activity at home, we found that patients often cited fatigue or decreased energy due to weather, especially heat, in addition to pain and declining physical health. This aligns well with a comprehensive 2017 narrative review on multiple sclerosis, which identified issues such as disease-related limitations; personal beliefs; fatigue; the lack of understanding the benefits of exercise; and practical hurdles, including financial issues, the lack of support, and the lack of accessibility, as key barriers to physical activity [42].

This study has some limitations that need to be considered. Our study was exploratory and observational in nature, with a limited sample size. Patients were also recruited from the same study site, and health care professionals invited those they felt would benefit from the program. It is likely that this led to a selection bias, with a preference for patients who were already inclined to incorporate more physical activity into their daily lives. It is also difficult to disentangle the relative effects of the program and the activity sensors. Previous research has also shown that simply wearing an activity sensor can promote a healthier lifestyle by encouraging self-monitoring [40,43]. To truly distinguish between the individual effects of the program and activity sensors, future research would need to include control groups. For example, in a subsequent study, 1 group could be assigned to receive the Stay With It program in combination with an activity tracker, whereas a control group could receive only the activity tracker. Ideally, a third group would receive standard care, such as that offered in Valens, without the activity tracker, which would further help in assessing the effect. Because not using a sensor also precludes the collection of daily-life data, an alternative could be a blinded activity sensor with the display covered (eg, by tape). This approach was successfully used in the 2020 study by Bentley et al [44].

Consumer-grade activity sensors are affordable and easy to use, but their accuracy can be limited for patients with physical impairments. For example, patients with Parkinson disease often have shorter, shuffling steps and slower walking speeds (for an overview, refer to the study by Wendel et al [45]). Because many of the step-counting algorithms in such sensors are based on data from healthy individuals, their accuracy may be reduced for unique movement patterns. Slower speeds may not trigger the algorithm to detect the steps. This observation is consistent with the findings that consumer-grade sensors show reduced accuracy in detecting shorter stride lengths, particularly in patients with musculoskeletal and neuromuscular conditions in a rehabilitation setting [41].

In terms of lessons learned, the findings from our study indicate that patients perceive a benefit from preparing for their return home in a detailed fashion, and the group setting was generally well received. This approach is particularly effective in settings where groups include patients with similar impairments and motivations, facilitating a mutual exchange of ideas. However, challenges arise when patients vary widely in their health conditions, as such disparities can create tensions: patients who are more severely impaired may feel frustrated when confronted with the severity of their disease, whereas patients who are less impaired may not only miss out on benefits but also feel anxious about their future health trajectory. Another lesson learned is that some patients have difficulty setting specific, realistic activity goals. In addition to being measurable, these goals also need to be motivating, as overly ambitious activity goals can lead to a constant sense of failure. In some cases, activity goals may be best defined in a one-on-one setting rather than in a group setting. Another key lesson from our research is the importance of involving health care professionals in the implementation process. In our study, these professionals were adept at identifying organizational shortcomings and were attuned to group dynamics. However, overburdening them with additional tasks is likely to be a hindrance to implementation and, in the long run, may drain their commitment. The study team assisted with setting up the activity sensor and documenting responses to free-text questions, which may have increased patient motivation. In future implementations, ongoing communication between patients and their primary therapists regarding the program and the activity sensor may continuously strengthen their motivation.

Conclusions

In conclusion, our research shows that integrating aftercare programs into routine care holds promise for supporting patients to stay active at home. Activity sensors may facilitate these efforts, although their potential may vary depending on individual health status, the level of impairment, and personal preferences. In particular, remote physical activity monitoring can also open new avenues for postrehabilitation care by signaling windows for timely interventions, for instance, through motivational messages in the event of a physical activity decline. Finally, our research also highlights that considering health care professionals' perspective is critical when implementing novel measurement devices (such as activity sensors) and novel interventions into routine care.

Acknowledgments

The authors are most grateful to all the patients and health care professionals at the Valens Rehabilitation Centre who dedicated their time to supporting this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of the 4 modules from the Valens Rehabilitation Centre's Stay With It program.

[PDF File (Adobe PDF File), 25 KB - mhealth_v11i1e50729_app1.pdf]

Multimedia Appendix 2

Free-text questions posed to patients and how they correspond to the 4 normalization process theory constructs and the study's 3 overarching research aims.

[PDF File (Adobe PDF File), 48 KB - mhealth_v11i1e50729_app2.pdf]

Multimedia Appendix 3

Free-text questions posed to health care professionals and how they correspond to the 4 normalization process theory constructs and the study's 3 overarching research aims.

[PDF File (Adobe PDF File), 766 KB - mhealth_v11i1e50729_app3.pdf]

Multimedia Appendix 4

Health conditions of patients participating in the study.

[PDF File (Adobe PDF File), 21 KB - mhealth_v11i1e50729_app4.pdf]

References

1. Scrutinio D, Giardini A, Chiovato L, Spanevello A, Vitacca M, Melazzini M, et al. The new frontiers of rehabilitation medicine in people with chronic disabling illnesses. Eur J Intern Med 2019 Mar;61:1-8 [FREE Full text] [doi: 10.1016/j.ejim.2018.10.019] [Medline: 30389274]
2. Sîrbu CA, Thompson DC, Plesa FC, Vasile TM, Jianu DC, Mitrica M, et al. Neurorehabilitation in multiple sclerosis-a review of present approaches and future considerations. J Clin Med 2022 Nov 27;11(23):7003 [FREE Full text] [doi: 10.3390/jcm11237003] [Medline: 36498578]
3. Vister E, Tijsma ME, Hoang PD, Lord SR. Fatigue, physical activity, quality of life, and fall risk in people with multiple sclerosis. Int J MS Care 2017 Mar 1;19(2):91-98 [FREE Full text] [doi: 10.7224/1537-2073.2015-077] [Medline: 32607067]
4. Marquez DX, Aguiñaga S, Vásquez PM, Conroy DE, Erickson KI, Hillman C, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med 2020 Oct 12;10(5):1098-1109 [FREE Full text] [doi: 10.1093/tbm/ibz198] [Medline: 33044541]
5. Izquierdo M, Merchant RA, Morley JE, Anker SD, Aprahamian I, Arai H, et al. International exercise recommendations in older adults (ICFSR): expert consensus guidelines. J Nutr Health Aging 2021;25(7):824-853. [doi: 10.1007/s12603-021-1665-8] [Medline: 34409961]
6. Faíl LB, Marinho DA, Marques EA, Costa MJ, Santos CC, Marques MC, et al. Benefits of aquatic exercise in adults with and without chronic disease-a systematic review with meta-analysis. Scand J Med Sci Sports 2022 Mar;32(3):465-486. [doi: 10.1111/sms.14112] [Medline: 34913530]

7. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, Del Coso J, Leyton-Román M, Luque-Casado A, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023 [FREE Full text] [doi: 10.3390/ijerph18042023] [Medline: 33669679]
8. Kalb R, Brown TR, Coote S, Costello K, Dalgas U, Garmon E, et al. Exercise and lifestyle physical activity recommendations for people with multiple sclerosis throughout the disease course. Mult Scler 2020 Oct 23;26(12):1459-1469. [doi: 10.1177/1352458520915629] [Medline: 32323606]
9. Halabchi F, Alizadeh Z, Sahraian MA, Abolhasani M. Exercise prescription for patients with multiple sclerosis; potential benefits and practical recommendations. BMC Neurol 2017 Sep 16;17(1):185 [FREE Full text] [doi: 10.1186/s12883-017-0960-9] [Medline: 28915856]
10. Fasczewski KS, Garner LM, Clark LA, Michels HS, Migliarese SJ. Medical therapeutic yoga for multiple sclerosis: examining self-efficacy for physical activity, motivation for physical activity, and quality of life outcomes. Disabil Rehabil 2022 Jan 12;44(1):106-113. [doi: 10.1080/09638288.2020.1760364] [Medline: 32393075]
11. Fasczewski KS, Gill DL, Rothberger SM. Physical activity motivation and benefits in people with multiple sclerosis. Disabil Rehabil 2018 Jun 14;40(13):1517-1523. [doi: 10.1080/09638288.2017.1300946] [Medline: 28291956]
12. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
13. Fiuza-Luces C, Santos-Lozano A, Joyner M, Carrera-Bastos P, Picazo O, Zugaza JL, et al. Exercise benefits in cardiovascular disease: beyond attenuation of traditional risk factors. Nat Rev Cardiol 2018 Dec 16;15(12):731-743. [doi: 10.1038/s41569-018-0065-1] [Medline: 30115967]
14. Heine M, van de Port I, Rietberg MB, van Wegen EE, Kwakkel G. Exercise therapy for fatigue in multiple sclerosis. Cochrane Database Syst Rev 2015 Sep 11;2015(9):CD009956 [FREE Full text] [doi: 10.1002/14651858.CD009956.pub2] [Medline: 26358158]
15. Block VJ, Bove R, Zhao C, Garcha P, Graves J, Romeo AR, et al. Association of continuous assessment of step count by remote monitoring with disability progression among adults with multiple sclerosis. JAMA Netw Open 2019 Mar 01;2(3):e190570 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0570] [Medline: 30874777]
16. Charron S, McKay KA, Tremlett H. Physical activity and disability outcomes in multiple sclerosis: a systematic review (2011-2016). Mult Scler Relat Disord 2018 Feb;20:169-177. [doi: 10.1016/j.msard.2018.01.021] [Medline: 29414293]
17. Adams V, Linke A. Impact of exercise training on cardiovascular disease and risk. Biochim Biophys Acta Mol Basis Dis 2019 Apr 01;1865(4):728-734 [FREE Full text] [doi: 10.1016/j.bbadis.2018.08.019] [Medline: 30837069]
18. Chiu C, Bishop M, Pionke JJ, Strauser D, Santens RL. Barriers to the accessibility and continuity of health-care services in people with multiple sclerosis: a literature review. Int J MS Care 2017;19(6):313-321 [FREE Full text] [doi: 10.7224/1537-2073.2016-016] [Medline: 29270089]
19. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
20. Pratap A, Grant D, Vegesna A, Tummalacherla M, Cohan S, Deshpande C, et al. Evaluating the utility of smartphone-based sensor assessments in persons with multiple sclerosis in the real-world using an app (elevateMS): observational, prospective pilot digital health study. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e22108 [FREE Full text] [doi: 10.2196/22108] [Medline: 33107827]
21. Drigny J, Joussain C, Gremeaux V, Morello R, Van Truc PH, Stapley P, et al. Development and validation of a questionnaire to assess barriers to physical activity after stroke: the barriers to physical activity after stroke scale. Arch Phys Med Rehabil 2019 Sep;100(9):1672-1679. [doi: 10.1016/j.apmr.2018.12.034] [Medline: 30684487]
22. Vissers M, van den Berg-Emons R, Sluis T, Bergen M, Stam H, Bussmann H. Barriers to and facilitators of everyday physical activity in persons with a spinal cord injury after discharge from the rehabilitation centre. J Rehabil Med 2008 Jun;40(6):461-467 [FREE Full text] [doi: 10.2340/16501977-0191] [Medline: 18509562]
23. Fleury J, Lee SM, Matteson B, Belyea M. Barriers to physical activity maintenance after cardiac rehabilitation. J Cardiopulm Rehabil 2004;24(5):296-305; quiz 306. [doi: 10.1097/00008483-200409000-00002] [Medline: 15602147]
24. Griffin N, Kehoe M. A questionnaire study to explore the views of people with multiple sclerosis of using smartphone technology for health care purposes. Disabil Rehabil 2018 Jun;40(12):1434-1442. [doi: 10.1080/09638288.2017.1300332] [Medline: 28322588]
25. Silveira SL, Motl RW. Activity monitor use among persons with multiple sclerosis: report on rate, pattern, and association with physical activity levels. Mult Scler J Exp Transl Clin 2019 Nov 09;5(4):2055217319887986 [FREE Full text] [doi: 10.1177/2055217319887986] [Medline: 31741742]
26. Wendrich K, van Oirschot P, Martens MB, Heerings M, Jongen PJ, Krabbenborg L. Toward digital self-monitoring of multiple sclerosis: investigating first experiences, needs, and wishes of people with MS. Int J MS Care 2019 Nov;21(6):282-291. [doi: 10.7224/1537-2073.2018-083] [Medline: 31889935]
27. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]

28. Sieber C, Haag C, Polhemus A, Sylvester R, Kool J, Gonzenbach R, et al. Feasibility and scalability of a fitness tracker study: results from a longitudinal analysis of persons with multiple sclerosis. Front Digit Health 2023 Feb 28;5:1006932 [FREE Full text] [doi: 10.3389/fdgth.2023.1006932] [Medline: 36926468]
29. Birken SA, Rohweder CL, Powell BJ, Shea CM, Scott J, Leeman J, et al. T-CaST: an implementation theory comparison and selection tool. Implement Sci 2018 Nov 22;13(1):143 [FREE Full text] [doi: 10.1186/s13012-018-0836-4] [Medline: 30466450]
30. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Jun 07;13(1):80. [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
31. Ross J, Stevenson F, Dack C, Pal K, May C, Michie S, et al. Developing an implementation strategy for a digital health intervention: an example in routine healthcare. BMC Health Serv Res 2018 Oct 19;18(1):794 [FREE Full text] [doi: 10.1186/s12913-018-3615-7] [Medline: 30340639]
32. May C, Rapley T, Mair FS, Treweek S, Murray E, Ballini L, et al. Normalization process theory on-line users' manual, toolkit and NoMAD instrument. Normalization Process Theory. 2015. URL: https://normalization-process-theory.northumbria.ac.uk/ [accessed 2021-11-04]
33. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
34. Bovend'Eerdt TJ, Botell RE, Wade DT. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. Clin Rehabil 2009 Apr;23(4):352-361. [doi: 10.1177/0269215508101741] [Medline: 19237435]
35. Löllgen H. Das Anstrengungsempfinden (RPE, Borg-Skala). Deutsche Zeitschrift für Sportmedizin 2004;55(11):299-300 [FREE Full text]
36. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
37. PROMIS global health (PROMIS 10). CODE Technology. 2015 Sep 16. URL: https://www.codetechnology.com/promis-global-10/ [accessed 2021-12-20]
38. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Sep 19;18(7):873-880 [FREE Full text] [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
39. Naef R, Kläusler-Troxler M, Ernst J, Huber S, Dinten-Schmid B, Karen T, et al. Translating family systems care into neonatology practice: a mixed method study of practitioners' attitudes, practice skills and implementation experience. Int J Nurs Stud 2020 Feb;102:103448. [doi: 10.1016/j.ijnurstu.2019.103448] [Medline: 31726312]
40. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
41. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
42. Ploughman M. Breaking down the barriers to physical activity among people with multiple sclerosis – a narrative review. Phys Ther Rev 2017 Apr 19;22(3-4):124-132. [doi: 10.1080/10833196.2017.1315212]
43. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
44. Bentley CL, Powell L, Potter S, Parker J, Mountain GA, Bartlett YK, et al. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16203 [FREE Full text] [doi: 10.2196/16203] [Medline: 32490838]
45. Wendel N, Macpherson CE, Webber K, Hendron K, DeAngelis T, Colon-Semenza C, et al. Accuracy of activity trackers in Parkinson disease: should we prescribe them? Phys Ther 2018 Aug 01;98(8):705-714. [doi: 10.1093/ptj/pzy054] [Medline: 29718452]

Abbreviations

BASEC: Business Administration System for Ethics Committees
NPT: normalization process theory
SMART: specific, measurable, achievable, realistic, time-bound

Edited by L Buis, G Eysenbach; submitted 14.07.23; peer-reviewed by C Goncalves, D Allen; comments to author 03.08.23; revised version received 18.09.23; accepted 09.11.23; published 08.12.23.

Please cite as:
Lu Z, Signer T, Sylvester R, Gonzenbach R, von Wyl V, Haag C
Implementation of Remote Activity Sensing to Support a Rehabilitation Aftercare Program: Observational Mixed Methods Study With Patients and Health Care Professionals
JMIR Mhealth Uhealth 2023;11:e50729
URL: https://mhealth.jmir.org/2023/1/e50729
doi:10.2196/50729
PMID:38064263

Physical Activity Pattern of Adults With Metabolic Syndrome Risk Factors: Time-Series Cluster Analysis

Junhyoung Kim[1], MPH; Jin-Young Choi[1], MPH; Hana Kim[1], MPH; Taeksang Lee[1], MPH; Jaeyoung Ha[1], MPH; Sangyi Lee[1], MPH; Jungmi Park[1], MPH; Gyeong-Suk Jeon[2], PhD; Sung-il Cho[1,3], PhD

1

2

3

Corresponding Author:
Sung-il Cho, PhD

Abstract

Background: Physical activity plays a crucial role in maintaining a healthy lifestyle, and wrist-worn wearables, such as smartwatches and smart bands, have become popular tools for measuring activity levels in daily life. However, studies on physical activity using wearable devices have limitations; for example, these studies often rely on a single device model or use improper clustering methods to analyze the wearable data that are extracted from wearable devices.

Objective: This study aimed to identify methods suitable for analyzing wearable data and determining daily physical activity patterns. This study also explored the association between these physical activity patterns and health risk factors.

Methods: People aged >30 years who had metabolic syndrome risk factors and were using their own wrist-worn devices were included in this study. We collected personal health data through a web-based survey and measured physical activity levels using wrist-worn wearables over the course of 1 week. The Time-Series Anytime Density Peak (TADPole) clustering method, which is a novel time-series method proposed recently, was used to identify the physical activity patterns of study participants. Additionally, we defined physical activity pattern groups based on the similarity of physical activity patterns between weekdays and weekends. We used the χ^2 or Fisher exact test for categorical variables and the 2-tailed *t* test for numerical variables to find significant differences between physical activity pattern groups. Logistic regression models were used to analyze the relationship between activity patterns and health risk factors.

Results: A total of 47 participants were included in the analysis, generating a total of 329 person-days of data. We identified 2 different types of physical activity patterns (early bird pattern and night owl pattern) for weekdays and weekends. The physical activity levels of early birds were less than that of night owls on both weekdays and weekends. Additionally, participants were categorized into stable and shifting groups based on the similarity of physical activity patterns between weekdays and weekends. The physical activity pattern groups showed significant differences depending on age (*P*=.004) and daily energy expenditure (*P*<.001 for weekdays; *P*=.003 for weekends). Logistic regression analysis revealed a significant association between older age (≥40 y) and shifting physical activity patterns (odds ratio 8.68, 95% CI 1.95-48.85; *P*=.007).

Conclusions: This study overcomes the limitations of previous studies by using various models of wrist-worn wearables and a novel time-series clustering method. Our findings suggested that age significantly influenced physical activity patterns. It also suggests a potential role of the TADPole clustering method in the analysis of large and multidimensional data, such as wearable data.

(JMIR Mhealth Uhealth 2023;11:e50663) doi:10.2196/50663

KEYWORDS

wrist-worn wearable; wearable data; physical activity pattern; TADPole clustering; TADPole; cluster; clustering; wearable; wearables; wrist-worn; physical activity; pattern; patterns; data analysis; data analytics; regression; risk; risks; time series; Time-Series Anytime Density Peak

Introduction

Physical activity has been linked to numerous health benefits. A cross-sectional study conducted in Japan reported that inactive individuals with <23 metabolic equivalent (MET) hours per week had more than double the risk of metabolic syndrome compared to active individuals (≥23 MET h/wk) [1]. Another study used X-means clustering to identify intensity and temporal activity patterns and demonstrated that inactive individuals had a 3-fold higher risk of cardiovascular disease compared to active individuals [2]. A systematic review also suggested that increased physical activity correlates with improved health status [3]. Additionally, a meta-analysis by Pearce et al [4] found that adults who achieved the recommended physical activity level (4.4 marginal MET h/wk) had a 25% lower risk of depression compared to inactive adults.

Wrist-worn wearables, such as smartwatches and smart bands equipped with computers and sensors, have become popular tools for measuring physical activity [5,6]. Most individuals opt to wear wrist-worn wearables for several reasons, including affordability, functionality, and stylish design [7]. This has enabled the measurement of physical activity in daily life rather than being limited to the laboratory setting. There have also been notable improvements in the accuracy of measurements obtained from wrist-worn wearables [8,9]. As a result, an increasing number of studies are focusing on measuring and analyzing physical activity using wrist-worn devices [7].

Several studies are currently exploring different aspects of physical activity using wrist-worn devices. These include investigations of the accuracy of these devices [8,9], the relationship between physical activity and personal characteristics [2,10], the impact of interventions using wrist-worn wearables [11,12], and behavior prediction [13-15]. Some studies have also sought to identify physical activity patterns using data collected from wrist-worn wearables [2,16-20]. However, it is important to note that these studies have certain limitations.

The diversity of wearable device models poses a challenge for observational studies using wearables within a population. Most previous studies either provided the participants with a specific device model or restricted participation to individuals using a particular model [21-23]. However, wearable device models are continuously evolving to cater to individual preferences. Furthermore, each model has its own app, which extracts data in a specific format. Consequently, there is a need for flexible methods that can effectively analyze essential information derived from diverse forms of wearable data.

Grouping methods, such as principal component analysis [16-18] and *k*-means clustering [2,19,20], have commonly been used to identify similarities among participants and summarize activity patterns within groups. Time-series analysis methods can also be used to classify daily activity patterns. However, previous studies using popular clustering methods have shown sensitivity to minor variations in data formats, resulting in inconsistent outcomes.

The *k*-means clustering method is a kind of partitional clustering method [24]. This clustering method is easy to implement and successfully distinguishes clusters using data from all participants, with low computational cost [20,24]. Therefore, it can be applied to large and multidimensional data. However, the number of clusters (parameter *k*) had to be predefined, because the parameter *k* is not commonly known; therefore, iterative analysis is required to get the optimal number of clusters. The X-means clustering mentioned above is also a type of *k*-means clustering and is a clustering method that automatically finds the number of clusters by taking the disadvantage of *k*-means into consideration [2]. In addition, the *k*-means clustering method provides unstable results due to its random selection of the initial centroid [24].

The hierarchical clustering method and Density-Based Spatial Clustering of Applications With Noise (DBSCAN) are also popular time-series clustering methods [19,20]. The hierarchical clustering method is a method of classifying clusters based on the hierarchical structure of data and basically considers 1 time series as 1 cluster [20,24]. The hierarchical clustering method has the advantage of visualizing the hierarchical structure of data, because it shows the hierarchical structure as a tree (ie, a dendrogram). However, its computational cost is high, and a significant number of data points must be excluded from the analysis to obtain the desired number of clusters, raising uncertainties about the accuracy of the resulting clusters [24]. DBSCAN is a density-based clustering method that calculates the density of data based on the Euclidean distance calculation method and excludes data considered as noise from clustering. However, as shown in a study by Dobbins and Rawassizadeh [20], DBSCAN has the highest computational cost, and the Euclidean method applied to DBSCAN is not suitable for multidimensional data. Thus, the hierarchical clustering method and DBSCAN are not feasible for large or multidimensional data.

A more flexible time-series clustering method called Time-Series Anytime Density Peak (TADPole) clustering has recently been proposed [24,25]. This is an algorithm that can perform fast clustering by reducing the distance calculation process in the Density Peak clustering method. This method uses dynamic time warping to calculate distances between series. Unlike the Euclidean method, which matches data at the same points across a series, this method identifies optimal warping paths between series to identify better point-to-point matches except for the first and last points. The prototype for TADPole clustering is partition around medoid clustering, which creates clusters by minimizing the sum of distances calculated based on an arbitrary series (ie, medoid) [24]. TADPole clustering classifies series as neighbors if the distance between them is below a certain cutoff value [24,25]. Theoretical and technical details are readily accessible in the previous literature [24,25] and thus are not repeated here.

The TADPole clustering algorithm can cluster multidimensional data measured by wearable devices as well as large data [25]. In this study, we chose the TADPole clustering method to test whether it can be effectively applied to wearable data, which was not addressed in its published paper [25]. To the best of our knowledge, this novel approach has not yet been applied to the study of health indicators measured using wearable devices. Therefore, this study assessed the feasibility of using a time-series clustering method to analyze wearable data for daily physical activity patterns and explored the association between these patterns and health risk factors.

Methods

Study Participants

This study examined physical activity patterns among at-risk individuals using wrist-worn wearables. Step counts, distances, and energy expenditure (EE) were measured over 1 week in a real-life setting between November 22, 2021, and December 2, 2021. Participants aged >30 years who had risk factors based on metabolic syndrome diagnostic criteria and were currently using wrist-worn devices (eg, smartwatches and smart bands) were included. The risk factors included blood pressure ≥130/85 mm Hg, fasting blood sugar ≥100 mg/dL, triglyceride levels ≥150 mg/dL, high-density lipoprotein level <40 or <50 mg/dL (for male and female individuals, respectively), and waist circumference ≥90 or ≥85 cm (for male and female individuals, respectively). The number of study participants was selected based on the analysis results using G*Power (Heinrich Heine Universität Düsseldorf) and previous similar research cases. First, we used G*Power to perform an ANOVA because the data of the study participants would be measured repeatedly. The effect size (Cohen *f*) was set to be 0.25, and the significance level (α) and power (1 – β) were assumed to be 5% and 95%, respectively. As a result, a total sample size of 36 was calculated. Next, we considered the previous work by Huh et al [26], which applied wearable technology to patients with metabolic syndrome and recruited a total of 53 people. However, 33 patients dropped out during the 12-week study period due to the withdrawal of consent, device malfunction, and the loss of follow-up. We finally decided to recruit 60 study participants to achieve a sufficient effect. We used the Seoul National University mailing system to recruit research participants. Starting on November 3, 2021, we sent 2 emails to all members of the university; this lasted until November 19, 2021, when the recruitment was completed. The purpose of the study, eligibility criteria for study participants, and research procedures were provided via email. A detailed explanation of the study was provided in a web-based meeting after all written informed consent was obtained. Before physical activity measurements, all participants completed a web-based questionnaire through Google Forms, which collected personal health data and details about their physical activity (including type, intensity, and duration). The participants wore their own wrist-worn wearables for ≥10 hours per day with the physical activity measurement function activated for 1 week.

Following the 1-week period of physical activity measurements, the participants were asked to complete another web-based questionnaire to assess user experience, including cognition, context, applicability, and behavioral changes. Of the 60 participants initially included in the study, 13 were excluded due to missing or limited baseline data (n=3) or the unavailability of physical activity data from the database (n=10). Consequently, the analysis involved 47 participants who met the inclusion criteria and had data available for analysis.

Data Collection

During the 7-day measurement period, the participants activated the measurement function on their wrist-worn wearables, generating a total of 329 person-days of data. The study included wrist-worn wearables from the Apple Watch, Samsung Galaxy Watch, and Xiaomi Mi Band series (Multimedia Appendix 1). Data were collected continuously throughout the measurement period, with each device automatically storing individual data and synchronizing it with the participants' cell phones.

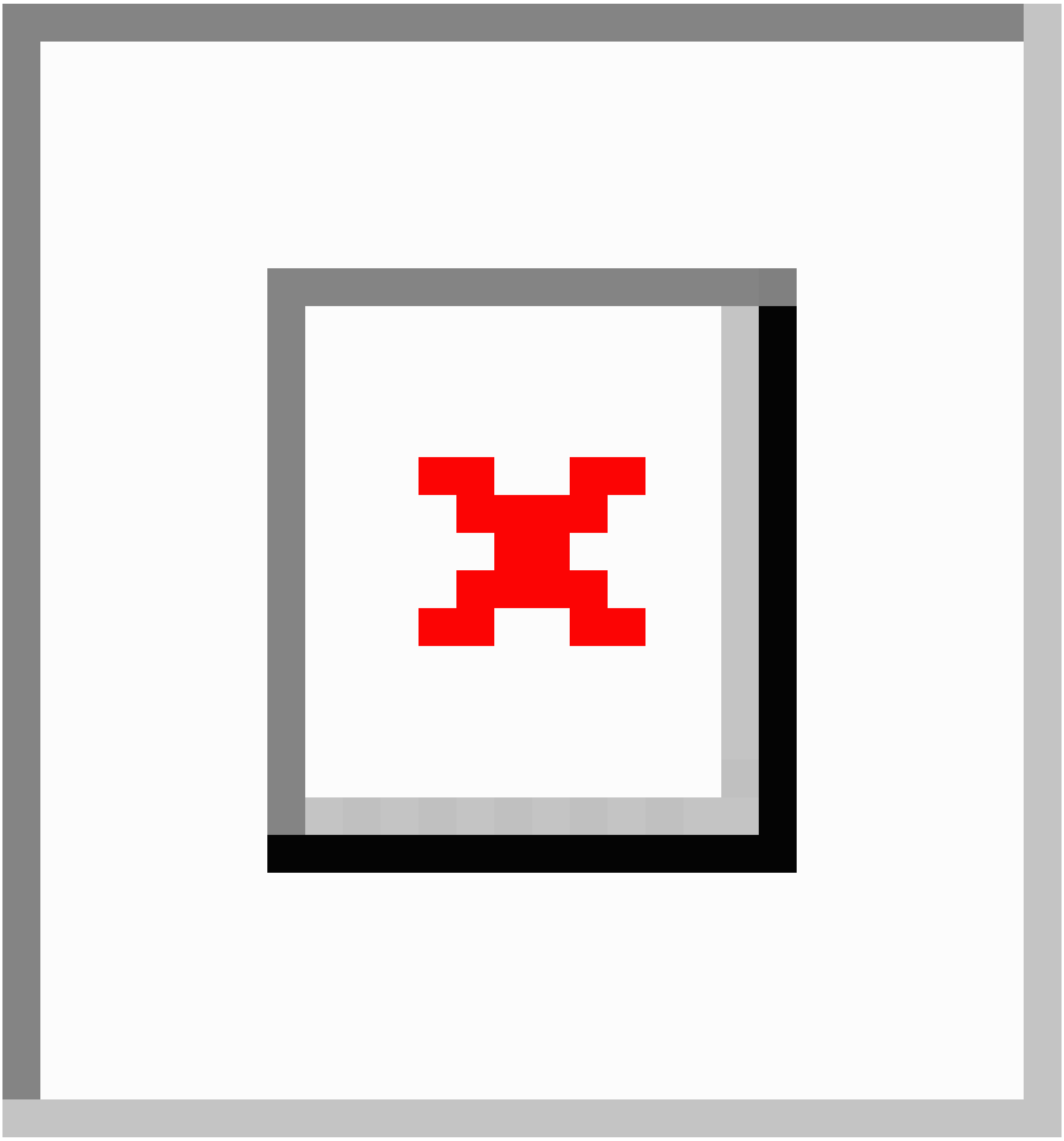
Upon the completion of the measurement period, each participant exported their individual data through the official data export system and submitted the data via email. We provided the participants with detailed instructions specific to the manufacturer, version, and brand of their wrist-worn wearables. We then decompressed the data files and preprocessed the data to extract the selected variables, including step count, distance, EE, and duration with the start and end points. These variables were then merged into a unified data format for analysis.

Time-Series Clustering

The time-series clustering method was applied in 3 steps to identify clusters representing the physical activity patterns of the study participants. First, due to variations in data recording formats among the different wrist-worn device brands, the data were edited to ensure consistency. Data from different brands were standardized into the same format. For instance, the Samsung Galaxy Watch series record EE as "kcal per minute" during wearing, whereas the Apple Watch and Xiaomi Mi Band series record EE as "cal" and "kcal," respectively, for each distinct activity. We converted these data into 10-minute "kcal" EE values. For activities performed for >10 minutes while wearing the Apple Watch or Xiaomi Mi Band series, the activity duration was divided into 10-minute intervals, assuming that a consistent amount of energy was expended during each interval.

Second, the data were divided into weekdays (Monday to Friday; 235 person-days) and weekends (Saturday and Sunday; 94 person-days), with the clustering method applied separately to each group. We calculated the average EE for weekdays and weekends based on the daily 10-minute EE values. It was assumed that the EE was 0 when the participant was not wearing the device. Figure 1 illustrates an example of physical activity measurements for 7 person-days, with each data point representing the EE in kcal per 10 minutes over the course of 1 week.

Figure 1. Physical activity measurement for 1 of the participants (7 person-days).



Third, time-series clustering was conducted through TADPole clustering, a recently developed technique that allows faster clustering by implementing a cutoff value to determine clusters [24,25]. To determine the optimal clustering model, we analyzed the expected number of clusters (parameter *k*) and the cutoff value. For the third step, we used the *dtwclust* package in RStudio [27]. This package offers a range of functions for conducting time-series clustering, including the TADPole clustering method. We had to specify certain parameters, including the cluster type, the number of clusters, the cutoff value, and the window size. Since we chose TADPole as the cluster type, we did not need to specify the distance parameter. As for the cutoff value and window size, we adjusted them based on the volume of data. Given that our study used 144 data points, we selected values that fell below this threshold. As the optimal number of clusters and cutoff values were unknown, cluster evaluation was performed using the silhouette index, which is a popular cluster validity index. Based on the cluster evaluation, the model with the highest silhouette index was selected as the optimal clustering model.

Statistical Analysis

The demographic characteristics of the participants, including sex, age, work type (sitting, standing, etc), daily EE (weekdays and weekends), physical activity changes after using wrist-worn wearables, weekly physical activity patterns, and number of risk factors (1 or >1), were recorded. For categorical variables, the number and proportion for each category were presented, along with the *P* value calculated using the χ^2 or Fisher exact test for variables with counts <5. For numerical variables, mean and SD with *P* values were calculated using the 2-tailed *t* test.

The association between weekly physical activity patterns and participant characteristics was analyzed using a logistic regression model. The regression model was evaluated in terms of pseudo-R^2, accuracy, Hosmer-Lemeshow goodness of fit, and the receiver operating characteristic curve. The results of the logistic regression were presented as odds ratios with 95% CIs and the corresponding *P* values. Statistical significance was taken as $P\leq.05$. The statistical analyses were performed using RStudio (version 2022.07.2+576; Posit) [28].

Ethical Considerations

The study was approved by the Institutional Review Board of Mokpo National University (approval MNUIRB-210625-SB-014-01). All participants provided informed consent before study participation. The submitted data were anonymized before analysis. Participants who provided data and finished the web-based survey received a compensation of ₩100,000 (US $77.65).

Results

General Participant Characteristics

Among the 47 participants, 23 (49%) were male and 24 (51%) were female (Table 1). In terms of age, 30 (64%) participants were aged <40 years, whereas 17 (36%) were aged ≥40 years. The majority (n=42, 89%) of the participants had a sedentary job. The average EE during weekdays was 223 (SD 175) kcal, whereas that on weekends was 191 (SD 164) kcal. After using wrist-worn wearables, 29 (62%) participants reported a decrease or no change in physical activity, whereas 18 (38%) participants reported an increase. In terms of health risk factors, 25 (53%) participants had only 1 risk factor, whereas 22 (47%) had >1 risk factors.

Table . General participant characteristics.

Variable and level		Value (N=47)
Sex, n (%)		
	Male	23 (49)
	Female	24 (51)
Age group (y), n (%)		
	<40	30 (64)
	≥40	17 (36)
Work type, n (%)		
	Sitting	42 (89)
	Other	5 (11)
Daily EE [a] (kcal), mean (SD)		
	Weekdays	223 (175)
	Weekends	191 (164)
Change in PA [b], n (%)		
	No change or decrease	29 (62)
	Increase	18 (38)
PA pattern group, n (%)		
	Stable	35 (75)
	Shifting	12 (25)
Number of health risk factors, n (%)		
	1	25 (53)
	>1	22 (47)

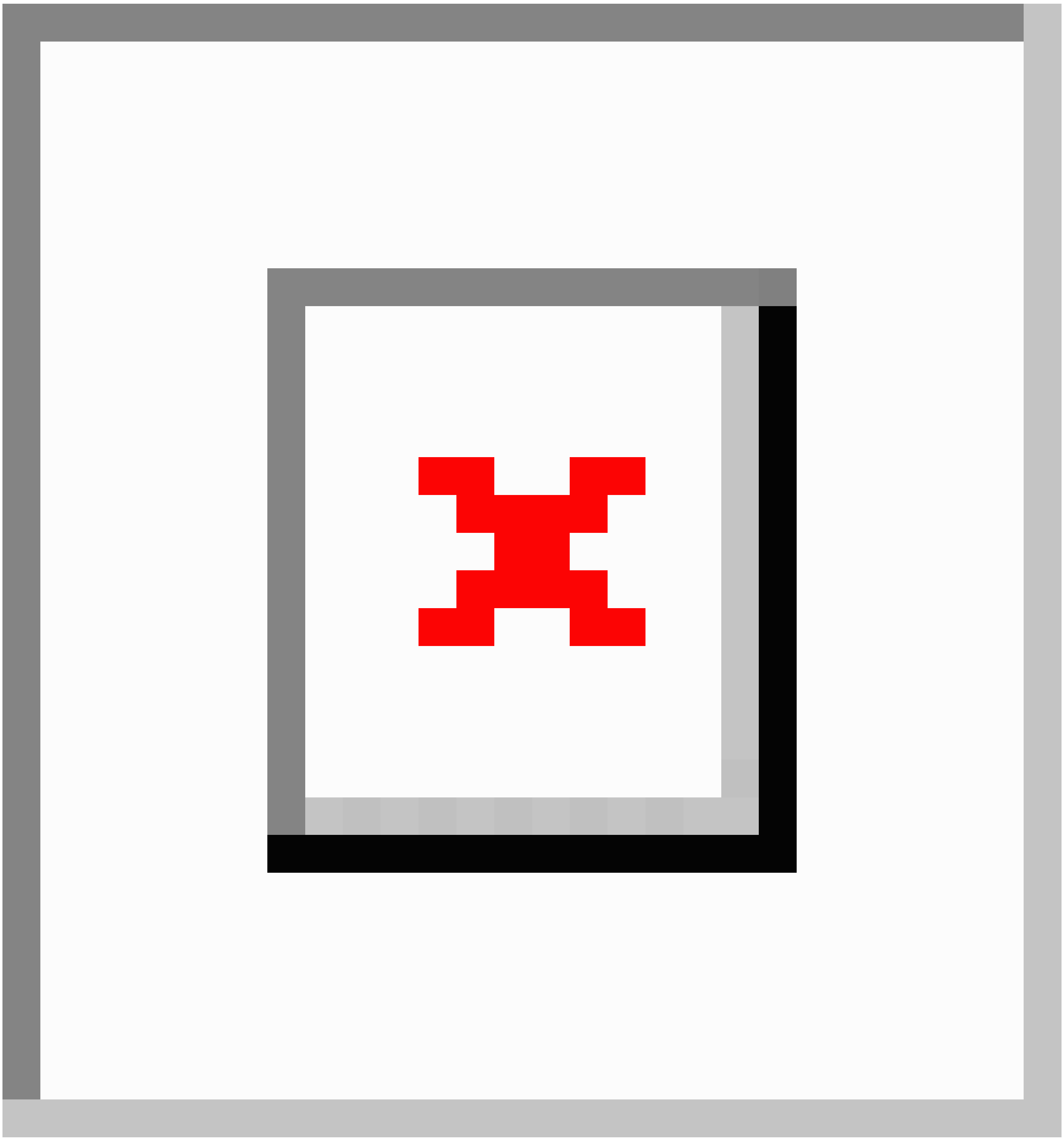
[a]EE: energy expenditure.

[b]PA: physical activity.

Physical Activity Patterns

The time-series cluster analysis resulted in the highest silhouette index when there were 2 clusters (*k*=2) for both weekdays and weekends. Therefore, 2 clusters each were distinguished for weekdays and weekends (Figure 2). The left and right columns of Figure 2 represent the weekday and weekend clusters, respectively. Each cluster included data from at least 2 participants, and 2 distinct cluster types were distinguished with different starting times for physical activity. The “early bird” type (represented by blue dots in Figure 2) initiated physical activity after 6 AM, whereas the “night owl” type (represented by orange dots in Figure 2) began physical activity before 6 AM.

Figure 2. Physical activity clusters on weekdays and weekends: (A) cluster 1: “early birds” on weekdays, (B) cluster 2: “night owls” on weekdays, (C) cluster 3: “early birds” on weekends, and (D) cluster 4: “night owls” on weekends.



Among the 37 early birds on weekdays, 57% (n=21) were female and 76% (n=28) were in their 30s, with female participants in their 30s accounting for the highest proportion (n=18, 48%). Among the 10 night owls on weekdays, 70% (n=7) were male and 80% (n=8) were in their aged ≥40 years, and 40% (n=4) were male individuals aged ≥40 years. Weekend physical activity patterns were mostly from early birds (n=45, 96%) regardless of sex or age. Out of the 45 early birds on weekends, 51% (n=23) were female and 64% (n=29) were in their 30s, and 40% (n=18) were female individuals aged <40 years. There was 1 participant per sex and age group who was a night owl on weekends, and there was no one who was a night owl on both weekdays and weekends.

Figure 2A shows the physical activity patterns of the 37 (79%) out of 47 participants belonging to cluster 1 (early birds) on weekdays. Physical activity occurred between 6 AM and 8 PM, with most activities being <10 kcal (mean 2.98 kcal). Meanwhile, cluster 2 included 10 (21%) “night owls” on weekdays (Figure 2B). Physical activity for cluster 2 typically started at midnight and ended before 4 PM. Cluster 2 also exhibited greater EE, with an average of 9.51 kcal.

During the weekends, the majority (45/47, 96%) of participants were early birds (cluster 3; Figure 2C). Cluster 3 had an average EE of 2.94 kcal. Cluster 4 (night owls) included only 2 (4%) participants on weekends (Figure 2D) and demonstrated a higher EE (average 8.60 kcal).

Although the average EE was similar between early birds and night owls, their physical activity patterns differed on weekdays and weekends. Regardless of the cluster type, physical activity tended to be shorter in duration on weekdays (Figure 2A and B), becoming longer and more continuous on weekends. Cluster 3, representing early birds on weekends, exhibited up to 10 consecutive physical activity periods, which is equivalent to 100 minutes (4-8 AM; Figure 2C). Cluster 4, representing the night owls on weekends, had the highest total EE of 507.59 kcal and up to 16 consecutive physical activity periods, which is equivalent to 160 minutes (4-8 AM; Figure 2D).

Based on the analysis of physical activity patterns, 2 groups were identified: the stable group and the shifting group. The stable group included individuals who maintained the same physical activity pattern on weekdays and weekends, regardless of whether they were classified as early birds or night owls. Among the 47 participants, 35 (74%) belonged to the stable group, exhibiting early bird physical activity patterns consistently throughout the week. There were no participants belonging to both clusters 2 and 4. On the other hand, the shifting group included individuals whose physical activity patterns differed between weekdays and weekends. There were 12 (26%) participants in the shifting group; 10 (21%) participants displayed an early bird pattern during weekdays (cluster 1) but changed to the night owl pattern on weekends (cluster 4). The remaining 2 (4%) participants exhibited the opposite pattern, that is, a night owl pattern during weekdays (cluster 4) and an early bird pattern on weekends (cluster 1).

Demographic descriptive statistics for the physical activity pattern groups, including the results of the χ^2 test for categorical variables and the *t* test for continuous variables, are presented in Table 2. There were no significant differences between the physical activity pattern groups except in age (*P*=.001) and EE (*P*<.001 for weekdays; *P*=.003 for weekends).

Table . Weekly physical activity (PA) group characteristics.

Variable and level		PA pattern group		*P* value
		Stable (n=35)	Shifting (n=12)	
Sex, n (%)				.19[a]
	Male	15 (43)	8 (67)	
	Female	20 (57)	4 (33)	
Age group (y), n (%)				.004[a]
	<40	27 (77)	3 (25)	
Work type, n (%)				>.99[a]
	Sitting	31 (89)	11 (92)	
	Other	4 (11)	1 (8)	
Daily EE [b] (kcal), mean (SD)				
	Weekdays	169 (130)	383 (194)	<.001[c]
	Weekends	145 (100)	327 (230)	.003[c]
PA changes, n (%)				.74[a]
	No change or decrease	21 (60)	8 (67)	
	Increase	14 (40)	4 (33)	
Number of health risk factors, n (%)				.18[a]
	1	21 (60)	4 (33)	
	>1	14 (40)	8 (67)	

[a]Fisher exact test.

[b]EE: energy expenditure.

[c]*t* test.

Association Between Physical Activity Patterns and Health Risk Factors

A logistic regression model was used to examine the associations of sex, age, and the number of health risk factors with weekly physical activity patterns (Table 3). Logistic regression model accuracy and diagnostic results are presented in Multimedia Appendix 1. Sex (*P*=.45) and the number of health risk factors (*P*=.33) were not significantly associated with the physical activity pattern. In contrast, age showed a statistically significant association with physical activity patterns; the higher age group had higher odds of differences between weekday and weekend physical activity patterns (odds ratio 8.68, 95% CI 1.95-48.85; *P*=.007).

Table . Associations between physical activity patterns and health risk factors.

Variable and level (reference)	OR[a] (95% CI)	*P* value
Sex: female (vs male)	0.69 (0.13-3.64)	.45
Age group: ≥40 y (vs <40 y)	8.68 (1.95-48.85)	.007
Number of health risk factors: >1 (vs 1)	2.21 (0.45-11.92)	.33

[a]OR: odds ratio.

To account for the possibility of reverse causality, we conducted another logistic regression analysis with the number of health risk factors as the outcome variable. Despite this adjustment, there were no significant associations between physical activity patterns and health risk factors (*P*>.99; Multimedia Appendix 1).

Discussion

Principal Findings

In this study, we assessed the effectiveness of the TADPole clustering method for identifying physical activity patterns from wearable data. We also explored the association between these patterns and health risk factors. We found that physical activity patterns on weekdays and weekends were categorized as either daytime (early bird) or nighttime (night owl) patterns. Furthermore, 2 groups were distinguished: 1 with consistent physical activity patterns on weekdays and weekends (stable group) and the other with different patterns between weekdays and weekends (shifting group). Age significantly influenced physical activity patterns.

Comparison to Prior Works

We found that physical activity patterns on weekdays and weekends differed as age increased. Our findings shed light on previously unaddressed or overlooked associations between physical activity patterns and health risk factors. Many previous studies did not report the association between age and physical activity patterns [17,18], and this association was reported only in a few studies [29-32]. Some of these studies reported results consistent with our findings. A study by Caspersen et al [29], which analyzed physical activity patterns according to sex and age, found that not only inactivity but also vigorous activity increased with age. In their study, those aged 18-29 years showed a pattern with the lowest vigorous activity and the highest sustained physical activity, whereas those aged ≥75 years showed a pattern with the highest vigorous activity and the lowest sustained physical activity, indicating differences in physical activity patterns by age. Another study by Rossen et al [32], which analyzed the physical activity patterns of individuals with diabetes for 2 years, found that the younger the age, the more physical activity increased 2 years later, showing that physical activity patterns can vary depending on age.

The physical activity patterns identified in our study were similar to chronotypes, which categorize individuals into morning type (M-type), evening type (E-type), and intermediate type (N-type) based on their preferred timings of activities and sleep [33]. M-type individuals are typically early birds, whereas E-type individuals are night owls who prefer late activity and sleep schedules. N-type individuals do not fall strictly into either category, and most adults belong to this type [33]. Previous studies have suggested that E-type individuals have lower physical activity levels and a higher risk of metabolic syndrome [34,35]. However, this study found that the individuals with night owl tendencies, that is, E-type individuals, exhibited higher physical activity levels compared to early birds, that is, M-type individuals. It is important to note that most participants in our study reported engaging in sedentary work, indicating that the increased physical activity levels among night owls were likely due to leisure activities rather than occupational tasks. This suggests that physical activity patterns are not determined solely by chronotype and that other factors, such as health awareness, can have a significant impact. It is worth noting that an individual's chronotype and activity times may vary based on age and occupation, potentially leading to health issues if not addressed [36]. Therefore, it is crucial to make efforts to achieve the recommended level of physical activity regardless of the specific activity pattern or chronotype.

None of the participants in our study met the criteria for "weekend warriors," which refers to individuals engaging in 1 or 2 sessions of physical activity, particularly on weekends, per week, consuming at least 1000 kcal [37]. Although the combined EE for cluster 2 (weekday night owls) and cluster 4 (weekend night owls) was the closest to that of weekend warriors at 735 kcal, no participants were included in both clusters. In a study conducted by Jang et al [38] in South Korea, only 2.1% of the participants were classified as weekend warriors, but there was no significant difference in metabolic risk between weekend warriors and the regularly active group. The weekend warrior physical activity pattern, which is popular in the United Kingdom, the United States, and Latin America [39], is associated with several health benefits, including a lower risk of obesity [38] and all-cause mortality [37]. A study of Chinese adults found that the weekend warrior physical activity pattern was associated with a lower risk of metabolic syndrome, hypertension, and diabetes in both male and female individuals [40]. Promoting physical activity guidelines may increase the number of weekend warriors in South Korea, where sedentary jobs are common (Table 1).

The TADPole clustering method addresses the limitations of previously popular time-series clustering techniques such as the *k*-means and hierarchical clustering methods in data analysis. As we mentioned in the introduction, the *k*-means clustering method often produced unstable results with clusters changing each time, whereas the TADPole clustering method consistently provided reliable clustering outcomes. However, similar to *k*-means, we iteratively performed the analysis to obtain the optimal clustering results. The TADPole clustering method stands out by providing more reliable results compared to hierarchical clustering and DBSCAN, as it uses all available data without any loss or exclusion. Additionally, our study demonstrated the feasibility of the TADPole clustering method, showing its suitability for handling large and multidimensional data such as wearable data.

Strengths

A key advantage of our study was that we used multiple wrist-worn device models. Although wearables provide data in different formats depending on the model, we standardized the data into a single format to successfully conduct statistical analyses. Furthermore, the TADPole clustering method allowed us to overcome the limitations of hierarchical and *k*-means clustering, which are commonly used time-series clustering methods, resulting in robust and reliable findings.

Limitations

This study had several limitations. It only included 47 participants, which may not have been sufficient to generate meaningful results. The measurement period for physical activity was only 7 days, which may not have been representative of the daily physical activity. Additionally, we assumed that the EE for physical activities exceeding 10 minutes was consistent across 10-minute intervals. Although these assumptions may not perfectly reflect reality, they were considered reasonable given that the participants were going about their normal daily routines.

Conclusions

This study successfully performed time-series clustering using various wrist-worn device models and found TADPole clustering to be a suitable tool for analyzing the data. Physical activity patterns on weekdays and weekends could be categorized into “early birds” and “night owls,” and these patterns were significantly influenced by age. To address the limitations of our study, additional studies with larger sample sizes are required.

Acknowledgments

This research was supported by the Ministry of Science and ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC) support program (IITP-2020-0-01778) supervised by the Institute of Information and Communications Technology Planning and Evaluation (IITP).

Data Availability

The data sets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Included wearable models and the results of the regression model.

[DOCX File, 10165 KB - mhealth_v11i1e50663_app1.docx]

References

1. Kim J, Tanabe K, Yokoyama N, Zempo H, Kuno S. Association between physical activity and metabolic syndrome in middle-aged Japanese: a cross-sectional study. BMC Public Health 2011 Aug 5;11:624. [doi: 10.1186/1471-2458-11-624] [Medline: 21819591]
2. Niemelä M, Kangas M, Farrahi V, et al. Intensity and temporal patterns of physical activity and cardiovascular disease risk in midlife. Prev Med 2019 Jul;124:33-41. [doi: 10.1016/j.ypmed.2019.04.023] [Medline: 31051183]
3. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
4. Pearce M, Garcia L, Abbas A, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022 Jun 1;79(6):550-559. [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
5. Al-Eidan RM, Al-Khalifa H, Al-Salman AM. A review of wrist-worn wearable: sensors, models, and challenges. J Sens 2018 Dec 19;2018:1-20. [doi: 10.1155/2018/5853917]
6. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020 Jun 26;3:90. [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]

7. Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384. [doi: 10.2196/34384] [Medline: 35076409]
8. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. Improving energy expenditure estimates from wearable devices: a machine learning approach. J Sports Sci 2020 Jul;38(13):1496-1505. [doi: 10.1080/02640414.2020.1746088] [Medline: 32252598]
9. Davoudi A, Wanigatunga AA, Kheirkhahan M, et al. Accuracy of Samsung Gear S smartwatch for activity recognition: validation study. JMIR Mhealth Uhealth 2019 Feb 6;7(2):e11270. [doi: 10.2196/11270] [Medline: 30724739]
10. Biddle SJH, Bengoechea EG, Pedisic Z, Bennie J, Vergeer I, Wiesner G. Screen time, other sedentary behaviours, and obesity risk in adults: a review of reviews. Curr Obes Rep 2017 Jun;6(2):134-147. [doi: 10.1007/s13679-017-0256-9] [Medline: 28421472]
11. Pope ZC, Barr-Anderson DJ, Lewis BA, Pereira MA, Gao Z. Use of wearable technology and social media to improve physical activity and dietary behaviors among college students: a 12-week randomized pilot study. Int J Environ Res Public Health 2019 Sep 25;16(19):3579. [doi: 10.3390/ijerph16193579] [Medline: 31557812]
12. Gonzalez-Plaza E, Bellart J, Arranz Á, Luján-Barroso L, Mirasol EC, Seguranyes G. Effectiveness of a step counter smartband and midwife counseling intervention on gestational weight gain and physical activity in pregnant women with obesity (Pas and Pes study): randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 15;10(2):e28886. [doi: 10.2196/28886] [Medline: 35166684]
13. Fuller D, Anaraki JR, Simango B, et al. Predicting lying, sitting, walking and running using Apple Watch and Fitbit data. BMJ Open Sport Exerc Med 2021 Apr 8;7(1):e001004. [doi: 10.1136/bmjsem-2020-001004] [Medline: 33907628]
14. O'Driscoll R, Turicchi J, Hopkins M, et al. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: validation study. JMIR Mhealth Uhealth 2021 Aug 4;9(8):e23938. [doi: 10.2196/23938] [Medline: 34346890]
15. Sucerquia A, López JD, Vargas-Bonilla JF. Real-life/real-time elderly fall detection with a triaxial accelerometer. Sensors (Basel) 2018 Apr 5;18(4):1101. [doi: 10.3390/s18041101] [Medline: 29621156]
16. Mesquita R, Spina G, Pitta F, et al. Physical activity patterns and clusters in 1001 patients with COPD. Chron Respir Dis 2017 Aug;14(3):256-269. [doi: 10.1177/1479972316687207] [Medline: 28774199]
17. Stenholm S, Pulakka A, Leskinen T, et al. Daily physical activity patterns and their association with health-related physical fitness among aging workers-the Finnish Retirement and Aging Study. J Gerontol A Biol Sci Med Sci 2021 Jun 14;76(7):1242-1250. [doi: 10.1093/gerona/glaa193] [Medline: 32766774]
18. Albalak G, Stijntjes M, Wijsman CA, et al. Timing of objectively-collected physical activity in relation to body weight and metabolic health in sedentary older people: a cross-sectional and prospective analysis. Int J Obes (Lond) 2022 Mar;46(3):515-522. [doi: 10.1038/s41366-021-01018-7] [Medline: 34782736]
19. Jones PJ, Catt M, Davies MJ, et al. Feature selection for unsupervised machine learning of accelerometer data physical activity clusters - a systematic review. Gait Posture 2021 Oct;90:120-128. [doi: 10.1016/j.gaitpost.2021.08.007] [Medline: 34438293]
20. Dobbins C, Rawassizadeh R. Towards clustering of mobile and smartwatch accelerometer data for physical activity recognition. Informatics 2018;5(2):29. [doi: 10.3390/informatics5020029]
21. Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
22. Kim Y, Wijndaele K, Sharp SJ, et al. Specific physical activities, sedentary behaviours and sleep as long-term predictors of accelerometer-measured physical activity in 91,648 adults: a prospective cohort study. Int J Behav Nutr Phys Act 2019 May 7;16(1):41. [doi: 10.1186/s12966-019-0802-9] [Medline: 31064403]
23. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digit Health 2020 Feb;2(2):e85-e93. [doi: 10.1016/S2589-7500(19)30222-5] [Medline: 33334565]
24. Sardá-Espinosa A. Time-series clustering in R using the dtwclust package. The R Journal 2019;11(1):22. [doi: 10.32614/RJ-2019-023]
25. Begum N, Ulanova L, Dau HA, Wang J, Keogh EJ. A general framework for density based time series clustering exploiting a novel admissible pruning strategy. arXiv. Preprint posted online on Dec 2, 2016. [doi: 10.48550/arXiv.1612.00637]
26. Huh U, Tak YJ, Song S, et al. Feedback on physical activity through a wearable device connected to a mobile phone app in patients with metabolic syndrome: pilot study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e13381. [doi: 10.2196/13381] [Medline: 31215513]
27. Sardá-Espinosa A. Package 'dtwclust'. The Comprehensive R Archive Network. 2023 Feb 28. URL: cran.r-project.org/web/packages/dtwclust/dtwclust.pdf [accessed 2023-11-15]
28. RStudio. Posit. URL: posit.co/products/open-source/rstudio/ [accessed 2023-11-15]
29. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. Med Sci Sports Exerc 2000 Sep;32(9):1601-1609. [doi: 10.1097/00005768-200009000-00013] [Medline: 10994912]

30. To QG, Stanton R, Schoeppe S, Doering T, Vandelanotte C. Differences in physical activity between weekdays and weekend days among U.S. children and adults: cross-sectional analysis of NHANES 2011-2014 data. Prev Med Rep 2022 Aug;28:101892. [doi: 10.1016/j.pmedr.2022.101892] [Medline: 35813401]
31. Malone SK, Patterson F, Grunin L, et al. Habitual physical activity patterns in a nationally representative sample of U.S. adults. Transl Behav Med 2021 Mar 16;11(2):332-341. [doi: 10.1093/tbm/ibaa002] [Medline: 31985811]
32. Rossen J, Hagströmer M, Larsson K, Johansson UB, von Rosen P. Physical activity patterns among individuals with prediabetes or type 2 diabetes across two years-a longitudinal latent class analysis. Int J Environ Res Public Health 2022 Mar 19;19(6):3667. [doi: 10.3390/ijerph19063667] [Medline: 35329362]
33. Montaruli A, Castelli L, Mulè A, et al. Biological rhythm and chronotype: new perspectives in health. Biomolecules 2021 Mar 24;11(4):487. [doi: 10.3390/biom11040487] [Medline: 33804974]
34. Henson J, Rowlands AV, Baldry E, et al. Physical behaviors and chronotype in people with type 2 diabetes. BMJ Open Diabetes Res Care 2020 Jul;8(1):e001375. [doi: 10.1136/bmjdrc-2020-001375] [Medline: 32675292]
35. Lotti S, Pagliai G, Colombini B, Sofi F, Dinu M. Chronotype differences in energy intake, cardiometabolic risk parameters, cancer, and depression: a systematic review with meta-analysis of observational studies. Adv Nutr 2022 Feb 1;13(1):269-281. [doi: 10.1093/advances/nmab115] [Medline: 34549270]
36. Zhang R, Cai X, Lin C, et al. The association between metabolic parameters and evening chronotype and social jetlag in non-shift workers: a meta-analysis. Front Endocrinol (Lausanne) 2022 Nov 21;13:1008820. [doi: 10.3389/fendo.2022.1008820] [Medline: 36479212]
37. Lee IM, Sesso HD, Oguma Y, Paffenbarger RSJ. The “weekend warrior” and risk of mortality. Am J Epidemiol 2004 Oct 1;160(7):636-641. [doi: 10.1093/aje/kwh274] [Medline: 15383407]
38. Jang YS, Joo HJ, Jung YH, Park EC, Jang SY. Association of the “weekend warrior” and other physical activity patterns with metabolic syndrome in the South Korean population. Int J Environ Res Public Health 2022 Oct 18;19(20):13434. [doi: 10.3390/ijerph192013434] [Medline: 36294014]
39. O’Donovan G, Sarmiento OL, Hamer M. The rise of the "weekend warrior". J Orthop Sports Phys Ther 2018 Aug;48(8):604-606. [doi: 10.2519/jospt.2018.0611] [Medline: 30064334]
40. Xiao J, Chu M, Shen H, et al. Relationship of “weekend warrior” and regular physical activity patterns with metabolic syndrome and its associated diseases among Chinese rural adults. J Sports Sci 2018 Sep;36(17):1963-1971. [doi: 10.1080/02640414.2018.1428883] [Medline: 29345215]

Abbreviations

DBSCAN: Density-Based Spatial Clustering of Applications With Noise
E-type: evening type
EE: energy expenditure
M-type: morning type
MET: metabolic equivalent
N-type: intermediate type
TADPole: Time-Series Anytime Density Peak

Edited by L Buis; submitted 08.07.23; peer-reviewed by D Gartner, T Wah; revised version received 18.10.23; accepted 19.10.23; published 01.12.23.

Please cite as:
Kim J, Choi JY, Kim H, Lee T, Ha J, Lee S, Park J, Jeon GS, Cho SI
Physical Activity Pattern of Adults With Metabolic Syndrome Risk Factors: Time-Series Cluster Analysis
JMIR Mhealth Uhealth 2023;11:e50663
URL: https://mhealth.jmir.org/2023/1/e50663
doi:10.2196/50663

Original Paper

Smartphone-Tracked Digital Markers of Momentary Subjective Stress in College Students: Idiographic Machine Learning Analysis

George Aalbers[1,2], BSc, MSc; Andrew T Hendrickson[1], PhD; Mariek MP Vanden Abeele[2,3], PhD; Loes Keijsers[4], PhD

[1]Department of Cognitive Science & Artificial Intelligence, Tilburg University, Tilburg, Netherlands

[2]Department of Communication and Cognition, Tilburg University, Tilburg, Netherlands

[3]Media, Innovation and Communication Technologies, Department of Communication Sciences, Ghent University, Ghent, Belgium

[4]Clinical Child and Family Studies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

Corresponding Author:
George Aalbers, BSc, MSc
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Warandelaan 2
Tilburg, 5037 AB
Netherlands
Phone: 31 13 466 9111
Email: h.j.g.aalbers@tilburguniversity.edu

Abstract

Background: Stress is an important predictor of mental health problems such as burnout and depression. Acute stress is considered adaptive, whereas chronic stress is viewed as detrimental to well-being. To aid in the early detection of chronic stress, machine learning models are increasingly trained to learn the quantitative relation from digital footprints to self-reported stress. Prior studies have investigated general principles in population-wide studies, but the extent to which the findings apply to individuals is understudied.

Objective: We aimed to explore to what extent machine learning models can leverage features of smartphone app use log data to recognize momentary subjective stress in individuals, which of these features are most important for predicting stress and represent potential digital markers of stress, the nature of the relations between these digital markers and stress, and the degree to which these relations differ across people.

Methods: Student participants (N=224) self-reported momentary subjective stress 5 times per day up to 60 days in total (44,381 observations); in parallel, dedicated smartphone software continuously logged their smartphone app use. We extracted features from the log data (eg, time spent on app categories such as messenger apps and proxies for sleep duration and onset) and trained machine learning models to predict momentary subjective stress from these features using 2 approaches: modeling general relations at the group level (*nomothetic approach*) and modeling relations for each person separately (*idiographic approach*). To identify potential digital markers of momentary subjective stress, we applied explainable artificial intelligence methodology (ie, Shapley additive explanations). We evaluated model accuracy on a person-to-person basis in out-of-sample observations.

Results: We identified prolonged use of messenger and social network site apps and proxies for sleep duration and onset as the most important features across modeling approaches (nomothetic vs idiographic). The relations of these digital markers with momentary subjective stress differed from person to person, as did model accuracy. Sleep proxies, messenger, and social network use were heterogeneously related to stress (ie, negative in some and positive or zero in others). Model predictions correlated positively and statistically significantly with self-reported stress in most individuals (*median person-specific correlation*=0.15-0.19 for nomothetic models and *median person-specific correlation*=0.00-0.09 for idiographic models).

Conclusions: Our findings indicate that smartphone log data can be used for identifying digital markers of stress and also show that the relation between specific digital markers and stress differs from person to person. These findings warrant follow-up studies in other populations (eg, professionals and clinical populations) and pave the way for similar research using physiological measures of stress.

(JMIR Mhealth Uhealth 2023;11:e37469) doi:10.2196/37469

KEYWORDS

mobile health; mobile phone; digital phenotype; digital biomarker; machine learning; personalized models

Introduction

Background

Stress is an important predictor of mental health problems such as burnout [1] and depression [2]. How stress influences mental health depends on its duration. Stress with a duration of minutes to hours (*acute stress*) is commonly considered an adaptive psychophysiological response, whereas stress lasting for weeks to months or even years (*chronic stress*) is believed to have adverse psychological and physiological consequences [3]. Given its potential effects, early detection and treatment of chronic stress is important to prevent mental health problems.

As asking individuals to consistently self-monitor and self-report stress over extended periods is a difficult, costly, and time-intensive procedure [4], researchers have started developing algorithms to unobtrusively detect stress from passively logged data, such as smartphone app use log data [5]. If successful, such algorithms open opportunities for early detection of the tipping point where acute stress turns into chronic stress, possibly unlocking earlier possibilities for scalable interventions, such as smartphone-based cognitive behavioral therapy with chatbots [6].

Previous research suggests that smartphone use log data (among other passively logged data sources) might be used to recognize how stressed a person feels [5,7-10], with pioneering studies in college students using call and SMS text messaging log data (among others) as predictors [9,10]. However, following technological advances that made smartphones more powerful, use of these devices has evolved, and college students now typically access their smartphone for activities other than calling or SMS text messaging, such as using social media [11]. As a result, more research is required to identify whether stress might be recognized using smartphone app use patterns with high relevance to the current generation of students. We, therefore, extend previous work in the domain [5,7-10] to a large sample of contemporary college students and explore to what extent machine learning models can leverage features of smartphone app use to recognize momentary subjective stress and which of these app use features are most important for predicting momentary subjective stress and represent potential *digital markers* of stress. Inspired by recent findings in clinical psychology [11,12] and communication science [13-16], we then shed light on the nature of the relations between these potential digital markers and stress and the degree to which these relations differ from one person to another. Following related machine learning research on mood recognition [17], we also assess how important digital markers are relative to temporal features: time of day, day of the week, day of the month, and before COVID-19 versus during COVID-19 lockdown.

Digital Markers

Digital markers are digital footprints, such as features of smartphone use log data, that are related to psychological or biological states [18], such as stress. Such features might represent any quantification of raw log data of digital devices, ranging from simple (eg, time spent using a device) [19] to more complex (eg, daily life patterns derived from device use) [20]. In this study, we investigate two types of potential digital markers of stress: (1) use of different types of smartphone apps (eg, duration and frequency of social network, messenger, and video apps) and (2) sleep proxies derived from smartphone app log data.

Smartphone Use Behaviors as Potential Digital Markers

Smartphones enable individuals to perform a wide range of behaviors with profound psychological meaning and relevance to momentary subjective stress. Extant evidence, however, suggests a complex and nuanced relation between smartphone use and mental health, with different types of app use showing different patterns of association. For example, smartphone use behaviors with particular theoretical relevance to stress are calling, mobile messaging, and using social media, but recent work on passively logged data found depression relates negatively to calling [21], whereas it relates positively to social media use [22]. Moreover, research suggests that relations between smartphone app use and mental health differ from person to person [12-16]. Altogether, with respect to the association between smartphone use and mental health, these and other findings [23,24] indicate that smartphone behaviors could be informative about stress, but open questions remain.

Sleep Proxies as Potential Digital Markers

Smartphone log data not only captures smartphone behavior with relevance to stress but might also be used to quantify sleep, which is a universal behavior related to stress [25]. As the human sleep-wake cycle closely aligns with smartphone app use patterns, different disciplines have leveraged smartphone log data (eg, call records, screen on-off status, and screen taps) to estimate sleep onset, offset, and duration [20]. To explore if such sleep proxies might be useful for stress recognition, we applied a rule-based algorithm (similar but not identical to a recent study by Massar et al [20]) to extract proxies for sleep duration and sleep onset from smartphone app log data and included these as features in our models.

Explainable Artificial Intelligence

One potential avenue to identify digital markers is to (1) train machine learning models to find a mathematical mapping from digital markers to self-reports of momentary subjective stress and (2) apply explainable artificial intelligence (eg, Shapley values [26]) to clarify how these models make predictions. Applying explainable artificial intelligence is necessary because the structure of some powerful models (eg, random forest [RF]) preclude the straightforward interpretation of parameters that linear statistical models have.

By taking this approach, we aim (1) to identify digital markers by testing which features of smartphone use log data a model uses to predict momentary subjective stress and (2) to understand the nature of the relation between these digital markers and stress. For instance, time spent on messenger apps might be important for the prediction of stress and negatively related to stress, suggesting that individuals tend to spend less time on these apps when feeling stressed. This could potentially indicate that stress reduces social interaction or vice versa.

Nomothetic Versus Idiographic

When training machine learning models on human-subjects data, it is important to take into account that individuals differ from another. Hence, both machine learning [27] and behavioral scientists [28] have underlined the importance of ensuring that models of human-subjects data are applicable to the individuals they pertain to. This is certainly also important in the domain of digital markers of stress. Behavioral scientists have shown that relations between digital trace data and psychological self-reports differ across individuals [19]. In parallel, machine learning researchers have demonstrated that personalized models, which are known as idiographic models in behavioral science, tend to predict subjective stress more accurately than nonpersonalized models [27], which are also referred to as nomothetic models. Naturally, these findings go hand in hand: a personalized (idiographic) model will more adequately capture person-specific dependencies between digital trace data and stress and therefore should make more accurate predictions than a nonpersonalized (nomothetic) model. In this study, we implement, evaluate, and compare both approaches.

Objectives

This study has four complementary aims to explore: (1) to what extent machine learning models can leverage features of smartphone app use log data to recognize momentary subjective stress in individuals, (2) which of these features are most important for predicting stress and represent potential digital markers of stress, (3) the nature of the relations between these digital markers and stress, and (4) the degree to which these relations differ from one person to another.

Methods

Participants

We followed reporting guidelines recommended for experience sampling studies [29]. For a preregistered data collection [30], we used the university participant pool to recruit 247 student participants with an Android operating system on their primary phone, 224 (90.7%) of whom we included for analysis. Their average age was 21.97 (SD 3.04) years and the majority were female (125/224, 55.8%). We excluded (1) participants with operating systems other than Android on their primary phone and (2) participants with insufficient survey responses for training idiographic machine learning models (<6). Most participants (186/224, 83%) started participation before the first (reported) local infection of SARS-CoV-2 and a minority (38/224, 17%) after. With the exception of 2 participants, all participants had been active in our study before the first nation-wide lockdown. Throughout the study, the original (Wuhan) strain of SARS-CoV-2 was dominant.

Procedure

Ethics Approval

Ethics approval was issued by the Tilburg University Ethics Committee (approval code REDC 2019/94c).

Onboarding

Participants were recruited through the university participant pool. After receiving web-based information through Qualtrics (Qualtrics XM), having been offered a possibility to ask questions, and signing an informed consent form, participants followed web-based instructions to install 2 apps on their smartphone. After completing these instructions, the participants attended an onboarding session in which we provided additional information, offered further opportunity to ask questions, and motivated the participants to participate to the best of their ability.

Technology

All participants installed 2 apps on their Android device: Ethica Data [31] and mobileDNA [32]. Ethica Data is an app that prompts participants to complete brief surveys on their smartphone (ie, *experience sampling*). MobileDNA is an app that unobtrusively logs a person’s smartphone app use, smartphone notifications, and location (ie, *passive logging*).

Sampling Scheme

The data collected in this study are part of a larger research project with other questions that required a so-called measurement burst design (1 period of intensive data collection followed by a break followed by another period of intensive data collection) [33]. The important advantage of this procedure is that it reduces participant burden, while capturing information on a larger timescale. Hence, in a 4-month period, Ethica notified the participants 5 times a day for a maximum of 60 days (30 days in month 1 and 30 days in month 4) at pseudorandom times between 8:30 AM and 10:30 PM to complete a 10-item survey (approximately 1 minute to complete) on stress and other constructs (fatigue, procrastination, and mood), whereas mobileDNA continuously logged smartphone app use. Following an initial push notification, each survey was available to the participant for 50 minutes. After 45 minutes, they received a reminder notification. After 50 minutes, the survey expired. The participants were allowed to catch up on 1 missed survey per day by starting and completing a new survey.

Monitoring Protocol

We actively monitored participant compliance and motivated participants with weekly emails containing personalized feedback. When a participant failed to complete many consecutive surveys, we sent an email to inquire why they could not comply with the study protocol and how any issues might be resolved. In a limited number of cases, the participants did not respond to such emails, in which case we contacted them through a phone call. The participants were compensated with course credits for research participation and were entered into a raffle comprising 20 prizes of 15 euros (US $22.06).

Compliance

The participants (N=224) completed a total of 44,381 surveys (198 per person on average). Though data collection spanned the introduction of the COVID-19 pandemic lockdown, the median participant remained in the study across 4 months. In the sample we analyzed, the median participant completed 205 surveys (SD 87.42) and had 3072 hours of smartphone use log data. The median time difference between receiving the initial notification and completing the survey was 6 minutes (SD 14.26; refer to Figure 1 for a histogram). Reasons for noncompliance ranged from technical difficulties (eg, not receiving any notifications and broken or lost smartphone) to not being able to complete a survey (eg, waking up too late and receiving a notification during work or lecture) to personal reasons (eg, attrition because of COVID-19 pandemic–related personal problems or collecting sufficient course credits). Figure 2 visualizes how compliance rate changed from the first to the last day of the study.

Figure 1. Distribution of the time difference between receiving the initial notification and completing the survey (in minutes).

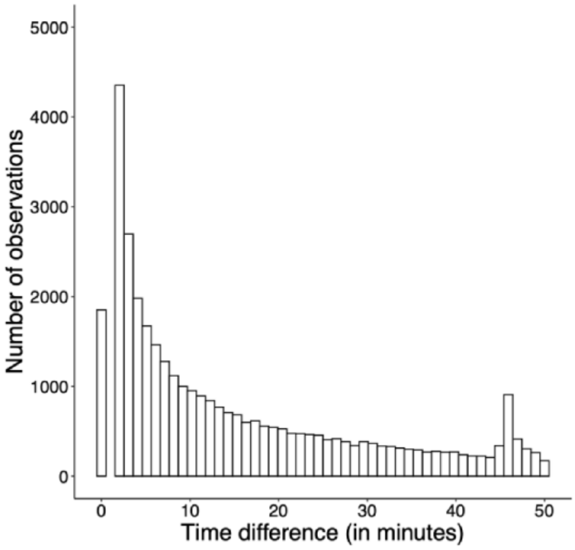
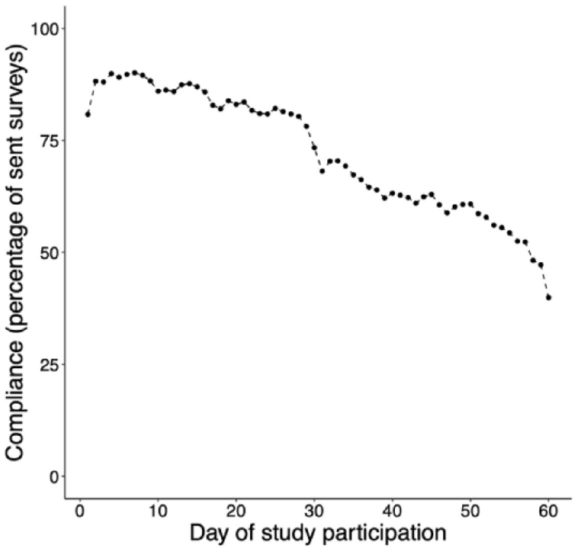


Figure 2. Daily compliance rate over time. To determine daily compliance rate, we computed the percentage of survey responses that were received on a given participation day relative to the total number of surveys to be sent each day based on our original sample size (224×5=1120). If, for instance, all participants completed 4 out of 5 surveys on their first day of participation, daily compliance equals 80%.



Measures

To measure participants' current (ie, in-the-moment) subjective level of stress, we used the Stress Experience Sampling Scale [34]. This scale consists of 2 items on a 7-point Likert scale ranging from 1 (not at all) to 4 (moderately) to 7 (very much): "Right now, I feel relaxed" and "Right now, I feel stressed (tense, restless, nervous or anxious)." The 2 items have an adequate intraclass coefficient (>50% of variance because of within-person fluctuations) and acceptable within-person reliability as assessed by within-person omega (ω=.71). We calculated an unweighted average of the 2 items and subtracted each participant's average level of stress from the resulting values (ie, within-person centering).

Textbox 1 provides an overview of the features included in this study. We analyzed three categories of features: (1) smartphone use behavior, (2) sleep, and (3) time. As the raw values of these features are on vastly different scales, which can dramatically impact model performance, we scaled all features to a range between 0 and 1 using MinMaxScaler in *sklearn* [35], based on the minimum and maximum values in the training data.

To extract time spent on different smartphone app categories, we first categorized apps using a coding scheme that maps app names (eg Whatsapp [36]) to 1 of the 18 major categories (Textbox 2; these core categories represent 4,046,581/5,277,494, 76.68% of all app events). Then, during the 60 minutes before each self-report of stress, we calculated (1) the total time spent on all apps in a category (*duration*) and (2) the total number of times apps in a category had been accessed (*frequency*).

To extract sleep duration and sleep onset proxies from the raw smartphone app log timestamps, we used an algorithm similar (but not identical) to a previously validated rule-based algorithm [11] (refer to Multimedia Appendix 1 for a description of our approach), an algorithm that was found to be strongly associated with actigraphy-based and self-reported sleep duration and onset.

To extract time features (ie, hour of the day, day of the week, day of the month, and lockdown status) from the raw self-report timestamps, we used pandas.datetime in Python 3.9.9 [37]. In the interest of model simplicity, we recoded day of the week into a binary variable (weekday=0 and weekend=1) rather than treating this variable as a categorical variable. We further coded lockdown status as a binary variable (before COVID-19 lockdown=0 and during COVID-19 lockdown=1) based on the lockdown timing in the Netherlands.

Textbox 1. Overview of the features included in models.

Smartphone use behavior

- Time (seconds) spent on smartphone app category X in the past 60 minutes
- Frequency (count) of opening smartphone app category X in the past 60 minutes

Sleep

- Sleep onset (hours and postmidnight hours >24)
- Sleep duration (hours)

Time

- Hour of the day (0 to 23 and starting at midnight)
- Day of the week (0=weekday and 1=weekend)
- Day of the month (0 to 31)
- COVID-19 (before lockdown=0 and during lockdown=1)

Textbox 2. Smartphone app categories and examples of apps per category.

- Browser: Chrome and Opera
- Calling: default dial apps
- Camera: default camera apps
- Dating: Tinder and Grindr
- Email: Gmail and Outlook
- Exercise: RunKeeper
- Food and drink: UberEATS
- Gallery: default gallery apps
- Game: CandyCrush
- Messenger: WhatsApp
- Music and audio: Spotify
- Productivity: Microsoft Word
- Shared transportation: 9292OV (Dutch public transport)
- Social network: Facebook, Instagram, and Twitter
- Tracker: pedometer apps
- Video: YouTube and Netflix
- Weather: default weather apps
- Work: StudentJob and EmployeeApp

Stress Recognition Models

Model Types

We trained 3 different types of machine learning models: least absolute shrinkage and selection operator (LASSO) regression, support vector regression (SVR), and RF regression. We trained all models using a nomothetic and an idiographic approach (explained in the section Model Cross-Validation). For brevity and clarity, we prepend an "N" to the abbreviations of our nomothetic models (ie, N-LASSO, N-SVR, and N-RF) and an "I" to the abbreviations of our idiographic models (ie, I-LASSO, I-SVR, and I-RF).

Model Cross-Validation

Generally, when we perform cross-validation (CV), we (1) split the data into train and test data, (2) specify a range of values for a model's different hyperparameters (ie, researcher-specified parameters that control model complexity), (3) identify the best hyperparameters per model using *k*-fold CV within the training set, and (4) evaluate each model based on predictive accuracy for the test data. In what follows, we outline how we cross-validated nomothetic and idiographic models.

To train nomothetic models, we applied a user split to distinguish between train and test data. We first used group *k*-fold CV (GroupKFold in *sklearn* [35]) to partition the data into 5 subsets. A total of 4 subsets contained data from 45 participants, and 1 subset contained data from 44 participants. We then selected 4 subsets (*train data*) to train a model. For training the model, we used 5-fold grid search CV (GridSearchCV in *sklearn* [35]) to minimize each model's default *sklearn* error metric (refer to Multimedia Appendix 2 for tuned hyperparameters and minimized error metrics). After training the model, we let the model make predictions on the data subset we did not include in training (*test data*; ie, all observations of participants excluded from training). Finally, we evaluated the accuracy of these predictions. We repeated this process until each subset of the data had been left out of training once and did so for all models.

To train idiographic models, we applied a time split to distinguish between train and test data. We iteratively selected 1 participant's data to train and test models only on these data. For each person, we assigned each participant's first 80% observations to a train data set and their final 20% observations to a test data set and trained each model (SVR, RF, and LASSO). We applied 5-fold grid search CV to each participant's train data to optimize hyperparameters. As the number of idiographic models to train is much larger, we applied 5-fold randomized search CV rather than grid search CV for training RFs, which are more computationally expensive than LASSO and SVR. Grid search CV always uses a larger number of hyperparameters than randomized search CV because the latter trains models on a random subset of all the hyperparameter settings used by the former. Randomized search CV considerably speeds up training time and makes training a large number of RF models more feasible. Finally, we let trained models make predictions on the individual's test data (ie, final 20% observations) and evaluated the accuracy of these predictions.

Model Evaluation

To evaluate the accuracy of models, we used Spearman rho rank-order correlation and mean absolute error (MAE) as evaluation metrics. Spearman rho rank-order correlation indicates whether a model tends to predict greater values when an individual feels more stressed without assuming a linear relation. We consider a model to perform above chance for a given individual if the Spearman ρ between predictions and self-reports between predictions and self-reports has a *P* value below .05.

Lower MAE values indicate a more accurate model. We consider the MAE to be an intuitive metric to assess predictive accuracy on a target variable measured on a 7-point Likert scale, as it allows us to make statements such as "on average, the model mispredicts momentary subjective stress by ±0.80 points on a 7-point Likert scale)." To evaluate if our models perform better than random guessing, we compare against the MAEs of a "naive" but person-specific baseline model. This "naive" baseline model always predicts an individual's average level of stress.

Model Explanation

One of the challenges of machine learning approaches is to understand the results, as they are more complex and therefore less intuitive than standard statistical approaches. To explain models, we use the Shapley additive explanations (SHAP) library [26] implemented in Python 3.9.9 [37] to calculate and visualize Shapley values. Shapley values can be used to determine (1) which features are most important in a model and (2) how features are related to model predictions.

Results

Recognizing Momentary Subjective Stress

Table 1 provides an overview of how accurately nomothetic models predict out-of-sample data. Model predictions correlated positively and significantly with self-reports in a majority of participants for N-LASSO (116/224, 51.3%), N-SVR (120/224, 53.6%) and N-RF, with N-RF performing best in terms of percent significant results (124/224, 55.4%). The median correlation between predictions and self-reports was weak for all models (between 0.15 and 0.19). Correlations also differed between participants: in 60 participants, the positive correlation was moderate or larger (ρ>0.3), whereas it was negative (range *P*<.001 to *P*=.049) in 2.2% (5/224) of people. In the median participant, models on average mispredicted momentary subjective stress by approximately 0.8 points on a 7-point Likert scale (MAE of 0.84; scale range: 1="Not at all," 4="Moderately," and 7="Very much"). The MAE of nomothetic models varied across individuals, but for 89.3% (200/224) of participants, the person-specific baseline outperformed all nomothetic models (refer to the sixth column in Table 1), followed by N-LASSO (20/224, 8.9%) and N-SVR (4/224, 1.8%). N-RF did not outperform the other models in terms of MAE.

Thus, nomothetic models make predictions that weakly and positively correlate with actual stress self-reports for the majority of participants (up to 124/224, 55.3%). This means that when these participants, who were not included in the training data set, feel stressed, the model tends to output a higher value, and when a participant does not feel stressed, the model tends to output a lower value. However, these models are not highly accurate, as they often do not outperform a person-specific baseline. For most participants, the association between predictions and actual self-reported stress is positive, but it is significantly negative (range $P<.001$ to $P=.049$) for a very small group (5/224, 2.2%). Thus, for a very small group of participants, the model tends to detect subjective stress when the person does not feel stressed, and vice versa.

We then evaluated how well idiographic models recognize momentary subjective stress in out-of-sample observations. For all models, model predictions of momentary subjective stress correlated positively and significantly with self-reports of momentary subjective stress, but only in a minority of participants. In 60 participants, this correlation was moderate or larger (ρ>0.3), whereas the overall median correlation was absent to weak (range of median correlations per model, ρ=0.00-0.09). In a minority (11/224, 4.9%), model predictions of stress and self-reported stress were significantly negatively associated. Similar to the nomothetic models, the median person-specific MAE for each idiographic model was slightly >0.8 points. These MAE scores are compared with a person-specific baseline based on the person-specific average level of stress in the participant's train data (ie, first 80%) rather than all their data to prevent data leakage from train to test data. MAE varied from individual to individual, but for 80.4% (180/224) of participants at least one of the idiographic models outperformed the person-specific baseline. The I-SVR most frequently outperformed all other models (including person-specific baseline; 92/224, 41.1%), followed by I-RF (47/224, 20.9%) and I-LASSO (43/224, 19.2%).

Table 1. Central tendency and range of out-of-sample predictive accuracy (Spearman ρ correlation and mean absolute error [MAE]) for nomothetic and idiographic models on a person-by-person basis.

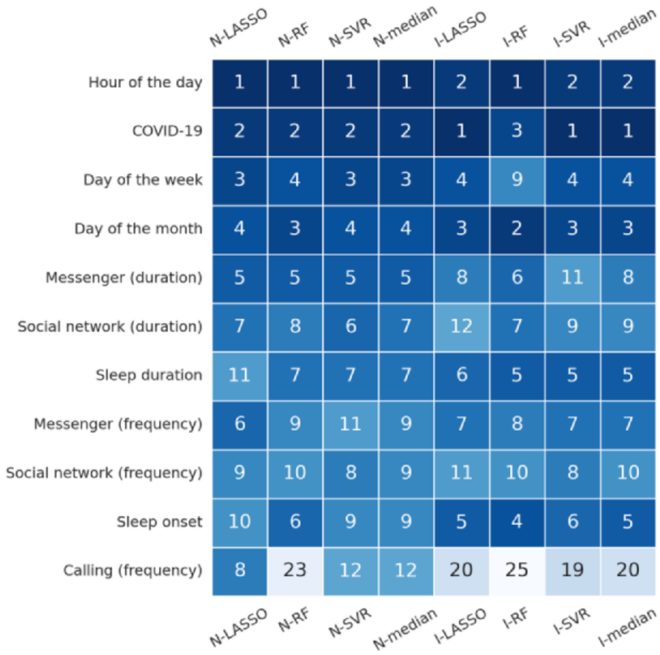
Model	Spearman ρ rank-order correlation		MAE		
	Median (range)	Percent significant	Median (range)	Better than baseline (%)	Best model (%)
N[a]-baseline	N/A[b]	N/A	*0.83*[c] (0 to 3.39)	N/A	*89.29*
N-LASSO[d]	0.15 (−0.45 to 0.65)	51.34	0.84 (0.11 to 2.04)	10.27	8.93
N-SVR[e]	0.16 (−0.31 to 0.64)	53.57	0.84 (0.17 to 2.04)	4.91	1.79
N-RF[f]	*0.19* (−0.37 to 0.57)	*55.34*	0.84 (0.26 to 2.04)	3.57	0
I[g]-baseline	N/A	N/A	*0.83* (0 to 3.40)	N/A	19.64
I-LASSO	0.00 (−1 to 1)	13.84	0.87 (0 to 3.36)	34.37	18.75
I-SVR	*0.09* (−0.79 to 1)	*23.21*	0.84 (0 to 3.49)	*50.45*	*41.07*
I-RF	0.08 (−1 to 1)	19.64	0.84 (0 to 3.27)	39.73	20.98

[a]N: nomothetic.
[b]N/A: not applicable.
[c]Italicized values represent best performance.
[d]LASSO: least absolute shrinkage and selection operator.
[e]SVR: support vector regression.
[f]RF: random forest.
[g]I: idiographic.

Identifying Digital Markers of Momentary Subjective Stress

To demonstrate which aspects of the model have the strongest predictive value, Figure 3 provides an overview of the 10 most important features per nomothetic and idiographic model. In all 6 models, temporal features, COVID-19 lockdown status, messenger and social network use, and sleep proxies were most important to the prediction stress. The largest disagreement between nomothetic and idiographic models was the importance of messenger and social network use relative to sleep duration and onset. The former features were more important in nomothetic models, whereas the latter were more important in idiographic models. Feature importance was relatively consistent for models across data splits (refer to Multimedia Appendix 3 for beeswarm plots per model per data split).

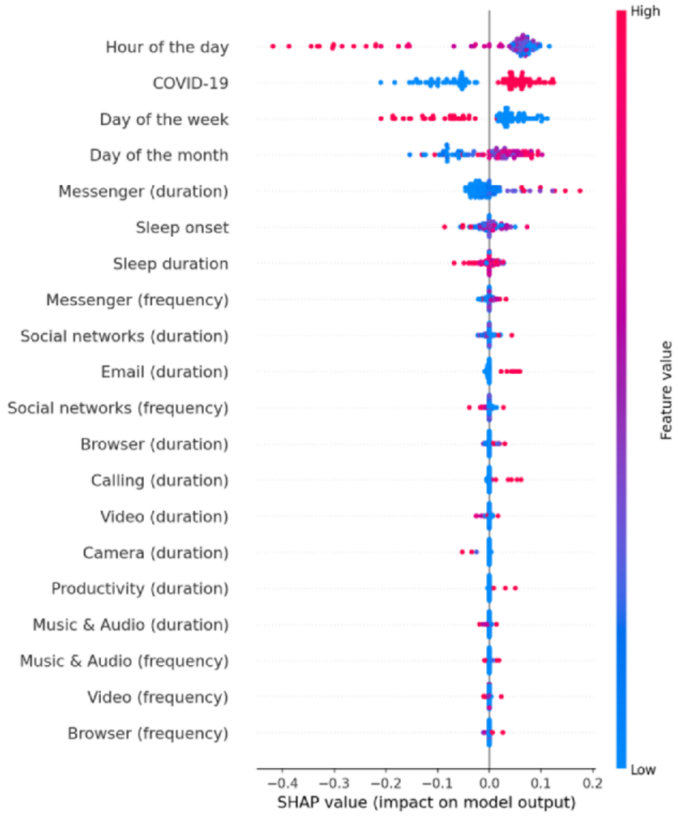
Figure 3. Feature importance ranking for each nomothetic (N) and idiographic (I) model, containing features that appeared in the top 10 features of any model. Numeric values represent the ranking of 1 feature for 1 model. Dark (top) cells represent more important features. N-median and I-median represent the median ranking of a feature across N and I models, respectively, where double values indicate a tie between 2 features. Features are ordered by N-median scores. LASSO: least absolute shrinkage and selection operator; RF: random forest; SVR: support vector regression.



Understanding Digital Markers of Momentary Subjective Stress

Feature importance provides relevant information about which features contribute most to predictions. However, it does not tell us whether small or large values of a feature are indicative of momentary subjective stress. For instance, although we have identified hour of the day as a relatively important predictor of momentary subjective stress, it is still unclear whether people tend to feel more stressed in earlier or later hours of the day. To clarify potential relations between features and momentary subjective stress, we present a beeswarm plot for the nomothetic RF (Figure 4; refer to Multimedia Appendix 3 for all beeswarm plots) to visualize how different features are related to model predictions in 1 split of the data. In this figure, for each feature (listed on the y-axis in decreasing order of importance), a point represents 1 test trial, the color of a point indicates the value of the feature (red for higher values [eg, later hour of the day] and blue for low values [eg, earlier hour of the day]), and the position along the x-axis indicates the SHAP value (positive values correspond with higher stress predictions; larger magnitude values indicate stronger impact). For any 1 feature, red points on the left side of the plot (high feature values and negative SHAP values) indicate a relation where increasing the feature value results in the model predicting lower outcome values (eg, higher hour predicts lower stress), whereas red points on the right side of the plot (high feature values and positive SHAP values) indicate a positive relation (eg, COVID-19 lockdown predicts higher stress). For instance, Figure 4 shows that N-RF predicts greater momentary subjective stress values (1) in earlier hours of the day (indicated in blue), (2) during COVID-19 lockdown, (3) on weekdays, and (4) later in the month. Similarly, when individuals spend more time on messenger (red), these models output greater values for momentary subjective stress.

Figure 4. Shapley additive explanations (SHAP) beeswarm plot indicating the relative importance of each feature and the relation between feature values and model prediction for the nomothetic random forest model. COVID-19 is coded as 0=before lockdown and 1=during lockdown. Weekday is coded as 0=weekday and 1=weekend.

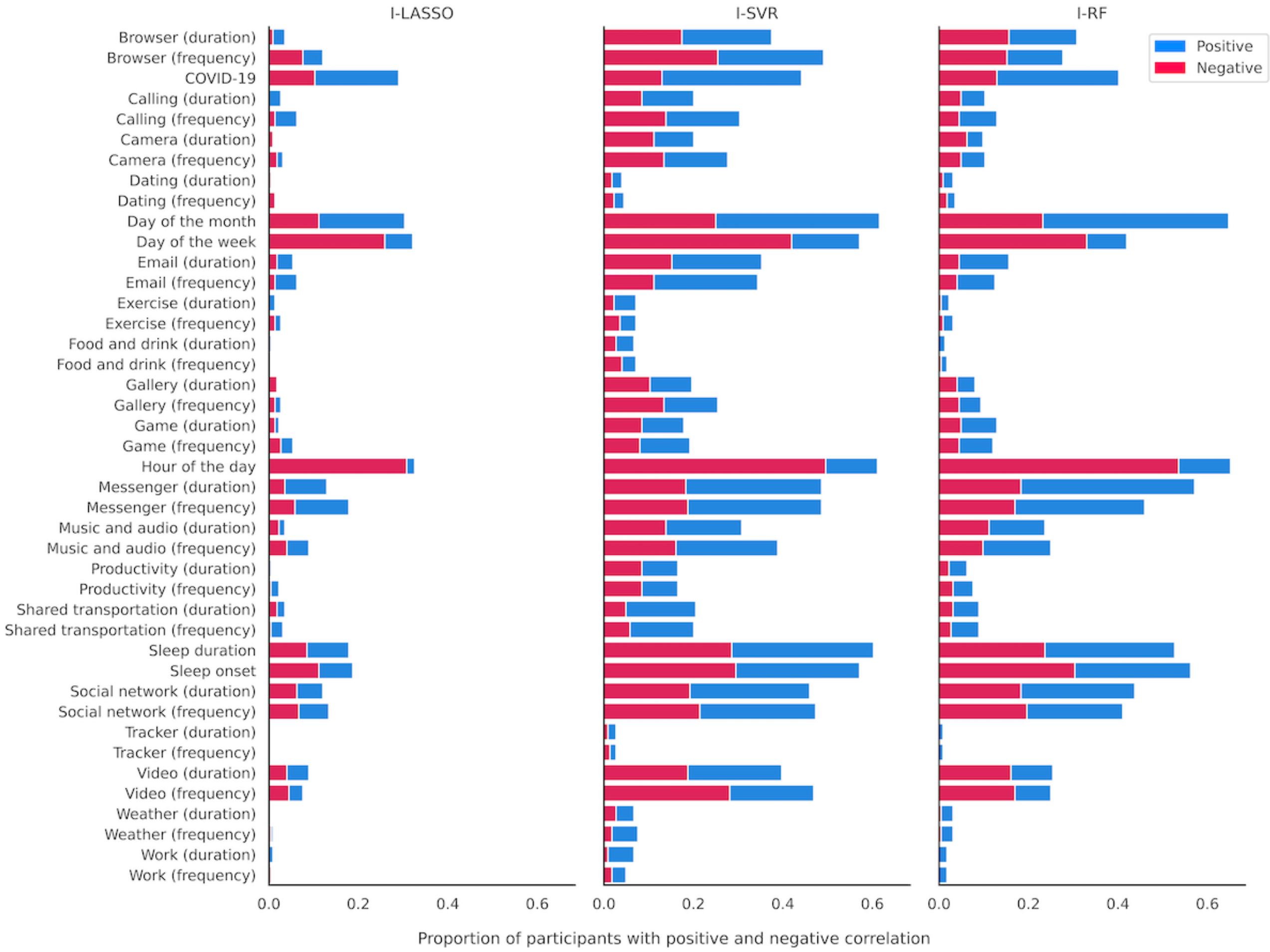


Interindividual Differences in the Relation Between Digital Markers and Stress

We also investigated whether the nature of relations between features and stress differed from person to person. As the zero-order Spearman rank-order correlations between features and stress are relatively weak (Multimedia Appendix 4), we instead calculated these correlations between feature values and SHAP values for each idiographic model. A positive correlation between feature and SHAP value indicates that when this feature has a higher value, the model predicts that an individual feels more stressed. SHAP values of more complex models have the potential benefit of capturing the nonlinear and interactive relations learned by the model. Correlations with *P* values of >.05 are not included.

Figure 5 shows the frequency of significant positive and negative correlations between digital markers and SHAP values for the prediction of stress for each idiographic model. Interestingly, most bars are both red and blue, which indicates a relatively heterogeneous relation across people between that feature and stress for most features. For instance, there is much heterogeneity in the relation between stress and some of the most important features (Figure 3), including social network use and sleep duration. Some bars are mostly red or mostly blue, suggesting a relatively homogeneous relation across people between these features and stress. For instance, the relation between stress and hour of the day is mostly negative and the relation between stress and shared transportation app use is mostly positive. A few features rarely show any correlation with predicted stress, and this is likely because that model has learned to disregard those features or because the correlation between feature values and SHAP values is either not reliable or monotonic.

Figure 5. Stacked bar plots representing the proportion of participants showing a significant positive (blue bar) or negative (red bar) correlation between feature values and Shapley additive explanations (SHAP) values for each idiographic model. A positive correlation indicates that when a given feature has a higher value, then the model predicts that an individual feels more stressed. A negative correlation indicates that when a given feature has a higher value, then the model predicts that an individual feels less stressed. For instance, the idiographic random forest (I-RF) predicts a lower level of stress at later hours of the day in 53.4% (120/224) of individuals. I-LASSO: idiographic least absolute shrinkage and selection operator; I-SVR: idiographic support vector regression.



Discussion

Principal Findings

The aims of our study were to explore (1) to what extent machine learning models can leverage features of smartphone app use log data to recognize momentary subjective stress, (2) which smartphone app use features are most important for predicting stress and represent potential digital markers of stress, (3) the nature of the relations between smartphone app use features and stress, and (4) the degree to which these relations differ from one person to another. We found that when individuals were more stressed, the best performing nomothetic models tended to predict a higher level of stress (and vice versa) in the majority of individuals (up to 124/224, 55.3%). However, they generally did not predict stress with greater accuracy than a naive baseline model, which always predicted that a person was experiencing their average level of stress. We found these results to be similar for idiographic models, although these models predicted stress with greater accuracy than a naive baseline model. Although performance should be improved to make clinical application feasible, this study does suggest that, in the absence of self-report or physiological data, digital markers can be used to recognize momentary subjective stress on a person-by-person basis in out-of-sample data.

Using explainable artificial intelligence, we found that temporal features, prolonged messenger and social network app use, and smartphone-tracked sleep proxies were the most important features of the best-performing nomothetic and idiographic models. These models consistently ordered these features, with temporal features as most important and app use as less important drivers of stress predictions. The largest disagreement between nomothetic and idiographic models is the importance of app use duration. In nomothetic models, these are more important than sleep features, but in idiographic models, the order is opposite. In sum, though, prolonged use of messenger and social network apps and sleep proxies might be valid digital markers of stress.

Our results suggest that, for an average person, in the earlier hours of the day, on weekdays, on later days of the month, and during COVID-19 lockdown (compared with before COVID-19 lockdown), individuals felt more stressed than they would usually do. Individuals also felt more stressed when spending more time on social apps (ie, messenger and social network apps). The relation between temporal features and (SHAP values for the prediction of) stress was rather consistent across individuals, suggesting these could represent universal principles for this population. These may even be explained by biological mechanisms (eg, cortisol awakening response [38]) that universally affect adolescents and young adults. Such features are likely important to include in passive tracking in the context of stress-related mental health problems such as depression and burnout.

One of the unique features of this study was to compare nomothetic and idiographic models in light of increasingly idiographic research practices in behavioral science [29]. Our findings show that model personalization is warranted especially when smartphone app use features are added to the model. That is, the relation between social app use and stress was negative for some individuals and positive or absent for others (Figure 5). We, therefore, find evidence that the idiographic approach of digital phenotyping research, that is, "the moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices, in particular smartphones" [39], is warranted in the context of stress.

In clinical practice, the added value of digital phenotyping of stress is that it might help us to understand how, for a given individual, stress is related to temporal features, how sleep impacts their stress levels, and how their stress relates to their smartphone app use. Identifying the latter relations might serve as a probe for qualitative investigation of how they respond to stress. For instance, if an individual spends less time on messenger apps when stressed, this could suggest that they avoid social contact, whereas seeking social support might be beneficial. If individuals are not aware of this pattern, personalized prediction models could provide novel clinical insights and could potentially help to improve therapy outcomes (eg, within personalized treatment modules [40]).

Finally, our exploratory findings provide important directions for confirmatory research. Contrary to the approach taken here, which is useful for discovering potential digital markers, a confirmatory approach would be to test a small number of (preregistered) hypotheses that make explicit what relation we expect between specific digital markers and momentary subjective stress. Confirmatory research is required to test the robustness of (1) the relation between stress and the potential digital markers identified here and (2) the interindividual variability in this relation.

Limitations

This study should be viewed in light of the following limitations. First, our findings have constraints on generality, as they are based on a sample of students at a small university in the Netherlands and, therefore, are more likely to generalize to student than nonstudent populations. Furthermore, we measured these individuals during the COVID-19 pandemic, when the original (Wuhan) strain of the SARS-CoV-2 virus was dominant. As results suggest that the COVID-19 pandemic and resulting lockdown affected participant stress (generally increasing stress but decreasing stress in some individuals), potentially, these results might have been different had there not been a pandemic. We encourage future research to test if our results replicate in other populations (eg, working adults or individuals diagnosed with mental disorders) and during a period with a limited SARS-CoV-2 infection rate.

Second, it is conceivable that not all students self-reported stress accurately at every assessment. A significant proportion of variance might therefore represent noise that cannot be explained by any variable or model, irrespective of modeling decisions. This is especially an issue for idiographic models that rely on the participants' final observations, which might be observations of lower quality because of study fatigue [41].

Third, we applied within-person mean centering to the self-reported stress. Although this corrects for differences in how people use a scale, it only allows nomothetic models to predict whether a person is currently experiencing more or less stress than usual and prevents models from predicting whether this person's stress level is very low or high relative to other people's stress level.

Fourth, we forced nomothetic models to learn one mapping function from features to outcome for all individuals in our sample. This is problematic because such models learn 1 set of parameters that might be accurate for some individuals but highly inaccurate for others (ie, *one-size-fits-all* fallacy) [42]. Truly idiographic models, which we also trained in this study, do not have this issue by default. However, this comes at the cost of strongly reduced sample size, which limits model complexity and may lead to overfitting. As collecting more self-report data per individual is not feasible for samples of this size, future studies could (1) focus on smaller samples with exceptionally motivated participants for a longer sampling period, (2) use wearables to measure psychophysiological signals of stress (eg, CortiWatch [43]), or (3) train machine learning models on a full data set without losing sight of interindividual differences in feature-outcome relations (eg, using transfer learning [44]).

Conclusions

Our exploratory study has 3 main conclusions. First, temporal features, sleep proxies, and prolonged use of messenger and social network apps are consistently identified as the most important digital markers for predicting momentary subjective stress. Second, for most people (200/224, 89.3%), these markers are not sufficiently informative to recognize momentary subjective stress with appreciable improvements in accuracy over a baseline model but do produce predictions that correlate with subjective stress. Third, the utility and relation with stress of (some) digital markers varies from person to person. On the one hand, 1 digital marker may be relevant to momentary subjective stress in one individual but not in another. On the other hand, the increase of 1 digital marker might imply lower stress for one individual and higher stress for another. Our study thus provides evidence for phenotypic heterogeneity in the relation between how we feel and the digital traces we leave behind. These findings are relevant for the implementation of algorithms in mobile health apps to prevent, monitor, and treat stress-related mental health problems.

Acknowledgments

The authors would like to thank the Tilburg Experience Sampling Center and the mobileDNA team at imec-mict-UGent for their assistance in setting up the digital phenotyping study. The authors thank their colleagues Ghaith Al Seirawan, Andrei Oprea, Ethel Pruss, Marieke van der Pol, and Kyle van Gaeveren for their outstanding help during data collection.

Funding was received from the Tilburg University IMPACT program, and a VIDI grant (NWO VIDI grant 452.17.011) was awarded to LK.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the algorithm used to estimate sleep duration and onset from smartphone app use log data.
[DOCX File , 16 KB - mhealth_v11i1e37469_app1.docx]

Multimedia Appendix 2
List of hyperparameters and the grid of hyperparameter values that were searched to optimize models in cross-validation.
[DOCX File , 17 KB - mhealth_v11i1e37469_app2.docx]

Multimedia Appendix 3
Shapley additive explanations (SHAP) beeswarm plots for all model types and splits of the data.
[DOCX File , 2793 KB - mhealth_v11i1e37469_app3.docx]

Multimedia Appendix 4
Stacked bar plot of zero-order correlations between features and stress on a person-by-person basis.
[DOCX File , 167 KB - mhealth_v11i1e37469_app4.docx]

References

1. Yates SW. Physician stress and burnout. Am J Med 2020 Mar;133(2):160-164 [FREE Full text] [doi: 10.1016/j.amjmed.2019.08.034] [Medline: 31520624]
2. Maddock C, Pariante CM. How does stress affect you? An overview of stress, immunity, depression and disease. Epidemiol Psichiatr Soc 2001 Jul;10(3):153-162. [doi: 10.1017/s1121189x00005285] [Medline: 11787449]
3. Rohleder N. Stress and inflammation - the need to address the gap in the transition between acute and chronic stress effects. Psychoneuroendocrinology 2019 Jul;105:164-171. [doi: 10.1016/j.psyneuen.2019.02.021] [Medline: 30826163]
4. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 2019 May;93:103151 [FREE Full text] [doi: 10.1016/j.jbi.2019.103151] [Medline: 30880254]
5. Ferdous R, Osmani V, Mayora O. Smartphone app usage as a predictor of perceived stress levels at workplace. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. 2015 Presented at: PervasiveHealth '15; May 20-23, 2015; Istanbul, Turkey p. 225-228 URL: https://ieeexplore.ieee.org/abstract/document/7349403 [doi: 10.4108/icst.pervasivehealth.2015.260192]
6. Williams R, Hopkins S, Frampton C, Holt-Quick C, Merry SN, Stasiak K. 21-day stress detox: open trial of a universal well-being chatbot for young adults. Soc Sci 2021 Oct 30;10(11):416 [FREE Full text] [doi: 10.3390/socsci10110416]
7. Maxhuni A, Hernandez-Leal P, Morales EF, Sucar LE, Osmani V, Mayora O. Unobtrusive stress assessment using smartphones. IEEE Trans Mob Comput 2021 Jun 01;20(6):2313-2325 [FREE Full text] [doi: 10.1109/tmc.2020.2974834]
8. Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: ACII' 13; September 2-5, 2013; Geneva, Switzerland p. 671-676 URL: https://affect.media.mit.edu/pdfs/13.Sano-etal-acii.pdf [doi: 10.1109/acii.2013.117]
9. Jaques N, Taylor S, Sano A, Picard R. Multi-task, multi-kernel learning for estimating individual wellbeing. In: Proceedings of the 2015 Neural Information Processing Systems Workshop on Multimodal Machine Learning. 2015 Presented at: NeurIPS '15; December 7-10, 2015; Montreal, Canada URL: https://affect.media.mit.edu/pdfs/15.Jaques-etal-NIPSMMML.pdf
10. Yu H, Klerman E, Picard RW, Sano A. Personalized wellbeing prediction using behavioral, physiological and weather data. In: Proceedings of the 2019 IEEE EMBS International Conference on Biomedical & Health Informatics. 2019 Presented at: BHI' 19; May 19-22, 2019; Chicago, IL, USA p. 1-4. [doi: 10.1109/bhi.2019.8834456]
11. Aalbers G, McNally RJ, Heeren A, de Wit S, Fried EI. Social media and depression symptoms: a network perspective. J Exp Psychol Gen 2019 Aug;148(8):1454-1462 [FREE Full text] [doi: 10.1037/xge0000528] [Medline: 30507215]

12. Rodriguez M, Aalbers G, McNally RJ. Idiographic network models of social media use and depression symptoms. Cogn Ther Res 2021 May 22;46(1):124-132 [FREE Full text] [doi: 10.1007/s10608-021-10236-2]
13. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. The effect of social media on well-being differs from adolescent to adolescent. Sci Rep 2020 Jul 01;10(1):10763 [FREE Full text] [doi: 10.1038/s41598-020-67727-7]
14. Siebers T, Beyens I, Pouwels JL, Valkenburg PM. Social media and distraction: an experience sampling study among adolescents. Media Psychol 2022;25(3):343-366 [FREE Full text] [doi: 10.1080/15213269.2021.1959350]
15. Verbeij T, Pouwels JL, Beyens I, Valkenburg PM. Experience sampling self-reports of social media use have comparable predictive validity to digital trace measures. Sci Rep 2022 May 09;12(1):7611 [FREE Full text] [doi: 10.1038/s41598-022-11510-3] [Medline: 35534600]
16. Pouwels JL, Valkenburg PM, Beyens I, van Driel II, Keijsers L. Some socially poor but also some socially rich adolescents feel closer to their friends after using social media. Sci Rep 2021 Oct 27;11(1):21176 [FREE Full text] [doi: 10.1038/s41598-021-99034-0] [Medline: 34707197]
17. Bachmann A, Klebsattel C, Budde M, Riedel T, Beigl M, Reichert M, et al. How to use smartphones for less obtrusive ambulatory mood assessment and mood recognition. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015 Presented at: UbiComp/ISWC '15 Adjunct; September 7-11, 2015; Osaka, Japan p. 693-702. [doi: 10.1145/2800835.2804394]
18. Montag C, Elhai JD, Dagum P. On blurry boundaries when defining digital biomarkers: how much biology needs to be in a digital biomarker? Front Psychiatry 2021 Sep 30;12:740292 [FREE Full text] [doi: 10.3389/fpsyt.2021.740292] [Medline: 34658973]
19. Aalbers G, vanden Abeele MM, Hendrickson AT, de Marez L, Keijsers L. Caught in the moment: are there person-specific associations between momentary procrastination and passively measured smartphone use? Mob Media Commun 2022 Jan;10(1):115-135 [FREE Full text] [doi: 10.1177/2050157921993896]
20. Massar SA, Chua XY, Soon CS, Ng AS, Ong JL, Chee NI, et al. Trait-like nocturnal sleep behavior identified by combining wearable, phone-use, and self-report data. NPJ Digit Med 2021 Jun 02;4(1):90 [FREE Full text] [doi: 10.1038/s41746-021-00466-9] [Medline: 34079043]
21. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. JMIR Ment Health 2020 Jan 24;7(1):e14045 [FREE Full text] [doi: 10.2196/14045] [Medline: 32012072]
22. Wang Y, Ren X, Liu X, Zhu T. Examining the correlation between depression and social behavior on smartphones through usage metadata: empirical study. JMIR Mhealth Uhealth 2021 Jan 06;9(1):e19046 [FREE Full text] [doi: 10.2196/19046] [Medline: 33404512]
23. Vahedi Z, Saiphoo A. The association between smartphone use, stress, and anxiety: a meta-analytic review. Stress Health 2018 Aug;34(3):347-358 [FREE Full text] [doi: 10.1002/smi.2805] [Medline: 29673047]
24. Gilbert A, Baumgartner SE, Reinecke L. Situational boundary conditions of digital stress: goal conflict and autonomy frustration make smartphone use more stressful. Mob Media Commun (forthcoming) 2022 Nov 24:205015792211380. [doi: 10.1177/20501579221138017]
25. Fischer D, McHill AW, Sano A, Picard RW, Barger LK, Czeisler CA, et al. Irregular sleep and event schedules are associated with poorer self-reported well-being in US college students. Sleep 2020 Jun 15;43(6):zsz300 [FREE Full text] [doi: 10.1093/sleep/zsz300] [Medline: 31837266]
26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NeurIPS'17; December 4-9, 2017; Long Beach, CA, USA p. 4768-4777 URL: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
27. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213 [FREE Full text] [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]
28. Bülow A, Keijsers L, Boele S, van Roekel E, Denissen JJ. Parenting adolescents in times of a pandemic: changes in relationship quality, autonomy support, and parental control? Dev Psychol 2021 Oct;57(10):1582-1596 [FREE Full text] [doi: 10.1037/dev0001208] [Medline: 34807682]
29. van Roekel E, Keijsers L, Chung JM. A review of current ambulatory assessment studies in adolescent samples and practical recommendations. J Res Adolesc 2019 Sep;29(3):560-577 [FREE Full text] [doi: 10.1111/jora.12471] [Medline: 31573762]
30. Aalbers G, Keijsers L, Abeele MV, Hendrickson AT. Preregistration of data collection. Open Science Framework. 2020 Jul 9. URL: https://osf.io/6fs92/ [accessed 2023-02-15]
31. Ethica data. Ethica. 2022. URL: https://www.ethicadata.com/ [accessed 2022-02-08]
32. Anrijs S, Bombeke K, Durnez W, Van Damme K, Vanhaelewyn B, Conradie P, et al. MobileDNA: relating physiological stress measurements to smartphone usage to assess the effect of a digital detox. In: Proceedings of the 20th International Conference on Human-Computer Interaction – Posters' Extended Abstracts. 2018 Presented at: HCI '18; July 15-20, 2018; Las Vegas, NV, USA p. 356-363. [doi: 10.1007/978-3-319-92279-9_48]

33. Nesselroade JR. The warp and the woof of the developmental fabric. In: Downs RM, Liben LS, Palermo DS, editors. Visions of Aesthetics, the Environment & Development: The Legacy of Joachim F. Wohlwill. New York, NY, USA: Psychology Press; Jul 31, 1991:213-240.
34. Aalbers G, Keijsers L, Abeele MV, Hendrickson AT. Experience Sampling Measures of Stress (StrESS-2), Fatigue (tiredness-3), and Procrastination (GPS-ESM-3): A Validation Study. Open Science Framework. 2020 Jul 9. URL: https://osf.io/qvj8g/ [accessed 2023-02-15]
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
36. WhatsApp. URL: https://web.whatsapp.com/ [accessed 2023-02-15]
37. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA, USA: CreateSpace; 2009.
38. Steptoe A, Serwinski B. Cortisol awakening response. In: Fink G, editor. Stress: Concepts, Cognition, Emotion, and Behavior. Cambridge, MA, USA: Academic Press; 2016:277-283.
39. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
40. Bos FM, von Klipstein L, Emerencia AC, Veermans E, Verhage T, Snippe E, et al. A web-based application for personalized ecological momentary assessment in psychiatric care: user-centered development of the PETRA application. JMIR Ment Health 2022 Aug 9;9(8):e36430 [FREE Full text] [doi: 10.2196/36430] [Medline: 35943762]
41. van Berkel N, Goncalves J, Koval P, Hosio S, Dingler T, Ferreira D, et al. Context-informed scheduling and analysis: improving accuracy of mobile self-reports. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 51. [doi: 10.1145/3290605.3300281]
42. Keijsers L, van Roekel E. Longitudinal methods in adolescent psychology: where could we go from here? And should we? In: Hendry LB, Kloep M, editors. Reframing Adolescent Research. London, UK: Routledge; 2018:56-77.
43. Rice P, Upasham S, Jagannath B, Manuel R, Pali M, Prasad S. CortiWatch: watch-based cortisol tracker. Future Sci OA 2019 Sep 26;5(9):FSO416 [FREE Full text] [doi: 10.2144/fsoa-2019-0061] [Medline: 31608155]
44. Woodward K, Kanjo E, Brown DJ, McGinnity TM. Towards personalised mental wellbeing recognition on-device using transfer learning “in the wild”. In: Proceedings of the 2021 IEEE International Smart Cities Conference. 2021 Presented at: ISC2 '21; September 07-10, 2021; Manchester, UK p. 1-7 URL: https://ieeexplore.ieee.org/document/9562774 [doi: 10.1109/ISC253183.2021.9562774]

Abbreviations

CV: cross-validation
LASSO: least absolute shrinkage and selection operator
MAE: mean absolute error
RF: random forest
SHAP: Shapley additive explanations
SVR: support vector regression

Edited by L Buis; submitted 22.02.22; peer-reviewed by D Gard, CI Sartorão Filho; comments to author 12.07.22; revised version received 01.09.22; accepted 21.09.22; published 23.03.23.

Please cite as:
Aalbers G, Hendrickson AT, Vanden Abeele MMP, Keijsers L
Smartphone-Tracked Digital Markers of Momentary Subjective Stress in College Students: Idiographic Machine Learning Analysis
JMIR Mhealth Uhealth 2023;11:e37469
URL: https://mhealth.jmir.org/2023/1/e37469
doi:10.2196/37469
PMID:36951924

Original Paper

Smartwatch-Based Maximum Oxygen Consumption Measurement for Predicting Acute Mountain Sickness: Diagnostic Accuracy Evaluation Study

Xiaowei Ye[1*], MD; Mengjia Sun[1*], MD; Shiyong Yu[1*], PhD; Jie Yang[1], PhD; Zhen Liu[1], MD; Hailin Lv[1], MD; Boji Wu[1], MD; Jingyu He[1], MD; Xuhong Wang[1], MD; Lan Huang[1], PhD

Institute of Cardiovascular Diseases of People's Liberation Army, The Second Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, China

[*]these authors contributed equally

Corresponding Author:
Lan Huang, PhD
Institute of Cardiovascular Diseases of People's Liberation Army
The Second Affiliated Hospital
Army Medical University (Third Military Medical University)
No 183, Xinqiao Street, Shapingba District
Chongqing, 400037
China
Phone: 86 23 68755601
Email: huanglan260@126.com

Abstract

Background: Cardiorespiratory fitness plays an important role in coping with hypoxic stress at high altitudes. However, the association of cardiorespiratory fitness with the development of acute mountain sickness (AMS) has not yet been evaluated. Wearable technology devices provide a feasible assessment of cardiorespiratory fitness, which is quantifiable as maximum oxygen consumption (VO_2max) and may contribute to AMS prediction.

Objective: We aimed to determine the validity of VO_2max estimated by the smartwatch test (SWT), which can be self-administered, in order to overcome the limitations of clinical VO_2max measurements. We also aimed to evaluate the performance of a VO_2max-SWT–based model in predicting susceptibility to AMS.

Methods: Both SWT and cardiopulmonary exercise test (CPET) were performed for VO_2max measurements in 46 healthy participants at low altitude (300 m) and in 41 of them at high altitude (3900 m). The characteristics of the red blood cells and hemoglobin levels in all the participants were analyzed by routine blood examination before the exercise tests. The Bland-Altman method was used for bias and precision assessment. Multivariate logistic regression was performed to analyze the correlation between AMS and the candidate variables. A receiver operating characteristic curve was used to evaluate the efficacy of VO_2max in predicting AMS.

Results: VO_2max decreased after acute high altitude exposure, as measured by CPET (25.20 [SD 6.46] vs 30.17 [SD 5.01] at low altitude; $P<.001$) and SWT (26.17 [SD 6.71] vs 31.28 [SD 5.17] at low altitude; $P<.001$). Both at low and high altitudes, VO_2max was slightly overestimated by SWT but had considerable accuracy as the mean absolute percentage error (<7%) and mean absolute error (<2 mL·kg^{-1}·min^{-1}), with a relatively small bias compared with VO_2max-CPET. Twenty of the 46 participants developed AMS at 3900 m, and their VO_2max was significantly lower than that of those without AMS (CPET: 27.80 [SD 4.55] vs 32.00 [SD 4.64], respectively; P=.004; SWT: 28.00 [IQR 25.25-32.00] vs 32.00 [IQR 30.00-37.00], respectively; P=.001). VO_2max-CPET, VO_2max-SWT, and red blood cell distribution width-coefficient of variation (RDW-CV) were found to be independent predictors of AMS. To increase the prediction accuracy, we used combination models. The combination of VO_2max-SWT and RDW-CV showed the largest area under the curve for all parameters and models, which increased the area under the curve from 0.785 for VO_2max-SWT alone to 0.839.

Conclusions: Our study demonstrates that the smartwatch device can be a feasible approach for estimating VO_2max. In both low and high altitudes, VO_2max-SWT showed a systematic bias toward a calibration point, slightly overestimating the proper

VO_2max when investigated in healthy participants. The SWT-based VO_2max at low altitude is an effective indicator of AMS and helps to better identify susceptible individuals following acute high-altitude exposure, particularly by combining the RDW-CV at low altitude.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200059900; https://www.chictr.org.cn/showproj.html?proj=170253

(JMIR Mhealth Uhealth 2023;11:e43340) doi:10.2196/43340

KEYWORDS

VO2max; maximum oxygen consumption; smartwatch; cardiopulmonary exercise test; acute mountain sickness

Introduction

In recent years, mountain climbing has become a popular activity for pleasure, work, and athletic competitions. However, inadequate acclimatization to hypobaric hypoxia results in a series of symptoms known as acute mountain sickness (AMS). AMS is relatively common among new travelers, affecting >30% of individuals ascending to 3500 m and >70% of those ascending above 6000 m [1]. AMS is characterized by the presence of headache in combination with other symptoms, including dizziness, fatigue, loss of appetite, and insomnia [2]. Although younger age, female gender, rapid ascent, low oxygen saturation (SpO_2), and abnormal ventilatory response to exercise have been previously associated with AMS and its severity [3-5], susceptible individuals still need to be further identified, especially with more accuracy and practicality.

Maximum oxygen consumption (VO_2max) is defined as the maximum capacity of the cardiovascular, respiratory, and muscular systems to deliver and utilize oxygen, which is reflected by an individual's cardiorespiratory fitness [6-9]. VO_2max is accurately measured by the cardiopulmonary exercise test (CPET) during a maximal graded exercise until exhaustion, which is considered the gold standard for cardiorespiratory functional assessment [10,11]. However, the use of direct measurements is limited, particularly at high altitude, as it is time-consuming and requires infrastructure and specialized personnel to conduct exercise assessments. Therefore, indirect measurement methods of VO_2max (Firstbeat fitness test [FFT]) have been developed and have become advantageous due to the popularity of smart wearable devices [12,13]. Previous studies have reported that VO_2max estimated by the FFT method is accurate and suitable for athletes owing to its lower exercise intensity [14]. However, for those who require to face the challenge of extreme high-altitude environments over a short period, the maximum intensity of exercise should also be avoided so as not to affect the acclimatization process. Previous studies have shown that the error of the FFT method is less than 5% at low altitudes [14]; however, it remains controversial whether it underestimates the true VO_2max. Additionally, its performance at high altitudes has not been evaluated and compared with that of the gold standard.

VO_2max decreases during acute or chronic exposure to high altitudes, which is mainly attributed to the reduction of PO_2 [15,16]. Moreover, in terms of limiting VO_2max, in addition to environmental factors, more attention is focused on the oxygen delivery pathway, central circulation [17], maximal cardiac output [18], oxygen-carrying capacity of the blood, ability to distribute that blood into the contracting muscles, and finally, the ability of the muscles to consume oxygen [19]. In other words, the above physiological processes and related indices involving oxygen transport may also be potential predictors of AMS [20]. Therefore, this study aims to compare the accuracy and consistency of VO_2max obtained from CPET and smartwatch test (SWT) at different atmospheric pressures and to determine whether VO_2max at low altitudes is correlated with AMS. Further, we tested the hypothesis that the combination of VO_2max-SWT and red blood cell (RBC) distribution width-coefficient of variation (RDW-CV) may be more efficient in predicting AMS.

Methods

Participant Recruitment

We recruited 46 healthy adults (27 women and 19 men, age range 22-54 years) from Chongqing, China, based on the inclusion and exclusion criteria. All participants had lived at low altitudes (<500 m) for at least 10 years and had no recent history of high-altitude (>2500 m) exposure (in the last 6 months). Participants with any one of the following conditions were excluded: respiratory and cardiovascular diseases, malignant tumors, liver and kidney dysfunctions, and psychiatric disorders or neuroses that would not allow them to complete the questionnaires.

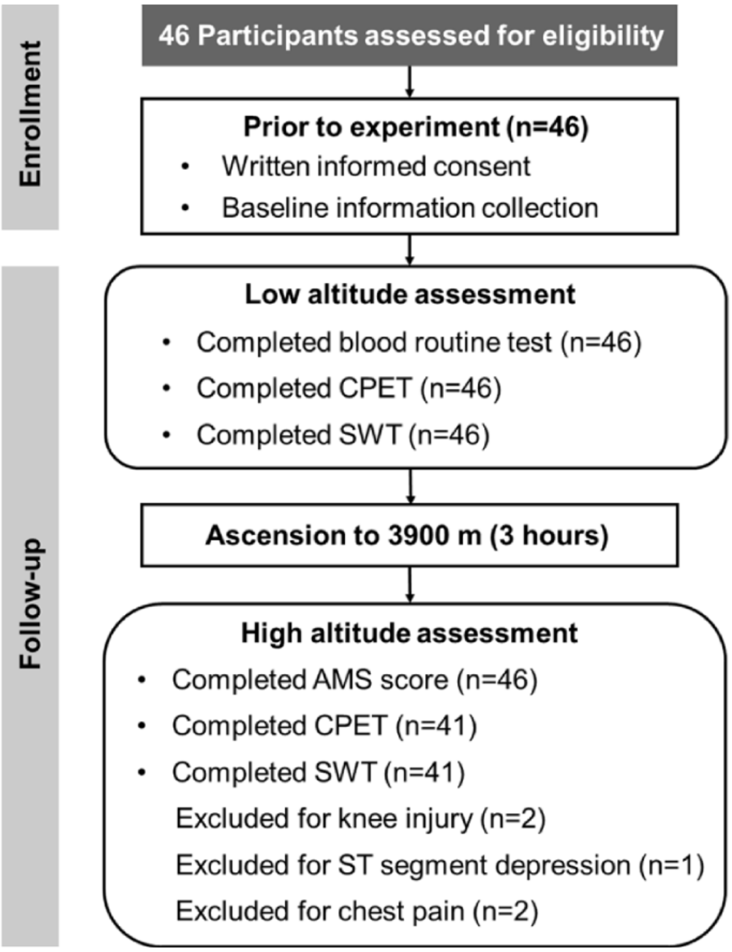
Ethics Approval

The study protocol (ChiCTR2200059900) complied with the Declaration of Helsinki and was approved by the ethics committee of Xinqiao Hospital of Army Medical University (approval: 2022-研第-060-01). Written informed consent was obtained from all the participants after the study details, procedures, benefits, and risks were explained.

Procedures

This study consisted of 2 exercise tests at low and high altitudes (Figure 1). The participants were instructed to avoid heavy load training 7 days before the tests and during the recovery days and to abstain from caffeine and alcohol for 24 hours before testing. On the first day of the study, each participant underwent a routine blood test before the SWT. After a 24-hour break, CPET was performed. The 2 tests (SWT and CPET) were performed at a similar time of day (SD 30 minutes) and were completed in 2 days. After resting for 3 days, the participants ascended to a high altitude (3900 m, Shigatse, China) in 3 hours by plane from a low altitude (300 m, Chongqing, China). On the second day at high altitude, they took 1 day off to complete the 2018 Lake Louise score assessment. Unfortunately, 2 individuals had knee injury because of the trip; therefore, they did not undergo exercise tests. Besides, 1 participant had an ST segment depression in the electrocardiogram and 2 participants had chest pain; therefore, they could not make it to the end. Finally, the remaining 41 participants repeated the exercise tests completely.

Figure 1. Cohort development diagram for this study. AMS, acute mountain sickness; CPET, cardiopulmonary exercise test; SWT, smartwatch test.



Blood Routine Examination

The participants were required to avoid eating or drinking anything (fasting) apart from water for up to 12 hours. Approximately 5 mL of intravenous blood was collected from the inside of the elbow and mixed with 1 mL of dipotassium ethylenediaminetetraacetic acid anticoagulant by using a tight band (tourniquet). Blood samples were analyzed using a BC-3000 plus automated hematology corpuscle analyzer (Mindray). The details of the 19 different parameters are presented in Multimedia Appendix 1. Blood tests at low altitude were performed between 7 AM and 9 AM on the same day before the exercise tests. All biochemical parameters were measured in the blood samples at the Clinical Laboratory of Cardiology Science of Xinqiao Hospital, Army Medical University.

CPET Analysis

The CPET was performed on an electronically braked cycle ergometer (EC3000e, Customed) in an erect position with breath-by-breath measurements through a tightly fitted face mask of minute ventilation, O_2 uptake, and CO_2 output by using a cardiopulmonary exercise testing system (Metalyzer 3B, Cortex). Before performing CPET, the baseline physiological measures for all devices used in this study were measured for 5 minutes in a resting state and subsequently in a standing position. After the baseline measurement, the test was conducted immediately. The cycle ergometry test protocol included 3 minutes of free-wheel cycling and subsequently proceeded with a continual increase in resistance by 25 W/min (according to the prior known exercise capacity [21], so that the test would last 10-12 minutes) until test completion or exhaustion. VO_2max was defined as the highest 30-s average value within the last minute of exercise until the first 15 s of recovery at peak exercise [22]. Standard 12-lead electrocardiogram, blood pressure, and SpO_2 were obtained at rest, every minute during exercise, and for ≥4 minutes during the recovery phase throughout the procedure using a 12-lead connection (custo-Cardio 3000BT-A, Cortex) in real time, blood pressure cuffs (Suntech Tango M2, Cortex) in the upper arm, and a finger clip portable oximeter (Nonin wristOx2), respectively.

SWT Analysis

We provided participants with a smartwatch (Huawei Watch GT Runner) and instructed them to wear it correctly on the left wrist, which enables reliable and persistent measurement of running speed, distance, and heart rate. Therefore, these measurements could be monitored continuously and automatically during each running activity, stored on the participant's mobile device (Huawei MatePad 11 DBY-W09), and regularly transmitted to a secure cloud server, which was later transferred to the Huawei Health Center software through Bluetooth. Specifically, VO_2max estimation steps were as follows: (1) the personal background information (age, height, and weight) of the participant was logged in and the exercise type (running outdoors) was selected; (2) the participant started to run with a smartwatch that measured the heart rate and speed on level ground; (3) the start and end points were in the same place, and the smartwatch was stopped by the researchers uniformly with a timely click; (4) the researchers subsequently saved the participants' running data to an album on the pad, facilitating further statistical analysis; and (5) the smartwatch and mobile device were formatted to prepare for the next test.

The signal processing by Huawei Watch GT Runner is licensed by the Firstbeat Technology's Fitness Test, which is based on intelligent detection for both data reliability and exercise pattern during successive recording [13]. Briefly, the moving average filter was applied to both heart rate and physical activity data. After filtering the data, only data points at which both heart rate and physical activity data increased were selected as a period of physical activity. This was conducted by differentiating the data and selecting where both differentiated data were positive. The situations where the data series were excluded are listed below: (1) significant heart rate decreases and exceptional striding pattern (identified as a situation of running on a very steep downhill or soft surface automatically), (2) significant heart rate increases while the velocity remained 0 (identified as stopping suddenly in the middle), and (3) a short duration of highly increasing intensity (identified as insufficient effort level). After exclusions, the selected data series were further segmented as different heart rate zones according to the effort levels. Of them, the reliable data segments that belong to a long series of successive heartbeat intervals (in generally 20 s-10 minutes and preferably 30 s-4 minutes) and with a small heart rate change level were recognized as sufficient effort and used to calculate the VO_2max. In these reliable segments, speed was measured based on acceleration measurements by using a satellite navigation system. VO_2max estimates were made for each reliable segment by using the following theoretical VO_2 equation: theoretical VO_2 ($mL \cdot kg^{-1} \cdot min^{-1}$) = 3.5 * speed (km/h). The obtained VO_2max for each data segment was weighed and subsequently utilized to make a linear equation for calculating the final VO_2max (the detailed rule of weighting is shown in patent US9237868B2).

Lake Louise Consensus Scoring System and AMS

The presence of AMS at high altitude was assessed using the Lake Louise consensus scoring system 2018 version [23]. According to the 4 main symptoms, namely, headache, gastrointestinal symptoms, fatigue/weakness, and dizziness/vertigo, the scores were 0, 1, 2, and 3 in the order of none, mild, moderate, and severe, respectively. A total score of ≥3 combined with headache can be diagnosed as AMS.

Statistical Analyses

Categorical variables were described as numbers and percentages. Descriptive statistics were presented as mean (SD) for variables with skewed distribution and median (IQR) for variables with normal distribution. The Mann–Wilcoxon rank-sum, independent-sample *t* test (2-sided test), Pearson chi-square test, and Fisher exact tests were used to compare the continuous and categorical variables statistically. The correlation magnitude and coefficient of determination between VO_2max-CPET and VO_2max-SWT were assessed using Pearson correlation. The intraclass correlation coefficient and paired-sample *t* tests (2-sided test) were performed to determine the agreement between VO_2max-CPET and VO_2max-SWT at low and high altitudes, respectively. We calculated the mean absolute error and mean absolute percentage error (MAPE) to evaluate the accuracy of the estimation. Furthermore, we used a Bland–Altman plot to investigate the level of agreement with 95% limits of agreement [24].

The relationship between the variables and AMS was examined by binomial logistic regression analysis with univariate analyses. The relationship between VO_2max-CPET, VO_2max-SWT, RDW-CV, and AMS was further examined by multivariate analyses. In the preliminary screening, we considered the variable with $P<.05$ as a potential risk factor, and an adjusted binary logistic regression model subjected the variable to identify the independent risk factors for AMS after the adjustment. Receiver operating characteristic (ROC) curves were constructed, and the Youden index was calculated. The optimal cutoff of variables for diagnosing AMS was determined at the point where the Youden index was maximum on the ROC analysis. We also compared the ROC curves of VO_2max-CPET, VO_2max-SWT, and RDW-CV alone or in combination. Differences were considered statistically significant at $P<.05$. Statistical analyses were performed using the SPSS Statistics software (IBM Corp) for Windows (version 26) and MedCalc software for Windows.

Results

Participant Characteristics

A total of 46 participants were recruited for this study, of whom 20 (44%) participants developed AMS. The clinical characteristics of participants with AMS and without AMS are presented in Table 1. There were no differences in age, sex, BMI, baseline heart rate, SpO_2, and blood pressure between the 2 groups. RBC count; hemoglobin, hematocrit, and mean corpuscular hemoglobin levels; mean corpuscular hemoglobin concentration; and RDW-SD did not differ significantly between the participants in the 2 groups, whereas the AMS group had higher RDW-CV at low altitude than the non-AMS group (14.25 [IQR 12.75-21.03] vs 12.70 [IQR 12.25-13.53], respectively; P=.02). In addition, the AMS group had lower VO_2max both measured by CPET (27.80 [SD 4.55] vs 32.00 [SD 4.04], respectively; P=.004) and estimated by SWT (28.00 [SD 6.75] vs 32.00 [IQR 30.00-37.00], respectively; P=.001) than the non-AMS group.

Table 1. Baseline characteristics, blood routine test, and maximum oxygen consumption of the participants at low altitudes.

Variables	Total (n=46)	Participants with AMS[a] (n=20)	Participants without AMS (n=26)	*P* value[b]
Baseline characteristics				
Age (years), mean (SD)	33.33 (7.80)	33.85 (8.41)	32.92 (7.44)	.70
Gender, n (%)				.17
Female	27 (59)	14 (70)	13 (50)	
Male	19 (41)	6 (30)	13 (50)	
BMI (kg/m^2), median (IQR)	22.19 (20.22-23.64)	21.89 (20.15-23.44)	22.40 (20.22-24.01)	.78
Alcohol use, n (%)				.18
Current drinker or ex-drinker	10 (22)	2 (10)	8 (31)	
Never	36 (78)	18 (90)	18 (69)	
Smoking status, n (%)				.92
Current smoker or ex-smoker	6 (13)	2 (10)	4 (15)	
Nonsmoker	40 (87)	18 (90)	22 (85)	
HR[c] (beats/min), mean (SD)	78.93 (9.69)	81.25 (9.72)	77.15 (9.46)	.16
SpO_2[d] (%), median (IQR)	97 (96-98)	97 (96-98.75)	97 (96-98)	.72
SBP[e] (mm Hg), median (IQR)	112.00 (103.75-121.25)	112.00 (105.00-126.50)	112.00 (102.75-118.75)	.92
DBP[f] (mm Hg), mean (SD)	74.11 (11.11)	72.95 (14.24)	75.00 (8.12)	.54
Blood routine test				
RBC[g] (10^{-9}/L), median (IQR)	4.51 (4.30-4.95)	4.40 (4.25-4.89)	4.71 (4.33-5.08)	.19
HGB[h] (g/L), mean (SD)	131.59 (10.03)	128.85 (11.45)	133.69 (8.42)	.11
HCT[i] (%), mean (SD)	44.10 (4.32)	43.31 (4.28)	44.70 (4.34)	.28
MCV[j] (fL), median (IQR)	94.10 (90.90-96.60)	94.35 (96.50-96.58)	94.00 (91.73-96.78)	.89
MCH[k] (pg), median (IQR)	28.77 (27.46-29.82)	28.54 (27.18-29.77)	28.82 (27.58-29.90)	.71
MCHC[l] (g/L), mean (SD)	304.62 (15.21)	304.46 (14.72)	304.75 (15.86)	.95
RDW-CV[m] (%), median (IQR)	13.10 (12.30-14.83)	14.25 (12.75-21.03)	12.70 (12.25-13.53)	.02
RDW-SD[n] (fL), median (IQR)	44.40 (41.48-46.68)	44.40 (41.00-46.83)	44.55 (41.48-46.70)	.84
Cardiorespiratory fitness				
VO_2max-CPET[o] (mL·kg^{-1}·min^{-1}), mean (SD)	30.17 (5.01)	27.80 (4.55)	32.00 (4.64)	.004
VO_2max-SWT[p] (mL·kg^{-1}·min^{-1}), median (IQR)	30.50 (27.75-34.25)	28.00 (25.25-32.00)	32.00 (30.00-37.00)	.001

[a]AMS: acute mountain sickness.

[b]Differences were considered statistically significant if $P<.05$.

[c]HR: heart rate.

[d]SpO_2: oxygen saturation.

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]RBC: red blood cell.

[h]HGB: hemoglobin.

[i]HCT: hematocrit.

[j]MCV: mean corpuscular volume.

[k]MCH: mean corpuscular hemoglobin.

[l]MCHC: mean corpuscular hemoglobin concentration.

[m]RDW-CV: red blood cell distribution width-coefficient of variation.

[n]RDW-SD: red blood cell distribution width-standard deviation.

[o]VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test.

[p]VO_2max-SWT: maximum oxygen consumption estimated by smartwatch test.

Accuracy and Consistency Analyses of VO2max Estimation in the SWT at Low and High Altitudes

Table 2 shows the VO_2max in the SWT and CPET at low and high altitudes. The values of VO_2max-SWT were significantly overestimated at both low (constant error=1.11 [SD 1.73] mL·kg^{-1}·min^{-1}; t_{45}=4.35; P<.001) and high (constant error=0.98 [SD 1.54] mL·kg^{-1}·min^{-1}; t_{40}=4.05; P<.001) altitudes. A 6% MAPE (mean absolute error=1.761 mL·kg^{-1}·min^{-1}) at low altitude and a 6.8% MAPE (mean absolute error=1.610 mL·kg^{-1}·min^{-1}) at high altitude were observed, indicating a low average deviation between the 2 methods (Table 2). Furthermore, a strong correlation was found between VO_2max-SWT and VO_2max-CPET values (low altitude: R^2=0.889; P<.001; high altitude: R^2=0.947; P<.001; Figure 2). The results of the intraclass correlation coefficient revealed that VO_2max-SWT had a good level of agreement with the directly measured VO_2max-CPET at low (0.942; P<.001) and high (0.973; P<.001) altitudes. Additionally, the Bland–Altman plots demonstrated a small bias of the VO_2max-SWT values compared to the VO_2max-CPET at low (bias=1.11 mL·kg^{-1}·min^{-1}, Figure 3A) and high (bias=1.00 mL·kg^{-1}·min^{-1}, Figure 3B) altitudes. VO_2max-SWT showed even a lower range of bias at high altitudes than at low altitudes (upper to lower limits of agreement: 6.0 mL·kg^{-1}·min^{-1} vs 6.8 mL·kg^{-1}·min^{-1}, respectively).

Table 2. Correlations and differences between the estimated maximum oxygen consumption in the smartwatch test and the measured maximum oxygen consumption in the cardiopulmonary exercise test.

	VO_2max-CPET[a] (mL·kg^{-1}·min^{-1}), mean (SD)	VO_2max-SWT[b] (mL·kg^{-1}·min^{-1}), mean (SD)	CE[c] (mL·kg^{-1}·min^{-1}), mean (SD)	t (df)	r	Intraclass correlation coefficient	Mean absolute error (mL·kg^{-1}·min^{-1})	Mean absolute percentage error (%)
Low altitude (n=46)	30.17 (5.01)	31.28 (5.17)	1.11 (1.73)	4.35 (45) (P[d]<.001)	0.943 (P<.001)	0.942 (P<.001)	1.761	6
High altitude (n=41)	25.20 (6.46)	26.17 (6.71)	0.98 (1.54)	4.05 (40) (P<.001)	0.973 (P<.001)	0.973 (P<.001)	1.610	6.80

[a]VO_2max-CPET: maximum oxygen consumption measured by the cardiopulmonary exercise test.

[b]VO_2max-SWT: maximum oxygen consumption estimated by the smartwatch test.

[c]CE: constant error (arithmetic mean of the difference between estimated and measured VO_2max).

[d]Differences were considered statistically significant if P<.05.

Figure 2. Linear regression plots between the estimated maximum oxygen consumption measured by smartwatch test and maximum oxygen consumption measured by cardiopulmonary exercise testing. Pearson correlation between the maximum oxygen consumption estimated by smartwatch and measured by cardiopulmonary exercise testing at low altitude (A) and at high altitude (B). The coefficient of determination (R2) and 95% CI bounds (dotted line) are depicted for the regression lines (solid). VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise testing; VO_2max-SWT: estimated maximum oxygen consumption by smartwatch test.

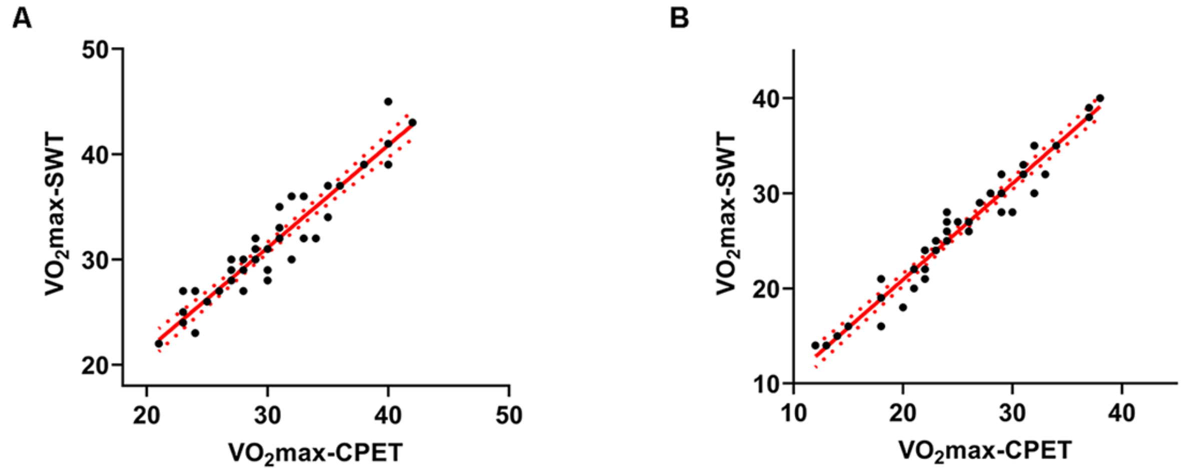
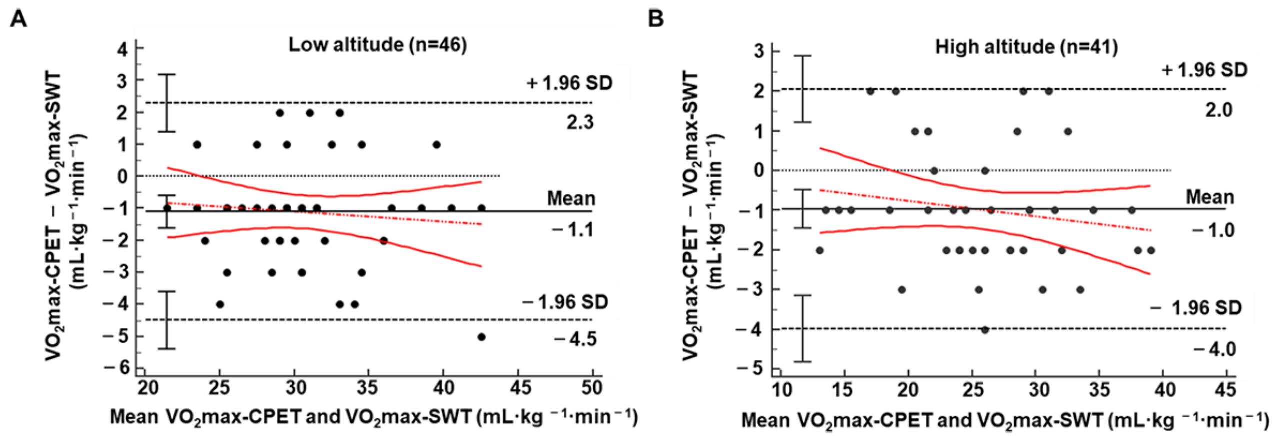


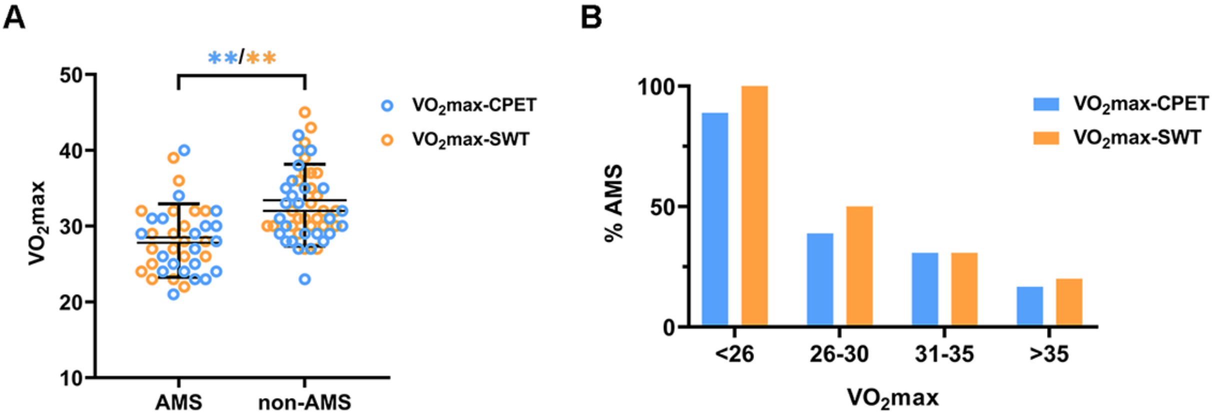
Figure 3. Bland-Altman plots between the estimated maximum oxygen consumption by smartwatch test and maximum oxygen consumption measured by cardiopulmonary exercise test at low altitude (A) and at high altitude (B). Mean biases (solid line), 95% limits of agreement (dashed line), and equality (dotted line) are also depicted. VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test; VO_2max-SWT: estimated maximum oxygen consumption by smartwatch test.



Distribution of VO_2max and Incidence of AMS

There was a significant difference in the VO_2max values at low altitude between the participants with and without AMS. VO_2max-CPET in the AMS group was lower than that in the non-AMS group (27.80 [SD 4.55] vs 32.00 [SD 4.64], respectively; *P*=.004), and a similar result was revealed by SWT (28.00 [IQR 25.25-32.00] vs 2.00 [IQR 30.00-37.00], respectively; *P*=.001; Figure 4A). The distribution of the VO_2max values based on the AMS results is shown in Figure 4B. Approximately 90% (16/17) of the participants with VO_2max values <26 mL·kg^{-1}·min^{-1} developed AMS compared with approximately 20% (3/15) of the participants who developed AMS with a VO_2max>35 mL·kg^{-1}·min^{-1}. Patients with AMS seemed to have a lower VO_2max, regardless of whether it was directly measured by CPET or estimated by SWT.

Figure 4. (A) Distribution of the estimated maximum oxygen consumption measured by smartwatch test and cardiopulmonary exercise test based on the diagnosis of acute mountain sickness. (B) Diagram of the probability of acute mountain sickness occurrence for different ranges of the maximum oxygen consumption value at low altitude (blue: maximum oxygen consumption measured by cardiopulmonary exercise test; orange: maximum oxygen consumption measured by smartwatch test). **Significantly different between acute mountain sickness and non–acute mountain sickness at *P*<.01. AMS: acute mountain sickness; VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test; VO_2max-SWT: estimated maximum oxygen consumption by smartwatch test.



Univariate and Multivariate Logistic Regression Analyses for AMS

To further explore the association between VO_2max and AMS, a univariate analysis was performed. Table 3 shows that VO_2max-CPET (odds ratio [OR] 0.807, 95% CI 0.686-0.949; *P*=.01) and VO_2max-SWT (OR 0.765, 95% CI 0.635-0.922; *P*=.005) at low altitude and baseline RDW-CV (OR 1.177, 95% CI 0.999-1.386; *P*=.05) were potentially associated with AMS occurrence. Multivariate regression analysis identified VO_2max-CPET (OR 0.770, 95% CI 0.640-0.926; *P*=.006) and RDW-CV (OR 1.263, 95% CI 1.028-1.553; *P*=.03) as well as VO_2max-SWT (OR 0.720, 95% CI 0.578-0.898; *P*=.004) and RDW-CV (OR 1.273, 95% CI 1.027-1.577; *P*=.03) as independent factors associated with the development of AMS at high altitude.

Table 3. Binomial logistic regression analysis of factors related to acute mountain sickness.

Variables	Univariable		Multivariable		Multivariable	
	OR[a] (95% CI)	*P* value[b]	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)	1.016 (0.942-1.096)	.69	N/A[c]	N/A	N/A	N/A
Male (Y/N[d])	2.333 (0.684-7.960)	.18	N/A	N/A	N/A	N/A
BMI (kg/m^2)	1.037 (0.852-1.262)	.72	N/A	N/A	N/A	N/A
Tobacco (Y/N)	1.636 (0.268-9.980)	.59	N/A	N/A	N/A	N/A
Alcohol (Y/N)	4.000 (0.744-21.496)	.11	N/A	N/A	N/A	N/A
HR[e] (beats/min)	1.048 (0.982-1.118)	.16	N/A	N/A	N/A	N/A
SpO_2[f] (%)	0.907 (0.558-1.472)	.69	N/A	N/A	N/A	N/A
SBP[g] (mm Hg)	1.002 (0.965-1.040)	.91	N/A	N/A	N/A	N/A
DBP[h] (mm Hg)	0.983 (0.932-1.037)	.53	N/A	N/A	N/A	N/A
VO_2max-CPET[i] ($mL·kg^{-1}·min^{-1}$)	0.807 (0.686-0.949)	.01	0.770 (0.640-0.926)	.006	N/A	N/A
VO_2max-SWT[j] ($mL·kg^{-1}·min^{-1}$)	0.765 (0.635-0.922)	.005	N/A	N/A	0.720 (0.578-0.898)	.004
RBC[k] (10^{-9}/L)	0.711 (0.219-2.308)	.57	N/A	N/A	N/A	N/A
HGB[l] (g/L)	0.949 (0.889-1.012)	.11	N/A	N/A	N/A	N/A
HCT[m] (%)	0.923 (0.800-1.066)	.28	N/A	N/A	N/A	N/A
MCV[n] (fL)	0.966 (0.876-1.064)	.48	N/A	N/A	N/A	N/A
MCH[o] (pg)	0.892 (0.660-1.205)	.46	N/A	N/A	N/A	N/A
MCHC[p] (g/L)	0.999 (0.961-1.038)	.95	N/A	N/A	N/A	N/A
RDW-CV[q] (%)	1.177 (0.999-1.386)	.05	1.263 (1.028-1.553)	.03	1.273 (1.027-1.577)	.03
RDW-SD[r] (fL)	0.985 (0.850-1.141)	.84	N/A	N/A	N/A	N/A

[a]OR: odds ratio.

[b]Differences were considered statistically significant if $P<.05$.

[c]N/A: not applicable.

[d]Y/N: yes/no.

[e]HR: heart rate.

[f]SpO_2: oxygen saturation.

[g]SBP: systolic blood pressure.

[h]DBP: diastolic blood pressure.

[i]VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test.

[j]VO_2max-SWT: maximum oxygen consumption estimated by smartwatch test.

[k]RBC: red blood cell.

[l]HGB: hemoglobin.

[m]HCT: hematocrit.

[n]MCV: mean corpuscular volume.

[o]MCH: mean corpuscular hemoglobin.

[p]MCHC: mean corpuscular hemoglobin concentration.

[q]RDW-CV: red blood cell distribution width-coefficient of variation.

[r]RDW-SD: red blood cell distribution width-standard deviation.

VO_2max-Based Model for Predicting AMS

As both VO_2max and RDW-CV were closely related to AMS, we constructed the combined predictive models for AMS. Table 4 and Figure 5 show the area under the curve (AUC) of VO_2max-CPET (AUC 0.743, 95% CI 0.597-0.889), VO_2max-SWT (AUC 0.785, 95% CI 0.646-0.923), and RDW-CV (AUC 0.708, 95% CI 0.547-0.868).

Either the AUC of the VO_2max-CPET or VO_2max-SWT was higher than that of RDW-CV (both P>.05, Table 4). For the VO_2max-SWT, a sensitivity of 65% and a specificity of 88.46% were observed at the optimal cutoff value of 29.5 $mL \cdot kg^{-1} \cdot min^{-1}$, with a higher positive predictive value of 81.25% and a negative predictive value of 76.67%. However, no significant difference was found in AUC when compared to the VO_2max-CPET (0.785 vs 0.743, respectively; P=.25). Although the independent indicators were effective and significant, this combined predictive model was more accurate when VO_2max and RDW-CV were combined. In other words, the combined model 2 enhanced the diagnostic power with modest AUC gains of 0.03-0.06, although not statistically different from model 1 (0.839 vs 0.804, respectively; P=.27) or VO_2max-SWT alone (0.839 vs 0.785, respectively; P=.28). The combined model 2 also improved the prediction of AMS by increasing sensitivity from 65% to 80% compared with VO_2max-SWT alone while attaining high specificity.

Table 4. Receiver operating characteristic curves to assess the performance of the maximum oxygen consumption measured by cardiopulmonary exercise test and the maximum oxygen consumption estimated by smartwatch test in predicting acute mountain sickness.[a]

	AUC[b] (95% CI)	Optimal cutoff values ($mL \cdot kg^{-1} \cdot min^{-1}$ or fL)	Sensitivity (%)	Specificity (%)	PPV[c] (%)	NPV[d] (%)
VO_2max-CPET[e]	0.743 (0.597-0.889)	26.50	45	96.15	90	69.44
VO_2max-SWT[f]	0.785 (0.646-0.923)	29.50	65	88.46	81.25	76.67
RDW-CV[g]	0.708 (0.547-0.868)	13.10	75	69.23	65.22	78.26
Model 1: VO_2max-CPET + RDW-CV	0.804 (0.675-0.933)	N/A[h]	65	92.31	87.50	77.42
Model 2: VO_2max-SWT + RDW-CV	0.839 (0.720-0.959)	N/A	80	84.62	80	84.62

[a]Comparison of area under the curve: maximum oxygen consumption measured by cardiopulmonary exercise test versus maximum oxygen consumption estimated by smartwatch test (P=.25); Model 1 versus maximum oxygen consumption measured by cardiopulmonary exercise test (P=.22); Model 2 versus maximum oxygen consumption estimated by smartwatch test (P=.28); Model 1 versus Model 2 (P=.27).

[b]AUC: area under the curve.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

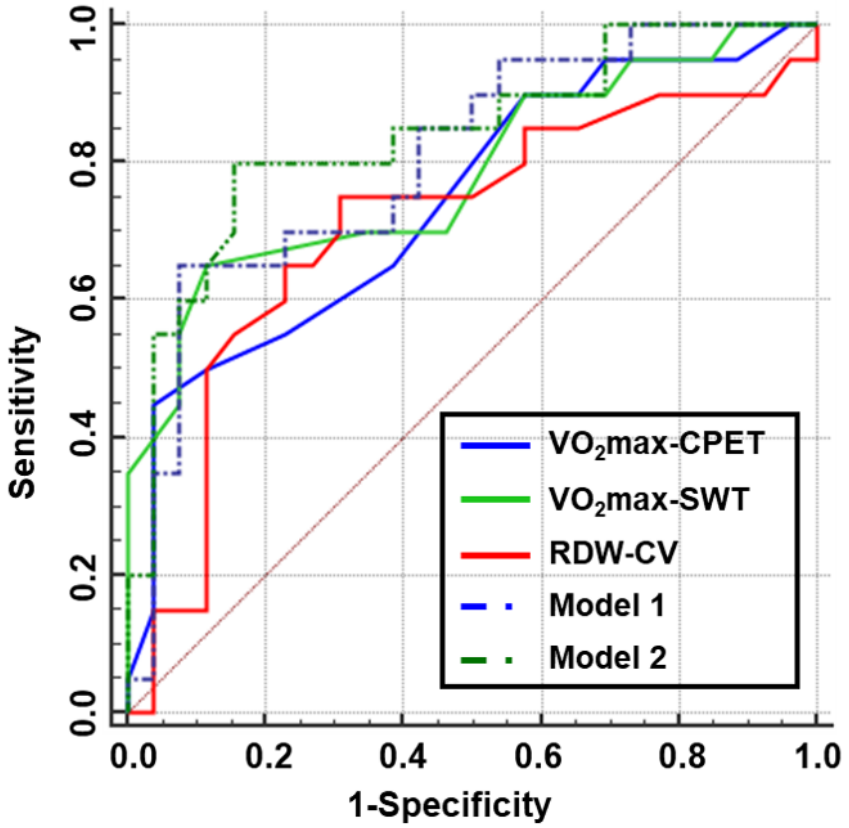
[e]VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test.

[f]VO_2max-SWT: maximum oxygen consumption estimated by smartwatch test.

[g]RDW-CV: red blood cell distribution width-coefficient of variation.

[h]N/A: not applicable.

Figure 5. Receiver operating characteristic curves for maximum oxygen consumption measured by the cardiopulmonary exercise test (blue solid line), estimated maximum oxygen consumption by the smartwatch test (green solid line), and red blood cell distribution width-coefficient of variation (red solid line), and for Model 1 (blue dotted line: combination of maximum oxygen consumption measured by cardiopulmonary exercise test and red blood cell distribution width-coefficient of variation) and Model 2 (green dotted line: combination of estimated maximum oxygen consumption by smartwatch test and red blood cell distribution width-coefficient of variation) in predicting acute mountain sickness. VO_2max-CPET: maximum oxygen consumption measured by cardiopulmonary exercise test; VO_2max-SWT: estimated maximum oxygen consumption by smartwatch test; RDW-CV: red blood cell distribution width-coefficient of variation.



Discussion

Principal Results

Our study comparatively evaluated the VO_2max (CPET vs SWT) of individuals at a low altitude and subsequently at a high altitude. We demonstrated that the smartwatch device was a feasible and accurate tool for assessing cardiorespiratory fitness at both altitudes. We also proposed a novel model based on smartwatch-derived VO_2max with good performance in predicting AMS. Our easy-to-use approach for estimating VO_2max can be more widely applied for screening individuals susceptible to AMS on a large scale.

Previous clinical trials [25-27] have shown that VO_2max ranged from 20 mL·kg^{-1}·min^{-1} to 50 mL·kg^{-1}·min^{-1} according to the sex, age, or ethnicity of participants. In our study, the value of VO_2max measured either in CPET (low altitude: 30.17 [SD 5.01] vs high altitude: 25.20 [SD 6.46]; P<.001) or SWT (low altitude: 31.28 [SD 5.17] vs high altitude: 26.17 [SD 6.71]; P<.001) was within a fair level range, reflecting the sedentary fitness of this cohort. The 10 pairs of participants' VO_2max at low altitude were nearly the same. The high overlap rate of our data may be because the measurement output from CPET and SWT are integers, which reduce the numerical difference. In other words, we consider that when the difference between the 2 measurements is less than 1, this difference may not show up. Such a high overlap rate outcome has been reported in a previous study [12]. Besides, repeated measures may help reduce the repetitive rate.

Similar to that shown in a previous report, VO_2max was lower by approximately 16.5% at 3900 m evaluated by both CPET and SWT compared to that at low altitude in our trial, which can be attributed to the reduction of atmospheric PO_2 at high altitude [28]. Interestingly, the VO_2max-SWT was slightly higher than the VO_2max measured by the breath-to-breath method at both low and high altitudes. Therefore, we considered that the following issues may be associated with the differences: (1) the exercise mode in SWT was accelerative running, which requires a larger number of muscle groups, whereas the mode of CPET was a cycle ergometer, mostly relying on endurance, and (2) CPET required maximal performance; however, SWT only required submaximal exercise intensity. In addition, the overestimated VO_2max of the SWT was more evident at high altitude, as highlighted by the higher MAPE (6.80% vs 6%, respectively) and wider 95% limits of agreement criterion. Interestingly, an increase in altitude did not significantly affect the R^2 and intraclass correlation coefficient values of VO_2max assessed by the SWT, suggesting its high compatibility for hypobaric hypoxic conditions. To keep up with the exercise pattern used in the SWT, it is more rigorous to perform the CPET program on a treadmill than on a cycle ergometer to minimize the impact of the different muscular factors. However, there are several distinct differences between a treadmill and a bicycle ergometer (ie, space, safety, and costs)—all of which are conducive to a bicycle ergometer. It is only possible to change the slope and speed in the treadmill exercise. However, a cycle ergometer attains the linearity in workload increment well [29]. Thus, a cycle ergometer program should be preferred over a treadmill program at high altitudes. Hence, we finally opted to use a cycle ergometer for CPET. Nevertheless, the VO_2max measured by the 2 methods has good correlation and consistency.

VO_2max evaluation using CPET is inconvenient in practice. In the past few decades, several new methods for estimating VO_2max have been investigated through a submaximal exercise protocol, including the Queen college step test [30], 20-m shuttle run test [31], and PWC170 [32,33]. Although they are easy to perform, the accuracy of the indirect method in estimating VO_2max remains controversial. Thus, more variables such as basic parameters (age, sex, BMI) [34] and exercise indicators (maximal heart rate, speed, and covered distance) [35,36] were utilized in discrepant equations for more accuracy in subsequent studies. For instance, Marsh [37] found a 4-stage incremental running program estimating VO_2max well (standard error of estimate=3.98-4.08 $mL\cdot kg^{-1}\cdot min^{-1}$; r=0.642-0.646). However, this equation cannot be applied to the general population because correlation data were obtained for male athletes [37]. Instead of simply substituting variables into the equation, the smartwatch employed an algorithm called *Firstbeat* to evaluate VO_2max in daily life. One of the key features of this patented technology is monitoring the running speed along with the heart rate continuously during each workout and automatically excluding the data without a linear relationship. A white paper of *Firstbeat* claimed that based on a database of 2690 freely performed runs by 79 individuals, its accuracy was up to 95% (MAPE<5%) and the error was below 3.5 $mL\cdot kg^{-1}\cdot min^{-1}$ [14]. For perspective, it is superior to most other indirect submaximal tests (10%-15%) and approaches the direct laboratory test (approximately 5%). Thus, the FFT method can be commonly used to estimate VO_2max when high-intensity exercise is limited or laboratory equipment is unavailable. Düking et al [38] found a coefficient of variation of 4% between Garmin watch and the criterion measure over the VO_2 peak range from 38 $mL\cdot kg^{-1}\cdot min^{-1}$ to 61 $mL\cdot kg^{-1}\cdot min^{-1}$ through a small sample study. In this regard, the smartwatch with FFT is a portable device with satisfactory accuracy for estimating VO_2max in the submaximal exercise protocol. Importantly, they also reported that the MAPE between the smartwatch and criterion measure was 7.1% when analyzing VO_2 peak below 45 $mL\cdot kg^{-1}\cdot min^{-1}$ whose discrepancy was less inaccurate in our study. The MAPE in this study was 6% and 6.8% at low and high altitudes, respectively. In [18], the 23 participants were of Caucasian origin, while in our study, the 46 participants were Chinese healthy adults. Hence, different race, gender, age, height, body weight and physical activity level of individuals may well explain differences in VO_2max, both between the other reports [39] and this study.

VO_2max represents the maximum oxygen utilization capacity of an individual. In normoxic conditions, a higher VO_2max indicates greater exercise capacity and better cardiorespiratory fitness [40-42]. However, due to PO_2 decrease (approximately 63% of low altitude at 3700 m) at high altitude, arterial SpO_2 and the amount of oxygen that eventually reaches the tissue and organ are reduced [43]. Thus, aerobic metabolic capacity and VO_2max are significantly inhibited. To cope with hypoxia at high altitude, the body undergoes a series of physiological compensations in the cardiovascular and respiratory systems, such as increased heart rate [44,45], blood pressure [46], and respiratory rate [47]. Although these compensatory responses can compensate for oxygen insufficiency in a short period, long-term exposure may lead to irreversible changes such as pulmonary hypertension, chronic pulmonary disease, and heart failure [48,49]. VO_2max has been an effective predictive indicator of mortality and rehospitalization in patients with chronic cardiovascular and respiratory diseases [50-52]. However, it has rarely been reported that VO_2max contributes to the prognosis and rehabilitation of acute and chronic mountain illnesses [50]. In this study, we present the first evidence of VO_2max as a predictor of AMS in a clinical trial.

It can be concluded from our results that individuals with a higher VO_2max are unlikely to develop AMS. ROC analysis demonstrated that VO_2max-CPET and VO_2max-SWT showed a similar predictive value, particularly for VO_2max-SWT, with a specificity up to 88.46% for a cutoff value of 29.5 $mL\cdot kg^{-1}\cdot min^{-1}$. The high specificity ensures a low incidence of AMS when VO_2max is below the cutoff value of lowlanders unsuitable for acute high-altitude exposure. In addition, we found that RDW-CV was more closely related to AMS than other routine blood parameters. We believe that the low RDW-CV represents a uniform distribution of erythrocyte developmental states, indicating a more effective compensation response of RBC under acute hypoxia stress [53,54]. Although RDW-CV was an independent predictor for AMS and RDW-CV combined with VO_2max showed a higher AUC, it is more convenient for individuals to obtain information on the potential suffering probability by using a smartwatch in daily life. Moreover, models 1 and 2 in Table 4 showed no statistical difference in AMS prediction compared to the single VO_2max model. Therefore, SWT-based VO_2max estimation can conveniently identify AMS-susceptible individuals and help evaluate the cardiorespiratory function and working capacity, which can benefit high-altitude travelers and workers and reduce the consumption of medical resources at high altitude. Although we cannot extend our validity claims to the entire population due to the small sample size and insufficiently diverse population characteristics, the conclusion that people with lower oxygen intake are more likely to develop AMS is well-founded and has been recently reported [55].

The VO_2max estimated by SWT and measured by CPET has high consistency, indicating that smartwatches may replace the

CPET system to obtain VO_2max accurately and objectively by monitoring the common physical activities with a portable, low-cost system. After each exercise, the smartwatch can measure and record the exercise information, integrate the calculation, and update the VO_2max value. In addition, VO_2max measured at low altitudes is highly correlated with the occurrence of AMS, with satisfactory prediction performance. Therefore, it is feasible to use a smartwatch to measure VO_2max at low altitudes to evaluate the possibility of AMS. In the future, this will benefit tourists, temporary workers, and other individuals who plan to travel at high altitudes and will help in identifying participants susceptible to AMS before high-altitude exposure.

To advance this field, several measures are required. First, there is an urgent need for validation standards for smartwatch devices to enable standardized research. Second, the open disclosure of commercial validation studies can enable better resource usage, as studies will not have to be repeated unnecessarily. Third, further development of smartwatch devices will allow new possibilities in the field of VO_2max monitoring. Finally, subsequent trials should continue to focus on validating these devices compared to conventional standards and broaden their use and demonstrate new possibilities for accurate VO_2max monitoring.

Limitations

This study had some limitations. Notably, the Lake Louise consensus scoring system (2018 version) is subjective; therefore, we described each symptom as clearly as possible and provided necessary instructions before the participants completed the questionnaire to deal with the subjectivity. Second, running exercise instead of cycling should be implemented by CPET to minimize the inconsistencies in VO_2max with SWT, which was not applied here due to the great safety risk at high altitude. Further studies and repeated measures are required to develop and investigate the predictive models of the SWT method based on submaximal running programs in terms of validity and reliability [11,56]. The present predictive models of the SWT method cannot be extended to the entire population. In the long term, conducting long-term validation research within a large and representative population is the scope of future studies, as the smartwatch will be extensively used by people. In addition, this study has insufficient reliability, as repeated measures analysis was not performed. It is contradictory to perform repeated measurements because measuring on the same day is limited by physical strength. Repeated measurements on different days also affect the assessment of AMS.

Conclusions

Our findings demonstrate that VO_2max estimated by SWT and CPET have good accuracy and agreement at both low and high altitudes. Importantly, smartwatch-based VO_2max at low altitudes was a convenient and effective approach to predict AMS and to identify susceptible individuals following acute high-altitude exposure, particularly by combining the RDW-CV at low altitudes.

Acknowledgments

We appreciate all the volunteers in this clinical trial and the Shigatse hospital in Tibet. This study was supported by grants from the National Natural Science Foundation of China (81730054) and research project of People's Liberation Army (BLJ18J007).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original data of the cardiopulmonary exercise test and smartwatch test in this study.
[7Z File , 224520 KB - mhealth_v11i1e43340_app1.7z]

References

1. Meier D, Collet T, Locatelli I, Cornuz J, Kayser B, Simel DL, et al. Does this patient have acute mountain sickness? The rational clinical examination systematic review. JAMA 2017 Nov 14;318(18):1810-1819. [doi: 10.1001/jama.2017.16192] [Medline: 29136449]
2. Hackett PH, Roach RC. High-altitude illness. N Engl J Med 2001 Jul 12;345(2):107-114. [doi: 10.1056/NEJM200107123450206] [Medline: 11450659]
3. Honigman B, Theis MK, Koziol-McLain J, Roach R, Yip R, Houston C, et al. Acute mountain sickness in a general tourist population at moderate altitudes. Ann Intern Med 1993 Apr 15;118(8):587-592. [doi: 10.7326/0003-4819-118-8-199304150-00003] [Medline: 8452324]
4. MacInnis MJ, Carter EA, Freeman MG, Pandit BP, Siwakoti A, Subedi A, et al. A prospective epidemiological study of acute mountain sickness in Nepalese pilgrims ascending to high altitude (4380 m). PLoS One 2013;8(10):e75644 [FREE Full text] [doi: 10.1371/journal.pone.0075644] [Medline: 24130729]

5. Richalet J, Larmignat P, Poitrine E, Letournel M, Canouï-Poitrine F. Physiological risk factors for severe high-altitude illness: a prospective cohort study. Am J Respir Crit Care Med 2012 Jan 15;185(2):192-198. [doi: 10.1164/rccm.201108-1396OC] [Medline: 22071330]
6. Yang J, Zhang L, Liu C, Zhang J, Yu S, Yu J, et al. Trimetazidine attenuates high-altitude fatigue and cardiorespiratory fitness impairment: A randomized double-blinded placebo-controlled clinical trial. Biomed Pharmacother 2019 Aug;116:109003 [FREE Full text] [doi: 10.1016/j.biopha.2019.109003] [Medline: 31125823]
7. Herdy AH, Ritt LEF, Stein R, Araújo CGSD, Milani M, Meneghelo RS, et al. Cardiopulmonary exercise test: background, applicability and interpretation. Arq Bras Cardiol 2016 Nov;107(5):467-481 [FREE Full text] [doi: 10.5935/abc.20160171] [Medline: 27982272]
8. Rissanen AE, Rottensteiner M, Kujala UM, Kurkela JLO, Wikgren J, Laukkanen JA. Cardiorespiratory fitness estimation based on heart rate and body acceleration in adults with cardiovascular risk factors: validation study. JMIR Cardio 2022 Oct 25;6(2):e35796 [FREE Full text] [doi: 10.2196/35796] [Medline: 36282560]
9. Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022 Jul 01;24(7):e36690 [FREE Full text] [doi: 10.2196/36690] [Medline: 35776492]
10. Beltz NM, Gibson AL, Janot JM, Kravitz L, Mermier CM, Dalleck LC. Graded exercise testing protocols for the determination of VOMax: historical perspectives, progress, and future considerations. J Sports Med (Hindawi Publ Corp) 2016;2016:3968393 [FREE Full text] [doi: 10.1155/2016/3968393] [Medline: 28116349]
11. Webster DE, Tummalacherla M, Higgins M, Wing D, Ashley E, Kelly VE, et al. Smartphone-based VO2max measurement with heart snapshot in clinical and real-world settings with a diverse population: validation study. JMIR Mhealth Uhealth 2021 Jun 04;9(6):e26006 [FREE Full text] [doi: 10.2196/26006] [Medline: 34085945]
12. Helgerud J, Haglo H, Hoff J. Prediction of VO2max from submaximal exercise using the smartphone application MyWorkout go: validation study of a digital health method. JMIR Cardio 2022 Aug 04;6(2):e38570 [FREE Full text] [doi: 10.2196/38570] [Medline: 35925653]
13. Mikko S, Aki P, Veli-Pekka K, Sami S, Joni K. Fitness test. USPTO. 2011 Feb 17. URL: https://image-ppubs.uspto.gov/dirsearch-public/print/downloadPdf/20110040193 [accessed 2022-10-14]
14. Gao WD, Nuuttila O, Fang HB, Chen Q, Chen X. A new fitness test of estimating VO in well-trained rowing athletes. Front Physiol 2021;12:701541 [FREE Full text] [doi: 10.3389/fphys.2021.701541] [Medline: 34276423]
15. MacInnis MJ, Nugent SF, MacLeod KE, Lohse KR. Methods to estimate V˙O2max upon acute hypoxia exposure. Med Sci Sports Exerc 2015 Sep;47(9):1869-1876. [doi: 10.1249/MSS.0000000000000628] [Medline: 25668403]
16. Roi GS, Giacometti M, von Duvillard SP. Marathons in altitude. Med Sci Sports Exerc 1999 May;31(5):723-728. [doi: 10.1097/00005768-199905000-00016] [Medline: 10331894]
17. Gonzalez- Alonso J, Mortensen SP. Comments of point:counterpoint: maximal oxygen uptake is/is not limited by a central nervous system governor. J Appl Physiol (1985) 2009 Jan;106(1):344-345 [FREE Full text] [doi: 10.1152/japplphysiol.zdg-8326.pcpcomm.2008] [Medline: 19202591]
18. Elliott AD, Skowno J, Prabhu M, Noakes TD, Ansley L. Evidence of cardiac functional reserve upon exhaustion during incremental exercise to determine VO2max. Br J Sports Med 2015 Jan;49(2):128-132. [doi: 10.1136/bjsports-2012-091752] [Medline: 23293009]
19. Otto JM, Montgomery HE, Richards T. Haemoglobin concentration and mass as determinants of exercise performance and of surgical outcome. Extrem Physiol Med 2013 Nov 26;2(1):33 [FREE Full text] [doi: 10.1186/2046-7648-2-33] [Medline: 24280034]
20. Yang J, Tan H, Sun M, Chen R, Zhang J, Liu C, et al. Prediction of high-altitude cardiorespiratory fitness impairment using a combination of physiological parameters during exercise at sea level and genetic information in an integrated risk model. Front Cardiovasc Med 2021;8:719776 [FREE Full text] [doi: 10.3389/fcvm.2021.719776] [Medline: 35071338]
21. Claes J, Cornelissen V, McDermott C, Moyna N, Pattyn N, Cornelis N, et al. Feasibility, acceptability, and clinical effectiveness of a technology-enabled cardiac rehabilitation platform (physical activity toward health-i): randomized controlled trial. J Med Internet Res 2020 Feb 04;22(2):e14221 [FREE Full text] [doi: 10.2196/14221] [Medline: 32014842]
22. Ganesananthan S, Rajkumar CA, Foley M, Thompson D, Nowbar AN, Seligman H, et al. Cardiopulmonary exercise testing and efficacy of percutaneous coronary intervention: a substudy of the ORBITA trial. Eur Heart J 2022 Sep 01;43(33):3132-3145 [FREE Full text] [doi: 10.1093/eurheartj/ehac260] [Medline: 35639660]
23. Roach RC, Hackett PH, Oelz O, Bärtsch P, Luks AM, MacInnis MJ, Lake Louise AMS Score Consensus Committee. The 2018 Lake Louise acute mountain sickness score. High Alt Med Biol 2018 Mar;19(1):4-6 [FREE Full text] [doi: 10.1089/ham.2017.0164] [Medline: 29583031]
24. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007;17(4):571-582 [FREE Full text] [doi: 10.1080/10543400701329422] [Medline: 17613642]
25. Koons NJ, Suresh MR, Schlotman TE, Convertino VA. Interrelationship between sex, age, blood volume, and VO. Aerosp Med Hum Perform 2019 Apr 01;90(4):362-368. [doi: 10.3357/AMHP.5255.2019] [Medline: 30922423]

26. O'Donovan G, Bakrania K, Ghouri N, Yates T, Gray LJ, Hamer M, et al. Nonexercise equations to estimate fitness in White European and South Asian men. Med Sci Sports Exerc 2016 May;48(5):854-859. [doi: 10.1249/MSS.0000000000000836] [Medline: 26694847]
27. Ogawa T, Spina RJ, Martin WH, Kohrt WM, Schechtman KB, Holloszy JO, et al. Effects of aging, sex, and physical training on cardiovascular responses to exercise. Circulation 1992 Aug;86(2):494-503. [doi: 10.1161/01.cir.86.2.494] [Medline: 1638717]
28. Mollard P, Woorons X, Letournel M, Lamberto C, Favret F, Pichon A, et al. Determinants of maximal oxygen uptake in moderate acute hypoxia in endurance athletes. Eur J Appl Physiol 2007 Aug;100(6):663-673. [doi: 10.1007/s00421-007-0457-0] [Medline: 17534646]
29. Agostoni P, Dumitrescu D. How to perform and report a cardiopulmonary exercise test in patients with chronic heart failure. Int J Cardiol 2019 Aug 01;288:107-113. [doi: 10.1016/j.ijcard.2019.04.053] [Medline: 31047701]
30. Chatterjee S, Chatterjee P, Mukherjee PS, Bandyopadhyay A. Validity of Queen's College step test for use with young Indian men. Br J Sports Med 2004 Jun;38(3):289-291 [FREE Full text] [doi: 10.1136/bjsm.2002.002212] [Medline: 15155428]
31. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci 1988;6(2):93-101. [doi: 10.1080/02640418808729800] [Medline: 3184250]
32. Rowland TW, Rambusch JM, Staab JS, Unnithan VB, Siconolfi SF. Accuracy of physical working capacity (PWC170) in estimating aerobic fitness in children. J Sports Med Phys Fitness 1993 Jun;33(2):184-188. [Medline: 8412055]
33. Yang T, Li X, Qin J, Li S, Yu J, Zhang J, et al. High altitude-induced borderline pulmonary hypertension impaired cardiorespiratory fitness in healthy young men. Int J Cardiol 2015 Feb 15;181:382-388. [doi: 10.1016/j.ijcard.2014.12.044] [Medline: 25555284]
34. Kline GM, Porcari JP, Hintermeister R, Freedson PS, Ward A, McCarron RF, et al. Estimation of VO2max from a one-mile track walk, gender, age, and body weight. Med Sci Sports Exerc 1987 Jun;19(3):253-259. [Medline: 3600239]
35. Cooper KH. A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. JAMA 1968 Jan 15;203(3):201-204. [Medline: 5694044]
36. Weller IM, Thomas SG, Gledhill N, Paterson D, Quinney A. A study to validate the modified Canadian Aerobic Fitness Test. Can J Appl Physiol 1995 Jun;20(2):211-221. [doi: 10.1139/h95-015] [Medline: 7640647]
37. Marsh CE. Evaluation of the American College of Sports Medicine submaximal treadmill running test for predicting VO2max. J Strength Cond Res 2012 Feb;26(2):548-554. [doi: 10.1519/JSC.0b013e3181bac56e] [Medline: 22262016]
38. Düking P, Van Hooren B, Sperlich B. Assessment of peak oxygen uptake with a smartwatch and its usefulness for training of runners. Int J Sports Med 2022 Jun;43(7):642-647 [FREE Full text] [doi: 10.1055/a-1686-9068] [Medline: 35094376]
39. Black NE, Vehrs PR, Fellingham GW, George JD, Hager R. Prediction of VO2max in children and adolescents using exercise testing and physical activity questionnaire data. Res Q Exerc Sport 2016;87(1):89-100. [doi: 10.1080/02701367.2015.1124969] [Medline: 26889586]
40. Lee D, Artero EG, Sui X, Blair SN. Mortality trends in the general population: the importance of cardiorespiratory fitness. J Psychopharmacol 2010 Nov;24(4 Suppl):27-35 [FREE Full text] [doi: 10.1177/1359786810382057] [Medline: 20923918]
41. Bacon AP, Carter RE, Ogle EA, Joyner MJ. VO2max trainability and high intensity interval training in humans: a meta-analysis. PLoS One 2013;8(9):e73182 [FREE Full text] [doi: 10.1371/journal.pone.0073182] [Medline: 24066036]
42. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. Med Sci Sports Exerc 1995 Sep;27(9):1292-1301. [Medline: 8531628]
43. Wagner PD. The physiological basis of reduced VO2max in Operation Everest II. High Alt Med Biol 2010;11(3):209-215. [doi: 10.1089/ham.2009.1058] [Medline: 20919887]
44. Richalet J. Physiological and clinical implications of adrenergic pathways at high altitude. Adv Exp Med Biol 2016;903:343-356. [doi: 10.1007/978-1-4899-7678-9_23] [Medline: 27343107]
45. White M, Yanowitz F, Gilbert EM, Larrabee P, O'Connell JB, Anderson JL, et al. Role of beta-adrenergic receptor downregulation in the peak exercise response in patients with heart failure due to idiopathic dilated cardiomyopathy. Am J Cardiol 1995 Dec 15;76(17):1271-1276. [doi: 10.1016/s0002-9149(99)80355-5] [Medline: 7503009]
46. Liu Y, Zhang J, Gao X, Wu X, Yu J, Chen J, et al. Correlation between blood pressure changes and AMS, sleeping quality and exercise upon high-altitude exposure in young Chinese men. Mil Med Res 2014;1:19 [FREE Full text] [doi: 10.1186/2054-9369-1-19] [Medline: 25722875]
47. Lipsitz LA, Hashimoto F, Lubowsky LP, Mietus J, Moody GB, Appenzeller O, et al. Heart rate and respiratory rhythm dynamics on ascent to high altitude. Br Heart J 1995 Oct;74(4):390-396 [FREE Full text] [doi: 10.1136/hrt.74.4.390] [Medline: 7488453]
48. Lichtblau M, Saxer S, Furian M, Mayer L, Bader PR, Scheiwiller PM, et al. Cardiac function and pulmonary hypertension in Central Asian highlanders at 3250 m. Eur Respir J 2020 Aug;56(2):1902474 [FREE Full text] [doi: 10.1183/13993003.02474-2019] [Medline: 32430419]
49. Penaloza D, Arias-Stella J. The heart and pulmonary circulation at high altitudes: healthy highlanders and chronic mountain sickness. Circulation 2007 Mar 06;115(9):1132-1146. [doi: 10.1161/CIRCULATIONAHA.106.624544] [Medline: 17339571]

50. De Cannière H, Smeets CJP, Schoutteten M, Varon C, Van Hoof C, Van Huffel S, et al. Using biosensors and digital biomarkers to assess response to cardiac rehabilitation: observational study. J Med Internet Res 2020 May 20;22(5):e17326 [FREE Full text] [doi: 10.2196/17326] [Medline: 32432552]
51. Clausen JSR, Marott JL, Holtermann A, Gyntelberg F, Jensen MT. Midlife cardiorespiratory fitness and the long-term risk of mortality: 46 years of follow-up. J Am Coll Cardiol 2018 Aug 28;72(9):987-995 [FREE Full text] [doi: 10.1016/j.jacc.2018.06.045] [Medline: 30139444]
52. Howden EJ, Sarma S, Lawley JS, Opondo M, Cornwell W, Stoller D, et al. Reversing the cardiac effects of sedentary aging in middle age-a randomized controlled trial: implications for heart failure prevention. Circulation 2018 Apr 10;137(15):1549-1560 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.030617] [Medline: 29311053]
53. Storz JF, Bautista NM. Altitude acclimatization, hemoglobin-oxygen affinity, and circulatory oxygen transport in hypoxia. Mol Aspects Med 2022 Apr;84:101052 [FREE Full text] [doi: 10.1016/j.mam.2021.101052] [Medline: 34879970]
54. Agarwal S. Red cell distribution width, inflammatory markers and cardiorespiratory fitness: results from the National Health and Nutrition Examination Survey. Indian Heart J 2012;64(4):380-387 [FREE Full text] [doi: 10.1016/j.ihj.2012.06.006] [Medline: 22929821]
55. Seiler T, Nakas CT, Brill A, Hefti U, Hilty MP, Perret-Hoigné E, et al. Do cardiopulmonary exercise tests predict summit success and acute mountain sickness? A prospective observational field study at extreme altitude. Br J Sports Med. Online ahead of print 2023 Mar 10. [doi: 10.1136/bjsports-2022-106211] [Medline: 36898769]
56. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: cross-sectional study. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e11889 [FREE Full text] [doi: 10.2196/11889] [Medline: 30888332]

Abbreviations

AMS: acute mountain sickness
AUC: area under the curve
CPET: cardiopulmonary exercise test
FFT: Firstbeat fitness test
MAPE: mean absolute percentage error
OR: odds ratio
RBC: red blood cell
RDW-CV: red blood cell distribution width-coefficient of variation
ROC: receiver operating characteristic
SpO2: oxygen saturation
SWT: smartwatch test
VO_2max: maximum oxygen consumption

Edited by G Eysenbach, L Buis; submitted 09.10.22; peer-reviewed by Z Zhang, 家 余; comments to author 29.11.22; revised version received 11.12.22; accepted 09.06.23; published 06.07.23.

Please cite as:
Ye X, Sun M, Yu S, Yang J, Liu Z, Lv H, Wu B, He J, Wang X, Huang L
Smartwatch-Based Maximum Oxygen Consumption Measurement for Predicting Acute Mountain Sickness: Diagnostic Accuracy Evaluation Study
JMIR Mhealth Uhealth 2023;11:e43340
URL: https://mhealth.jmir.org/2023/1/e43340
doi:10.2196/43340
PMID:37410528

Original Paper

Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study

Kiran K G Ravindran[1,2], PhD; Ciro della Monica[1,2], PhD; Giuseppe Atzori[1,2], MSc, RPGST; Damion Lambert[1,2], MSc; Hana Hassanin[2,3,4], MBBS, Dr med; Victoria Revell[1,2], PhD; Derk-Jan Dijk[1,2], PhD

[1]Surrey Sleep Research Centre, School of Biosciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[2]UK Dementia Research Institute, Care Research and Technology Centre at Imperial College, London and the University of Surrey, Guildford, United Kingdom

[3]Surrey Clinical Research Facility, School of Biosciences, Faculty of Health and Medical Sciences, Guildford, United Kingdom

[4]National Institute for Health Research - Royal Surrey Clinical Research Facility, Guildford, United Kingdom

Corresponding Author:
Kiran K G Ravindran, PhD
Surrey Sleep Research Centre, School of Biosciences, Faculty of Health and Medical Sciences
University of Surrey
Surrey Clinical Research Building
Egerton Road
Guildford, GU27XP
United Kingdom
Phone: 44 01483683709
Email: k.guruswamyravindran@surrey.ac.uk

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2023/1/e54856

Abstract

Background: Contactless sleep technologies (CSTs) hold promise for longitudinal, unobtrusive sleep monitoring in the community and at scale. They may be particularly useful in older populations wherein sleep disturbance, which may be indicative of the deterioration of physical and mental health, is highly prevalent. However, few CSTs have been evaluated in older people.

Objective: This study evaluated the performance of 3 CSTs compared to polysomnography (PSG) and actigraphy in an older population.

Methods: Overall, 35 older men and women (age: mean 70.8, SD 4.9 y; women: n=14, 40%), several of whom had comorbidities, including sleep apnea, participated in the study. Sleep was recorded simultaneously using a bedside radar (Somnofy [Vital Things]: n=17), 2 undermattress devices (Withings sleep analyzer [WSA; Withings Inc]: n=35; Emfit-QS [Emfit; Emfit Ltd]: n=17), PSG (n=35), and actigraphy (Actiwatch Spectrum [Philips Respironics]: n=18) during the first night in a 10-hour time-in-bed protocol conducted in a sleep laboratory. The devices were evaluated through performance metrics for summary measures and epoch-by-epoch classification. PSG served as the gold standard.

Results: The protocol induced mild sleep disturbance with a mean sleep efficiency (SEFF) of 70.9% (SD 10.4%; range 52.27%-92.60%). All 3 CSTs overestimated the total sleep time (TST; bias: >90 min) and SEFF (bias: >13%) and underestimated wake after sleep onset (bias: >50 min). Sleep onset latency was accurately detected by the bedside radar (bias: <6 min) but overestimated by the undermattress devices (bias: >16 min). CSTs did not perform as well as actigraphy in estimating the all-night sleep summary measures. In an epoch-by-epoch concordance analysis, the bedside radar performed better in discriminating sleep versus wake (Matthew correlation coefficient [MCC]: mean 0.63, SD 0.12, 95% CI 0.57-0.69) than the undermattress devices (MCC of WSA: mean 0.41, SD 0.15, 95% CI 0.36-0.46; MCC of Emfit: mean 0.35, SD 0.16, 95% CI 0.26-0.43). The accuracy of identifying rapid eye movement and light sleep was poor across all CSTs, whereas deep sleep (ie, slow wave sleep) was predicted with moderate accuracy (MCC: >0.45) by both Somnofy and WSA. The deep sleep duration estimates of Somnofy correlated (r^2=0.60; P<.01) with electroencephalography slow wave activity (0.75-4.5 Hz) derived from PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.58; Emfit: r^2=0.11, P=.21).

Conclusions: These CSTs overestimated the TST, and sleep stage prediction was unsatisfactory in this group of older people in whom SEFF was relatively low. Although it was previously shown that CSTs provide useful information on bed occupancy, which may be useful for particular use cases, the performance of these CSTs with respect to the TST and sleep stage estimation requires improvement before they can serve as an alternative to PSG in estimating most sleep variables in older individuals.

(JMIR Mhealth Uhealth 2023;11:e46338) doi:10.2196/46338

KEYWORDS

contactless sleep technologies; evaluation; nearables; polysomnography; older adults; sleep; Withings sleep analyzer; Emfit; Somnofy

Introduction

Background

Sleep is a major determinant of the quality of life, and sleep disturbances are a risk factor for a variety of health conditions. Longitudinal monitoring of objective sleep measures using unobtrusive, low-cost technologies will allow data-driven identification of clinical biomarkers and covariates of sleep and health and facilitate better monitoring of the efficacy of sleep interventions [1,2]. This is particularly relevant in aging, as sleep disturbances increase with age and are a risk factor for several health conditions prevalent in older adults, including dementia.

Polysomnography (PSG) is considered the gold standard for evaluating sleep, but longitudinal implementation of PSG at scale is not feasible because of the cost and burden it imposes on the user. Rest-activity monitoring (actigraphy) through a wrist-worn wearable device is currently the most widely used alternative for monitoring sleep in real-world settings. However, actigraphy is limited to a binary classification (sleep vs wake) and underestimates wake, especially in disrupted sleep and sleep disorders [3]. Furthermore, similar to PSG, clinical-grade actigraphy devices are of relatively high cost.

Consumer-grade low-cost wearable devices (wearables) are a potential alternative to clinical-grade actigraphy devices. However, wearables are not an ideal solution for longitudinal monitoring because they still pose a burden to the user and rely on portable battery technology (need to be recharged); therefore, their acceptability, especially in older people and people living with dementia, may be low [4-7].

Contactless Sleep Technology

Contactless sleep technologies (CSTs), also known as nearables, are of great interest for conducting longitudinal sleep recordings in general and in older individuals and those with mild cognitive impairment or dementia in particular. This is because, unlike wearables, CSTs do not have to be worn, are inconspicuous (embedded into the living environment), do not impose any burden on the user, and are not constrained by limited battery life because they are wired to the mains electricity. Furthermore, CSTs are designed within the context of digital platforms that allow data to be collected, relayed to cloud storage facilities, and processed to provide objective sleep and vital sign measures.

With CSTs demonstrating acceptable accuracy compared with the existing gold standards, they could play an integral part in creating a "bedroom of the future," enabling long-term, continuous, and real-time remote monitoring of sleep and physiology at night by clinicians and health care providers. They may also support and monitor the effectiveness of interventions for improving sleep, which may ultimately facilitate longer independent living for older adults [8,9].

Bedside radars and undermattress devices are the 2 most commonly used types of CSTs. Bedside radars use radar technology (commonly ultrawideband radio frequency), whereas the various undermattress devices use several sensing technologies, such as pneumatic, piezoelectric, and electromechanical films. Although contactless devices use a variety of sensing approaches, the main information acquired by these devices is based on the ballistographic signal. This is a composite signal containing a wealth of information on body movements, breathing, and cardiac activity, which can be used to estimate wake and sleep stages [10].

The plethora of low-cost CSTs opens up a wide range of possibilities for longitudinal sleep monitoring at scale in home settings. However, only a handful of these consumer-grade devices have been evaluated against PSG or actigraphy [8,11]. Without rigorous evaluation, there are also potential risks of the multimodal capabilities and clinical value of CSTs not being used to their full potential and the data being misinterpreted [12]. Several of these validation studies have important limitations. CST sleep prediction algorithms are usually trained on healthy participants, which limits the heterogeneity of the training data and potentially affects real-world performance in more heterogeneous populations. Most studies that intend to validate these devices and their algorithms, which are preferably referred to as evaluation studies rather than validation studies [13], are conducted in young healthy adults without sleep disturbances rather than in older participants with health problems and sleep disturbance. Furthermore, most evaluation studies do not offer interdevice comparisons, thus limiting conclusions on the relative performance of the devices [14-17]. Another important but often overlooked issue in evaluating CSTs is the consideration of the period over which "sleep" is analyzed. In PSG studies conducted in sleep laboratories, the analysis period (AP) is defined as the interval between lights off and lights on. For actigraphy studies at home, the American Academy of Sleep Medicine (AASM) [18] recommendation is that the AP be derived from a sleep diary [19]. The sleep measures are then estimated using this AP. By contrast, consumer sleep technologies rely on the automatic estimation of the rest (or time-in-bed) period. In many evaluation studies, the AP is nevertheless set to the AP according to the PSG, but this may yield biased performance metrics that are not relevant to the use of these devices in the real world. A final consideration for evaluation studies is the choice of performance measures. Traditional measures such as sensitivity and accuracy may not be appropriate in cases in which there are imbalances in the number of observations across classes, such as in sleep-wake classification [20,21].

Evaluating CSTs against gold-standard PSG and actigraphy in older adults to understand their accuracy and reliability requires evaluation protocols that address the above-discussed issues. We have previously shown that CSTs provide accurate information on “bed occupancy” in older people in a home setting [22]. Here, the performance of 3 contactless devices, namely a bedside radar (Somnofy [Vital Things]) and 2 undermattress devices (Withings sleep analyzer [WSA; Withings Inc] and Emfit-QS [Emfit; Emfit Ltd]), in estimating sleep parameters was evaluated in comparison with those of gold-standard PSG and actigraphy (Actiwatch Spectrum [AWS; Philips Respironics]) in a heterogeneous population of older men and women. We made use of the “first night effect,” which refers to the reduced quality of sleep when participants are sleeping in a novel environment [23]. In addition, an extended (10-hour) time-in-bed period was imposed to better mimic sleep patterns in dementia. We addressed the pitfalls in evaluating CSTs and quantified the agreement between these CSTs and PSG and between these CSTs and actigraphy (AWS) using different performance measures and AP definitions. The usefulness of CSTs may vary across use cases. For some use cases, a simple all-night estimate of the total sleep time (TST) may be sufficient, whereas for other use cases, it is necessary to estimate sleep stages and epoch-by-epoch (EBE) concordance. The performance of these devices was, therefore, evaluated both with respect to all-night summary measures or EBE concordance and at various levels of characterization of the sleep-wake phenotype, that is, from a simple 2 category classification, namely wakefulness and sleep, to 4 category classification, namely wake, light sleep (LS), deep sleep (DS), and rapid eye movement (REM) sleep.

Methods

Study Population

Participants were recruited to the study via a targeted search of the Surrey Clinical Research Facility participant database, followed by telephone screening and self-reported assessment of health. During an in-person screening visit, an array of assessments and clinical procedures was performed to determine suitability for the study. Participants were considered eligible if they met the inclusion criteria (being aged 65 to 85 y, living independently, being a nonsmoker, having self-declared stable medical conditions, and consuming <28 units of alcohol per wk). Data were collected from 2 cohorts: cohort 1 with 18 participants (January to March 2020) and cohort 2 with 17 participants (June to November 2021).

Study Protocol

Participants came to the sleep laboratory (Surrey Sleep Research Centre, Guildford, United Kingdom) for 1 overnight sleep recording after they had been using the devices for 8 to 14 days at home (the analyses of the home data have been reported elsewhere [22]). We deliberately included no adaptation night in this protocol and used the first night effect to create conditions of mild sleep disturbance [23]. During the overnight in-laboratory study, participants were provided with a time-in-bed period of 10 hours. Full PSG was recorded according to the AASM guidelines using the SomnoHD PSG system (SOMNOmedics GmbH). Sleep (sleep stages: wake, REM, stage N1 of non-REM sleep [N1], stage N2 of non-REM sleep [N2], and stage N3 of non-REM sleep [N3]) was scored at 30-second intervals in the DOMINO software environment (SOMNOmedics GmbH) by 2 independent scorers, and a consensus hypnogram was generated. All recordings were visually inspected, and artifacts were removed. We estimated 30-second epoch-wise slow wave activity (SWA) power (0.75-4.5 Hz) from the electroencephalography (EEG) spectrogram created by the fast Fourier transform applied to 4-second epochs after tapering with a hamming window. In general, the left frontal referenced to right mastoid (F3-M2) channel derivation was used for the computation of SWA unless the quality of this channel was poor, in which case the right frontal referenced to left mastoid (F4-M1) channel derivation was used.

The PSG sleep summary estimates were computed for the interval from lights off to lights on, as per the AASM guidelines. The apnea hypopnea index (AHI) was obtained by applying >3% oxygen desaturation or a respiratory event accompanied by an arousal as a criterion for identifying hypopnea [18]. Other relevant study population measures such as the Mini-Mental State Examination, Pittsburgh Sleep Quality Index (PSQI), and Epworth Sleepiness Scale (ESS) were collected using questionnaires.

Contactless Sleep Trackers and Actigraphy Device Evaluated

This study was conducted in 2 cohorts. In cohort 1 (n=18), we evaluated the WSA and AWS. The devices evaluated in cohort 2 (n=17) were the WSA, Emfit, and Somnofy bedside radar. The contactless device and AWS data were simultaneously collected along with PSG. Time synchronization was assessed and achieved across the devices (including PSG) by connecting them to the same secure network.

The WSA is a pneumatic undermattress device, and the Emfit is an electromechanical film undermattress device. Both devices were placed underneath the mattress and adjacent to each other at the thoracic level [16,24]. The bedside radar (Somnofy) device uses low-power ultrawideband radar and was placed on a bedside table pointing toward the participant’s thorax [15]. The devices were set up as per the manufacturer’s guidelines. All 3 devices detect activity and physiological parameters to identify sleep stages. The bedside radar (Somnofy) is the only device in the study that performs active sensing using radio waves (via the Doppler radar technique) and collects environmental variables such as light, sound, and particulate pollution.

AWS is a standard clinical-grade actigraphy device that has been widely deployed for home monitoring and evaluated against PSG in the sleep literature [25,26]. Actiware (version 6.0.7; Philips Respironics) was used to set up the device and download and analyze the collected data. AWS was set up to output sleep and activity labels at 60-second epoch and synchronized to the clock time by the Actiware during configuration. The device estimates the wearer's activity in terms of activity counts per epoch and assigns sleep or wake labels using a preset threshold. The activity threshold was set to medium threshold (epochs with an activity count over 40 were labeled by the device as wake) [22].

Collected Data

The WSA and Somnofy data were downloaded using the application programming interface provided by the respective manufacturer, whereas the Emfit device data were downloaded from a simple web interface. The Emfit data consisted of "csv" files, whereas the WSA and Somnofy data consisted of "json" files. A further description of device deployment is provided in *Contactless Sleep Technology* section and Table S1 in Multimedia Appendix 1.

The contactless devices generated a 4-stage hypnogram with DS (assumed to be equivalent to N3 sleep), LS (assumed to be equivalent to N1 or N2 sleep), REM, and wake labels. The temporal resolution of the hypnograms generated by Emfit and Somnofy was 30 seconds, whereas for WSA and AWS, the temporal resolution was 60 seconds. Because the analysis was performed on clock time, we made the relevant daylight-saving correction to the device Coordinated Universal Time series. The Emfit and Somnofy time series were further synchronized to PSG using the activity or movement data from the respective devices through cross-correlation [15]. PSG was available for all the participants, but owing to the differences in device deployment between cohorts 1 and 2, erroneous automated summary, and data loss, the number of nights differed between devices. The total number of nights of data used for each device was as follows: WSA: n=35 (34 for the automated device-selected AP), AWS: n=18, Somnofy: n=17, and Emfit: n=16.

All the 3 contactless devices automatically detected the overnight in-bed periods and generated sleep-wake summary estimates for these periods, including TST, sleep onset latency (SOL), wake after sleep onset (WASO), sleep efficiency (SEFF), and sleep stage (DS, LS, REM, and wake) duration estimates. We combined the LS and DS estimates to derive the non-REM (NREM) duration. The AWS data were downloaded, and sleep-wake time series and sleep summary measures were generated using the Actiware. The APs or rest intervals for sleep summary generation can be determined automatically by the Actiware or set manually by the user.

Performance Assessment Categories and APs

Performance analysis can be broadly separated into 2 categories: sleep summary measures, which provide the cumulative estimates of nocturnal sleep, and EBE concordance, resulting in detailed information on the sleep architecture detection accuracy of the devices under evaluation.

All the evaluated CSTs detected the overnight-in-bed periods based on their respective proprietary algorithms and generated sleep or wake summary measures. The sleep summary estimates associated with this automated device-selected AP are referred to as the "analysis period–automatic (AP-A)." Further, the sleep summary estimates of the devices were calculated using the sleep stage time series during the period from lights off to lights on (referred to as "analysis period–manual [AP-M]"). AP-M allowed us to compare the devices against PSG for the same AP. For the sleep summary agreement evaluation, the above-mentioned AP-A and AP-M estimates were compared with PSG sleep summary estimates.

EBE agreement analysis was performed for the total recording interval of PSG (from the start of PSG recording to the end of PSG recording) and for the period from lights off to lights on for completeness. The performance of each device was compared with that of the gold-standard PSG for all available nights. All the data analyses reported here were performed using MATLAB (version 2021b; Math Works).

To determine the satisfactory level of agreement in sleep summary estimates and EBE concordance between the device and PSG estimates, we used the interscorer difference estimates available in the literature. The metrics used and their agreement thresholds are discussed in the following sections.

The number of participants for whom data were available was not the same for all the CSTs owing to the varied deployment in cohorts 1 and 2 and errors in data collection. For each comparison, the maximum number of available participants was used for performance assessment.

Sleep Summary Agreement Assessment Approach

The sleep or wake summary measures available across all devices were TST (total time spent in sleep during the "lights off" period as assessed by PSG), SOL (time elapsed from lights off to the first incidence of sleep), WASO (time spent in wake during the "lights off" period as assessed by PSG), and SEFF (ratio of the TST to the total recording time (TRT) of PSG expressed as percentage). Hence, these summary measures were used for comparisons across all the devices and APs (AP-A and AP-M). In addition, for the contactless devices, but not AWS, sleep stage duration measures such as LS, DS, REM sleep, and wake durations were also compared. The results of the AP-M analysis are presented in Multimedia Appendix 1. Bland-Altman plots were created to understand the agreement between the device-estimated measures and the gold-standard PSG measures. The Shapiro-Wilk test was performed to check the normality of the differences [27-30]. We found that for most sleep parameters, the differences passed the normality test. The exceptions were the AP of AWS automatic analysis and Somnofy and SOL of Somnofy. These deviations from normality were deemed to be related to the small sample size and outliers, and no corrections were made.

Apart from the bias, limits of agreement, and the associated 95% CIs, we estimated the minimum detectable change and

Pearson correlation (ρ) and assessed the reliability using consistency intraclass correlation (ICC) with 2-way random effects, and effect size for the magnitude of differences (Cohen *d* or standardized difference here) [27].

Sleep summary measures have values ranging from few minutes to hundreds of minutes (SEFF is a percentage measure). Owing to this difference in measurement range and units across the different sleep summary measures, the above-mentioned traditional metrics do not allow direct intermetric comparison across the different devices or the ranking of device performance. To address this problem, we used standardized metrics such as the symmetric mean absolute percentage error (SMAPE) [31] and standardized absolute difference (SAD) to quantify the bias and dispersion in different sleep measures (sleep-wake and sleep stage duration). Given that x is the reference device estimate (PSG) and y is the test device estimate with n simultaneous measurements, SMAPE and SAD are defined as follows:

(1)

(2)

Here, s^2 is the variance of the reference and test device measurements. These measures are directionless and unitless and hence allow for direct comparisons of the measurement agreement across devices and estimates. Standardized differences, Cohen *d*, and SAD were classified as follows: 0.1-<0.3=small; 0.3-<0.5=moderate; ≥0.5=large. SMAPE values ranged from 0% to 100%.

The estimates of the average agreement (ICC) between scorers and the average score reported by Younes et al [32] were used to define the satisfactory agreement level for the sleep summary measures. The ICC thresholds for the different duration estimates were as follows: 0.84 for WASO, 0.75 for REM, 0.65 for NREM, 0.67 for LS; and 0.63 for DS.

EBE Concordance Assessment Approach

The concordance between the sleep stage hypnogram time series automatically generated by each device and the PSG hypnogram was estimated for the total recording interval of PSG. The 5-stage PSG hypnogram was converted into a 4-stage hypnogram similar to the device hypnograms by combining N1 and N2 as LS (LS=N1 or N2) and assuming N3 as DS. The concordance analysis was performed at the level of PSG hypnogram resolution, which was in 30-second intervals. The 60-second WSA and AWS hypnograms were converted to 30-second resolution by imputing the unavailable 30-second epoch data with the label of the next adjacent minute. Epochs scored as artifacts in PSG and missing (WSA) and no presence (Somnofy and Emfit) epochs in the devices were excluded from the concordance analysis, that is, only valid or complete pairs of hypnogram labels between PSG and the devices were used for the analysis. Finally, to achieve an accurate EBE concordance assessment, PSG, AWS, and CST hypnograms were aligned via cross-correlation, and the lag within a 10-epoch window that provided the best alignment and concordance was used.

To investigate the changes in concordance with different sleep staging resolutions, the analysis was performed at the following three levels of the hypnogram: (1) two stages (sleep and wake), (2) three stages (NREM [LS or DS], REM, and wake), and (3) four stages (DS, LS, REM, and wake). In addition, we performed EBE concordance analysis for the lights off period (Multimedia Appendix 1). Sensitivity (sleep prediction accuracy) and specificity (wake prediction accuracy) are reported for sleep or wake EBE analysis. For the different sleep stage concordance analysis, sensitivity, specificity, accuracy, F_1-score, and Matthew correlation coefficient (MCC, which accounts for class imbalance and is a better alternative to the κ metric or its variants) are reported for completeness and consistency with the existing literature [20,21].

The estimates of the interrater reliability reported by Lee et al [33] were used to define the satisfactory agreement level for the EBE concordance. Because MCC and the κ metrics have almost identical values when both metrics are in the positive quadrant [20], we used the κ values reported by Lee et al [33] to define the MCC threshold for satisfactory EBE concordance. The MCC thresholds for the different sleep stages were as follows: 0.70 for sleep or wake, 0.69 for REM, 0.48 for NREM, 0.40 for LS, and 0.57 for DS.

Ethical Considerations

The study received a favorable opinion from the University of Surrey Ethics Committee (reference UEC 2019 065 FHMS) and was conducted in accordance with the Declaration of Helsinki, the Principles of Good Clinical Practice, and relevant guidelines and regulations of the University of Surrey. Potential participants were given detailed information about the study protocol, and they provided written informed consent before any study procedures were performed.

Results

Study Population Characteristics

The study involved a total of 35 participants (age: mean 70.8, SD 4.86; range 65-83 y; women: n=14, 40%; men: n=21, 60%) with no self-reported history of mental health or neurological problems. Cohorts 1 and 2 were similar with respect to demographics and PSG-assessed sleep parameters (Table 1). Approximately 60% (20/35) of the participants had ≥1 self-reported comorbidities, including type-2 diabetes (2/35, 6%), hypertension (2/35, 6%), obesity (BMI>30; 6/35, 17%), and arthritis (6/35, 17%). Reported comorbidities were stable and well controlled, with no recent medication changes or hospitalizations that interfered with the study conduct.

The values shown in Table 1 are mean, SD, and range. The significance of the difference between cohorts 1 and 2 is given along with the effect size (Cohen *d*). Significant differences (P<.05) are highlighted in italics. AP in PSG is the period from the lights off to lights on (AP-M).

None of the participants were below the cutoff (23) for clinically significant cognitive impairment as indexed by the Mini-Mental

State Examination. The PSQI scores were on average <5, with the highest score being 10, which indicates that the majority of the participants did not have clinically significant sleep disturbance (PSQI>5). None of the participants experienced excessive daytime sleepiness as indexed by the ESS (>10). Nevertheless, the clinical PSG recording revealed that 49% (17/35) of the participants had moderate (9/35, 26%; AHI: 15 to <30) or severe (8/35, 23%; AHI: >30) sleep apnea, whereas 46% (16/35) of the participants had mild apnea (AHI: 5 to <15). In view of the health status of the participants, we refer to this study population as "heterogeneous."

The sleep opportunity period (period available for sleep, described here as the AP), set to "lights off" period as per AASM guidelines [18], ranged from 466 to 586 minutes. TST ranged from 282 to 504 minutes, leading to sleep efficiencies ranging from 52.7% to 92.6%. SOL was, on average, within a normal range (normal reference SOL [age 65-79 y]: mean 19.5, 95% CI 15.2-23.8 min [34]), with the longest SOL being 49.5 minutes. All sleep stages were present in all participants. REM sleep duration was relatively short, and WASO was high, with considerable between-participant variation. The definition of the PSG sleep summary measures and a detailed summary of the device sleep summary characteristics of the participants can be found in Tables S2 and S3 in Multimedia Appendix 1.

Table 1. Study population characteristics.

Characteristics	Cohort 1 (n=18), mean (SD; range)	Cohort 2 (n=17), mean (SD; range)	*P* value	Effect size	Pooled (n=35), mean (SD; range)
Demographics					
Age (y)	69.67 (5.04; 65-80)	72 (4.49; 65-83)	.16	−0.50	70.8 (4.86; 65-83)
Gender (women), n (%)	8 (44)	6 (35)	N/A[a]	N/A	14 (40)
BMI (kg/m^2)	27.03 (4.84; 21.87-39.75)	26.42 (4.71; 20-36.8)	.71	0.13	26.73 (4.72; 20-39.75)
AHI[b] (events/h)	20.89 (17.46; 1.6-66.7)	18.96 (13.62; 4.2-58.8)	.72	0.13	19.95 (15.51; 1.6-66.7)
MMSE[c]	28.44 (1.46; 25-30)	28.88 (1.32; 25-30)	.36	−0.32	28.66 (1.39; 25-30)
PSQI[d]	4.22 (1.86: 1-7)	4 (2.35; 1-10)	.76	0.11	4.11 (2.08; 1-10)
ESS[e]	3.72 (2.67; 1-9)	3.47 (2.32; 0-8)	.77	0.10	3.6 (2.47; 0-9)
Polysomnography sleep measures					
TRT[f] (min)	600.03 (1.54; 596.11-603.35)	601.2 (9.95; 590.25-622.43)	.62	−0.17	600.6 (6.93; 590.25-622.43)
AP[g] (min)	571.75 (18.49; 523.78-586)	512.39 (22.79; 466-539.89)	*.001* [h]	*2.95*	542.91 (36.36; 466-586)
TST[i] (min)	422.47 (53.16; 326.5-504)	347.32 (55.46; 282-469)	*.001*	*2.95*	385.97 (65.67; 282-504)
SOL[j] (min)	11.56 (10.62; 0-40.5)	18.91 (13.71; 2-49.5)	.08	−0.62	15.13 (12.60; 0-49.5)
WASO[k] (min)	130.44 (47.56; 60-206.5)	146.09 (56.43; 30.5-240)	.38	−0.31	138.04 (51.89; 30.5-240)
SEFF[l] (%)	73.86 (8.82; 57.28-87.56)	67.92 (11.33; 52.27-92.6)	.09	0.60	70.99 (10.41; 52.27-92.6)
N3[m] (percentage of TST)	20.21 (5.94; 11.58-36.48)	19.74 (9.66; 3.08-34.93)	.86	0.06	19.98 (7.85; 3.08-36.48)
N2[n] (percentage of TST)	47.70 (8.08; 33.69-61.55)	47.27 (7.83; 36.5-61.1)	.87	0.05	47.49 (7.85; 33.69-61.55)
N1[o] (percentage of TST)	17.08 (7.12; 5.58-37.8)	19.22 (9; 7.46-36.89)	.001	−4.15	18.12 (8.04; 5.58-37.81)
REM[p] (percentage of TST)	15 (5.99; 5.49-23.96)	13.77 (4.93; 3.39-19.65)	.51	0.23	14.41 (5.46; 3.39-23.96)
NREM[q] (percentage of TST)	84.99 (5.99; 76.04-94.51)	86.23 (4.93; 80.35-96.6)	.52	−0.23	85.59 (5.46; 76.04-96.6)
Wake (min)	142 (50.83; 62-243.5)	165 (60.12; 37.5-257.5)	.23	−0.43	153.17 (55.94; 37.5-257.5)

[a]N/A: not applicable.
[b]AHI: apnea hypopnea index.
[c]MMSE: Mini-Mental State Examination.
[d]PSQI: Pittsburgh Sleep Quality Index.
[e]ESS: Epworth Sleepiness Scale.
[f]TRT: total recording time.
[g]AP: analysis period.
[h]Significant differences (*P*<.05) are italicized.
[i]TST: total sleep time.
[j]SOL: sleep onset latency.
[k]WASO: wake after sleep onset.
[l]SEFF: sleep efficiency.
[m]N3: stage N3 of non–rapid eye movement sleep.
[n]N2: stage N2 of non–rapid eye movement sleep.
[o]N1: stage N1 of non–rapid eye movement sleep.
[p]REM: rapid eye movement.
[q]NREM: non–rapid eye movement.

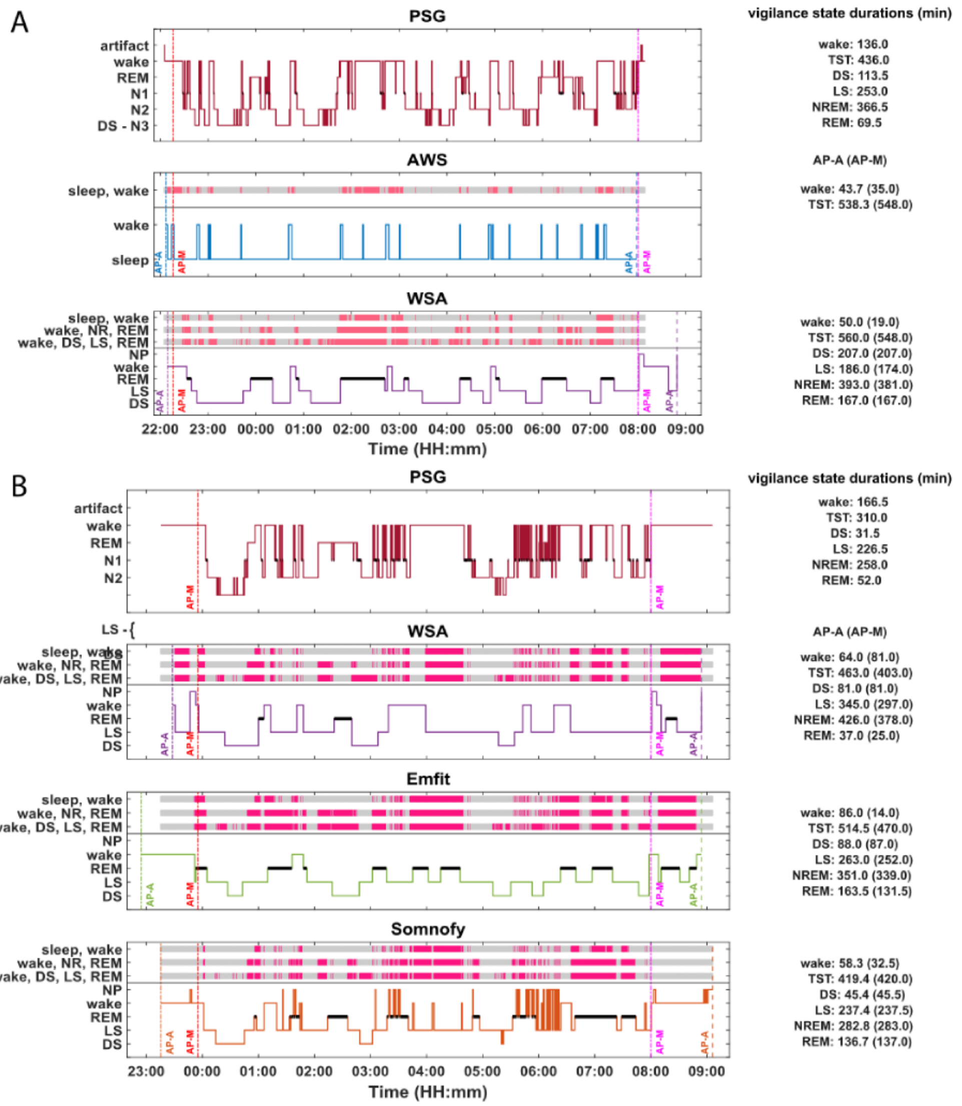
Examples of All-Night Recordings Obtained via PSG, AWS, and CSTs

Figure 1 shows example hypnograms for 2 individuals with concurrent PSG, AWS, and CST recordings, including their EBE concordance with the PSG as well as all-night sleep summary estimates (on the right). The time series of the sleep stages is displayed for the duration of the AP as determined automatically by the device (AP-A). The numbers on the right side of the panel summarize the sleep stage duration. Values within parentheses represent the AP-A (device-determined AP), whereas the values outside parentheses are based on the AP-M (AP set from lights off to lights on). REM epochs are depicted by thick black lines. The mismatches or misclassifications between PSG and the devices are depicted above the device hypnograms for different "resolutions" (2 stages [sleep and wake], 3 stages [NREM, REM and wake], and 4 stages [DS, LS, REM, and wake]). The gray and red bars indicate the match and mismatch epochs, respectively, between the devices and PSG. All the devices have an extra label that depicts no presence or artifacts.

In these examples, the AP determined by the device does not match the PSG light off period, and sometimes the device (eg, Figure 1B, WSA) identifies sleep before lights off (Figures 1A and 1B). In these examples, the AP determined by Somnofy was closest to the PSG AP, followed by WSA and Emfit. Figures 1A and 1B illustrate that SEFF was rather low in these participants and that AWS and CSTs detected far fewer awakenings than PSG.

The example EBE concordance of the device sleep prediction compared with PSG at 3 distinct levels of sleep stage prediction (2 classes: sleep and wake; 3 classes: NREM, REM, and wake; and 4 classes: LS, DS, REM, and wake) is depicted above the hypnograms in Figure 1. From this visualization, we observe that the PSG-device EBE concordance improves with a reduction in the number of classes considered.

Figure 1. Examples of 5-stage polysomnography (PSG) hypnograms, Actiwatch Spectrum (AWS) hypnograms, and the contactless device hypnograms. (A) Example of cohort 1 (AWS and Withings sleep analyzer [WSA]). (B) Example of cohort 2 (WSA, Emfit-QS [Emfit], and Somnofy). Data are plotted for the analysis period–automatic (AP-A) or analysis period–manual (AP-M), whichever is longer. The analysis periods are marked with vertical lines. Vertical lines (red and magenta) at the beginning and end of the night indicate lights off and lights on, respectively. DS: deep sleep; LS: light sleep; N1: stage N1 of non–rapid eye movement sleep; N2: stage N2 of non–rapid eye movement sleep; N3: stage N3 of non–rapid eye movement sleep; NREM: non–rapid eye movement; REM: rapid eye movement; TST: total sleep time.



AP Differences

First, we considered the period over which these devices estimate the various sleep parameters. For the CSTs, sleep summary agreement analysis was performed using both the AP automatically determined by the device (AP-A) and the summary estimate computed over the “lights off” period (AP-M). For AWS, the following three APs were considered: (1) the “automatic” AP (AP-A), (2) the “manual” AP based on lights off periods as entered in the sleep diary (sleep diary “lights off” period [AP-M1]), and (3) the “manual” AP based on the PSG “lights off” period (AP-M2). All available data were used for the agreement analysis, except for 1 WSA recording, for which the device generated a summary over a period of >24 hours.

The difference in AP depicted in Table 2 and Figure S1 and Table S4 in Multimedia Appendix 1 shows that the APs estimated by AWS (both AP-A and AP-M1) were close to the PSG AP, with no significant differences between AWS AP-A and AP-M2. Among the CSTs, on average, AP-A was longer than the lights off period by ≈70 minutes and ≈135 minutes by WSA and Emfit, respectively, whereas Somnofy AP-A was not significantly (P<.05) different from the lights off period (bias ≈13 min). It should be noted that because the AP-M of CSTs were manually set to correspond to the lights off period (AP-M2 in the case of AWS), the bias was 0 for this comparison and hence not reported (see Table S5 in Multimedia Appendix 1).

Metrics of agreement between the all-night sleep summary device estimates based on AP-A (AP determined by the device) and PSG estimates based on AP-M (AP set from lights off to lights on) are listed in Table 2. The values shown are mean and 95% CI.

Table 2. All-night sleep or wake summary measure agreement metrics.

Sleep measure	AWS-A[a] (n=18), mean (95% CI)	WSA-A[b] (n=34), mean (95% CI)	Emfit-A[c] (n=16), mean (95% CI)	Somnofy-A[d] (n=17), mean (95% CI)
AP[e] (min)				
Bias (LOA[f] down to LOA up)	−5.91 (−85.11 to 73.28)	69.99 (−40.21 to 180.19)	134.8 (61.02 to 208.58)	13.01 (−61.82 to 87.84)
SAD[g]	0.73 (0.24 to 1.21)	1.94 (1.6 to 2.29)	3.82 (3.31 to 4.34)	0.68 (0.18 to 1.18)
SMAPE[h]	2 (0 to 4)	7 (5 to 8)	12 (10 to 13)	2 (0 to 3)
ICC[i]	0.34 (0 to 0.75)	0.1 (0 to 0.55)	0.64 (0 to 0.87)	0.01 (0 to 0.64)
TST[j] (min)				
Bias (LOA down to LOA up)	−11.72 (−193.62 to 170.2)	120.51 (−48.42 to 289.45)	206.22 (49.6 to 362.84)	95.74 (−32.71 to 224.18)
SAD	0.93 (0.44 to 1.41)	2.05 (1.7 to 2.39)	3.98 (3.46 to 4.49)	1.85 (1.35 to 2.35)
SMAPE	9 (5 to13)	15 (12 to 18)	23 (18 to 28)	13 (10 to 17)
ICC	0.48 (0 to 0.8)	0.21 (0 to 0.61)	—[k]	0.52 (0 to 0.83)
SOL[l] (min)				
Bias (LOA down to LOA up)	7.44 (−31.0 to 45.9)	16.91 (−8.2 to 42.02)	27.19 (−12.37 to 66.74)	−5.65 (−45.84 to 34.53)
SAD	1.18 (0.7 to 1.67)	1.25 (0.91 to 1.59)	1.79 (1.27 to 2.30)	0.98 (0.48 to 1.48)
SMAPE	59 (42 to 76)	43 (35 to 52)	46 (31 to 61)	41 (27 to 55)
ICC	—	0.75 (0.49 to 0.87)	0.36 (0 to 0.78)	0.1 (0 to 0.67)
WASO[m] (min)				
Bias (LOA down to LOA up)	−0.25 (−197.54 to 197.04)	−72.65 (−198.4 to 53.11)	−52.06 (−183.06 to 78.94)	−87.95 (−205.12 to 29.22)
SAD	1.07 (0.6 to 1.56)	1.61 (1.27 to 1.96)	1.66 (1.15 to 2.18)	1.8 (1.3 to 2.3)
SMAPE	34 (25 to 43)	44 (35 to 52)	32 (23 to 40)	51 (37 to 65)
ICC	0.37 (0 to 0.76)	0.42 (0 to 0.71)	—	0.52 (0 to 0.83)
SEFF[n] (%)				
Bias (LOA down to LOA up)	−1.03 (−34.99 to 32.94)	12.92 (−9.96 to 35.8)	20.37 (−4.52 to 45.27)	16.66 (−8.61 to 41.94)
SAD	1.07 (0.58 to 1.55)	1.58 (1.23 to 1.92)	2.61 (2.09 to 3.12)	1.59 (1.09 to 2.09)
SMAPE	10 (7 to 14)	10 (8 to 12)	14 (10 to 18)	12 (8 to 16)
ICC	0.34 (0 to 0.75)	0.45 (0 to 0.73)	—	0.57 (0 to 0.84)

[a]AWS-A: Actiwatch Spectrum automatic analysis estimates.

[b]WSA-A: Withings sleep analyzer automatic analysis estimates.

[c]Emfit-A: Emfit-QS automatic analysis estimates.

[d]Somnofy-A: Somnofy automatic analysis estimates.

[e]AP: analysis period.

[f]LOA: limit of agreement.

[g]SAD: standardized absolute difference.

[h]SMAPE: symmetric mean absolute percentage error.

[i]ICC: intraclass correlation.

[j]TST: total sleep time

[k]Not available.

[l]SOL: sleep onset latency.

[m]WASO: wake after sleep onset.

[n]SEFF: sleep efficiency.

All-Night Sleep Summary Measures: TST, SOL, WASO, and SEFF

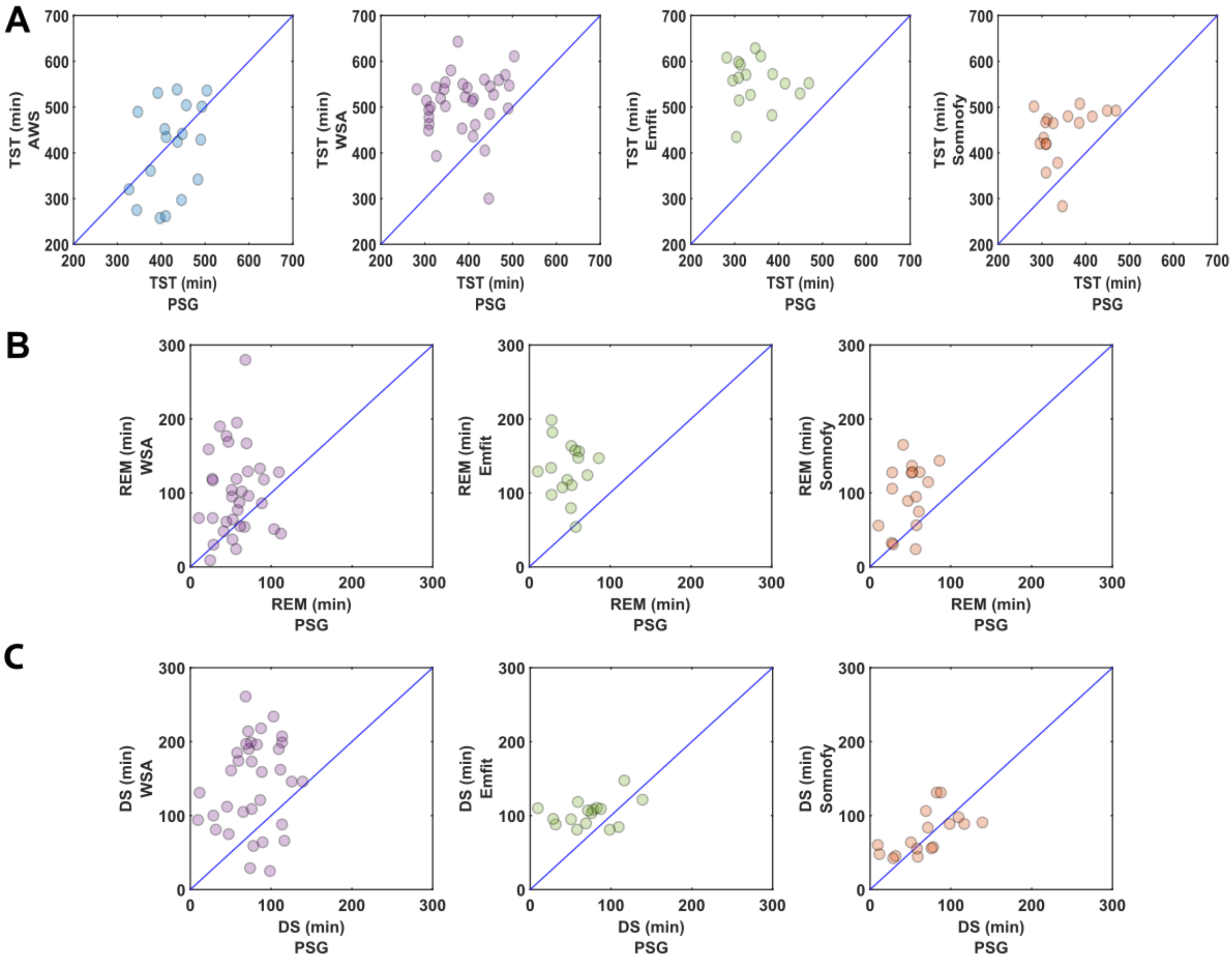
All-night sleep summary measures, namely TST, SOL, WASO and SEFF, computed for the automatically determined AP are presented in Table 2 and Figure S1 and Table S2 in Multimedia Appendix 1, whereas the estimates based on AP-M are presented in Table S5 in Multimedia Appendix 1. The all-night sleep summary estimates derived from AWS were, on average, very close to the PSG estimates, and the bias was not markedly different from 0 for TST, SOL, WASO, and SEFF. By contrast, most of the all-night sleep summary measures derived from the 3 CSTs deviated considerably from the PSG estimates for both AP-A and AP-M. All CSTs consistently overestimated TST (bias>90 min) and SEFF (bias>13%) and underestimated WASO (bias>50 min). WSA and Emfit overestimated SOL (bias>16 min), whereas the bias for SOL, as determined by Somnofy, was small.

The magnitude of dispersion of the difference (SAD) for AWS sleep summary estimates was considerable but was smaller than those for the CST sleep summary estimates. SAD values of the CSTs were large (>1.0) for AP and all sleep summary estimations (TST, SOL, WASO, and SEFF) apart from the Somnofy AP and SOL estimate. For all the devices, the absolute bias in the difference between the device and PSG estimates, quantified using SMAPE, was lowest for TST and SEFF (<25%).

The scatter plots in Figure 2A and Figure S2A in Multimedia Appendix 1 depict the overestimation, dispersion, and poor agreement of the CST TST estimates for AP-A and AP-M, respectively. They also show the considerable dispersion in AWS estimates.

For WASO and SOL, the bias was >30% for all CSTs. The results obtained from other measures of agreement such as Pearson correlation (ρ) and consistency ICC followed the Bland-Altman metrics such as bias, minimum detectable change, and dispersion measures (Tables S4 and S5 in Multimedia Appendix 1). The results of the AP-M estimates followed AP-A, except for SOL, which could be attributed to the AP that was set to the PSG lights off period.

Figure 2. Scatter plots of all-night sleep summary measures automatically generated by the devices versus manually scored polysomnography (PSG). (A) Total sleep time (TST). (B) Rapid eye movement (REM) sleep duration. (C) Deep sleep (DS) duration. The number of participants contributing to the data of each device is as follows: 18 for Actiwatch Spectrum (AWS); 34 for Withings sleep analyzer (WSA); 16 for Emfit-QS (Emfit); and 17 for Somnofy.



Sleep Stage Summary: LS, DS, REM, and NREM

The CSTs also provide a classification of sleep epochs as REM, LS, and DS. LS and DS can be combined to provide an estimate of NREM sleep. The differences in the sleep stage duration measures compared with PSG are depicted in Table 3 and Figure S1 and Tables S6 and S7 in Multimedia Appendix 1.

Metrics that depict the agreement between the sleep stage duration device estimates based on the AP-A (AP determined by the device) and PSG estimates based on AP-M (AP set from lights off to lights on) are shown in Table 3. The values shown are mean, and 95% CI.

All 3 CSTs overestimated both REM and NREM. WSA had the lowest bias (<11 min) for LS and did not show any consistent overestimation or underestimation. Somnofy (≈40 min) and Emfit (>90 min) overestimated LS, with the former having a lower bias than the latter. All CSTs had large dispersions (SAD>1.4) for REM, NREM, and LS. For DS, Somnofy had the lowest bias (<7 min) and dispersion (SAD<1) compared with the undermattress devices (bias>30 min and SAD>1.4). Among the undermattress devices, Emfit had a lower bias and dispersion.

The differences in the sleep stage duration measures were markedly, except for the LS estimates of WSA and DS estimates of Somnofy. The SMAPE was >30% for REM sleep duration across all the CSTs, and for NREM, LS, and DS, Somnofy had a lower SMAPE than the undermattress devices. The results obtained in the AP-M analysis (Table S8 in Multimedia Appendix 1) were similar to the AP-A results. Among the CSTs, only the DS duration estimates of Somnofy (both AP-A and AP-M) had satisfactory agreement (ICC>0.63).

The scatter plots in Figures 2B and 2C and Figures S2B and S2C in Multimedia Appendix 1 depict the discrepancy in the REM and DS duration estimates of the CSTs. The REM durations show overestimations and a large dispersion in the estimates across devices. For the DS duration estimate, a high level of agreement was observed for Somnofy, whereas the undermattress devices overestimated and showed large dispersion.

When the analysis was repeated for the sleep stages expressed as a percentage of TST (see Table S7 in Multimedia Appendix 1), the results were mixed. We found that WSA had the lowest percentage of bias (≈5%) and SAD for REM, followed by Somnofy and Emfit. For LS, Somnofy had the lowest bias (≈−5%) and SAD, followed by Emfit and WSA. For DS, Emfit had the lowest bias (≈−2 min), followed by Somnofy and WSA. The results were similar for the AP-M analysis (Table S9 in Multimedia Appendix 1).

Metrics that depict the agreement between the sleep stage duration device estimates based on the AP-A (AP determined by the device) and PSG estimates based on AP-M (AP set from lights off to lights on) are shown in Table 3. The values shown are mean, and 95% CI.

All 3 CSTs overestimated both REM and NREM. WSA had the lowest bias (<11 min) for LS and did not show any consistent overestimation or underestimation. Somnofy (≈40 min) and Emfit (>90 min) overestimated LS, with the former having a lower bias than the latter. All CSTs had large dispersions (SAD>1.4) for REM, NREM, and LS. For DS, Somnofy had the lowest bias (<7 min) and dispersion (SAD<1) compared with the undermattress devices (bias>30 min and SAD>1.4). Among the undermattress devices, Emfit had a lower bias and dispersion.

The differences in the sleep stage duration measures were markedly, except for the LS estimates of WSA and DS estimates of Somnofy. The SMAPE was >30% for REM sleep duration across all the CSTs, and for NREM, LS, and DS, Somnofy had a lower SMAPE than the undermattress devices. The results obtained in the AP-M analysis (Table S8 in Multimedia Appendix 1) were similar to the AP-A results. Among the CSTs, only the DS duration estimates of Somnofy (both AP-A and AP-M) had satisfactory agreement (ICC>0.63).

The scatter plots in Figures 2B and 2C and Figures S2B and S2C in Multimedia Appendix 1 depict the discrepancy in the REM and DS duration estimates of the CSTs. The REM durations show overestimations and a large dispersion in the estimates across devices. For the DS duration estimate, a high level of agreement was observed for Somnofy, whereas the undermattress devices overestimated and showed large dispersion.

When the analysis was repeated for the sleep stages expressed as a percentage of TST (see Table S7 in Multimedia Appendix 1), the results were mixed. We found that WSA had the lowest percentage of bias (≈5%) and SAD for REM, followed by Somnofy and Emfit. For LS, Somnofy had the lowest bias (≈−5%) and SAD, followed by Emfit and WSA. For DS, Emfit had the lowest bias (≈−2 min), followed by Somnofy and WSA. The results were similar for the AP-M analysis (Table S9 in Multimedia Appendix 1).

Table 3. Agreement metrics for sleep stage duration measures.

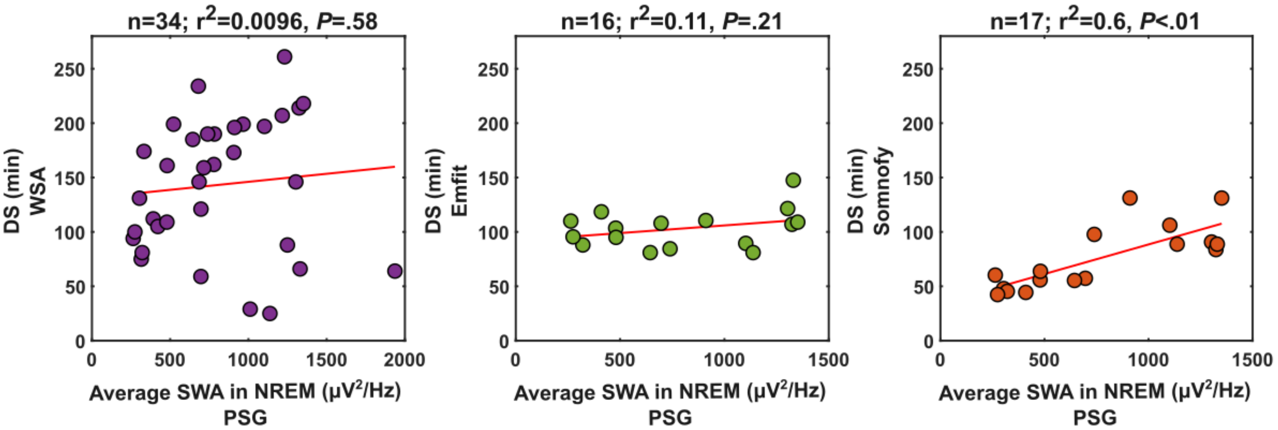
Sleep measure	WSA-A[a] (n=34), mean (95% CI)	Emfit-A[b] (n=16), mean (95% CI)	Somnofy-A[c] (n=17), mean (95% CI)
REM[d] (min)			
Bias (LOA[e] down to LOA up)	44.57 (−78.07 to 167.22)	84.12 (−2.01 to 170.26)	48.14 (−32.2 to 128.48)
SAD[f]	1.27 (0.93 to 1.62)	2.93 (2.41 to 3.44)	1.59 (1.09 to 2.09)
SMAPE[g]	33 (25 to 41)	46 (35 to 58)	34 (23 to 45)
ICC[h]	0.07 (−0.86 to 0.54)	—[i]	0.40 (0 to 0.78)
NREM[j] (min)			
Bias (LOA down to LOA up)	75.94 (−119.47 to 271.35)	122.09 (−0.74 to 244.94)	47.59 (−68.28 to 163.45)
SAD	1.82 (1.48 to 2.17)	2.85 (2.33 to 3.36)	1.39 (0.89 to 1.89)
SMAPE	14 (11 to 17)	17 (12 to 22)	10 (7 to 13)
ICC	—	—	0.33 (0 to 0.76)
Light sleep (min)			
Bias (LOA down to LOA up)	10.54 (−200.13 to 221.22)	91.78 (−22.06 to 205.63)	40.85 (−76.59 to 158.29)
SAD	1.43 (1.08 to 1.77)	2.3 (1.8 to 2.82)	1.4 (0.90 to 1.9)
SMAPE	17 (12 to 21)	18 (13 to 23)	12 (8 to 17)
ICC	—	0.18 (0 to 0.71)	0.12 (0 to 0.68)
Deep sleep (min)			
Bias (LOA down to LOA up)	65.4 (−60.25 to 191.04)	30.31 (−33.52 to 94.15)	6.73 (−51.02 to 64.49)
SAD	1.66 (1.31 to 2)	1.45 (0.93 to 1.96)	0.78 (0.29 to 1.28)
SMAPE	37 (30 to 44)	25 (14 to 36)	20 (11 to 30)
ICC	0.24 (0 to 0.62)	0.43 (0 to 0.8)	0.74 (0.3 to 0.91)

[a]WSA-A: Withings sleep analyzer automatic analysis estimates.
[b]Emfit-A: Emfit-QS automatic analysis estimates.
[c]Somnofy-A: Somnofy automatic analysis estimates.
[d]REM: rapid eye movement.
[e]SAD: standardized absolute difference.
[f]LOA: limit of agreement.
[g]SMAPE: symmetric mean absolute percentage error.
[h]ICC: intraclass correlation.
[i]Not available.
[j]NREM: non–rapid eye movement.

DS and EEG SWA in NREM Sleep

Visual scoring of DS is based on an amplitude and incidence criterion for slow waves such that a 30-second epoch is scored as N3 when ≥6 seconds of this epoch consists of slow waves with an amplitude equal to or greater than 75 μV [18]. Slow waves also occur in N2, and the amplitude of the slow waves declines with aging [35]. It has been repeatedly argued that DS should be quantified in a less arbitrary manner [36]. The most commonly used measure is SWA, defined as EEG power density in the range of 0.75 to 4.5 Hz, in NREM sleep. Therefore, we investigated whether DS as detected by the CSTs was associated with SWA. Somnofy DS duration was significantly correlated (r^2=0.6; P<.01) with the average SWA detected via PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.58; Emfit: r^2=0.11, P=.21; Figure 3).

Figure 3. Correlation between device estimates deep sleep (DS) duration and average slow wave activity (SWA; polysomnography [PSG]) in non–rapid eye movement [NREM]. The DS duration estimates are over the analysis period–automatic of the devices. The average SWA represents the average power in the 0.75-to-4.5-Hz band in NREM epochs. r2 is the coefficient of determination of the linear model, and P is the significance level. Emfit: Emfit-QS; WSA: Withings sleep analyzer.



EBE Concordance

The CSTs generate a 4-stage sleep time series with REM, LS, DS, and wake, whereas AWS generates a 2-stage sleep-wake time series. The pooled confusion matrices for each device are shown in Figure 4. The EBE concordance between PSG and the compared devices at 3 distinct levels of sleep stage prediction resolution computed over the TRT of PSG is depicted in Table 4 and Figure S3 in Multimedia Appendix 1. The sensitivity (sleep detection accuracy) of the undermattress devices was high (>0.9), with specificity similar to (WSA) or lower (Emfit) than that of AWS. The MCC, which, unlike accuracy and specificity, is robust against class imbalance, of AWS was comparable to those of WSA and Emfit. Somnofy outperformed the undermattress devices and AWS with a moderate MCC value (0.63, 0.57-0.69). With respect to individual sleep stage prediction concordance, Somnofy showed moderate performance across all sleep stages, whereas WSA had moderate concordance for wake and DS. Emfit had poor concordance for all sleep stages. In DS estimation, WSA (MCC: 0.47, 0.42-0.52) was marginally better than Somnofy (MCC: 0.46, 0.35-0.57). The EBE concordance analysis performed over the lights off period of PSG revealed similar results to those of the TRT analysis (see Table S8 and Figure S4 in Multimedia Appendix 1). The violin plots in Figures S3 and S4 in Multimedia Appendix 1 are used to show the distribution over the participants, and each dot within the violin corresponds to the performance measure for a single participant. Among the CSTs, only the NREM EBE concordance of Somnofy for TRT had satisfactory agreement (MCC: >0.48).

The values shown in Table 4 are mean, SD, and 95% CI. Here, NREM sleep denotes epochs with either DS or LS, and sleep or wake denotes the binary sleep stage prediction performance. The metrics were computed for the TRT (from the start to the end of PSG recording) of PSG. The number of participants contributing to each device was as follows: 18 for AWS; 35 for WSA; 16 for Emfit; and 17 for Somnofy.

We further explored the EBE concordance between the CSTs and PSG using an alternate assumption of LS being equivalent to N1 and DS being equivalent to both N2 and N3, and the results are provided in Figure S5 and Table S11 in Multimedia Appendix 1. We found that the accuracy (measured through MCC) of LS detection was significantly reduced across the 3 devices compared with the original ground-truth label (LS=N1 or N2 and DS=N3), disproving the alternate assumption. This also reaffirms the ground truth that LS predicted by the CSTs is equivalent to both N1 and N2, and DS is equivalent to N3.

Figure 4. Pooled confusion matrices. The pooled confusion matrices are derived by summing participant-wise epoch-by-epoch concordance confusion matrices. The panels on the left indicate the matrices computed over the total recording time, and the panels on the right indicate the lights off period. Total number of epochs for each device for the total recording time is as follows: 21,323 for Actiwatch Spectrum (AWS); 40,923 for Withings sleep analyzer (WSA); 18,809 for Emfit-QS (Emfit); and 20,278 for Somnofy. Total number of epochs for each device for the lights off period is as follows: 20,319 for AWS; 37,502 for WSA; 16,322 for Emfit; and 17,322 for Somnofy. The number of participants contributing to the data of each device is as follows: 18 for AWS; 35 for WSA; 16 for Emfit; and 17 for Somnofy.

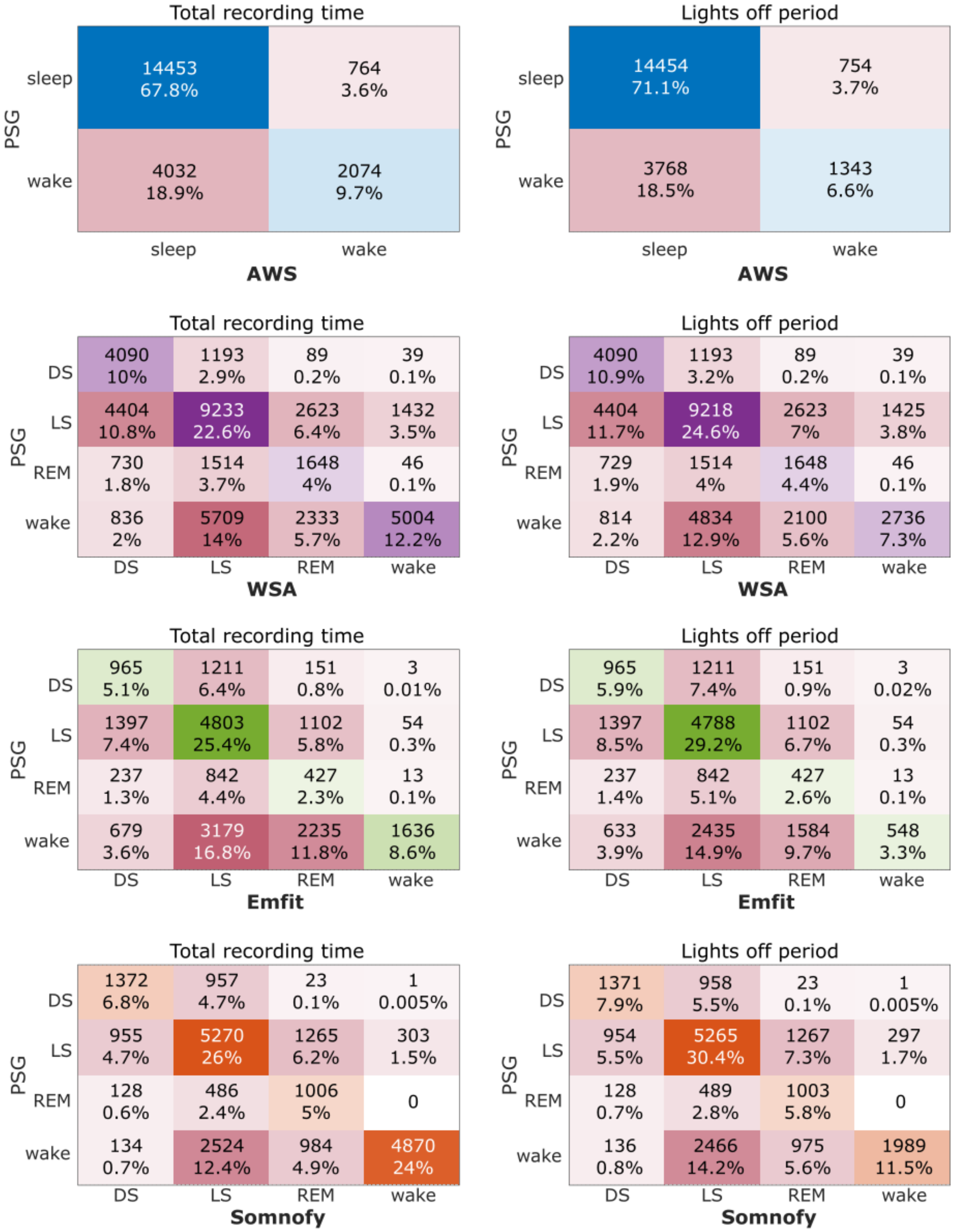


Table 4. Epoch-by-epoch agreement.

Sleep stage	Sensitivity, mean (SD; 95% CI)	Specificity, mean (SD; 95% CI)	Accuracy, mean (SD; 95% CI)	MCC[a], mean (SD; 95% CI)	F_1-score, mean (SD; 95% CI)
Sleep or wake					
AWS[b]	0.95 (0.02; 0.94-0.96)	0.34 (0.12; 0.28-0.4)	0.78 (0.06; 0.74-0.81)	0.37 (0.11; 0.32-0.43)	0.85 (0.05; 0.83-0.88)
WSA[c]	0.95 (0.09; 0.92-0.98)	0.37 (0.16; 0.31-0.42)	0.75 (0.09; 0.71-0.78)	0.41 (0.15; 0.36-0.46)	0.83 (0.07; 0.8-0.85)
Emfit[d]	0.99 (0.03; 0.97-1)	0.22 (0.14; 0.15-0.3)	0.67 (0.11; 0.61-0.73)	0.35 (0.16; 0.26-0.43)	0.78 (0.08; 0.73-0.82)
Somnofy	0.97 (0.06; 0.94-1)	0.58 (0.17; 0.5-0.67)	0.81 (0.08; 0.76-0.85)	0.63 (0.12; 0.57-0.69)	0.85 (0.07; 0.82-0.88)
REM[e]					
WSA	0.4 (0.22; 0.32-0.48)	0.86 (0.1; 0.83-0.89)	0.82 (0.09; 0.79-0.85)	0.24 (0.16; 0.18-0.3)	0.32 (0.16; 0.27-0.38)
Emfit	0.25 (0.19; 0.15-0.35)	0.8 (0.06; 0.77-0.83)	0.76 (0.06; 0.73-0.79)	0.12 (0.08; 0.07-0.16)	0.18 (0.11; 0.13-0.24)
Somnofy	0.62 (0.25; 0.49-0.74)	0.88 (0.08; 0.84-0.92)	0.86 (0.08; 0.82-0.9)	0.39 (0.18; 0.3-0.49)	0.42 (0.18; 0.32-0.51)
NREM[f]					
WSA	0.83 (0.12; 0.79-0.87)	0.52 (0.13; 0.47-0.56)	0.68 (0.07; 0.66-0.71)	0.38 (0.13; 0.33-0.42)	0.74 (0.06; 0.72-0.76)
Emfit	0.83 (0.11; 0.78-0.89)	0.44 (0.13; 0.37-0.51)	0.64 (0.11; 0.58-0.7)	0.35 (0.14; 0.27-0.42)	0.7 (0.1; 0.65-0.76)
Somnofy	0.84 (0.07; 0.81-0.88)	0.69 (0.12; 0.63-0.75)	0.76 (0.07; 0.72-0.8)	0.53 (0.13; 0.47-0.6)	0.77 (0.08; 0.73-0.81)
Light sleep					
WSA	0.54 (0.13; 0.49-0.58)	0.65 (0.13; 0.61-0.7)	0.59 (0.06; 0.57-0.61)	0.2 (0.11; 0.16-0.24)	0.52 (0.08; 0.49-0.55)
Emfit	0.63 (0.09; 0.58-0.67)	0.54 (0.1; 0.48-0.59)	0.57 (0.08; 0.53-0.61)	0.17 (0.14; 0.1-0.24)	0.53 (0.1; 0.47-0.58)
Somnofy	0.67 (0.1; 0.62-0.72)	0.69 (0.09; 0.64-0.73)	0.68 (0.06; 0.65-0.71)	0.35 (0.14; 0.28-0.42)	0.61 (0.11; 0.55-0.66)
Deep sleep					
WSA	0.79 (0.23; 0.71-0.87)	0.83 (0.09; 0.8-0.86)	0.82 (0.06; 0.8-0.84)	0.47 (0.15; 0.42-0.52)	0.51 (0.16; 0.45-0.56)
Emfit	0.39 (0.24; 0.26-0.52)	0.85 (0.03; 0.84-0.87)	0.79 (0.05; 0.77-0.82)	0.21 (0.16; 0.13-0.29)	0.28 (0.17; 0.19-0.37)
Somnofy	0.54 (0.21; 0.43-0.65)	0.93 (0.04; 0.91-0.95)	0.89 (0.04; 0.87-0.91)	0.46 (0.21; 0.35-0.57)	0.51 (0.21; 0.4-0.62)

[a]MCC: Matthew correlation coefficient.

[b]AWS: Actiwatch Spectrum.

[c]WSA: Withings sleep analyzer.

[d]Emfit: Emfit-QS.

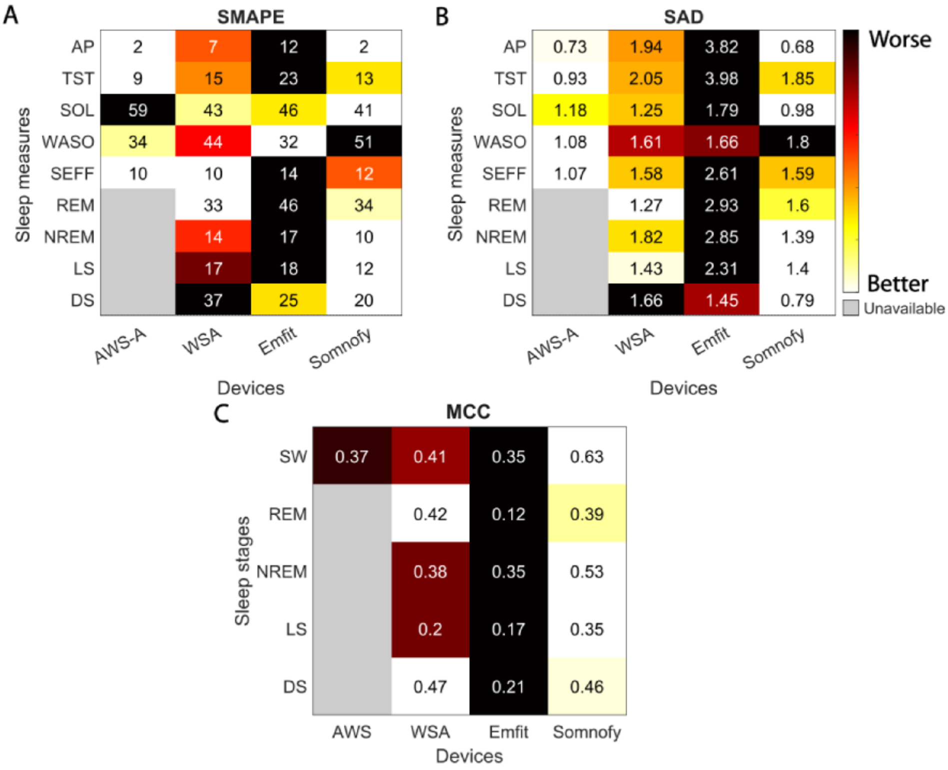
[e]REM: rapid eye movement.

[f]NREM: non–rapid eye movement.

Summarizing the Sleep Summary Measures and EBE Agreement

To provide an effective way to visualize the differences between the evaluated devices, we created heatmaps of SMAPE, SAD, and MCC (Figure 5 and Figure S6 in Multimedia Appendix 1). From these heatmaps, it appears that Somnofy is the best-performing CST, and Emfit is the worst-performing CST. We also noticed that AWS consistently outperformed the CSTs in the all-night sleep summary measure estimation, except for SOL, which was best estimated by Somnofy. We also ranked the devices using SMAPE and SAD for the all-night sleep measures and sleep stage duration estimates and MCC for EBE concordance (Tables S12 and S13 in Multimedia Appendix 1).

Figure 5. Agreement matrices depicting the sleep summary and epoch-by-epoch concordance. (A) Symmetric mean absolute percentage error (SMAPE). (B) Standardized absolute difference (SAD). The sleep summary measures are computed over the analysis period (AP)-automatic. (C) Matthew correlation coefficient (MCC, computed over the total recording time of polysomnography). The number of participants contributing to the data of each device is as follows: 18 for Actiwatch Spectrum (AWS); 34 for Withings sleep analyzer (WSA; 35 for MCC); 16 for Emfit-QS (Emfit); and 17 for Somnofy. The color code of all the agreement matrices is scaled across each row. AWS-A: Actiwatch Spectrum automatic analysis estimates; DS: deep sleep; LS: light sleep; NREM: non–rapid eye movement; REM: rapid eye movement; SEFF: sleep efficiency; SOL: sleep onset latency; SW: slow wave; TST: total sleep time; WASO: wake after sleep onset.



Discussion

Principal Findings

Overview

A comparison of 3 consumer CSTs against PSG and actigraphy in older people revealed that for all-night sleep summary measures of binary classification (sleep vs wake), CSTs did not perform as well as actigraphy. This is in line with our evaluation of CSTs against sleep diary–assisted actigraphy in an at-home setting [22]. In sleep stage classification (DS, LS, and REM), the bedside radar (Somnofy) outperformed the undermattress devices (WSA and Emfit) and had satisfactory agreement with PSG for DS duration estimate. For LS and REM estimates, the agreement was unsatisfactory for all the devices. The data were acquired from older adults with a variety of health conditions, including sleep apnea, during their first night in a sleep laboratory, and the average SEFF was only 71%. The protocol also resulted in a large interindividual variation in polysomnographic sleep-wake parameters, which contributes to the relevance of this evaluation for the intended real-world implementation of these "digital health" devices. The data imply that these contactless sleep-tracking devices provide some useful information on sleep in older people. Indeed, we have previously evaluated the CSTs against sleep diary–assisted actigraphy and have shown that the CSTs provide reliable estimates of bed occupancy in older people living at home [22]. Given their multimodal capabilities, improvements to their sleep detection algorithms to generalize results across populations could potentially lead to reliable sleep measures on par with PSG in the near future [8,37].

AP Differences

One of the characteristics of the quantification of sleep by contactless devices in real-world implementation is that the period over which the analysis is performed is automatically determined by the device. This contrasts with the standard polysomnographic assessments in the sleep laboratory, in which the AP is manually determined and usually set to the period from lights off to lights on. The automatically determined APs (AP-A) of AWS and Somnofy were close to the PSG lights off period (AP-M), whereas for the other devices, the AP was often very different from the PSG-based AP. These differences in performance may be related to whether the devices measure ambient light. Whereas the Somnofy bedside radar uses both changes in ambient light and bed presence information to determine the AP, the undermattress devices (WSA and Emfit) use only bed occupancy information.

Inaccurate estimation of the AP is a contributor to the relatively poor performance of WSA and Emfit with respect to the estimation of the latency to sleep onset, as these estimates improved significantly when the AP was set to the light off period. Other sleep parameters are less affected by the AP. Nevertheless, our analyses suggest that ambient light information may be useful in improving SOL estimations in CSTs and emphasize that the first target for performance improvement of these technologies is the AP.

Sleep Summary

The CSTs overestimate TST and SEFF and hence underestimate WASO for all AP settings. Compared with PSG, the performance of the CSTs in estimating sleep stage duration was poor, with SMAPE ranging from 10% to 46% across all devices and sleep stages. The CSTs performed less well than the wearable actigraphy device (AWS) even when the AWS AP was automatically determined. The estimation of REM sleep was particularly poor across all the devices. The DS duration estimates of Somnofy were closer to the PSG estimates, whereas the undermattress devices (WSA and Emfit) performed poorly.

DS and EEG SWA in NREM Sleep

The observation that the undermattress devices' estimates of DS duration did not correlate with SWA but DS duration as detected by the Somnofy bedside radar did correlate with SWA is somewhat puzzling because all the devices use contactless ballistographic signals. Nevertheless, the superior performance of Somnofy in assessing this neurophysiological characteristic of sleep is of interest because SWA has often been proclaimed to be of particular importance for the recovery value of sleep [35].

EBE Concordance

The CSTs offered EBE sleep stage predictions compared with the simple sleep-wake prediction time series available in AWS. The EBE concordance of the CSTs with PSG varied across sleep stages. Overall, Somnofy had the best performance across all sleep stage predictions and satisfactory EBE concordance for NREM compared with PSG. Among the undermattress devices, WSA had a better performance than Emfit, which performed worse than AWS, even in sleep or wake discrimination [38].

Prior Works

Overall, the all-night sleep summary and REM sleep stage duration estimation results of the CSTs in our study were in line with the observations reported by others [16,38-41]. The similar EBE sleep or wake concordance of the undermattress devices with AWS is in line with the results reported by Chinoy et al [38] for contactless devices.

WSA

To the best of our knowledge, there are no performance evaluation studies comparing WSA with PSG for objective sleep estimation in older adults. In a recent evaluation study by Edouard et al [24] (n=118; age: mean 49.3, SD 12.1 y), WSA overestimated TST and underestimated WASO compared with PSG, which is in line with the results obtained in our study. A notable difference is that the evaluation conducted by Edouard et al [24] was limited to TST, SEFF, and WASO estimates of WSA.

Somnofy

The performance of Somnofy in the estimation of TST, SEFF, and WASO in our cohorts of older people is poorer than that in a study of 71 nights by Toften et al [15] in participants without sleep disorders, whereas the DS duration estimate is similar between the 2 studies. The EBE concordance of Somnofy agrees with the overall sensitivity (0.97) estimate reported by Toften et al [15], whereas the specificity is lower in our cohort (specificity in Toften et al [15]: 0.72; specificity in our study: 0.34).

Emfit

The results of the Emfit evaluation in our study were congruent with those in a study by Kholghi et al [16], in which the TST was overestimated, WASO underestimated, and all sleep stage duration estimates were poor. A notable difference is that although the EBE sleep or wake discrimination sensitivity is similar (0.99), the specificity of Emfit in our study is low (0.22) but higher than that reported by Kholghi et al [16] (0.10).

It should be noted that Toften et al [15] and Kholghi et al [16] evaluated the respective devices in younger (age: mean 28.9, SD 9.7 y) and middle-aged (age: mean 53.7, SD 16.5 y) populations, respectively.

Limitations

The primary limitations of this study are the small sample size (<20 for AWS and Emfit) and the fact that not all devices were concurrently implemented in all participants. Given the variety of confounding factors in our cohort, including age, comorbidities, and sleep disorders, the small sample size reduced the statistical power of the performance measures used, with larger error margins in the estimates and increased sensitivity to outliers. Another limitation of the study is that, owing to the proprietary nature of the devices, the data synchronization process was based on clock times and the best alignment of the device and PSG activity or movement data (for Somnofy and Emfit) and hypnograms, which is not an ideal approach. However, because all the devices were synchronized to a common network and epochs were of 30-second and 1-minute intervals, we did not find any significant synchronization issues in the study data. The final limitation of this study is the lack of transparency in the algorithms used by the different CSTs for sleep prediction and summary generation. The limited information available on the data processing pipelines involved and training set used hinders the interpretability of the evaluation results.

Conclusions

Our inclusion or exclusion criteria were chosen such that even though the participants were in a stable health condition, several of them had comorbidities that are common in older adults [42,43]. This, together with the first night effect, resulted in mildly disturbed sleep. Our chosen population and extended period in bed contribute to the relevance of this evaluation study for assessing sleep in the real world and target populations such as people living with dementia [42]. Some of the accidental medical findings in this study, such as a case of arrhythmia (Figure S7 in Multimedia Appendix 1), provided further insights into the performance of these devices. Because the algorithms used by the CSTs for sleep staging rely on the contactless ballistographic signal, which is primarily composed of activity, breathing, and heart rate, any condition that affects cardiopulmonary function can potentially affect the performance of the algorithms. These findings point to some of the limitations of the CSTs for real-world deployment.

The study revealed that the standard actigraphy device (AWS) provides fairly accurate estimates of all-night sleep summary measures compared with PSG, but the interparticipant measurement errors were still large. From the ranking created using the various performance measures, among the CSTs, it may be concluded that overall, Somnofy outperforms the undermattress sensors. However, how useful a device depends not only on its performance but also on the use case, costs, scalability, acceptability, etc. For example, for some use cases, an estimate of the approximate TST may be sufficient, whereas for other use cases, a good EBE concordance may be important.

Overall, it can be concluded that contactless sleep-tracking devices provide some useful information on sleep behavior, but their estimates of sleep stages are not very accurate. Owing to their unintrusive nature and higher user acceptability, CSTs may offer the opportunity for clinical digital phenotyping of sleep, behavior, and health in older adults at scale in their own homes. However, our assessment underscores the clear need for improvement in the performance of CSTs across all sleep estimation domains (summary and EBE sleep) and relevant populations before they can be effectively deployed in the real world.

Acknowledgments

This work was supported by the UK Dementia Research Institute (UK DRI), Care Research and Technology Centre at Imperial College, London, and the University of Surrey, Guildford, United Kingdom, which receives its funding from the UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK. The authors thank the staff members of the Clinical Research Facility for their help with data collection. They would also like to extend their thanks to Prof Anne Skeldon, University of Surrey, for her comments on the manuscript and their colleagues at the UK Dementia Research Institute, Care Research and Technology Centre, for their helpful discussions.

Authors' Contributions

KKGR curated, visualized, and analyzed the data and drafted the manuscript. DJD conceived the study, supervised the analysis, and contributed to the writing of the manuscript. CdM, GA, HH, and VR contributed to the design of the study and were responsible for participant recruitment and screening and study conduct. The devices were set up and data sets were downloaded and curated by DL, GA, CdM, and VR. GA and CdM scored the polysomnography data. All the authors contributed to the data collection and writing of the manuscript and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no competing nonfinancial interests but the following competing financial interests. The Withings and Emfit devices used in this study were purchased from the manufacturers without any price reductions. The Somnofy devices used in this study were provided by Vital Things, Norway, at no cost. The manufacturers of these devices were not involved in the design or conduct of this study, analysis and interpretation of the data, or preparation of the manuscript.

Multimedia Appendix 1
Supplemental material.
[DOC File , 22031 KB - mhealth_v11i1e46338_app1.doc]

References

1. Galambos C, Skubic M, Wang S, Rantz M. Management of dementia and depression utilizing in- home passive sensor data. Gerontechnology 2013 Jan 01;11(3):457-468. [doi: 10.4017/gt.2013.11.3.004.00] [Medline: 24049513]
2. Spira AP, Covinsky K, Rebok GW, Stone KL, Redline S, Yaffe K. Objectively measured sleep quality and nursing home placement in older women. J Am Geriatr Soc 2012 Jul 15;60(7):1237-1243 [FREE Full text] [doi: 10.1111/j.1532-5415.2012.04044.x] [Medline: 22702839]
3. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 01;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
4. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 03;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
5. Bloom HG, Ahmed I, Alessi CA, Ancoli-Israel S, Buysse DJ, Kryger MH, et al. Evidence-based recommendations for the assessment and management of sleep disorders in older persons. J Am Geriatr Soc 2009 May;57(5):761-789 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02220.x] [Medline: 19484833]

6. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
7. Baumert M, Cowie MR, Redline S, Mehra R, Arzt M, Pépin JL, et al. Sleep characterization with smart wearable devices: a call for standardization and consensus recommendations. Sleep 2022 Dec 12;45(12):zsac183. [doi: 10.1093/sleep/zsac183] [Medline: 35913733]
8. Schütz N, Saner H, Botros A, Pais B, Santschi V, Buluschek P, et al. Contactless sleep monitoring for early detection of health deteriorations in community-dwelling older adults: exploratory study. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e24666 [FREE Full text] [doi: 10.2196/24666] [Medline: 34114966]
9. Hussain Z, Sheng QZ, Zhang WE, Ortiz J, Pouriyeh S. Non-invasive techniques for monitoring different aspects of sleep: a comprehensive review. ACM Trans Comput Healthc 2022 Mar 03;3(2):1-26. [doi: 10.1145/3491245]
10. Sadek I, Biswas J, Abdulrazak B. Ballistocardiogram signal processing: a review. Health Inf Sci Syst 2019 Dec;7(1):10 [FREE Full text] [doi: 10.1007/s13755-019-0071-7] [Medline: 31114676]
11. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph Beddit sleep tracker: a validation study. J Clin Sleep Med 2019 Mar 15;15(3):483-487 [FREE Full text] [doi: 10.5664/jcsm.7682] [Medline: 30853052]
12. Perez-Pozuelo I, Zhai B, Palotti J, Mall R, Aupetit M, Garcia-Gomez JM, et al. The future of sleep health: a data-driven revolution in sleep science and medicine. NPJ Digit Med 2020 Mar 23;3:42 [FREE Full text] [doi: 10.1038/s41746-020-0244-4] [Medline: 32219183]
13. de Zambotti M, Menghini L, Grandner MA, Redline S, Zhang Y, Wallace ML, et al. Rigorous performance evaluation (previously, "validation") for informed use of new technologies for sleep health measurement. Sleep Health 2022 Jun;8(3):263-269 [FREE Full text] [doi: 10.1016/j.sleh.2022.02.006] [Medline: 35513978]
14. Chinoy ED, Cuellar JA, Jameson JT, Markwald RR. Performance of four commercial wearable sleep-tracking devices tested under unrestricted conditions at home in healthy young adults. Nat Sci Sleep 2022;14:493-516 [FREE Full text] [doi: 10.2147/NSS.S348795] [Medline: 35345630]
15. Toften S, Pallesen S, Hrozanova M, Moen F, Grønli J. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). Sleep Med 2020 Nov;75:54-61. [doi: 10.1016/j.sleep.2020.02.022] [Medline: 32853919]
16. Kholghi M, Szollosi I, Hollamby M, Bradford D, Zhang Q. A validation study of a ballistocardiograph sleep tracker against polysomnography. J Clin Sleep Med 2022 Apr 01;18(4):1203-1210 [FREE Full text] [doi: 10.5664/jcsm.9754] [Medline: 34705630]
17. Tal A, Shinar Z, Shaki D, Codish S, Goldbart A. Validation of contact-free sleep monitoring device with comparison to polysomnography. J Clin Sleep Med 2017 Mar 15;13(3):517-522 [FREE Full text] [doi: 10.5664/jcsm.6514] [Medline: 27998378]
18. Troester MM, Quan SF, Berry RB, Plante DT, Abreu AR, Alzoubaidi M, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. In: American Academy of Sleep Medicine. Darien, IL: American Academy of Sleep Medicine; 2023.
19. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American academy of sleep medicine clinical practice guideline. J Clin Sleep Med 2018 Jul 15;14(7):1231-1237 [FREE Full text] [doi: 10.5664/jcsm.7230] [Medline: 29991437]
20. Chicco D, Warrens MJ, Jurman G. The Matthews Correlation Coefficient (MCC) is more informative than Cohen's Kappa and brier score in binary classification assessment. IEEE Access 2021;9:78368-78381. [doi: 10.1109/access.2021.3084050]
21. Delgado R, Tibau XA. Why Cohen's Kappa should be avoided as performance measure in classification. PLoS One 2019;14(9):e0222916 [FREE Full text] [doi: 10.1371/journal.pone.0222916] [Medline: 31557204]
22. G Ravindran KK, Della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, et al. Contactless and longitudinal monitoring of nocturnal sleep and daytime naps in older men and women: a digital health technology evaluation study. Sleep 2023 Jul 20:zsad194 (forthcoming). [doi: 10.1093/sleep/zsad194] [Medline: 37471049]
23. Ding L, Chen B, Dai Y, Li Y. A meta-analysis of the first-night effect in healthy individuals for the full age spectrum. Sleep Med 2022 Jan;89:159-165. [doi: 10.1016/j.sleep.2021.12.007] [Medline: 34998093]
24. Edouard P, Campo D, Bartet P, Yang RY, Bruyneel M, Roisman G, et al. Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. J Clin Sleep Med 2021 Jun 01;17(6):1217-1227 [FREE Full text] [doi: 10.5664/jcsm.9168] [Medline: 33590821]
25. Markwald RR, Bessman SC, Reini SA, Drummond SP. Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. J Clin Sleep Med 2016 Jan;12(1):95-103 [FREE Full text] [doi: 10.5664/jcsm.5404] [Medline: 26285110]
26. Cheung J, Leary EB, Lu H, Zeitzer JM, Mignot E. PSG Validation of minute-to-minute scoring for sleep and wake periods in a consumer wearable device. PLoS One 2020 Sep 17;15(9):e0238464 [FREE Full text] [doi: 10.1371/journal.pone.0238464] [Medline: 32941498]

27. Haghayegh S, Kang HA, Khoshnevis S, Smolensky MH, Diller KR. A comprehensive guideline for Bland-Altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings. Physiol Meas 2020 Jun 15;41(5):055012. [doi: 10.1088/1361-6579/ab86d6] [Medline: 32252039]
28. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
29. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965 Dec 01;52(3-4):591-611. [doi: 10.1093/biomet/52.3-4.591]
30. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
31. Makridakis S. Accuracy measures: theoretical and practical concerns. Int J Forecast 1993 Dec;9(4):527-529. [doi: 10.1016/0169-2070(93)90079-3]
32. Younes M, Kuna ST, Pack AI, Walsh JK, Kushida CA, Staley B, et al. Reliability of the American academy of sleep medicine rules for assessing sleep depth in clinical practice. J Clin Sleep Med 2018 Feb 15;14(2):205-213 [FREE Full text] [doi: 10.5664/jcsm.6934] [Medline: 29351821]
33. Lee YJ, Lee JY, Cho JH, Choi JH. Interrater reliability of sleep stage scoring: a meta-analysis. J Clin Sleep Med 2022 Jan 01;18(1):193-202 [FREE Full text] [doi: 10.5664/jcsm.9538] [Medline: 34310277]
34. Boulos MI, Jairam T, Kendzerska T, Im J, Mekhael A, Murray BJ. Normal polysomnography parameters in healthy adults: a systematic review and meta-analysis. Lancet Respir Med 2019 Jun;7(6):533-543. [doi: 10.1016/s2213-2600(19)30057-8]
35. Dijk DJ, Groeger JA, Stanley N, Deacon S. Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep. Sleep 2010 Feb;33(2):211-223 [FREE Full text] [doi: 10.1093/sleep/33.2.211] [Medline: 20175405]
36. Borbély AA. Refining sleep homeostasis in the two-process model. J Sleep Res 2009 Mar;18(1):1-2. [doi: 10.1111/j.1365-2869.2009.00750.x] [Medline: 19250170]
37. Boiko A, Martínez Madrid N, Seepold R. Contactless technologies, sensors, and systems for cardiac and respiratory measurement during sleep: a systematic review. Sensors (Basel) 2023 May 24;23(11):5038 [FREE Full text] [doi: 10.3390/s23115038] [Medline: 37299762]
38. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291 [FREE Full text] [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
39. Kahawage P, Jumabhoy R, Hamill K, de Zambotti M, Drummond SP. Validity, potential clinical utility, and comparison of consumer and research-grade activity trackers in Insomnia Disorder I: in-lab validation against polysomnography. J Sleep Res 2020 Feb;29(1):e12931. [doi: 10.1111/jsr.12931] [Medline: 31626361]
40. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiol Int 2015;32(7):1024-1028 [FREE Full text] [doi: 10.3109/07420528.2015.1054395] [Medline: 26158542]
41. Danzig R, Wang M, Shah A, Trotti LM. The wrist is not the brain: estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. J Sleep Res 2020 Feb;29(1):e12926 [FREE Full text] [doi: 10.1111/jsr.12926] [Medline: 31621129]
42. Browne J, Edwards DA, Rhodes KM, Brimicombe DJ, Payne RA. Association of comorbidity and health service usage among patients with dementia in the UK: a population-based study. BMJ Open 2017 Mar 09;7(3):e012546 [FREE Full text] [doi: 10.1136/bmjopen-2016-012546] [Medline: 28279992]
43. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]

Abbreviations

AASM: American Academy of Sleep Medicine
AHI: apnea hypopnea index
AP: analysis period
AP-A: analysis period–automatic
AP-M: analysis period–manual
AP-M1: sleep diary "lights off" period
AP-M2: polysomnography "lights off" period
AWS: Actiwatch Spectrum
CST: contactless sleep technology
DS: deep sleep
EBE: epoch-by-epoch
EEG: electroencephalography
Emfit: Emfit-QS

ESS: Epworth Sleepiness Scale
F3-M2: left frontal referenced to right mastoid
F4-M1: right frontal referenced to left mastoid
ICC: intraclass correlation
LS: light sleep
MCC: Matthew correlation coefficient
N1: stage N1 of non–rapid eye movement sleep
N2: stage N2 of non–rapid eye movement sleep
N3: stage N3 of non–rapid eye movement sleep
NREM: non–rapid eye movement
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
REM: rapid eye movement
SAD: standardized absolute difference
SEFF: sleep efficiency
SMAPE: symmetric mean absolute percentage error
SOL: sleep onset latency
SWA: slow wave activity
TRT: total recording time
TST: total sleep time
WASO: wake after sleep onset
WSA: Withings sleep analyzer

Edited by L Buis; submitted 07.02.23; peer-reviewed by B Cole, T Vaughan; comments to author 24.05.23; revised version received 11.07.23; accepted 25.08.23; published 25.10.23.

Please cite as:
G Ravindran KK, della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, Dijk DJ
Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study
JMIR Mhealth Uhealth 2023;11:e46338
URL: https://mhealth.jmir.org/2023/1/e46338
doi:10.2196/46338
PMID:37878360

Original Paper

Accuracy of 11 Wearable, Nearable, and Airable Consumer Sleep Trackers: Prospective Multicenter Validation Study

Taeyoung Lee[1*], MS; Younghoon Cho[1,2*], MD; Kwang Su Cha[1], MS; Jinhwan Jung[1], PhD; Jungim Cho[1], BS; Hyunggug Kim[1], MS; Daewoo Kim[1], PhD; Joonki Hong[1], PhD; Dongheon Lee[1], MS; Moonsik Keum[2], BS; Clete A Kushida[3], MD, PhD; In-Young Yoon[4], MD, PhD; Jeong-Whun Kim[5], MD, PhD

[1]Asleep Co., Ltd., Seoul, Republic of Korea

[2]Clionic Lifecare Clinic, Seoul, Republic of Korea

[3]Department of Psychiatry and Behavioral Sciences, Stanford University Medical Center, Redwood City, CA, United States

[4]Department of Psychiatry, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[5]Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jeong-Whun Kim, MD, PhD
Department of Otorhinolaryngology-Head and Neck Surgery
Seoul National University Bundang Hospital
82, Gumi-ro 173beon-gil, Bundang-gu
Gyeonggi-do
Seongnam-si, 13620
Republic of Korea
Phone: 82 10 3079 7405
Email: kimemails7@gmail.com

Abstract

Background: Consumer sleep trackers (CSTs) have gained significant popularity because they enable individuals to conveniently monitor and analyze their sleep. However, limited studies have comprehensively validated the performance of widely used CSTs. Our study therefore investigated popular CSTs based on various biosignals and algorithms by assessing the agreement with polysomnography.

Objective: This study aimed to validate the accuracy of various types of CSTs through a comparison with in-lab polysomnography. Additionally, by including widely used CSTs and conducting a multicenter study with a large sample size, this study seeks to provide comprehensive insights into the performance and applicability of these CSTs for sleep monitoring in a hospital environment.

Methods: The study analyzed 11 commercially available CSTs, including 5 wearables (Google Pixel Watch, Galaxy Watch 5, Fitbit Sense 2, Apple Watch 8, and Oura Ring 3), 3 nearables (Withings Sleep Tracking Mat, Google Nest Hub 2, and Amazon Halo Rise), and 3 airables (SleepRoutine, SleepScore, and Pillow). The 11 CSTs were divided into 2 groups, ensuring maximum inclusion while avoiding interference between the CSTs within each group. Each group (comprising 8 CSTs) was also compared via polysomnography.

Results: The study enrolled 75 participants from a tertiary hospital and a primary sleep-specialized clinic in Korea. Across the 2 centers, we collected a total of 3890 hours of sleep sessions based on 11 CSTs, along with 543 hours of polysomnography recordings. Each CST sleep recording covered an average of 353 hours. We analyzed a total of 349,114 epochs from the 11 CSTs compared with polysomnography, where epoch-by-epoch agreement in sleep stage classification showed substantial performance variation. More specifically, the highest macro F1 score was 0.69, while the lowest macro F1 score was 0.26. Various sleep trackers exhibited diverse performances across sleep stages, with SleepRoutine excelling in the wake and rapid eye movement stages, and wearables like Google Pixel Watch and Fitbit Sense 2 showing superiority in the deep stage. There was a distinct trend in sleep measure estimation according to the type of device. Wearables showed high proportional bias in sleep efficiency, while nearables exhibited high proportional bias in sleep latency. Subgroup analyses of sleep trackers revealed variations in macro F1 scores based on factors, such as BMI, sleep efficiency, and apnea-hypopnea index, while the differences between male and female subgroups were minimal.

Conclusions: Our study showed that among the 11 CSTs examined, specific CSTs showed substantial agreement with polysomnography, indicating their potential application in sleep monitoring, while other CSTs were partially consistent with polysomnography. This study offers insights into the strengths of CSTs within the 3 different classes for individuals interested in wellness who wish to understand and proactively manage their own sleep.

(JMIR Mhealth Uhealth 2023;11:e50983) doi:10.2196/50983

KEYWORDS

consumer sleep trackers; wearables; nearables; airables; sleep monitoring; sleep stage; comparative study; polysomnography; multicenter study; deep learning; artificial intelligence; Fitbit Sense 2, Amazon Halo Rise, SleepRoutine

Introduction

With the growing recognition of the importance of sleep for overall health [1], there has been a significant rise in public interest in monitoring sleep patterns using consumer sleep trackers (CSTs) [2-4]. While the laboratory monitoring of sleep using the traditional sleep analysis tool polysomnography has several limitations associated with the need for cumbersome sensors [5], CSTs facilitate individual monitoring of sleep at home using minimal equipment. Recently, many big tech companies, including Apple, Samsung, and Google, as well as health care startups like Withings and Oura, have released their own CSTs. These companies have made significant contributions to enhancing the performance of CSTs by integrating deep learning algorithms and biosignal sensing technologies [6,7]. As a result, CSTs have emerged as accessible solutions for home sleep monitoring [2,6,8-11]. CSTs are widely used by not only individuals interested in wellness who wish to understand and proactively manage their own sleep, but also those who want to self-check and screen for sleep disorders.

This study classified CSTs into 3 types: wearables, nearables, and airables. Wearable devices or wearables, such as smartwatches and ring-shaped devices, are generally worn by users to track sleep using sensors like photoplethysmography sensors and accelerometers [6,12-16]. Nearable devices or nearables, placed near the body without direct contact, have radar or mattress pads to detect subtle movements during sleep [9,17]. Airable devices or airables use mobile phones to analyze sleep via built-in microphones or environmental sensors [2,3,8]. This classification is based on the measurement methods and biological signals used in each category.

Given the surge of diverse CSTs, it is necessary to conduct comprehensive and objective evaluations of the performance of these CSTs available in the market [4,15,18-21]. Some studies compared CSTs and alternative tools available for sleep analysis, such as electroencephalography headbands [4] or subjective sleep diaries [12] (without employing the gold standard polysomnography), which failed to validate the consistency between CSTs and polysomnography. Chinoy et al [11] compared the performance of 7 CSTs with polysomnography (Fatigue Science Readiband, Fitbit Alta HR, Garmin Fenix 5S, Garmin Vivosmart 3, EarlySense Live, ResMed S+, and SleepScore Max). However, this previous study showed limitations of recruitment from a single institution and exclusion of widely used CSTs available commercially.

To address these limitations, we conducted a multicenter study comparing widely used or newly released CSTs with in-lab polysomnography in a hospital setting. By simultaneously assessing multiple CSTs, we aimed to minimize bias and evaluate their performance across various metrics. Subgroup analysis was also performed to assess the impact of demographic factors on performance, including sex assigned at birth, apnea-hypopnea index (AHI), and BMI. By performing the most extensive simultaneous comparison of widely used CSTs and conducting a multicenter study with diverse demographic groups, this study offers comprehensive insights into the performance and applicability of these CSTs for sleep monitoring.

Methods

Participants

The demographic information of the study participants is presented in Table 1. A total of 75 individuals were recruited from Seoul National University Bundang Hospital (SNUBH) and Clionic Lifecare Clinic (CLC). Of these 75 individuals, 37 (27 males and 10 females) with scheduled polysomnography for sleep disorders were recruited from SNUBH and 38 (12 males and 26 females) were recruited through an online platform from CLC. Both institutions used the same inclusion criteria, including age between 19 and 70 years and presence of subjective sleep discomfort. Individuals with uncontrolled acute respiratory conditions were excluded. Participant demographics revealed that the sample consisted of 52% (39/75) males, with a mean age of 43.59 (SD 14.10) years and a mean BMI of 23.90 (SD 4.07) kg/m^2. Significant differences in sleep measures were observed between the 2 institutions, including time in bed, total sleep time, wake after sleep onset (WASO), and AHI. Multimedia Appendix 1 presents the number of measurements and data collection success rate for each CST.

Table 1. Comparative analysis of participant demographics.

Characteristic	Total (N=75)	SNUBH[a] (n=37)	CLC[b] (n=38)	*P* value[c]
Male, n (%)	39 (52)	27 (73)	12 (32)	<.001[d]
Age (years), mean (SD)	43.59 (14.10)	53.49 (11.96)	33.95 (8.07)	<.001[d]
BMI (kg/m^2), mean (SD)	23.90 (4.07)	24.64 (4.07)	23.18 (3.98)	.12
Time in bed (hours), mean (SD)	7.24 (0.92)	8.01 (0.31)	6.49 (0.66)	<.001[d]
Total sleep time (hours), mean (SD)	5.82 (1.33)	6.26 (1.2)	5.40 (1.33)	.005[d]
Sleep latency (hours), mean (SD)	0.27 (0.37)	0.27 (0.38)	0.26 (0.38)	.93
Wake after sleep onset (hours), mean (SD)	1.15 (1.2)	1.48 (1.22)	0.83 (1.23)	.02[d]
Sleep efficiency (%), mean (SD)	81.00 (17.5)	78.40 (15.54)	83.54 (19.1)	.20
Apnea-hypopnea index, mean (SD)	18.18 (20.39)	26.56 (24.25)	10.02 (10.99)	<.001[d]

[a]SNUBH: Seoul National University Bundang Hospital.

[b]CLC: Clionic Lifecare Clinic.

[c]All *P* values were obtained using 2-sample independent *t* tests. For the male category, Fisher exact test was applied.

[d]Statistical significance (*P*<.05).

Evaluation of CSTs

We evaluated 11 different CSTs in this study. Wearables included ring-type devices (Oura Ring 3, Oura) and watch-type devices (Apple Watch 8, Apple Inc; Galaxy Watch 5, Samsung Electronics Co, Ltd; Fitbit Sense 2, Fitbit Inc; and Google Pixel Watch, Google LLC). Nearables included pad-type devices (Withings Sleep Tracking Mat, Withings) and motion sensor devices (Amazon Halo Rise, Amazon Inc; and Google Nest Hub 2, Google LLC). Airables included mobile apps (SleepRoutine, Asleep; SleepScore App, SleepScore Labs; and Pillow, Neybox Digital Ltd) with iPhone 12s and Galaxy S21s. The selection of these devices was based on their popularity and availability in the market at the time of the study. The methods of usage and application for each sleep tracker were based on user instructions provided by the respective manufacturers. To mitigate a possible learning curve for each device, the researchers educated participants on how to use each device before measurements, and in the case of wearable CSTs, they ensured that the devices were properly fitted. During the study, software updates of all devices were performed on March 1, 2023, to ensure that they were up-to-date, and automatic updates were disabled.

Study Design

This was a prospective cross-sectional study conducted to investigate the accuracy of various CSTs and polysomnography in analyzing sleep stages. It was conducted at 2 independent medical institutions in South Korea, namely SNUBH, a tertiary care hospital, and CLC, a primary care clinic.

All participants were contacted by phone at least 2 days prior to participating in the polysomnography study and were provided with instructions. On the day before and the day of the test, they were advised to abstain from alcohol and caffeine consumption and refrain from engaging in strenuous exercise, and were informed of the designated test time. These measures were taken to standardize participant behaviors and minimize the influence of potential confounding factors. On the designated test days, participants visited the hospitals and received detailed explanations about the study. They provided written informed consent and underwent polysomnography at each institution. Polysomnography recordings were conducted in a controlled sleep laboratory environment in accordance with the guidelines recommended by the American Academy of Sleep Medicine (AASM) [22]. Two technicians independently interpreted the results, followed by a review by sleep physicians.

To address the issue of interference due to multiple CSTs sharing the same biosignals, the participants were divided into 2 groups in both medical institutions: multi-tracker group A and multi-tracker group B, as illustrated in Figure 1. The configurations of the CSTs are presented in Figure 2. Specifically, at SNUBH, multi-tracker group A consisted of 18 individuals and multi-tracker group B consisted of 19 individuals. Similarly, at CLC, each group included 19 individuals. Across both institutions, the demographic statistics for participants in multi-tracker groups A and B demonstrated no significant differences across all metrics, as presented in Multimedia Appendix 2. Each group included a combination of noninterfering CSTs. Specifically, the nearables Google Nest Hub 2 and Amazon Halo Rise, which use similar radar sensors to detect motion, were allocated to different groups. In the case of wearables, participants were allowed to simultaneously wear a maximum of 2 watch devices, which are a type of wearable, with 1 on each wrist. Consequently, Fitbit Sense 2 and Pixel Watch were assigned to multi-tracker group A, while Galaxy Watch 5 and Apple Watch 8 were assigned to multi-tracker group B. As a result, these devices were expected to yield approximately half of the intended measurements. Airables, which were available on both iOS and Android devices (SleepRoutine and SleepScore), were analyzed, with half of them on iOS and the other half on Android. We used Pillow on iOS, as it is not available on Android. The polysomnography and CST results were then compared and analyzed.

Figure 1. Flowchart outlining the experimental design. Experimental procedures involving subject enrollment, CST assignment, and experimental settings for simultaneous measurement involving both CSTs and PSG. CLC: Clionic Lifecare Clinic; CST: consumer sleep tracker; PSG: polysomnography; SNUBH: Seoul National University Bundang Hospital.

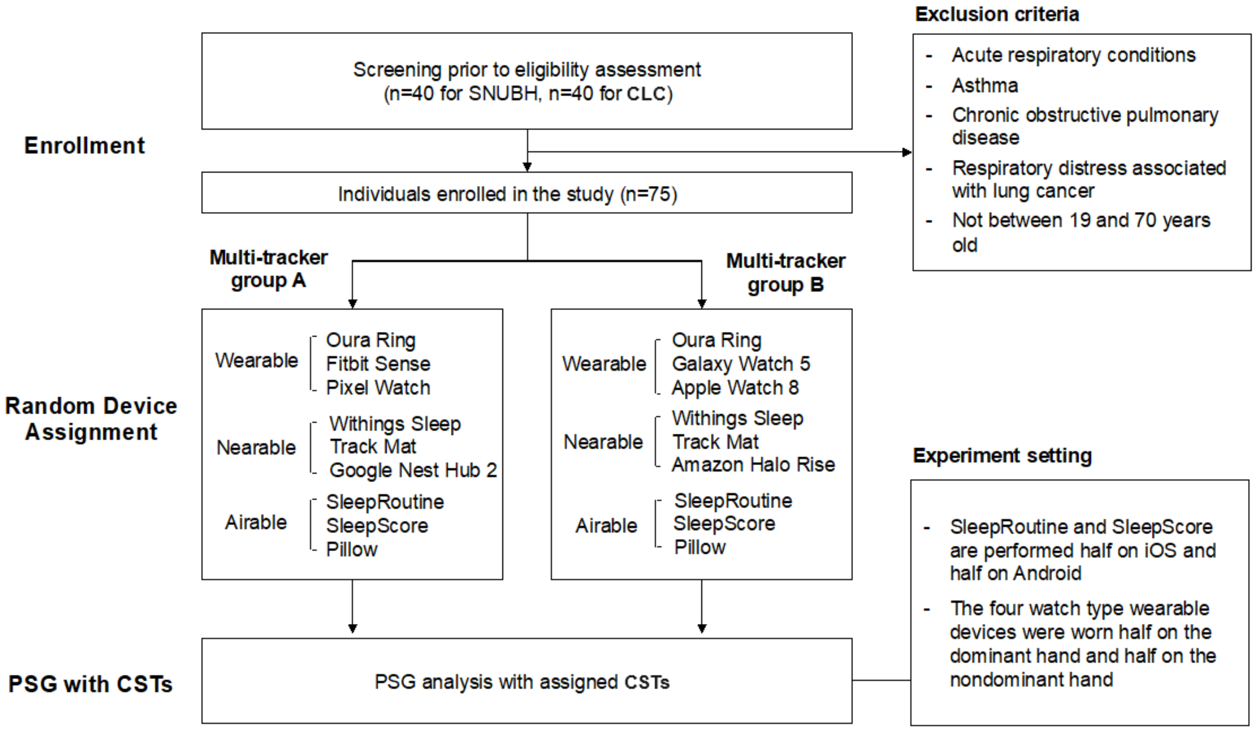
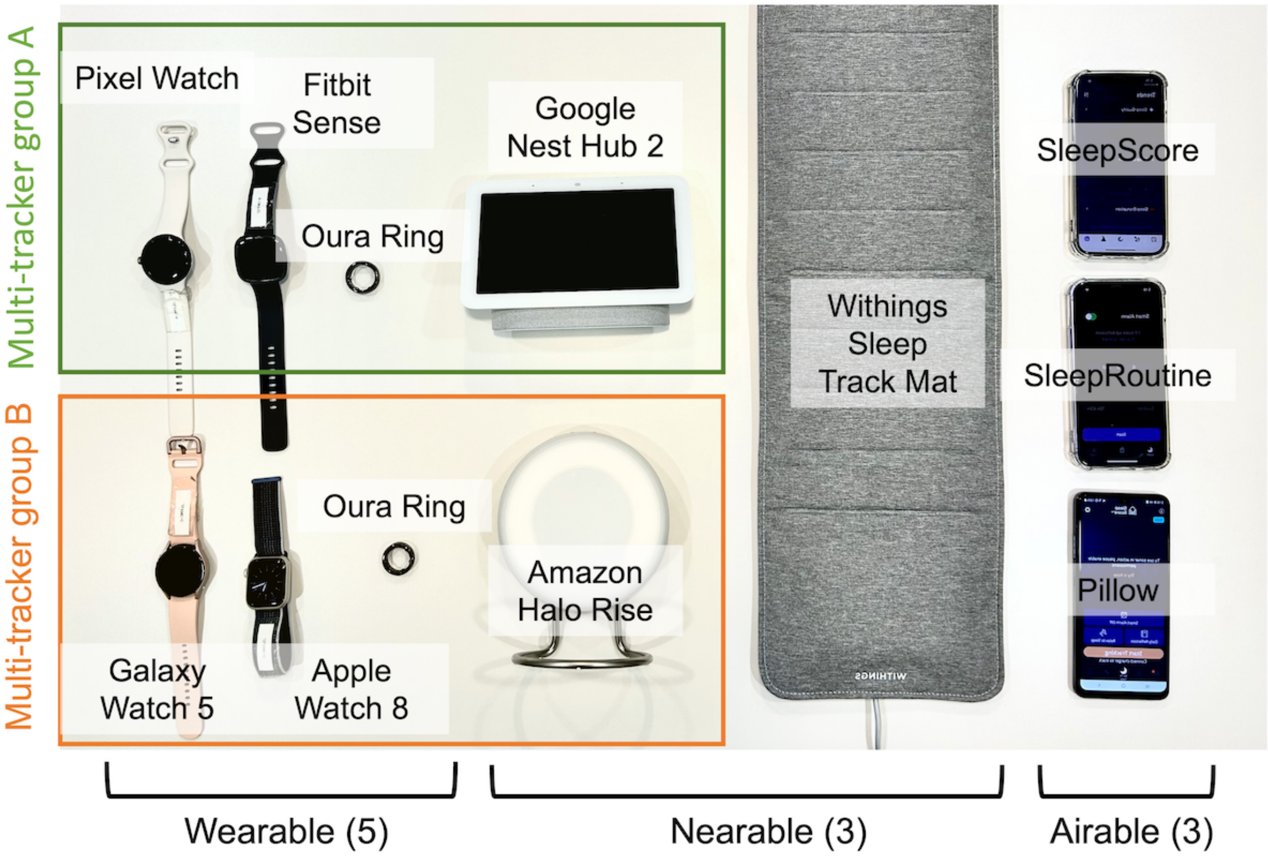


Figure 2. Configuration of consumer sleep trackers used in the experiment.



Ethics Approval

Ethics approval was obtained from the respective Institutional Review Board (IRB) of each institution (IRB number B-2302-908-301 from SNUBH and number P01-202302-01-048 from CLC).

Statistical Methods and Evaluation Metrics

Two-sample independent *t* tests were employed to compare demographic information (Table 1) and sleep measures, and significance was determined based on a *P* value of <.05. The average sleep measurements were compared, and any proportional bias was assessed using Pearson correlation and a Bland-Altman plot. The study used sensitivity, specificity, and F1 scores as evaluation metrics for sleep stage classification. Macro F1 scores, weighted F1 scores, and kappa values were used to summarize the results of the evaluation, considering the imbalance in data classes, such as sleep stages. All statistical analyses and visualizations were conducted using Python 3 (version 3.9.16) and used the scikit-learn, matplotlib, and scipy libraries.

Data Preprocessing

Three main steps were followed in the data processing stage. First, raw sleep score data were extracted from each CST device, either through direct download via the manufacturer's app or the web portal, or by requesting raw data from the SleepRoutine device manufacturer. The sleep score codes were standardized across devices, with the wake stage assigned 0, the light stage assigned 1, the deep stage assigned 2, and the rapid eye movement (REM) stage assigned 3. Apple Watch 8 used alternative expressions, such as "core sleep" instead of light sleep.

The extracted data were synchronized in time to compare CST results of sleep tracking with polysomnography accurately. The sleep stages measured by devices earlier than polysomnography scores were discarded. Conversely, for devices that started measuring after polysomnography scoring, the sleep stages were marked as the wake stage until the measurement began. The end point of all device measurements was aligned with the end of polysomnography, resulting in consistent measurement of total time in bed across the devices.

Because the sleep stages changed at every epoch, the results may be inaccurate if the start time of the 30-second epoch differed. Therefore, the results of sleep stages involving all devices, including polysomnography, were segmented into 1-second intervals and compared every second. This approach enabled a more precise comparison of sleep stage results between polysomnography and CST measurements, and eliminated potential bias.

Results

Epoch-by-Epoch Analysis: Overall Performance

Table 2 presents the results of epoch-by-epoch agreements between polysomnography and each of the 11 CSTs under the sleep stage classification. SleepRoutine (airable) demonstrated the highest macro F1 score of 0.6863, which was closely followed by Amazon Halo Rise (nearable), with a macro F1 score of 0.6242. In terms of Cohen κ, a measure of interrater agreement, 3 wearables (Google Pixel Watch, Galaxy Watch 5, and Fitbit Sense 2), 1 nearable (Amazon Halo Rise), and 1 airable (SleepRoutine) demonstrated moderate agreement with sleep stage classification (κ=0.4-0.6). On the other hand, 2 wearables (Apple Watch 8 and Oura Ring 3), 1 nearable (Withings Sleep Tracking Mat), and 1 airable (SleepScore) showed a fair level of agreement (κ=0.2-0.4). Finally, Google Nest Hub 2 (nearable) and Pillow (airable) exhibited only a slight level of agreement across sleep stage classifications. The performance of CSTs was assessed in 2 distinct institutions, where the macro F1 scores, averaged over all devices in each institution, were 0.4973 and 0.4876 at SNUBH and CLC, respectively. There was no significant difference in performance between these 2 locations. Among the 11 CSTs evaluated, 5 (Galaxy Watch 5, Apple Watch 8, Amazon Halo Rise, Pillow, and SleepRoutine) exhibited better performance at SNUBH, while the remaining CSTs demonstrated superior performance at CLC.

Table 2. Epoch-by-epoch agreement: classification of 4 sleep stages.

Variable	Overall				SNUBH[a]	CLC[b]
	Accuracy	Weighted F1	Cohen κ	Macro F1	Macro F1	Macro F1
Airable						
SleepRoutine (n=67)[c]	0.7106[d]	0.7166[d]	0.5565[d]	0.6863[d]	0.7188[d]	0.6551[d]
SleepScore (n=38)	0.4329	0.4472	0.2065	0.4049	0.4408	0.3094
Pillow (n=74)	0.2830	0.2906	0.0741	0.2588	0.2604	0.2564
Nearable						
Withings Sleep Tracking Mat (n=75)	0.4921	0.5007	0.2455	0.4496	0.4205	0.4837
Google Nest Hub 2 (n=33)	0.4121	0.4089	0.0644	0.3009	0.2676	0.3299
Amazon Halo Rise (n=28)	0.6634	0.6706	0.4807	0.6242	0.6231	0.6031
Wearable						
Google Pixel Watch (n=30)	0.6355	0.6143	0.4044	0.5669	0.5381	0.5925
Galaxy Watch 5 (n=22)	0.6494	0.6499	0.4177	0.5761	0.6261	0.5651
Fitbit Sense 2 (n=26)	0.6464	0.6296	0.4185	0.5814	0.5130	0.6268
Apple Watch 8 (n=26)	0.5640	0.5731	0.2976	0.4910	0.5436	0.4203
Oura Ring 3 (n=53)	0.5427	0.5518	0.3492	0.5186	0.5187	0.5211

[a]SNUBH: Seoul National University Bundang Hospital.

[b]CLC: Clionic Life Center.

[c]The number in parenthesis indicates the number of participants tested with each device.

[d]Top-performing consumer sleep tracker.

Epoch-by-Epoch Analysis: Performance According to Sleep Stages

The performance of various sleep trackers across different sleep stages is presented in Table 3. For the wake and REM stages, SleepRoutine (airable) achieved the highest macro F1 scores of 0.7065 and 0.7596, respectively. These scores substantially surpassed those of the second-best tracker, Amazon Halo Rise (nearable), by a margin of 0.1098 for the wake stage and 0.0313 for the REM stage. For the deep stage, Google Pixel Watch and Fitbit Sense 2, which are wearables, exhibited superior performance with macro F1 scores of 0.5933 and 0.5564, respectively. Google Pixel Watch achieved the highest performance with a substantial margin. It surpassed Fitbit Sense 2 by a margin of 0.0368 and outpaced SleepRoutine, which was the sleep tracker with the third highest score, with an even larger margin of 0.0567. For the light stage, an array of sleep trackers, including 3 wearables (Google Pixel Watch, Galaxy Watch 5, and Fitbit Sense 2), 1 nearable (Amazon Halo Rise), and 1 airable (SleepRoutine), demonstrated similarly high levels of performance, with a macro F1 score ranging from 0.7142 to 0.7436. Additional detailed assessments of sleep stage performance, including accuracy, weighted F1, and area under the receiver operating characteristic curve metrics, are presented in Multimedia Appendices 3-6.

Figure 3 presents the confusion matrices for the sleep stages of the 11 CSTs, providing a clear visual representation of prediction biases and misclassification. Multimedia Appendix 7 presents the mean and variance of predicted values across participants. Analysis of the average tendencies across all devices revealed a prediction bias toward the light sleep stage. Google Nest Hub 2 (nearable) showed the largest bias toward the light stage among all the devices. Unlike other devices, Pillow (airable) was highly biased toward the deep stage, predicting 59% of epochs as deep, whereas only 10.8% of epochs were deep based on the results of polysomnography. The confusion matrices also revealed distinct patterns of misclassification in sleep stage prediction for device types. Wearables primarily misclassified wake as light, while nearables strongly misclassified REM as light. Airables, on the other hand, demonstrated a relatively higher frequency of confusion between the light and deep stages. Figure 4 presents a comparison of hypnograms illustrating the epoch-by-epoch agreement at the individual level, which facilitated the evaluation of agreement between CSTs and polysomnography in a time-series format. Additional hypnograms are presented in Multimedia Appendix 8.

Regarding Figure 4, the division of groups was necessary owing to the limited number of watches worn simultaneously, as explained in the Methods section. As 9 devices were used simultaneously for each subject, the hypnograms for each device are presented, with the polysomnography result displayed at the top. As shown in Figure 4, SleepRoutine, Amazon Halo Rise, and Galaxy Watch 5 exhibited more frequent transition of stages and predicted wake in the middle of sleep more frequently, resulting in better estimation of WASO, as shown in the analysis of sleep parameters.

Table 3. Epoch-by-epoch agreement: classification for detecting individual sleep stage.

Variable	Wake stage[a]			Light stage[a]			Deep stage[a]			REM[b] stage[a]		
	F1	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1	Sensitivity	Specificity
Airable												
SleepRoutine (n=67)[c]	0.7065[d]	0.7246[d]	0.9269	0.7436[d]	0.7054	0.7665[d]	0.5355	0.6712	0.8973	0.7596[d]	0.7394	0.9609[d]
SleepScore (n=38)	0.4057	0.3665	0.8696	0.5147	0.4355	0.7272	0.3574	0.5247	0.8264	0.3418	0.4587	0.7895
Pillow (n=74)	0.2828	0.1934	0.9572	0.3409	0.2490	0.7534	0.2673	0.8594[d]	0.4449	0.1440	0.1140	0.9126
Nearable												
Withings Sleep Tracking Mat (n=75)	0.4419	0.4172	0.8854	0.5764	0.5328	0.6336	0.3800	0.5633	0.8270	0.4001	0.3964	0.8906
Google Nest Hub 2 (n=33)	0.3296	0.3068	0.8649	0.5619	0.5772	0.4518	0.1245	0.1308	0.8883	0.1876	0.1805	0.8514
Amazon Halo Rise (n=28)	0.5967	0.6612	0.8921	0.7142	0.6609	0.7484	0.4575	0.5467	0.9018	0.7283	0.7490[d]	0.9401
Wearable												
Google Pixel Watch (n=30)	0.3456	0.2277	0.9784[d]	0.7150	0.7657	0.5620	0.5922[d]	0.6937	0.9290	0.6146	0.6548	0.9029
Galaxy Watch 5 (n=22)	0.4755	0.4814	0.9104	0.7346	0.7280	0.6412	0.4963	0.4752	0.9481[d]	0.5982	0.6265	0.9058
Fitbit Sense 2 (n=26)	0.3807	0.2714	0.9602	0.7262	0.7734[d]	0.5727	0.5564	0.6710	0.9247	0.6623	0.6812	0.9297
Apple Watch 8 (n=26)	0.5493	0.4481	0.9624	0.6680	0.6649	0.5737	0.3073	0.4130	0.8412	0.4394	0.4276	0.9070
Oura Ring 3 (n=53)	0.4527	0.3822	0.9264	0.5953	0.5072	0.7630	0.4272	0.7784	0.7974	0.5993	0.7118	0.8716

[a]Individual sleep stage classification was used to categorize each class and the remaining classes.

[b]REM: rapid eye movement.

[c]The number in parenthesis indicates the number of participants tested with each device.

[d]Top-performing consumer sleep tracker.

Figure 3. Normalized confusion matrices for 11 consumer sleep trackers (CSTs). Four-stage sleep classification confusion matrices comparing CSTs. Each row in the confusion matrix is the sleep stage annotated by polysomnography, while each column represents the sleep stage annotated by the CST. REM: rapid eye movement.

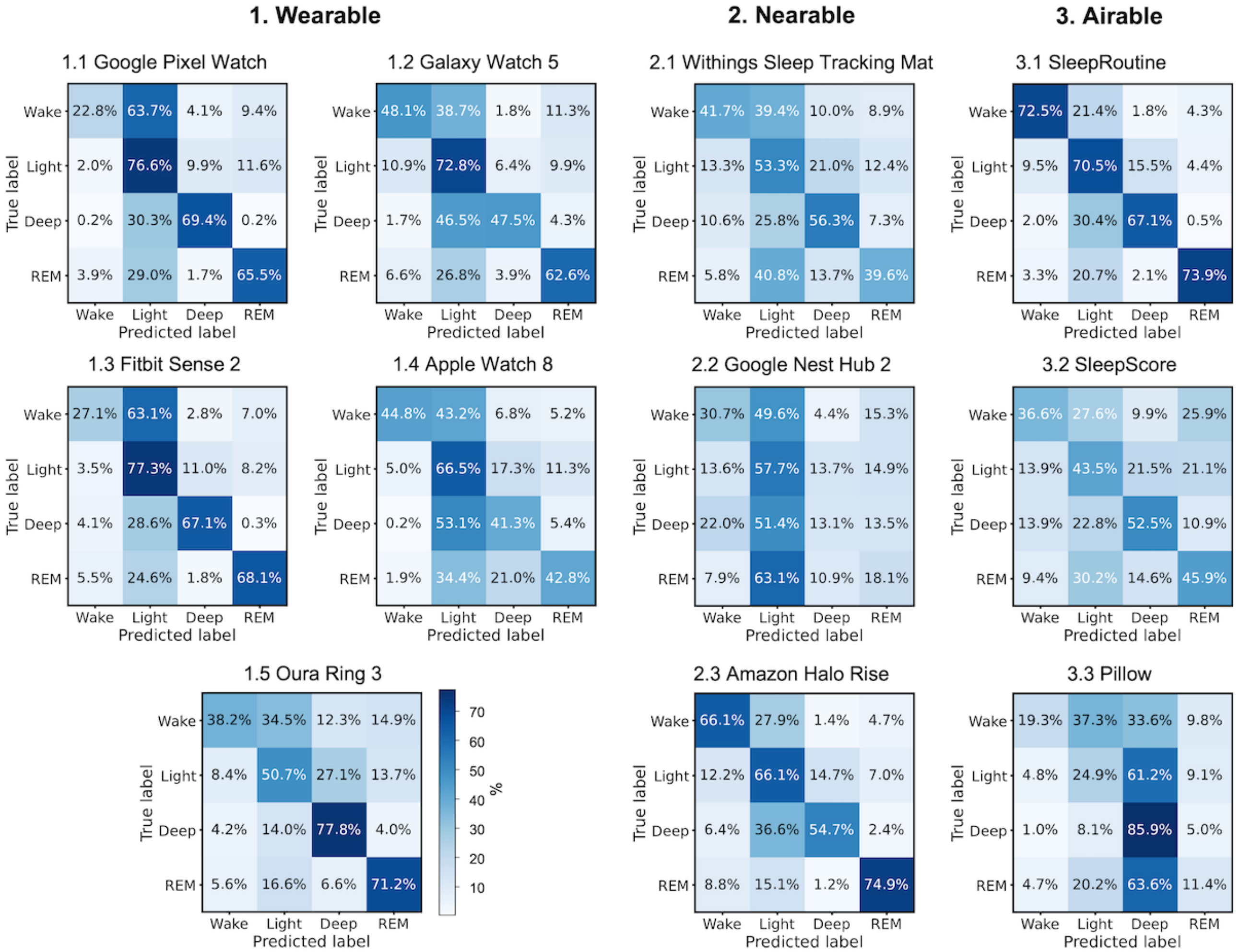
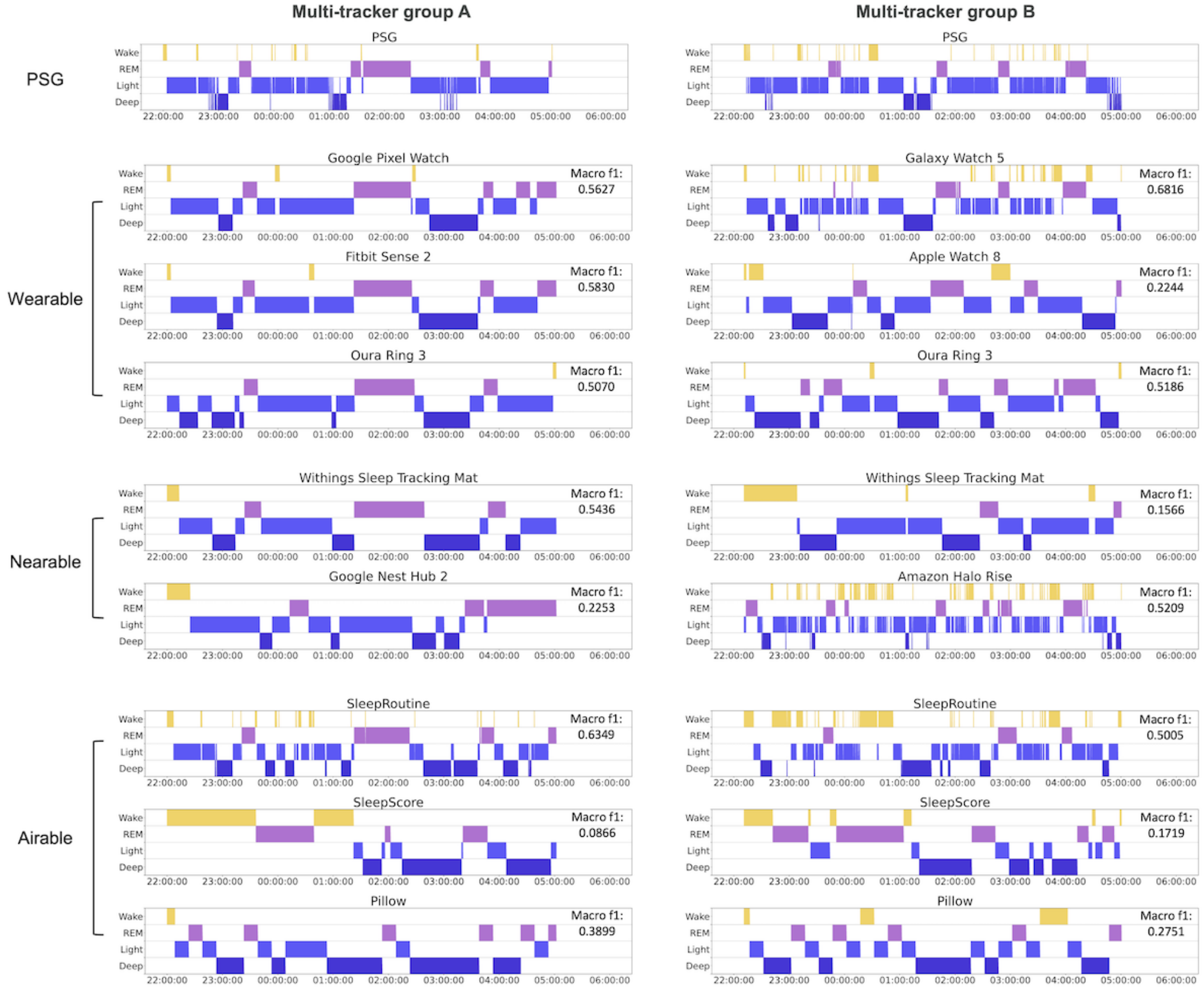


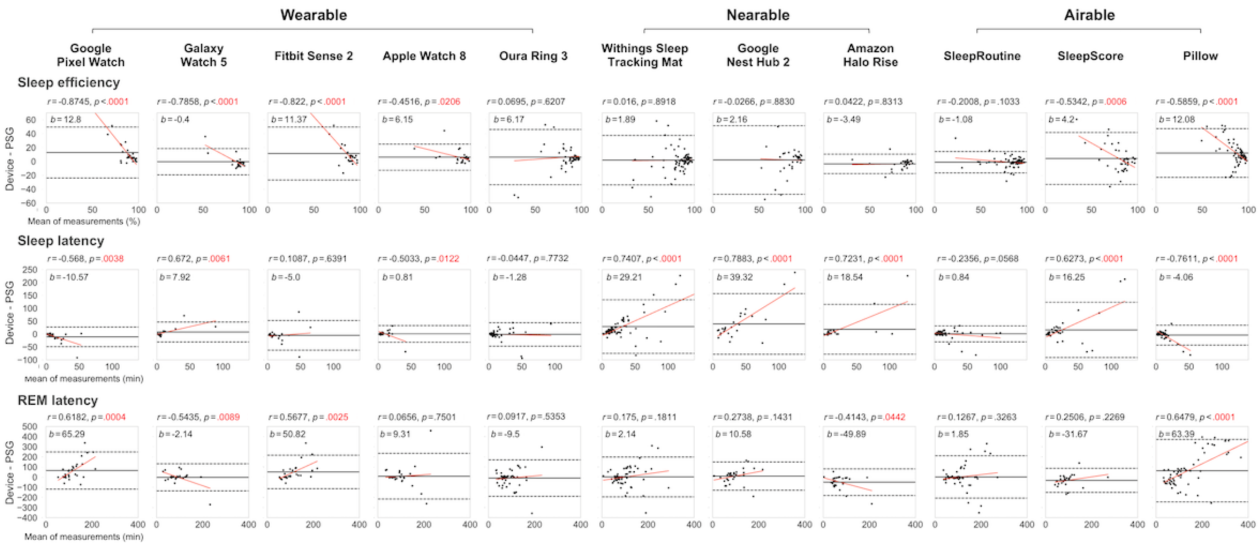
Figure 4. Sample hypnograms of 11 consumer sleep trackers (CSTs) involving 2 subjects in different groups. Hypnogram samples for each CST were selected based on the last measured subjects in multi-tracker group A (female; age, 35 years; BMI, 30.1; apnea-hypopnea index [AHI], 2.9) and multi-tracker group B (female; age, 26 years; BMI, 20; AHI, 3.5). PSG, polysomnography.



Sleep Measure Analysis

Figure 5 presents the Bland-Altman plots of CSTs, illustrating the performance of sleep measurements, including sleep efficiency, sleep latency, and REM latency, when compared with polysomnography. The average value of polysomnography sleep efficiency ranged from 77.57% to 86.05%, while the bias for each CST varied from −3.4909 percentage points (Amazon Halo Rise) to 12.8035 percentage points (Google Pixel Watch). Polysomnography values for sleep latency ranged from 10.80 minutes to 19.80 minutes, with CST biases ranging from −0.81 minutes (Apple Watch 8) to 39.42 minutes (Google Nest Hub 2). Polysomnography values for REM latency ranged from 87.00 minutes to 112.20 minutes, with CST biases ranging from −49.89 minutes (Amazon Halo Rise) to 65.29 minutes (Google Pixel Watch). The devices demonstrated distinct and best performances for each sleep metric. In terms of sleep efficiency, Galaxy Watch 5 (wearable) achieved a minimal bias of −0.4%. In the case of estimation of sleep latency, Apple Watch 8 (wearable) exhibited a bias of 0.81 minutes. Lastly, SleepRoutine (airable) demonstrated the best performance for REM latency with a bias of 1.85 minutes. The proportional bias, presented as “r” in Figure 5, indicates how consistent the mean bias was regardless of the sleep measure. Oura Ring and SleepRoutine showed no proportional bias (ie, no significant correlation in the Bland-Altman plot) for any sleep measure. The difference in mean values between polysomnography and each CST for each sleep measure is described in Multimedia Appendix 9. Additional information is provided in Multimedia Appendix 10.

Figure 5. Bland-Altman plots of consumer sleep trackers (CSTs) and polysomnography (PSG) for sleep efficiency, sleep latency, and rapid eye movement (REM) latency measurements. The plots present the mean bias (middle horizontal black solid line), and upper (upper horizontal black dashed line) and lower (lower horizontal black dashed line) limits of agreement. In the figure, “b” represents bias and “r” denotes the Pearson correlation coefficient between the mean of measurements and the difference between the CSTs and PSG. The correlation coefficient is displayed along with its corresponding *P* value. The red line indicates the estimated linear regression line.



Subgroup Analysis

Subgroup analyses were conducted for all devices, considering factors, including sex assigned at birth, AHI, sleep efficiency, and BMI. The macro F1 scores for each subgroup are presented in Table 4. Multimedia Appendix 11 presents the subgroup analysis results of epoch-by-epoch agreement for the AHI. The average performance of CSTs showed a comprehensive relationship between sleep tracker performance and these parameters. In terms of BMI, the average macro F1 score was 0.5043 for individuals with a BMI of ≤25 kg/m^2, whereas it dropped to 0.4790 for those with a BMI of >25 kg/m^2, indicating a gap of 0.0253. Similarly, for sleep efficiency, the scores were 0.4757 for individuals with a sleep efficiency of ≤85% and 0.4902 for those with a sleep efficiency of >85%, with a difference of 0.0145. In the case of the AHI, the scores were 0.4905 for an AHI of ≤15 and 0.5024 for an AHI of >15, resulting in a difference of 0.0119. In contrast, the difference between male and female subgroups was minimal, with a macro F1 score of 0.4926 for males and 0.4932 for females, resulting in a negligible difference of 0.0006. In each subgroup, the highest variations were observed with the airable SleepScore for AHI (difference: 0.0929), the nearable Google Pixel Watch for sleep efficiency (difference: 0.1067), the wearable Galaxy Watch 5 for BMI (difference: 0.0785), and the airable SleepScore for sex assigned at birth (difference: 0.0872). Multimedia Appendices 12 and 13 present the subgroup analysis results of epoch-by-epoch agreement in the institutions. Additionally, Multimedia Appendices 14 and 15 provide an overview of the average macro F1 scores individually calculated for each participant.

Table 4. Epoch-by-epoch agreement: subgroup analysis of the apnea-hypopnea index and demographic characteristics.

Variable	Apnea-hypopnea index		Sleep efficiency		BMI		Gender	
	≤15	>15	≤85%	>85%	≤25	>25	Male	Female
Airable								
SleepRoutine (n=67)[a]	0.6536[b]	0.7320[b]	0.6971[b]	0.6490[b]	0.6840[b]	0.6889[b]	0.7137[b]	0.6568[b]
SleepScore (n=38)	0.3636	0.4565	0.4107	0.3808	0.4118	0.3937	0.4431	0.3559
Pillow (n=74)	0.2602	0.2567	0.2472	0.2567	0.2601	0.2548	0.2670	0.2446
Nearable								
Withings Sleep Tracking Mat (n=75)	0.4644	0.4225	0.3766	0.4653	0.4760	0.3964	0.4587	0.4355
Google Nest Hub 2 (n=33)	0.3000	0.3059	0.3115	0.2762	0.3209	0.2517	0.2889	0.3059
Amazon Halo Rise (n=28)	0.6160	0.6389	0.6297	0.5857	0.6414	0.5801	0.6075	0.6491
Wearable								
Google Pixel Watch (n=30)	0.5670	0.5626	0.5035	0.6102	0.5653	0.5791	0.5235	0.5956
Galaxy Watch 5 (n=22)	0.5701	0.5790	0.6029	0.5547	0.5521	0.6306	0.5655	0.5867
Fitbit Sense 2 (n=26)	0.5839	0.5753	0.5325	0.6090	0.5910	0.5541	0.5320	0.6129
Apple Watch 8 (n=26)	0.4861	0.4950	0.4326	0.4804	0.5093	0.4561	0.5263	0.4414
Oura Ring 3 (n=53)	0.5302	0.5021	0.4882	0.5245	0.5354	0.4830	0.4926	0.5405

[a]The number in parenthesis indicates the number of participants tested with each device.

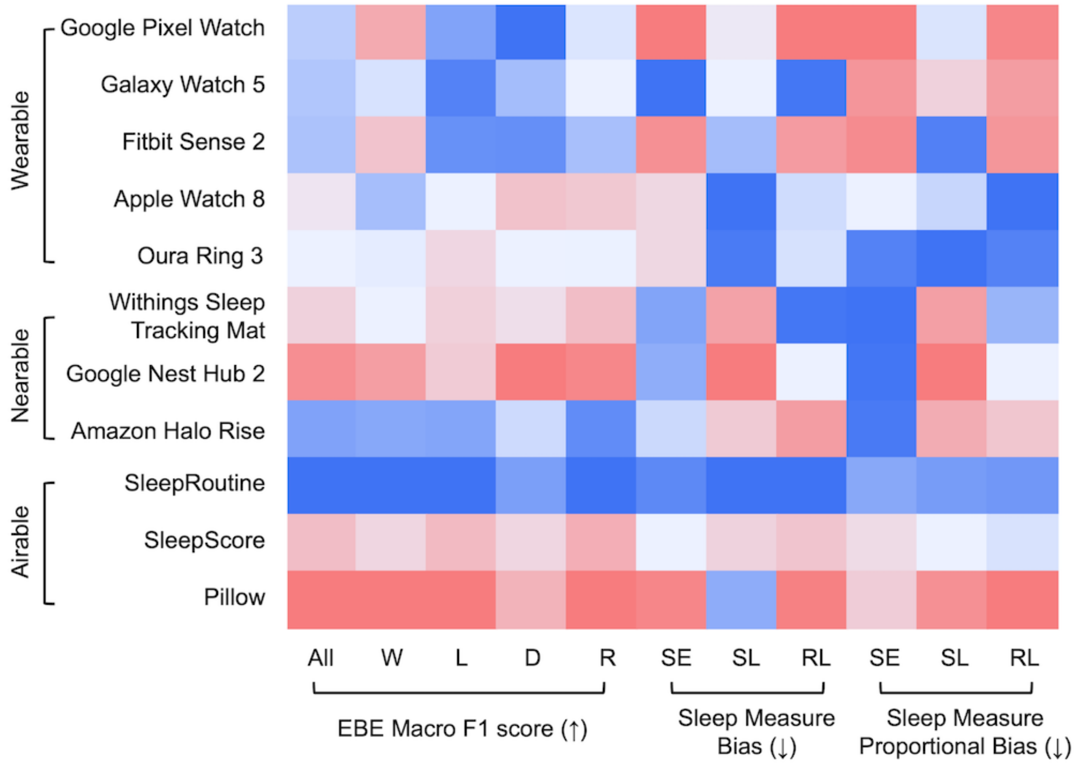
[b]Top-performing consumer sleep trackers.

Discussion

Principal Findings

We conducted an extensive analysis of 11 CSTs involving 75 subjects, which, to the best of our knowledge, represents the largest number of devices simultaneously evaluated in the literature [3,4,10,11,13,18]. The findings are illustrated in Figure 6, which presents the relative performances of the 11 CSTs in estimating sleep stages and sleep measures. Our findings revealed that Google Pixel Watch, Galaxy Watch 5, and Fitbit Sense 2 demonstrated competitive performance among wearables, while Amazon Halo Rise and SleepRoutine stood out among nearables and airables, respectively.

Figure 6. Relative performance rank heatmap of 11 consumer sleep trackers (CSTs). Heatmap of the relative performance of sleep stage and classification of sleep measures normalized to the highest and lowest macro F1 values in each CST. D: deep; EBE: epoch-by-epoch agreement; L: light; R: rapid eye movement; RL: rapid eye movement latency; SE: sleep efficiency; SL: sleep latency; W: wake.



Wearables

Wearables, including watch and ring-type sleep trackers, represent the most prevalent CSTs in the market [23,24]. They employ photoplethysmography sensors and accelerometer sensors to measure cardiac activity (eg, heart rate variability) and body movements. Given their reliance on similar biosignals for sleep tracking, wearables exhibit consistent patterns in estimating sleep stages. First, most wearables generally overestimate sleep by misclassifying wake stages, leading to a substantial negative proportional bias in estimating sleep efficiency, which results in worse performance for individuals with low sleep efficiency (Figure 5). This bias was specifically observed when actigraphy was used to measure sleep efficiency and WASO [5,11]. This can be attributed to the dependence of actigraphy and wearables on body movement to determine sleep-wake states. Insomniacs often lie still in bed while trying to sleep, even though they are actually awake [24]. As a result, these periods of wakefulness can be misinterpreted as sleep. Nevertheless, Oura Ring showed negligible proportional bias, potentially owing to its use of additional features, such as body temperature and circadian rhythm, for sleep staging [6]. Second, wearables comprising the top 3 CSTs demonstrated substantial alignment in the classification of deep stages. In particular, the results from Oura Ring 3 and Fitbit Sense 2 in this study showed improved accuracy in sleep stage detection compared to previous studies that focused on earlier versions of Oura Ring and Fitbit in assessing the accuracy of wearable sleep evaluations [25,26]. Thus, wearables may facilitate accurate detection of different stages of deep sleep, given their unique association with autonomic nervous system stabilization. Heart rate variability, a key indicator of autonomic nervous system activity, can be directly measured by photoplethysmography sensors [27]. Therefore, wearables are effective in monitoring deep sleep stages.

Nearables

Nearables, encompassing pad and motion sensor-type sleep trackers, use overall body movements and respiratory efforts (thoracic and abdominal) for sleep monitoring. Similar to wearables, nearables also exhibit aligned tendencies. First, all nearables tend to overestimate sleep onset latency, resulting in a significant mean bias (29.02 minutes for nearables, −2.71 minutes for wearables, and 4.34 minutes for airables) and a significant positive proportional bias in sleep latency measurement. This indicates that nearables may overestimate sleep latency, particularly in individuals with prolonged sleep latency. During extended periods of attempting to fall asleep in bed, users may experience increased restlessness and movement, which makes it challenging for nearables to estimate the sleep stage using radar-like sensors [28]. Second, unlike wearables, nearables demonstrated the least sensitivity in deep stage classification (as shown in Figure 3). Distinguishing stages of deep sleep from light sleep based on variations in respiratory patterns requires precise monitoring of respiratory activity. However, the radar-like sensors employed by nearables, while efficient at detecting larger body movements, have difficulty capturing smaller fidgeting movements, which represent a challenge in accurately identifying the stage of deep sleep.

Airables

Airables excel in terms of their accessibility by not requiring the purchase of additional hardware. However, their clinical validation is not well-established as noted in previous studies up until 2022, which highlighted the limited agreement between airables and polysomnography as a notable limitation [25,29]. In this study, our aim was to validate the latest airable CSTs. One distinguishing feature of airable CSTs is their use of diverse sensor types, and the variation in performance is substantially influenced by the specific sensor type and accompanying algorithm. Thus, we chose 3 types of airable CSTs considering diversity (microphone, ultrasound, and accelerometer-based applications). These methodological distinctions contribute to pronounced variations in the determination of sleep stage. Pillow requires placement on the mattress and uses the smartphone's accelerometer sensor to detect user movements through the mattress. Notably, Pillow showed a prediction bias toward the deep stage, suggesting that movement information during sleep was insufficient for the accurate determination of sleep stage. SleepScore uses a sonar biomotion sensor and directs the smartphone's speaker toward the chest area to emit ultrasonic signals above 18 kHz, tracking thoracic respiratory effort. Depending on the biosignal used, SleepScore shows similar tendencies with nearables, demonstrating a substantial mean bias and positive proportional bias in estimating sleep onset latency. SleepRoutine analyzes the sound recorded during sleep [2]. Sleep sounds provide a wealth of sleep-related information, including changes in breathing regularity linked to autonomic nervous system stabilization, changes in breathing sound characteristics (such as tone, pitch, and amplitude) due to altered respiratory muscle tension, and noise from body movements. Among all CSTs, SleepRoutine exhibited the highest accuracy in predicting the wake and REM stages.

REM Sleep Stage Estimation Performance

REM was the stage where most CSTs demonstrated relatively higher agreement with polysomnography compared with other stages. Among the top 5 CSTs with the highest macro F1 scores (SleepRoutine, Amazon Halo Rise, Fitbit Sense 2, Galaxy Watch 5, and Google Pixel Watch), the REM stage showed a substantially higher average F1 score of 0.672, compared with 0.501 for wake and 0.528 for deep sleep. This can be attributed to the unique characteristics of REM sleep, which include increased irregularity in heart rate and breathing, minimal muscle movement, and rapid variations in blood pressure and body temperature [30]. These features allow easy detection of different types of biosignals and accurate classification of REM sleep.

Cost-Effectiveness

We evaluated the costs of 11 sleep tracking technologies based on the costs analyzed in Multimedia Appendix 16. Wearables, with an average price of US $386, offer a wide range of functions, including messaging and various apps, beyond sleep tracking. Oura Ring, while lacking these supplementary features, provides a broad spectrum of health tracking functions. Nearables, with an average price of US $123, include a variety of features across different models. Google Nest Hub and Amazon Halo Rise offer extra features, such as an IoT hub and wake-up light, whereas Withings Sleep Mat is exclusively designed for sleep tracking. Airables, which are app-based technologies, harness smartphone sensors for sleep tracking, requiring only a subscription fee of US $53 and no additional hardware. This economical and flexible option, which can be easily canceled, represents a cost-effective solution for sleep tracking.

Standardized Validation and Data Transparency

Standardized methods of validation and data transparency are crucial for comparing sleep trackers [31], particularly due to the increasing use of deep learning algorithms whose inner workings are often opaque. In our study, we adhered to established frameworks for standardized validation [15,32], while also conducting multi-center evaluations based on diverse demographic factors. Regarding data transparency, we provided comprehensive details of validation; however, obtaining access to the training data of each CST was challenging. Transparency in both training and validation data is essential for building trustworthy artificial intelligence models and can also contribute to a better understanding of CSTs [33].

Limitations

It is important to note the limitations of our study. First, data collection rates significantly varied between the 2 institutions as the study was independently implemented. Issues, such as battery management, account management, and human errors, resulted in data omissions. Second, demographic differences were detected between the institutions, including disparities in time spent in bed and total sleep time. Operational issues led to slightly earlier waking of participants in CLC. Third, this study focused solely on the Korean population, with limited ability to analyze performance differences among various races. Future studies should incorporate multiracial comparisons and evaluate CST performance across diverse home environments for realistic assessments.

Conclusions

Our study represents a comprehensive and comparative analysis of 11 CSTs and their accuracy in tracking sleep in a sleep lab setting. The objective of this study was to gain insights into the performance and capabilities of these CSTs. Personalized sleep health management is necessary to enable individuals to make informed choices for monitoring and improving sleep quality. Further, our findings emphasize the importance of understanding the characteristics and limitations of these devices. It lays the foundation for guiding the development of sleep trackers in the future. Accordingly, future studies should focus on developing accurate sleep stage classification systems by integrating different types of biosignals in a home environment.

Acknowledgments

SleepRoutine, which is one of the consumer sleep trackers included in this study, was developed and operated by Asleep, and the deep learning algorithm used in SleepRoutine was trained using data from Seoul National University Bundang Hospital, to which some of the authors are affiliated. The entities played no role in study design, data collection, analysis and interpretation of data, or writing of this manuscript. The authors conducted the study independently. This work was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT; number: RS-2022-00156561).

Data Availability

Access to data sets from Seoul National University Bundang Hospital and Clionic Lifecare Clinic, used with permission for this study, should be requested directly from these institutions. Subject to ethical approval from the institutional review boards, the corresponding author agrees to share deidentified individual participant data, the study protocol, and the statistical analysis plan with academic researchers following completion of a data use agreement. Proposals should be directed to the corresponding author. All experimental and implementation details that can be shared are described in detail in this paper and the multimedia appendices.

Authors' Contributions

TL and YC contributed to conceptualization, data curation, formal analysis, investigation, methodology, writing the original draft, and review and editing. KSC and JJ performed software coding, validation, and visualization, and reviewed the draft. JJ, JC, HK, DK, JH, and DL contributed to conceptualization, software coding, and manuscript revision. MK conducted data curation and validation. CAK supervised the project, contributed to conceptualization, and provided guidance. IYY and JWK are the principal investigators and contributed to conceptualization, design, interpretation, writing, and project administration. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

YC is the director of Clionic Lifecare Clinic. Authors affiliated with Asleep and CAK hold stocks or stock options in Asleep.

Multimedia Appendix 1
The number of observations in each institution.
[PDF File (Adobe PDF File), 37 KB - mhealth_v11i1e50983_app1.pdf]

Multimedia Appendix 2
Comparative analysis of participant demographics across institutions (multi-tracker group A vs multi-tracker group B).
[PDF File (Adobe PDF File), 81 KB - mhealth_v11i1e50983_app2.pdf]

Multimedia Appendix 3
Epoch-by-epoch agreement: wake stage classification.
[PDF File (Adobe PDF File), 41 KB - mhealth_v11i1e50983_app3.pdf]

Multimedia Appendix 4
Epoch-by-epoch agreement: light stage classification.
[PDF File (Adobe PDF File), 41 KB - mhealth_v11i1e50983_app4.pdf]

Multimedia Appendix 5
Epoch-by-epoch agreement: deep stage classification.
[PDF File (Adobe PDF File), 41 KB - mhealth_v11i1e50983_app5.pdf]

Multimedia Appendix 6
Epoch-by-epoch agreement: rapid eye movement stage classification.
[PDF File (Adobe PDF File), 42 KB - mhealth_v11i1e50983_app6.pdf]

Multimedia Appendix 7
Normalized confusion matrices: mean and variance of predicted values across participants.
[PDF File (Adobe PDF File), 221 KB - mhealth_v11i1e50983_app7.pdf]

Multimedia Appendix 8
Additional hypnogram examples for 11 consumer sleep trackers of 2 subjects in different groups.
[PDF File (Adobe PDF File), 59 KB - mhealth_v11i1e50983_app8.pdf]

Multimedia Appendix 9
Comparison of sleep parameters with polysomnography.
[PDF File (Adobe PDF File), 62 KB - mhealth_v11i1e50983_app9.pdf]

Multimedia Appendix 10
Supplementary results of the sleep measure analysis.
[PDF File (Adobe PDF File), 62 KB - mhealth_v11i1e50983_app10.pdf]

Multimedia Appendix 11
Epoch-by-epoch agreement: subgroup analysis of the apnea-hypopnea index.
[PDF File (Adobe PDF File), 70 KB - mhealth_v11i1e50983_app11.pdf]

Multimedia Appendix 12
Epoch-by-epoch agreement: subgroup analysis of the apnea-hypopnea index and demographic characteristics in Seoul National University Bundang Hospital.
[PDF File (Adobe PDF File), 83 KB - mhealth_v11i1e50983_app12.pdf]

Multimedia Appendix 13
Epoch-by-epoch agreement: subgroup analysis of the apnea-hypopnea index and demographic characteristics in Clionic Lifecare Clinic.
[PDF File (Adobe PDF File), 83 KB - mhealth_v11i1e50983_app13.pdf]

Multimedia Appendix 14
Group-averaged macro F1 scores: subgroup analysis of the apnea-hypopnea index and demographic characteristics in Seoul National University Bundang Hospital.
[PDF File (Adobe PDF File), 150 KB - mhealth_v11i1e50983_app14.pdf]

Multimedia Appendix 15
Group-averaged macro F1 scores: subgroup analysis of the apnea-hypopnea index and demographic characteristics in Clionic Lifecare Clinic.
[PDF File (Adobe PDF File), 148 KB - mhealth_v11i1e50983_app15.pdf]

Multimedia Appendix 16
Cost-effectiveness of consumer sleep trackers.
[PDF File (Adobe PDF File), 102 KB - mhealth_v11i1e50983_app16.pdf]

References

1. Watson N, Badr M, Belenky G, Bliwise D, Buxton O, Buysse D, et al. Recommended amount of sleep for a healthy adult: a Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. Sleep 2015 Jun 01;38(6):843-844 [FREE Full text] [doi: 10.5665/sleep.4716] [Medline: 26039963]
2. Tran HH, Hong JK, Jang H, Jung J, Kim J, Hong J, et al. Prediction of sleep stages via deep learning using smartphone audio recordings in home environments: model development and validation. J Med Internet Res 2023 Jun 01;25:e46216. [doi: 10.2196/46216] [Medline: 37261889]
3. Fino E, Plazzi G, Filardi M, Marzocchi M, Pizza F, Vandi S, et al. (Not so) Smart sleep tracking through the phone: findings from a polysomnography study testing the reliability of four sleep applications. J Sleep Res 2020 Feb;29(1):e12935. [doi: 10.1111/jsr.12935] [Medline: 31674096]
4. Chinoy ED, Cuellar JA, Jameson JT, Markwald RR. Performance of four commercial wearable sleep-tracking devices tested under unrestricted conditions at home in healthy young adults. Nat Sci Sleep 2022 Mar;14:493-516. [doi: 10.2147/nss.s348795]
5. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 01;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]

6. Altini M, Kinnunen H. The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring. Sensors (Basel) 2021 Jun 23;21(13):4302 [FREE Full text] [doi: 10.3390/s21134302] [Medline: 34201861]
7. Hong JK, Lee T, Delos Reyes RD, Hong J, Tran HH, Lee D, et al. Confidence-based framework using deep learning for automated sleep stage scoring. Nat Sci Sleep 2021 Dec;13:2239-2250. [doi: 10.2147/nss.s333566]
8. Fino E, Mazzetti M. Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. Sleep Breath 2019 Mar 23;23(1):13-24. [doi: 10.1007/s11325-018-1661-3] [Medline: 29687190]
9. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep Parameter Assessment Accuracy of a Consumer Home Sleep Monitoring Ballistocardiograph Beddit Sleep Tracker: A Validation Study. J Clin Sleep Med 2019 Mar 15;15(3):483-487 [FREE Full text] [doi: 10.5664/jcsm.7682] [Medline: 30853052]
10. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
11. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291 [FREE Full text] [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
12. Klier K, Wagner M. Agreement of Sleep Measures-A Comparison between a Sleep Diary and Three Consumer Wearable Devices. Sensors (Basel) 2022 Aug 18;22(16):6189 [FREE Full text] [doi: 10.3390/s22166189] [Medline: 36015949]
13. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
14. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
15. Nguyen QNT, Le T, Huynh QBT, Setty A, Vo TV, Le TQ. Validation framework for sleep stage scoring in wearable sleep trackers and monitors with polysomnography ground truth. Clocks Sleep 2021 May 03;3(2):274-288 [FREE Full text] [doi: 10.3390/clockssleep3020017] [Medline: 34063579]
16. Menghini L, Yuksel D, Goldstone A, Baker FC, de Zambotti M. Performance of Fitbit Charge 3 against polysomnography in measuring sleep in adolescent boys and girls. Chronobiol Int 2021 Jul;38(7):1010-1022 [FREE Full text] [doi: 10.1080/07420528.2021.1903481] [Medline: 33792456]
17. Schade MM, Bauer CE, Murray BR, Gahan L, Doheny EP, Kilroy H, et al. Sleep Validity of a Non-Contact Bedside Movement and Respiration-Sensing Device. J Clin Sleep Med 2019 Jul 15;15(7):1051-1061 [FREE Full text] [doi: 10.5664/jcsm.7892] [Medline: 31383243]
18. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Rev Med Devices 2016 May;13(5):497-506. [doi: 10.1586/17434440.2016.1171708] [Medline: 27043070]
19. Robbins R, Seixas A, Masters LW, Chanko N, Diaby F, Vieira D, et al. Sleep tracking: A systematic review of the research using commercially available technology. Curr Sleep Med Rep 2019;5(3):156-163 [FREE Full text] [doi: 10.1007/s40675-019-00150-1] [Medline: 33134038]
20. Mouritzen NJ, Larsen LH, Lauritzen MH, Kjær T. Assessing the performance of a commercial multisensory sleep tracker. PLoS One 2020 Dec 11;15(12):e0243214 [FREE Full text] [doi: 10.1371/journal.pone.0243214] [Medline: 33306678]
21. Ananth S. Sleep apps: current limitations and challenges. Sleep Sci 2021;14(1):83-86 [FREE Full text] [doi: 10.5935/1984-0063.20200036] [Medline: 34104344]
22. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Neumosur. Darien, IL: American Academy of Sleep Medicine URL: https://www.neumosur.net/files/grupos-trabajo/suenio/AASM-Manual-2012.pdf [accessed 2023-10-06]
23. Global Market Insights. 2023. URL: https://www.gminsights.com/industry-analysis/sleep-tech-devices-market [accessed 2023-10-06]
24. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. J Clin Sleep Med 2008 Oct 15;4(5):487-504 [FREE Full text] [Medline: 18853708]
25. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557 [FREE Full text] [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
26. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
27. Vanoli E, Adamson PB, Ba-Lin, Pinna GD, Lazzara R, Orr WC. Heart rate variability during specific sleep stages. A comparison of healthy subjects with patients after myocardial infarction. Circulation 1995 Apr 01;91(7):1918-1922. [doi: 10.1161/01.cir.91.7.1918] [Medline: 7895347]

28. Quick Installation Guide - Withings Sleep Analyzer. Withings. URL: https://media.withings.com/kits/guides/2020/sleep-analyzer/qig_sleep-analyzer_en-fr-de-es-it-pt.pdf [accessed 2023-10-06]
29. Ibáñez V, Silva J, Navarro E, Cauli O. Sleep assessment devices: types, market analysis, and a critical view on accuracy and validation. Expert Rev Med Devices 2019 Dec;16(12):1041-1052. [doi: 10.1080/17434440.2019.1693890] [Medline: 31774330]
30. Shrivastava D, Jung S, Saadat M, Sirohi R, Crewson K. How to interpret the results of a sleep study. J Community Hosp Intern Med Perspect 2014;4(5):24983 [FREE Full text] [doi: 10.3402/jchimp.v4.24983] [Medline: 25432643]
31. Lehmann CU, Miller MR. Standardization and the practice of medicine. J Perinatol 2004 Mar 24;24(3):135-136. [doi: 10.1038/sj.jp.7211060] [Medline: 15044928]
32. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 2021 Feb 12;44(2):zsaa170 [FREE Full text] [doi: 10.1093/sleep/zsaa170] [Medline: 32882005]
33. Good Machine Learning Practice for Medical Device Development: Guiding Principles. US FDA. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles [accessed 2023-10-06]

Abbreviations

AHI: apnea-hypopnea index
CLC: Clionic Lifecare Clinic
CST: consumer sleep tracker
REM: rapid eye movement
SNUBH: Seoul National University Bundang Hospital
WASO: wake after sleep onset

Edited by L Buis; submitted 18.07.23; peer-reviewed by Z Liang, K Klier; comments to author 02.08.23; revised version received 08.08.23; accepted 20.09.23; published 02.11.23.

Please cite as:
Lee T, Cho Y, Cha KS, Jung J, Cho J, Kim H, Kim D, Hong J, Lee D, Keum M, Kushida CA, Yoon IY, Kim JW
Accuracy of 11 Wearable, Nearable, and Airable Consumer Sleep Trackers: Prospective Multicenter Validation Study
JMIR Mhealth Uhealth 2023;11:e50983
URL: https://mhealth.jmir.org/2023/1/e50983
doi:10.2196/50983
PMID:37917155

Original Paper

Assessing the Effect of Extreme Weather on Population Health Using Consumer-Grade Wearables in Rural Burkina Faso: Observational Panel Study

Mara Koch[1]; Ina Matzke[1]; Sophie Huhn[1]; Ali Sié[1,2], PhD; Valentin Boudo[2], MSc; Guillaume Compaoré[2], MSc; Martina Anna Maggioni[3,4], PhD, PD; Aditi Bunker[1,5], Dr sc hum; Till Bärnighausen[1,6,7], Prof Dr; Peter Dambach[1], PhD, PD; Sandra Barteit[1], Dr sc hum

[1]Heidelberg Institute of Global Health, Faculty of Medicine, University Hospital, Heidelberg University, Heidelberg, Germany

[2]Centre de Recherche en Santé, Nouna, Burkina Faso

[3]Charité - Universitätsmedizin Berlin, Institute of Physiology, Center for Space Medicine and Extreme Environments Berlin, Berlin, Germany

[4]Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy

[5]Center for Climate, Health, and the Global Environment, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[6]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[7]Africa Health Research Institute, KwaZulu-Natal, South Africa

Corresponding Author:
Sandra Barteit, Dr sc hum
Heidelberg Institute of Global Health
Faculty of Medicine, University Hospital
Heidelberg University
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 6221 5634030
Fax: 49 6221 565039
Email: barteit@uni-heidelberg.de

Abstract

Background: Extreme weather, including heat and extreme rainfall, is projected to increase owing to climate change, which can have adverse impacts on human health. In particular, rural populations in sub-Saharan Africa are at risk because of a high burden of climate-sensitive diseases and low adaptive capacities. However, there is a lack of data on the regions that are anticipated to be most exposed to climate change. Improved public health surveillance is essential for better decision-making and health prioritization and to identify risk groups and suitable adaptation measures. Digital technologies such as consumer-grade wearable devices (wearables) may generate objective measurements to guide data-driven decision-making.

Objective: The main objective of this observational study was to examine the impact of weather exposure on population health in rural Burkina Faso using wearables. Specifically, this study aimed to assess the relationship between individual daily activity (steps), sleep duration, and heart rate (HR), as estimated by wearables, and exposure to heat and heavy rainfall.

Methods: Overall, 143 participants from the Nouna health and demographic surveillance system in Burkina Faso wore the Withings Pulse HR wearable 24/7 for 11 months. We collected continuous weather data using 5 weather stations throughout the study region. The heat index and wet-bulb globe temperature (WBGT) were calculated as measures of heat. We used linear mixed-effects models to quantify the relationship between exposure to heat and rainfall and the wearable parameters. Participants kept activity journals and completed a questionnaire on their perception of and adaptation to heat and other weather exposure.

Results: Sleep duration decreased significantly (P<.001) with higher heat exposure, with approximately 15 minutes shorter sleep duration during heat stress nights with a heat index value of ≥25 °C. Many participants (55/137, 40.1%) reported that heat affected them the most at night. During the day, most participants (133/137, 97.1%) engaged in outdoor physical work such as farming, housework, or fetching water. During the rainy season, when WBGT was highest, daily activity was highest and increased when the daily maximum WBGT surpassed 30 °C during the rainiest month. In the hottest month, daily activity decreased per degree increase in WBGT for values >30 °C. Nighttime HR showed no significant correlation with heat exposure. Daytime HR data were insufficient for analysis. We found no negative health impact associated with heavy rainfall. With increasing rainfall, sleep duration increased, average nightly HR decreased, and activity decreased.

Conclusions: During the study period, participants were frequently exposed to heat and heavy rainfall. Heat was particularly associated with impaired sleep and daily activity. Essential tasks such as harvesting, fetching water, and caring for livestock expose this population to weather that likely has an adverse impact on their health. Further research is essential to guide interventions safeguarding vulnerable communities.

(JMIR Mhealth Uhealth 2023;11:e46980) doi:10.2196/46980

KEYWORDS

wearable; consumer-grade wearable; sleep; activity; heart rate; climate change; heat; rain; weather; sub-Saharan Africa; global health; public health; mobile phone

Introduction

Climate Change Effects on Human Health

There is growing scientific evidence that environmental conditions and extreme weather exposure associated with climate change are having negative effects on human health [1-3]. In particular, heat presents one of the most immediate health threats associated with climate change [2,4]. In combination with humidity, the health risks of heat can be exacerbated [5]. Previous studies have found that rising temperatures negatively affect sleep [6,7], which can cause, for example, decreased cognitive function [8], compromised immune function [9], and adverse cardiovascular outcomes [10]. Daily activity has also been found to be negatively affected by heat [11,12], which can cause reduced working capacity and productivity [13] in addition to adverse long-term effects on morbidity and mortality [14-16]. In rural populations in low-income countries, where people often rely on agriculture and livestock for their livelihood, higher exposure to extreme weather events may have adverse effects on nutrition and health. In addition, heart rate (HR) is another health parameter affected by heat. HR has been found to increase in hotter conditions [17,18], and increased HR is associated with numerous effects on the cardiovascular system [19].

Climate Change and Health in Sub-Saharan Africa

Sub-Saharan Africa, including Burkina Faso, is expected to be severely affected by climate change [20]. The average surface temperature across Africa is projected to increase at a higher rate than the global average [21]. Changes in rainfall patterns are already causing severe droughts and floods [22]. Despite the mounting scientific evidence, population-level health effects of climate change in Africa, especially in low-resource contexts in the sub-Saharan region, are still poorly understood because of a lack of objective measurements of health parameters in response to exposure to extreme weather events [2,23]. African populations are especially at risk because of a high burden of climate-sensitive diseases and low adaptive capacity [22,24].

Research Infrastructures in Low- and Middle-Income Countries

To inform public health actions and identify emerging health concerns and increased-risk groups, public health surveillance is crucial [25,26]. In general, there is a scarcity of continuous and spatially distributed data available in low- and middle-income countries, particularly in sub-Saharan Africa, to conduct research on the effects of climate change on population health. To that end, research infrastructures such as health and demographic surveillance systems (HDSSs), of which >50 have been implemented across Asia and Africa, are important ecosystems capable of surveilling geographically defined populations. HDSSs provide valid and reliable population-based data on population dynamics (birth, death, and in- and out-migration), particularly in areas with inadequate or nonexistent vital event registration and health information systems. The Nouna HDSS, located in northwestern Burkina Faso, has been collecting long-term data on the health and demographics of a population of >120,000 individuals since 1992 [27].

Previous research on climate change and health in the Nouna HDSS has shown the relationship between climate variations and nutritional outcomes in children aged <5 years [28] and the impact of varying climate and weather conditions on population mortality, with a high indication of excess burden of noncommunicable disease and mortality [29,30]. However, most HDSSs do not capture local-level weather parameters such as temperature and precipitation, which are particularly key indicators of exposure to extreme weather events [22]. They are also unable to produce more comprehensive heat measures such as the heat index (HI) or wet-bulb globe temperature (WBGT), which have been found to be better indexes for heat exposure than temperature alone [31,32]. HDSSs can also provide continuous real-world surveillance of an individual's health by obtaining objective measurements and highly resolved health data if new sensors such as consumer-grade wearable electronic devices (hereinafter referred to as *wearables*) are incorporated [33].

Consumer-Grade Wearables for Climate Change and Health Research

Consumer-grade wearables, including the Withings Pulse HR, were introduced for the first time in the Nouna HDSS and were found to be a feasible and acceptable method for continuous health surveillance of individuals, allowing for ecological momentary assessments [34]. A number of studies have used wearables in climate change and health research, but none have used wearables in low-income countries to assess the health effects of extreme weather [35]. To our knowledge, there is no population-wide effort to systematically monitor populations at an individual level to better quantify and characterize the effects of climate change on health. Data on the daily effects of climate change on people's lives and health are currently scarce and are crucial to tailor interventions and adaptation measures to vulnerable populations. Climate change adaptation is essential to protect vulnerable groups in low-resource environments from rising average temperatures and exposure to extreme weather.

The overarching objective of this observational study was to examine the impact of heat and heavy rainfall on the population's health in rural Burkina Faso using consumer-grade wearables. Specifically, we captured daily activity, sleep, and HR using a wearable in a sample of the Nouna HDSS population. Our primary objectives were to study the relationships between (1) daily activity and heat and heavy rain, (2) nighttime sleep duration and heat and heavy rain, and (3) HR and heat and heavy rain.

Our secondary research question focused on the stratification of these relationships according to month and different demographic subgroups, specifically for sex, age group, and BMI group.

Methods

Study Design

We conducted an observational panel study covering a population of 143 participants in the Nouna HDSS in northwestern Burkina Faso from August 2021 to June 2022. During the 11 months of data collection, study participants were equipped with wearables (Withings Pulse HR) that they wore continuously. Weather data were collected at 5 weather stations throughout the study area. This study is reported according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement: guidelines for reporting observational studies (Multimedia Appendix 1).

Study Setting and Population

The study was conducted in the Nouna HDSS in northwestern Burkina Faso, which is located approximately 40 km from the Mali border and 250 km from the capital, Ouagadougou. The region's tropical climate is defined by one rainy season, which typically lasts from June to September, and high temperatures throughout the year [36].

Individuals were eligible for study participation if they (1) were aged ≥16 years, (2) had no plans for long-term travel during the study period, and (3) consented to study participation.

We calculated a sample size of 150 participants based on an eligible population of 100,000 (total HDSS population aged >6 years), a confidence level of 95%, and an error margin of 8%. To ensure that each sex was equally represented in the study population, random sampling was stratified by sex (for details on the sampling, see the study protocol by Barteit et al [37]).

Ethics Approval

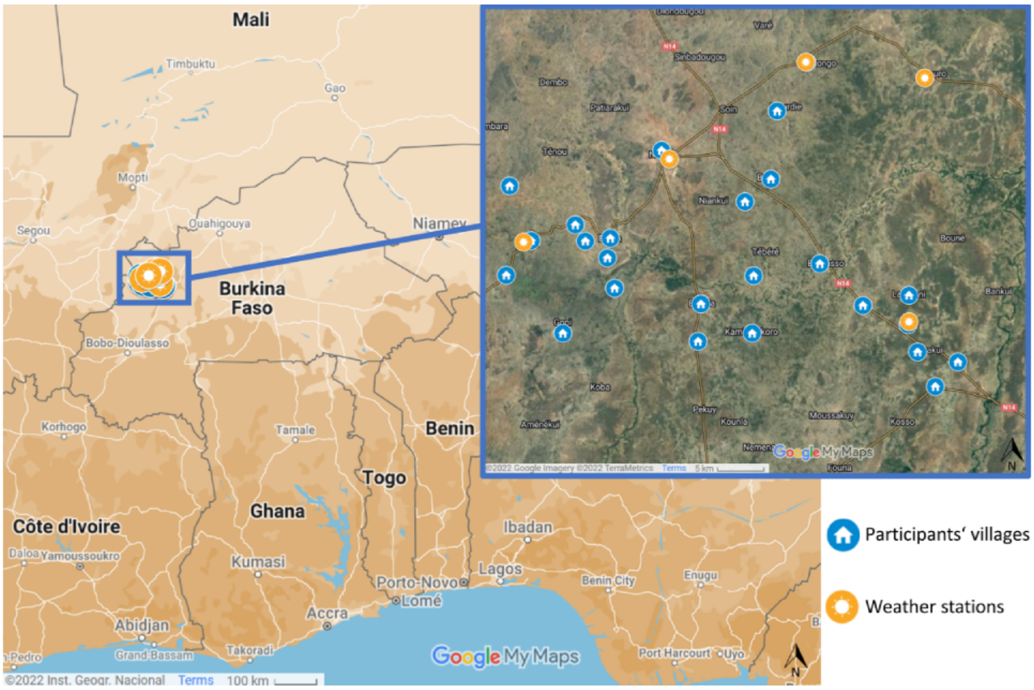
Ethics approval was granted by the Comité d'ethique pour la recherche en santé in Burkina Faso (approval date: March 13, 2020; 2020-3-041) and by the ethical committee of the Heidelberg University Hospital, Germany (approval date: May 6, 2019; S-294/2019).

Study Proceedings

Weather Data

In mid-2020, a total of 5 weather stations were set up to cover the spatial variability of different weather exposures across the study area. The nearest weather station was assigned to each of the 25 study villages (Figure 1). The distance was calculated as the shortest distance between 2 points on a WGS 84 ellipsoid using the distGeo function of the R package *geosphere* (version 1.5.18 [38]; R Foundation for Statistical Computing), which accounts for the ellipsoid shape of the Earth's surface.

Figure 1. Spatial distribution of weather stations across the Nouna health and demographic surveillance system area (image on the left) and the study area with the weather stations and the respective villages of participants involved in the study (image on the right). Map data: Google, Inst. Geogr. Nacional, created with Google MyMaps 2022.



Consumer-Grade Wearables

From August 2021 to June 2022, study participants were provided with consumer-grade wearable devices (Withings Pulse HR; Figure 2) and instructions on how to wear them correctly on their wrists (correct position and tightness). Participants wore the Withings Pulse HR for the entire study period. Considering that electricity is not available in all households in the study area, participants received a foldable solar panel with USB ports to charge the wearable and a smartphone to synchronize its data (for details, see the study protocol by Barteit et al [37]). Participants received weekly visits from a fieldworker. At each visit, the fieldworker checked the functionality of the wearable and smartphone, charged all devices with a portable power bank, and synchronized the data to the server for remote data web access. Wearables that were damaged during the study were replaced.

Figure 2. Withings Pulse HR consumer-grade wearable that participants wore during the study.

Activity Journals and Heat Questionnaires

The 21 most prevalent local activities were listed in a structured activity journal that study participants completed once a month according to their activities after getting up, in the morning, at noon, in the afternoon, in the evening, and at night (Multimedia Appendix 2). At the end of the study period, a questionnaire was administered evaluating the perception of heat with multiple-choice questions regarding whether, when, and how heat affected participants and their adaptation strategies (Multimedia Appendix 2).

Technical Measurement Details

Weather Stations

The weather stations were equipped with ADCON sensors and measured several variables, including air temperature, relative humidity, precipitation, wind speed and direction, and global radiation, at 15-minute intervals. The data were uploaded using the advantage Pro software (version 6.8; ADCON telemetry, OTT Hydromet GmbH).

Consumer-Grade Wearables

Consumer-grade wearables are waterproof and feature a 3-axis accelerometer that continuously estimates steps (identified by amplitude and periodic pattern) and sleep parameters (duration, time spent awake, sleep onset and offset, and sleep depth) through automatic data postprocessing using a manufacturer's algorithm. Steps have been frequently used in research to quantify daily activity [16]. Sleep duration specifies the total recorded time that a study participant spent asleep on a given night. Sleep interruptions were defined as wake time after sleep onset, whereas sleep onset was defined as the time of falling asleep and sleep offset was defined as the time of waking up. In addition, the device's photoplethysmography sensor estimates the pulse rate every 10 minutes or every 90 seconds in activity mode. For the purpose of this study, we deemed the pulse rate and HR to be equivalent. One charging cycle can last up to 21 days, and the internal storage can hold up to 5 days of recorded data. The data collected from the wearables in this study were wirelessly transmitted via Bluetooth to an app on the study participants' smartphone and, when an internet connection was available, uploaded from the mobile device to the Withings server (*health mate*).

Data Processing

Demographic Data

Data on weight, height, and date of birth and death were available for each study participant. Participants were categorized as young adults (aged <25 years), middle-aged adults (aged ≥25 and <65 years), and older adults (aged ≥65 years) based on their age at the beginning of the study period. BMI was calculated as weight (kg)/(height (m) × height (m) and grouped into 3 categories: underweight (<18.5 kg/m^2), healthy weight (≥18.5 and <25 kg/m^2), and overweight (≥25 kg/m^2).

Weather Data

We calculated WBGT estimates according to the formula provided by Carter et al [39] and used the heat.index function of the R package *weathermetrics* (version 1.2.2 [40]) to calculate the HI according to the US National Weather Service complex algorithm from temperature and relative humidity [40].

Weather extreme indexes were calculated for each day of the study period so that days with and without extreme weather could be compared. As weather extremes are often characterized relative to historical data (90th percentile of a 30-year reference period) [41] and we only had approximately 2 years of weather data from the 5 weather stations that were installed in the Nouna HDSS in 2020, we used the following climate extreme indexes developed by Climpact and recommended by the World Meteorological Organization Expert Team on Sector-specific Climate Indices: (1) number of days with heavy precipitation (count of days in which the daily precipitation was ≥20 mm), (2) number of tropical nights (count of nights in which the minimum temperature was >20 °C), and (3) number of hot days (count of days in which the daily maximum temperature was ≥35 °C).

In addition, we calculated weather extreme indexes based on WBGT and HI (also called apparent temperature) to provide a more accurate assessment of heat exposure [31,32]. We used WBGT as a heat parameter during the daytime and HI during the nighttime. The WBGT calculation includes global radiation and wind speed, and as the data are measured using weather stations outdoors, we found HI to be more applicable than WBGT for the assessment of heat exposure during the night when participants were mostly indoors. On the basis of findings of current climate and health research using WBGT [7,42,43] and the National Weather Service cutoffs for HI [44], we used the following cutoffs: (1) the number of heat stress days was defined as the count of days in which the daily maximum WBGT was ≥30 °C, and (2) the number of heat stress nights was defined as the count of nights in which the minimum HI was ≥25 °C.

For correlating weather data with wearable data, we divided the weather data into daytime and nighttime data. On the basis of median sleep onset and offset times (median sleep onset 10:09 PM; median sleep offset 5:55 AM), cutoffs for nighttime weather were determined, resulting in measurements between 10 PM and 6 AM the following day. Daytime weather was defined as measurements between 6 AM and 10 PM. For comparison between seasons, weather data were also categorized into rainy (June-September), cool dry (October-January), or hot dry (February-May) season according to the month.

Consumer-Grade Wearables

Data Completeness

To increase the representativeness of the measurements and address missing wearable data, we considered participants who had data coverage of 25% of the measurement days as complete cases, similar to the study by Minor et al [6]. Using a 50% criterion yielded comparable results (Multimedia Appendix 3); however, HR data had too little completeness for a higher cutoff. We did not impute data as data imputation was found not to be necessary for unbiased results when only the outcome variable was affected by missing data [45]. In addition, studies have found that multiple imputation did not increase precision when using linear mixed models for data analysis [46]. Furthermore, Jakobsen et al [47] did not recommend using data imputation for a high percentage of missing data.

Daily Activity

In line with previous studies, a filter of at least 10 hours of wear time per participant per day was used to exclude days with insufficient wear time [12,48,49]. We defined nonwear time as more than 1 hour between measurements. Values of 0 steps were excluded as, according to the manufacturer, nonwear time and a measurement of 0 steps cannot be distinguished. Duplicate measurements were removed. We aggregated the steps into 15-minute intervals to be consistent with weather data intervals and also as daily step counts as a measure of daily activity.

Sleep Data

We limited sleep length based on the onset and offset times to eliminate incorrect values. Adapted from the study by Minor et al [6], we defined nighttime sleep with a sleep onset of ≥5 PM and a sleep offset of ≤1 PM as limits, with a 2-hour adjustment to account for earlier bedtimes in this study population. If a participant had multiple sleep observations in a single night, they were summarized into one sleep observation by adding the sleep duration values and the duration between the first offset and second onset to wake time after sleep onset. If multiple sleep observations overlapped for one participant, they were excluded as error measurements. In accordance with the manufacturer's declaration that sleep detection measurements with a time difference of <3 hours between sleep onset and sleep offset are invalid measurements, we excluded those measurements.

HR Measurement

Duplicate HR measurements were removed, and values greater than the age-predicted maximal HR according to the equation by Tanaka et al [50] (208 – 0.7 × age) were excluded. HR measurements were rounded and aggregated (mean, minimum, and maximum) into 15-minute intervals to be consistent with weather data intervals. HR measurements were divided into daytime and nighttime HRs using the median sleep onset and offset as nighttime definitions (10 PM-6 AM). Measurements for 1 day were included when at least 2 hours of measurements were available; nighttime measurements for 1 night were included when at least 1 hour was covered by at least one measurement every 15 minutes. We chose the threshold of 15 minutes in accordance with previous research [34].

Statistical Analysis and Data Modeling

Descriptive Data Analysis

All data analyses were conducted using R (RStudio version 2022.07.02+576; Posit, PBC).

First, the weather and wearable data were descriptively analyzed (covering minimum, maximum, mean, and SD values). Activities reported in the activity diaries were thematically summarized into 9 categories and quantified relative to the total number of activity journal responses. Similarly, the heat questionnaire was analyzed, and responses were summarized based on the frequency of replies.

Mixed-Effects Models

Associations Between Weather Exposure and Daily Activity (Steps), Sleep Duration, and Nighttime HR

We conducted a linear mixed-effects analysis of the relationship between weather exposure and daily activity (steps), sleep duration, and nighttime HR using the lmer function of the *lme4* R package (version 1.1.34 [51]). We used the following formula:

$$Y_i = b_0 + b_1 Heat_i + b_2 Precipitation_i + b_z Z_i + \varepsilon_i \quad \textbf{(1)}$$

In this linear mixed-effects model, *i* indicates each study participant. The dependent variable Y_i sequentially represents sleep duration (in hours), daily activity (in steps), and average nighttime HR (in beats per minute [bpm]) of individual *i*. The independent variable of interest "Heat" represents the minimum nighttime HI (HI_{min}) for sleep and nighttime HR and maximum daytime WBGT ($WBGT_{max}$) for daily activity (steps). The independent variable "Precipitation" represents the total daily rainfall. Furthermore, we controlled for month, weekend or weekday, age group, sex, and BMI group by adding these variables stepwise as independent variables, represented as *Z*. Weekend was added as a possible confounder as we expected the participants' activities to vary between weekends and weekdays, also affecting nighttime sleep. Similarly, agricultural activity most likely varies between months. Furthermore, by adding month as an independent variable, we indirectly also accounted for differences in daylight hours. The demographic confounders sex, age, and BMI group have frequently been shown to be associated with sleep, HR, and daily activity and, therefore, were included in our models.

By adding study participants as a random term, the linear mixed-effects model accounted for the nonindependence of the repeated measures. In addition, linear mixed models were chosen as they can estimate parameters from existing data to handle missing data and independent variables can be on a continuous scale with differing times between measurement points. We used maximum likelihood for the estimation of the model parameters. In total, 3 separate models were constructed for sleep duration, daily activity, and average nighttime HR. Covariates were added through hierarchical model building, with removal based on the chi-square likelihood test (lower log-likelihood values were removed) and a significance level of 5%. We used leave-one-subject-out (LOSO) cross-validation to assess model performance. We compared the performance of the resulting models on the respective validation sets using R^2 and root mean square error as well as the average values for these 2 parameters.

If visual inspection of the residual plots revealed deviation from linearity, we introduced a quadratic term for the predictor variable using the following equation:

$$Y_i = b_0 + b_1 Heat_i + b_2 Heat^2{}_i + b_3 Precipitation_i + b_z Z_i + \varepsilon_i \quad \textbf{(2)}$$

We created an additional model for each of the 3 wearable parameters using temperature and daily rainfall as predictors to account for heat metrics other than WBGT and HI. Model selection was based on the Akaike information criterion, with a lower value indicating an improved model fit.

Differences in Subgroup Sensitivity to Heat

We evaluated the effect of heat exposure on daily activity, sleep duration, and nighttime HR between the different subgroups. This way, it was possible to assess the different heat sensitivity of each subgroup. Therefore, we included an interaction term between continuous measures for heat and a categorical variable *X*, which successively represented age group, BMI group, sex, or month. If adding the interaction term significantly improved the model, we ran the model for each subset separately to compare the effects of heat exposure. We interpreted the results for each subgroup relative to its corresponding reference category (middle-aged adults, healthy weight participants, male individuals, and coldest month).

$$Y_i = b_0 + b_1 Heat_i \times X + b_2 Precipitation_i + b_z Z_i + \varepsilon_i \quad \textbf{(3)}$$

Associations Between Weather Extremes and Daily Activity, Sleep Duration, and Nighttime HR

We built linear mixed-effects models using binary variables for heat (heat stress day or night) and heavy rainfall as independent variables in place of the continuous variables to evaluate the association between weather extremes and the wearable parameters. The model formula changed as follows:

$$Y_i = b_0 + b_1 Heatstress_i + b_2 HeavyPrecipitation_i + b_z Z_i + \varepsilon_i \quad \textbf{(4)}$$

Results

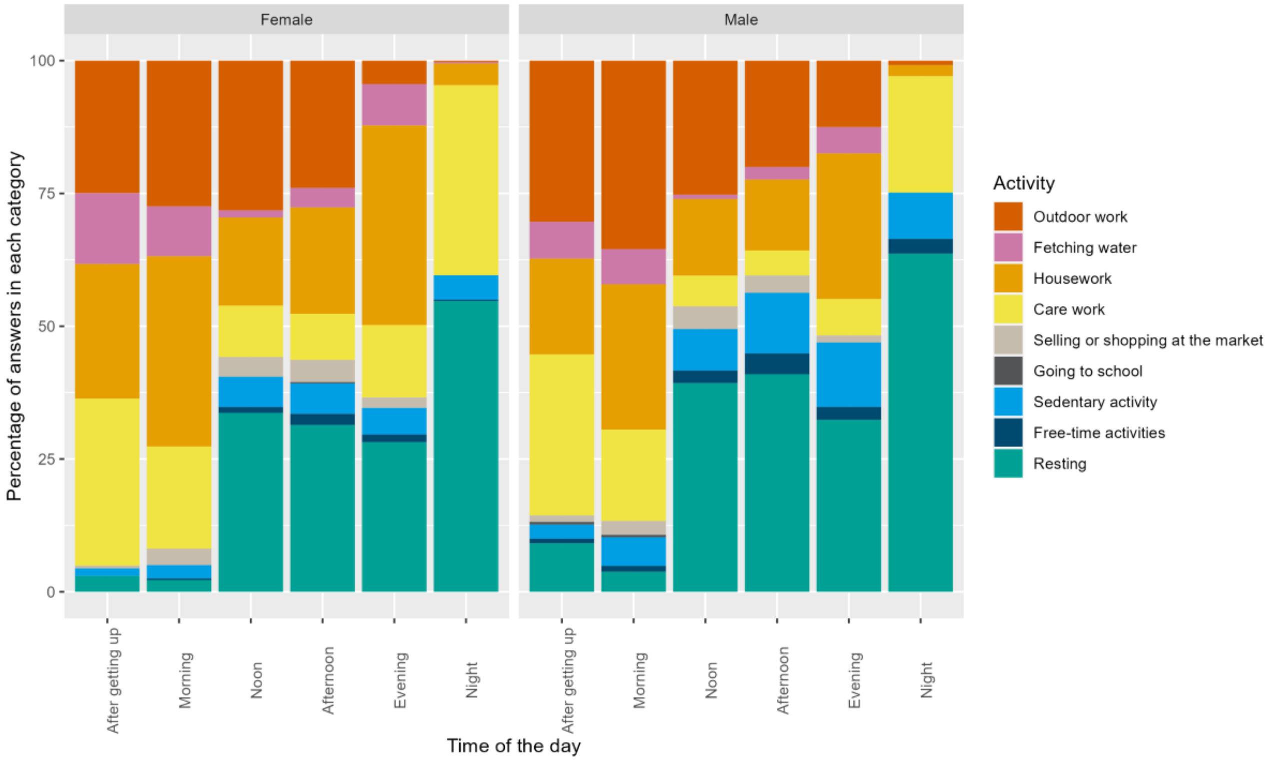
Participant Characteristics

We originally recruited 152 participants. During the study, of these 152 participants, 7 (4.6%) withdrew their consent, 1 (0.7%) died, and 1 (0.7%) migrated out of the study region. Therefore, the results report on a total of 143 study participants.

Approximately half (71/143, 49.7%) of the study population was female. The age of the study participants ranged from 16 to 79 years, with an average age of 43 (SD 13) years. Most (128/143, 89.5%) were middle-aged (aged 25-65 years). The average BMI of the participants was 22.3 (SD 2.7) kg/m^2, ranging from 16.6 to 32.42 kg/m^2, with most (112/143, 78.3%) falling into the healthy weight category (18.5≤BMI<25).

When looking at the distribution of self-reported activities throughout the day (Figure 3), we observed that after waking up and in the morning were the most active times of the day, when participants engaged in outdoor work, housework, and errands such as fetching water and care work. The rest of the day was often spent resting. Male participants reported more outdoor work, whereas female participants reported more care work and housework, especially in the evening and at night.

Figure 3. Self-reported daily activities of study participants stratified by sex and time of day (after getting up, in the morning, at noon, in the afternoon, in the evening, and at night).



Weather

The average distance between the villages and the closest weather station was 7.6 km (range 0.1-19.0, SD 14.4 km). For the entire study region (average of 4 weather stations), the temperature range was 10.9 to 44.0 °C with a mean of 28.5 °C, and WBGT ranged from 7.3 to 32.8 °C with a mean of 23.3 °C. The maximum daily precipitation measured was 101.5 mm with an average of 1.9 mm per day and an average cumulative sum over the entire study period of 630.5 mm. During the 1-year study period, there were, on average, 121 (range 110-131) days classified as heat stress days with maximum WBGT values of >30 °C and an average of 117 (range 102-135) nights classified as heat stress nights with minimum HI values of >25 °C. Furthermore, study participants were exposed to heavy rainfall, with ≥20 mm on average on 8 (range 6-11) study days and single measurements exceeding 50 mm of precipitation per day on 7 study days. Table 1 provides a detailed overview of weather data stratified by season.

Table 1. Weather parameters and extreme weather events for the study duration of 11 months by season (average of all 4 closest weather stations).

Weather exposure	Rainy season (June-September)	Cool dry season (October-January)	Hot dry season (February-May)
Mean air temperature (°C; SD)	27.6 (3.8)	26.8 (6.4)	31.1 (6.0)
Mean WBGT[a] estimate (°C; SD)	25.9 (2.5)	20.5 (5.6)	23.4 (4.9)
Number of days with maximum daily air temperature of ≥35 °C	24	73	104
Number of days with maximum daily WBGT estimate of ≥30 °C	46	33	42
Number of days with rainfall of ≥20 mm	8	0	1
Number of nights[b] with minimum air temperature of ≥20 °C	91	48	98
Number of nights[b] with minimum HI[c] of ≥25 °C	40	15	63

[a]WBGT: wet-bulb globe temperature.

[b]Nights were defined as 10 PM to 6 AM (median sleep onset–median sleep offset).

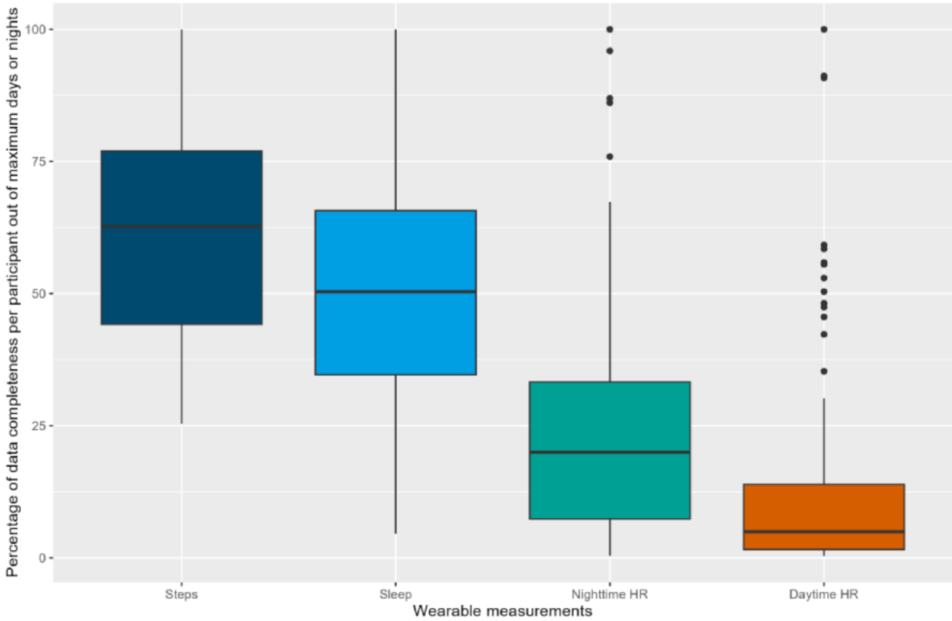
[c]HI: heat index.

Wearable Data

Data Completeness

The data completeness of wearable measurements of daily activity, sleep, and HR is shown in Figure 4. The average data completeness per participant was 166 (SD 66) days for daily activity; 153 (SD 69) nights for sleep; and 58 (SD 51) nights and 34 (SD 51) days for nighttime and daytime HR, respectively.

Figure 4. Box plot of data completeness per wearable parameter in percentage of the respective maximum number of days or nights with data for 1 participant. HR: heart rate.



Daily Activity Measurements

Overview

We had to remove 26.64% (2,438,464/9,152,256) of the data points, with the final data set including a total of 6,713,792 step measurements and 22,853 days from 90.9% (130/143) of the participants. The mean daily activity (in steps) was 9602 (SD 6499; minimum: 194; maximum: 47,522). Figure 5 shows that the average number of steps taken each day changed based on the season. In the rainy season, the average number of steps was 11,328 (SD 8049); in the cool dry season, it was 9532 (SD 6140); and, in the hot dry season, it was 8700 (SD 5681).

The distribution of steps over the course of the day also varied from season to season (Figure 6). There were activity peaks in the morning and late afternoon during all seasons. In the rainy season, the peak occurred at 11:15 AM with a mean of 310 (SD 367) steps per 15-minute interval, whereas in the hot dry season, it occurred at 9 AM with a mean of 231 (SD 269) steps per 15-minute interval.

Figure 5. Average daily steps (black dots) with IQR (gray area) and maximum daily wet-bulb globe temperature (WBGT; °C; orange line) for the duration of the study. The background color indicates the 3 seasons (cool dry, hot dry, and rainy).

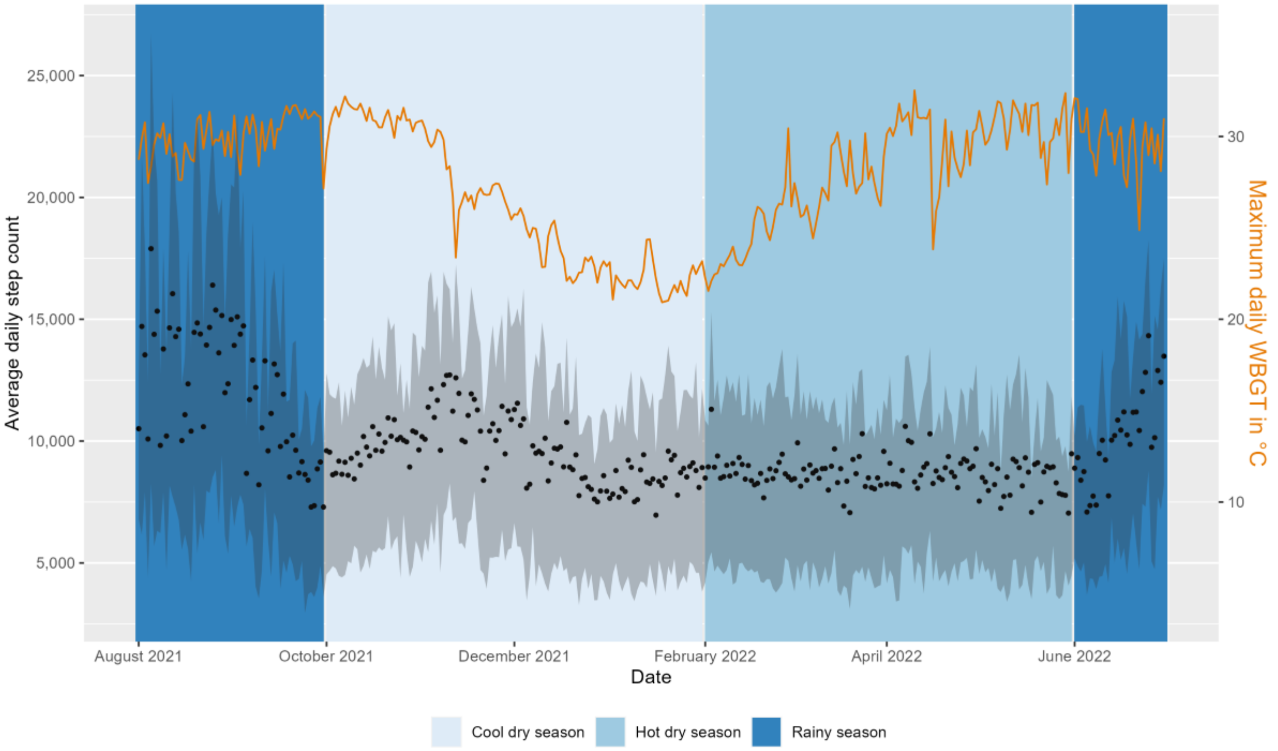
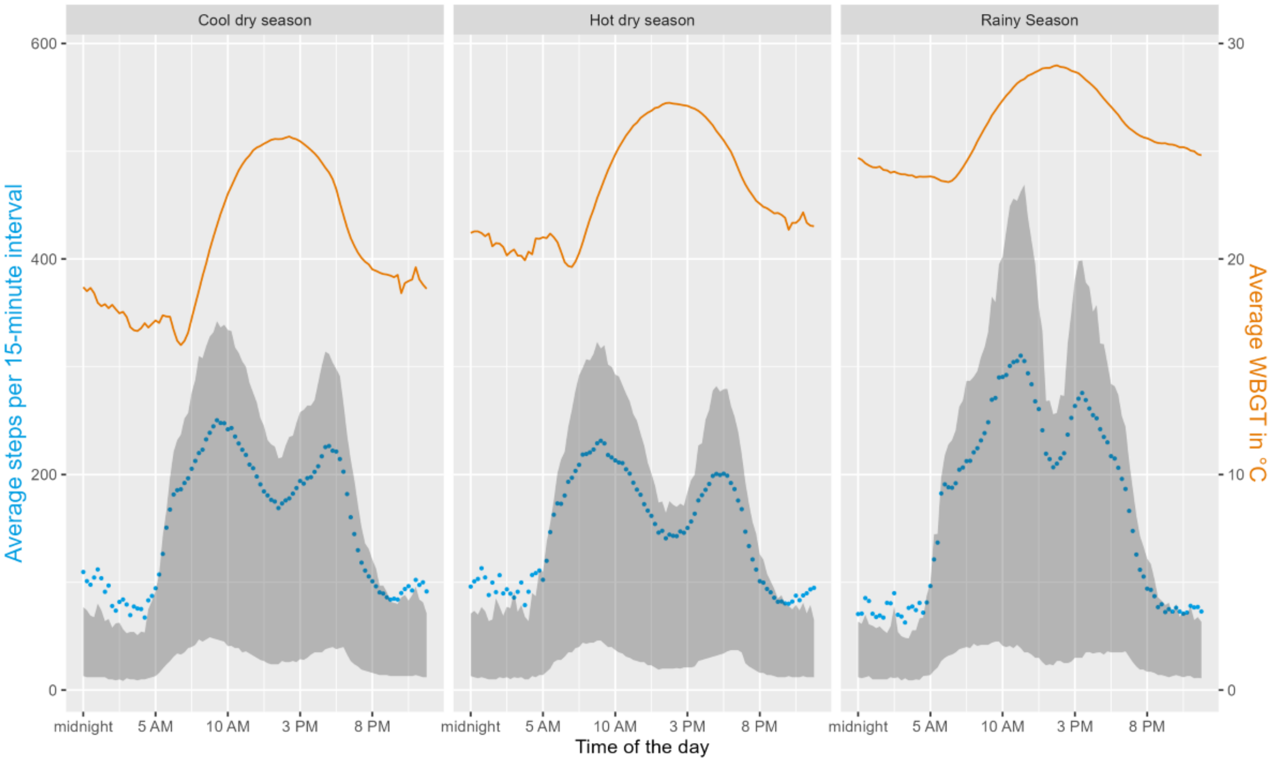


Figure 6. The distribution of the average number of steps per 15-minute interval (blue dots) with the IQR (gray area) over the course of a day and wet-bulb globe temperature (WBGT; measured every 15 minutes; orange line) for the 3 seasons in Burkina Faso (cool dry, hot dry, and rainy).



Weather Exposure and Daily Activity Measurements (Steps)

We fitted a linear mixed model with participants as a random effect. Daily steps showed significant variance in intercepts across participants (χ^2_1=9455.8; *P*<.001). Maximum daily WBGT ($WBGT_{max}$) and total daily precipitation were included as predictors. $WBGT_{max}$ as a fixed effect was highly significant (2-tailed *t* test, $t_{22,849}$=12.73; *P*<.001). However, as the residual plot of daily activity and $WBGT_{max}$ showed nonlinearity, we added a quadratic term for $WBGT_{max}$ ($WBGT_{max}^2$), which improved the model fit. Herein, we report the combined effects of $WBGT_{max}$ and $WBGT_{max}^2$. For a $WBGT_{max}$ of 20 °C, we found an increase of 146 steps for every additional degree of WBGT (°C), whereas for a $WBGT_{max}$ of 30 °C, we observed a decrease of 83 steps for every additional degree of WBGT (°C). The main effect of daily precipitation was also significant ($t_{22,725}$=−6.57; P<.001), and we found that daily activity decreased by an average of 39 steps for every 1-mm increase in total daily rainfall (95% CI −44 to −23). The final model included adjustments for age group, month, and weekend or weekday. We also developed a model using the maximum daily temperature in °C as a predictor instead of $WBGT_{max}$, which was shown to be a less accurate model. We cross-validated the model using the LOSO approach and calculated R^2 values of 0.45 for some participants as test data, with an average of 0.12.

Different Heat Sensitivity by Age, BMI, Sex, and Month

To assess the correlation between different subgroups and the effect of $WBGT_{max}$ on daily activity, we added an interaction term for age group, sex, or month and $WBGT_{max}$ and $WBGT_{max}^2$. If the addition of the interaction term improved the fit of the model, we ran the model independently for each subgroup to compare the effect of $WBGT_{max}$ on daily activity. Table 2 provides a summary of the per-degree effects of $WBGT_{max}$ on daily activity for each subgroup. For different sex and age groups, we did not find significant differences in heat sensitivity (P>.05). During the hottest month (April) and the coolest month (January), the effect of increasing $WBGT_{max}$ on daily activity was not significant. During August, the rainiest month, the effect of increasing $WBGT_{max}$ on daily activity was opposite to that during January and April, with a 5126-step decrease per degree increase in $WBGT_{max}$ at 20 °C (compared with an increase in January and April) and an increase of 678 steps (compared with a decrease in January and April) at 30 °C.

Table 2. Combined effect of 1-degree increases in maximum wet-bulb globe temperature (WBGTmax) and squared WBGTmax (WBGTmax2) on daily activity (in steps) for different age groups and sex and during different seasons.

Subgroup	Combined effect of $WBGT_{max}$ and $WBGT_{max}^2$ at 20 °C on steps	Combined effect of $WBGT_{max}$ and $WBGT_{max}^2$ at 30 °C on steps	Estimate for $WBGT_{max}$	P value for $WBGT_{max}$	Estimate for $WBGT_{max}^2$	P value for $WBGT_{max}^2$
All participants (n=130)	+146	−83	606	.03	−12	.02
Older adults (aged ≥65 years; n=9)	+118	−60	475	.38	−9	.37
Middle-aged adults (aged 25-65 years; n=114)	+151	−93	638	.03	−12	.02
Young adults (aged <25 years; n=7)	+207	−10	641	.94	−11	.61
Male individuals (n=65)	+142	−88	601	.13	−11	.11
Female individuals (n=65)	+138	−67	547	.14	−10	.13
Rainiest month (August; n=79)	−5126	678	−16,734	.03	290	.02
Coolest month (January; n=122)	+370.34	−1047.06	3205.14	.35	−71	.36
Hottest month (April; n=99)	+304	−18	965	.43	−16	.45

Weather Extremes and Daily Activity

We further assessed the differences in daily activity between exposure to weather extremes (heat stress days and heavy rainfall) and nonexposure. We created a linear mixed model with the weather extreme exposure as binary fixed effects instead of $WBGT_{max}$ and total daily rainfall. The final model was adjusted for age group, month, and weekend or weekday. We found no statistically significant difference in daily activity on heat stress days with $WBGT_{max}$ ≥30 °C compared with days with $WBGT_{max}$ <30 °C ($t_{22,725}$=−0.64; P=.52). On days with heavy rainfall compared with days without heavy rainfall ($t_{22,724}$=−3.25; P<.001), daily activity was significantly lower, with an estimate of −855 steps. The results for all 3 models can be found in Multimedia Appendix 3.

Sleep Measurements

Overview

After removing 6.89% (1523/22,095) of the sleep observations during data processing, the final data set comprised 20,572 nights covering the study duration of 334 nights from 83.2% (119/143) of the study participants. The average sleep duration was 6 hours, 49 minutes (SD 1 min, 48 s) per night, the time of sleep onset was 10:23 PM (SD 1 h, 47 min), and the time of sleep offset was 05:44 AM (SD 1 h, 28 min; Figure 7). On average, study participants were awake for 36 (SD 30) minutes after sleep onset. Nearly half (10,012/20,505, 48.83%) of the participant nights showed insufficient sleep according to the age group definitions of sufficient sleep duration by Hirshkowitz et al [52]. This definition states that sleep duration of <8 hours for those aged <18 years and <7 hours for those aged >18 years is insufficient. The average sleep duration was shorter throughout the study when the minimum nighttime HI was higher (Figure 8).

Figure 7. Histogram of the time that study participants fell asleep (orange colored) at night and woke up (blue colored) in the morning.

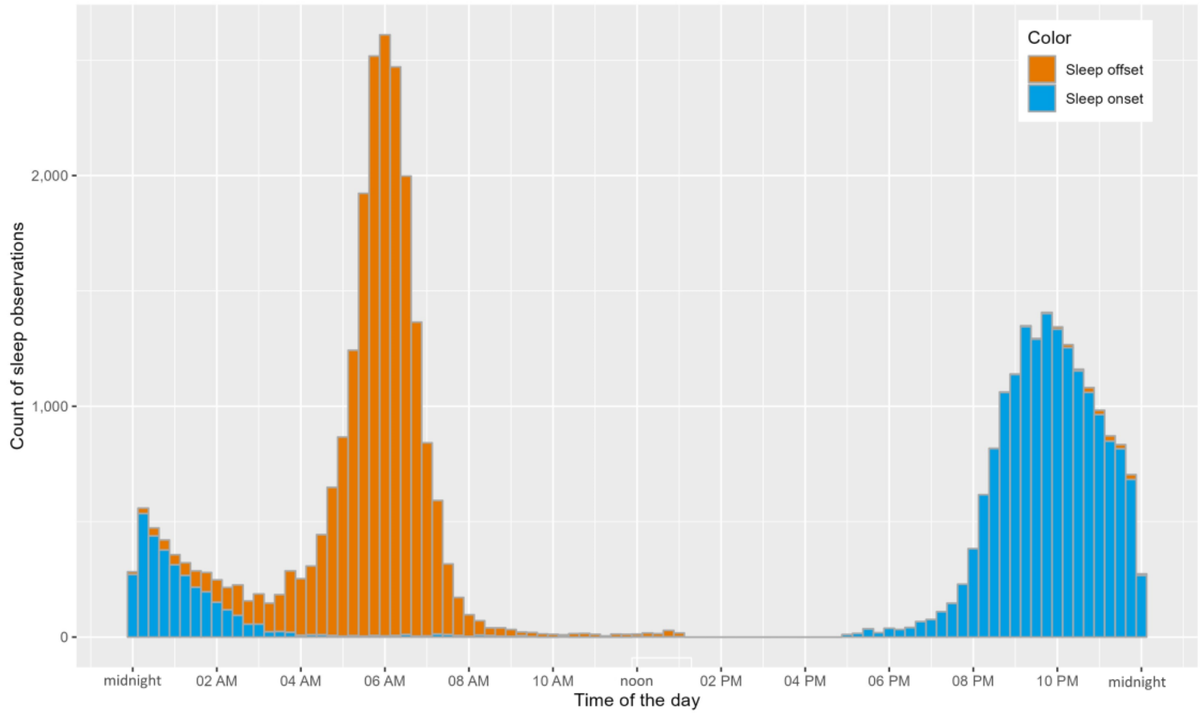
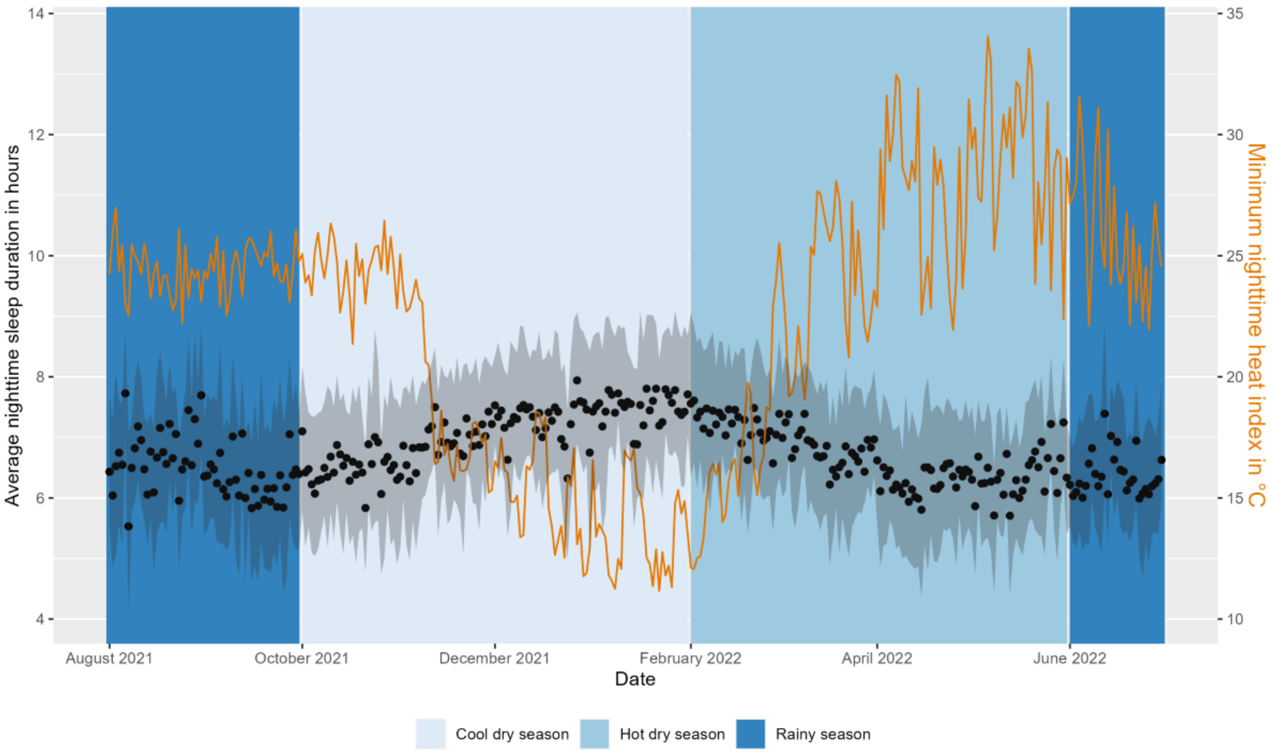


Figure 8. Average nighttime sleep duration (hours; black dots) with the IQR (gray area) and minimum heat index (°C; orange line; average of all 4 weather stations) throughout the study period. The background color indicates the 3 seasons (cool dry, hot dry, and rainy).



Weather Exposure and Sleep Measurements (Duration)

To provide a detailed analysis of the relationship between sleep duration and nighttime heat, we fitted a linear mixed-effects model with study participants as a random term and the minimum nighttime HI (HI_{min}) as a predictor of sleep duration. Sleep duration showed significant variance in intercepts across participants (χ^2_1=3596.9; *P*<.001). HI_{min} as a fixed effect was highly significant ($t_{20,520}$=−9.60; *P*<.001). For every 1-degree increase in HI_{min}, sleep duration decreased on average by 0.04 hours (2.4 minutes; 95% CI −0.074 to −0.065) in the fully adjusted model. We also included total daily rainfall as a predictor and found a significant ($t_{20,457}$=6.27; *P*<.001) but small main effect of an increase of 0.01 hours (36 seconds; 95% CI

0.007-0.013) in sleep duration per mm increase in rainfall. The final model included adjustments for age group, BMI group, month, and weekend or weekday. Sex was considered as a covariate but was removed after failing to reach statistical significance (P<.05). We cross-validated the model using the LOSO approach and calculated R^2 values of 0.34 for some participants as test data, with an average of 0.10. In addition, we developed a model using the minimum nighttime temperature in °C as a predictor of HI_{min} instead, which we found to be less accurate.

Different Heat Sensitivity by Age, BMI, Sex, and Month

By sequentially adding an interaction term for age group, BMI group, sex, and month with HI_{min}, we assessed whether these subgroups exhibited varying levels of heat impact regarding sleep duration. We found no statistically significant differences between age groups, BMI groups, or sex. In the hottest month (April), the impact of heat was lowest, with an estimate for the effect of HI_{min} of −0.02 (SE 0.01) on sleep duration. In the coolest and driest month (January), the estimate for the effect of HI_{min} was −0.04 (SE 0.02), and in the rainiest month (August), the impact of heat was highest, with an estimate of −0.125 (SE 0.04).

Weather Extremes and Sleep Duration

We further assessed the difference in sleep duration between exposure to nights with heat stress and heavy rainfall and nonexposure. We created a linear mixed model with the binary weather extreme factors as fixed effects. Adding heavy rainfall as a fixed effect did not improve the model fit. The final model was adjusted for month and weekend or weekday. We found significantly ($t_{20,470}$=−7.81; P<.001) shorter sleep duration on heat stress nights with HI_{min} ≥25 °C compared with nights with HI_{min} <25 °C by an estimated 0.24 hours (15 minutes; 95% CI −0.31 to −0.18). The results for all 3 models of sleep duration can be found in Multimedia Appendix 3.

HR Measurements

Overview

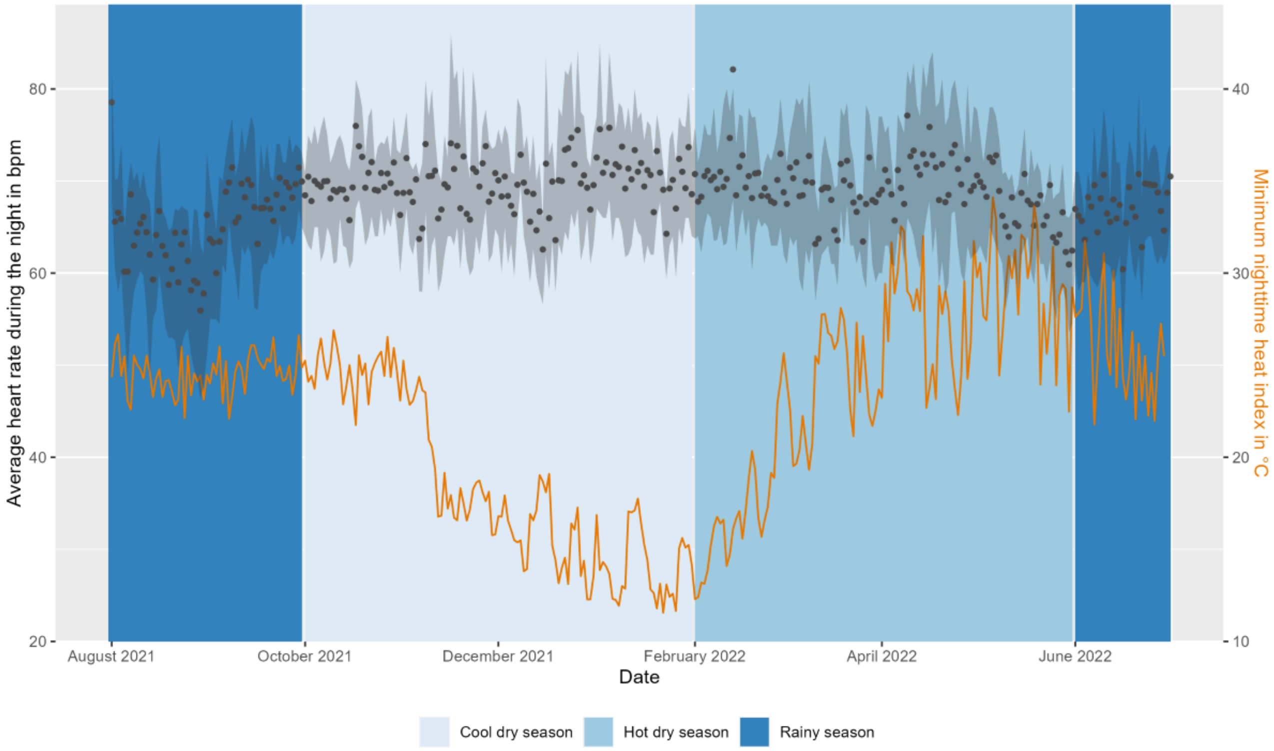
The data-cleaning process removed 88.55% (488,906/552,099) of the HR measurements. We split the measurements into daytime and nighttime HR. The final data set of daytime HR comprised a total of 3062 fifteen-minute intervals collected across 249 participant days from 11.9% (17/143) of the study participants spanning 190 study days. Nighttime HR data included 31,423 fifteen-minute intervals from 32.2% (46/143) of the participants collected across 3025 participant days.

During the day, the average HR was 90 (SD 23) bpm; during the night, the average HR was 69 (SD 13) bpm. We further only considered nighttime HR as an outcome variable as daytime measures had low data completeness for HR.

During the night, the average HR and HI decreased in parallel until the average HR increased again between 5 AM and 6 AM (Figure 9). This pattern was observed in all 3 seasons, although average HR values were lower during the rainy season, whereas HI values were similar between the hot, dry, and rainy seasons (Figure 10).

Figure 9. Average heart rate in beats per minute (bpm; blue dots) with the IQR (gray area) per 15-minute interval and average heat index (orange line) in °C during the night by season.

Figure 10. Average nighttime heart rate (beats per minute [bpm]; black dots) with the IQR (gray area) and minimum heat index (°C; orange line; average of all 4 weather stations) over the duration of the study. The background color indicates the 3 seasons (cool dry, hot dry, and rainy).



Weather Exposure and HR Measurements

Using a mixed-effects model, we explored the relationship between minimum nighttime HI values (HI_{min}) and average nighttime HR. Average nighttime HR showed significant variance in intercepts across participants (χ^2_1=1030.8, *P*<.001). As HI_{min} and nighttime HR did not show a linear relationship in the residual plots, we added a quadratic term for HI_{min} (HI_{min}^2) to improve the model fit. After adding HI_{min}, HI_{min}^2, and total daily precipitation as predictors, we found that neither HI_{min} (t_{2998}=–0.31; *P*=.75) nor HI_{min}^2 (t_{2998}=0.46; *P*=.65) as fixed effects were significant. However, the main effect of daily precipitation was significant (t_{2985}=–1.95; *P*=.009), and we found that, for every 1-mm increase in total daily rainfall, the average nighttime HR decreased by 0.04 bpm (95% CI −0.08 to −0.00). The final model was adjusted for sex, age group, BMI group, and month. Weekends failed to reach statistical significance (*P*<.05) as a covariate. We cross-validated the model using the LOSO approach and calculated R^2 values of up to 0.65, with an average of 0.14. In addition, we created a model using the minimum nighttime temperature in °C as a predictor instead of HI_{min}. When comparing both models, we found no better model fit for the minimum nighttime temperature model.

Different Heat Sensitivity by Age, BMI, Sex, and Month

We sequentially added an interaction term for age group, BMI group, sex, or month and HI_{min}. Interactions for sex, BMI group, and month did not improve the model fit. HR in relation to heat exposure did significantly differ between age groups. As there was only HR data from one young adult (aged <25 years), we only compared middle-aged and older adults. No significant association between HI_{min} and nighttime HR was found either for middle-aged adults (t_{2642}=–1.41; *P*=.16) or for older adults (t_{147}=−0.34; *P*=.73).

Weather Extremes and HR

To further compare nighttime HR between extreme and nonextreme weather exposures, we created a linear mixed model with the binary predicting factors heat stress night (HI_{min} ≥25 °C) or heavy rainfall (total daily rainfall of ≥20 mm). The models were adjusted for age group, sex, BMI group, and month. Heat stress nights did not have a significant effect on nighttime HR (t_{2993}=–0.69; *P*=.49). On nights with heavy rainfall, nighttime HR was estimated to be lower by 2 bpm (95% CI −3.81 to −0.21; t_{2987}=–2.19; *P*=.03) compared with nights with precipitation of <20 mm. The full results of all 3 models can be found in Multimedia Appendix 3.

Subjective Heat Perception and Adaptation

In response to our questionnaire (with a total response rate of 137/143, 95.8%), almost half (67/137, 48.9%) of the participants reported that heat had an impact on their daily lives. Of the study participants who reported an impact, most reported an impact primarily at night (55/67, 82%), in the afternoon (33/67, 49%), at noon (18/67, 27%), and during outdoor work (15/67, 22%) in the form of poorer sleep (44/67, 66%), sweating (44/67, 66%), exhaustion (27/67, 40%), and fatigue (14/67, 21%).

Most participants (89/137, 65%) stated that they worked outdoors, and some (44/137, 32.1%) worked both indoors and outdoors, whereas only a few (4/137, 2.9%) worked indoors only. When asked about their adaptative measures against heat when they were indoors, the participants said that they left windows or doors open for ventilation (93/137, 67.9%), slept in a cooler place or outdoors (42/137, 30.7%), drank more water (35/137, 25.5%), rested (29/137, 21.2%), did the most strenuous work during cooler times (15/137, 10.9%), or did not take any adaptive measures (23/137, 16.8%). When outdoors, they mostly rested in the shade (79/137, 57.7%), drank more water (77/137, 56.2%), wore loose clothing (37/137, 27%), rested (24/137, 17.5%), used a hat or sun protection (21/137, 15.3%), went inside (19/137, 13.9%), and did the most strenuous work during cooler times (15/137, 10.9%) to protect themselves against the heat. When asked if they were affected by other weather extremes, 21.9% (30/137) of the participants said yes, especially by heavy rains causing flooding of houses, yards, and fields (13/30, 43%); high humidity in the house (8/30, 27%); and destruction of houses (3/30, 10%). A total of 17% (5/30) of the participants mentioned that they were negatively affected by cold temperatures as well.

Discussion

Summary of Findings

To investigate the impact of weather on 143 individuals in the rural communities of the Nouna HDSS in Burkina Faso, we used consumer-grade wearables to collect data on their daily activity, sleep, and HR over the course of 11 months. We used weather indexes based on rainfall, temperature, HI, and WBGT estimates to quantify the weather extremes of heavy rainfall, nights with heat stress, and hot days. In addition, we conducted a questionnaire with study participants regarding their perceptions of weather exposure, as well as an activity diary in which study participants were asked to provide retrospective information on their daily activities. We found that sleep duration decreased with higher heat exposure, which corresponds with the participants' questionnaire responses indicating that they were most affected by the heat at night. In contrast, sleep duration increased with higher precipitation. During the day, most participants (89/137, 65%) worked outdoors doing physical labor in the form of, for example, farming, housework, or fetching water. With increasing WBGT values up to a threshold of approximately 30 °C, daily activity (steps) increased. Increased WBGT above this threshold was associated with decreased daily activity. However, this effect varied by month, with a decrease in daily activity (steps) per degree increase in WBGT during the rainiest month for a WBGT of 20 °C and an increase in activity for a WBGT of ≥30 °C. In addition, increasing precipitation was correlated with lower daily activity (steps). As daytime HR data were limited, we focused primarily on nighttime HR, for which we found that increasing HI had no significant impact. There was a small but statistically significant decrease in the average nighttime HR as rainfall levels rose.

Daily Activity

Similar to the findings of Edwards et al [53], daily activity in the form of steps increased with higher heat exposure up to a certain threshold. Our study population's threshold was at WBGT values of 30 °C during most months, which can be considered harmful to health in terms of heat exposure [42]. In general, people seemed to avoid heat exposure, especially during the hot dry season, by deferring outdoor activities until later in the day when temperatures had cooled down and resting during the hottest hours, as indicated by lower step counts during the hottest hours of the day and as reported by the participants in the self-perceived heat questionnaire. In accordance with Al-Mohannadi et al [11], we found that heat stress had no statistically different effects on people of different ages and sex even though women reported less outdoor work in their activity journals. The per-degree effect of WBGT on daily activity (steps) varied between months, with higher absolute step counts and an increase in daily activity observed at a 30 °C $WBGT_{max}$ in the rainiest month (August) compared with a decrease in the coolest (January) and hottest (April) months. A reason for this may be related to the season. Considering the agricultural calendar for Burkina Faso [54], most activities in the heat seemed to take place during usual times of harvest throughout the rainy season. Given the rural Burkinabé population's dependence on agriculture, many of whom are subsistence farmers who rely on harvest outcomes for their own nutrition and primary source of income [55], it appears that they cannot afford to suspend agricultural activities during extreme heat. Furthermore, higher precipitation was associated with decreased activity, which confirms previous findings [12,56-58].

Sleep

We found that sleep duration decreased with increasing heat exposure, similar to studies conducted in other countries, including middle-income countries with tropical climates [6,12,56,59]. However, contrary to what has been observed in several other studies, we did not find that overnight heat exposure led to poorer sleep either in older participants compared with younger participants or in female participants compared with male participants [6,59]. The estimated effect of nighttime heat on sleep duration was lowest during the hottest month (April). This might be an indication of people getting used to high air temperatures. In contrast, heat sensitivity was highest during the rainiest and most humid month (August), which matches the findings that humans are affected more by heat when relative humidity is higher compared with higher air temperature alone [5]. On half of the nights evaluated, sleep duration was insufficient, and overall averages were up to an hour lower than those observed in studies conducted in other countries [60]. Our estimates of sleep duration did not provide insights into changes in sleep physiology. However, it has been found that rapid eye movement sleep decreases with higher ambient temperatures in laboratory settings [61]. Moreover, the wearables did not capture sleep for <3 hours. As such, we could only evaluate nighttime sleep as opposed to shorter naps. As a result, it is unclear whether individuals were able to make up for nighttime sleep deficits by taking naps throughout the day. Increasing precipitation was associated with increasing sleep duration in accordance with the findings of the large-scale study by Minor et al [6].

HR Findings

In contrast to earlier findings [62,63], average HR values were lower in our study when individuals were exposed to increasing heat; however, this effect was not statistically significant.

Furthermore, it is important to consider that we looked at average nighttime HR, which reflects resting HR, and not HR during activities, as in other studies. It must also be noted that our analysis was based on a very small data set owing to the low data completeness for HR. With the large proportion of missing data, our study results should only be considered as hypothesis generating [47]. The lower HR values during nights with higher HI_{min} may indicate heat adaptation of study participants [64]. The nighttime HR was significantly influenced by the amount of rainfall that occurred each day. When the total amount of rainfall was greater, as it often is during severe rains, HR values were slightly lower. These results corroborate the earlier observation that more frequent and heavier rainfall is associated with longer periods of sleep, indicating better rest. As the cardiovascular system is severely affected by life-threatening events such as heatstrokes, it would be essential to conduct additional research on the cardiovascular effects of climate change and extreme weather exposure.

Weather Exposure

Although the highest temperatures often occurred during the dry months, WBGT and HI readings peaked during the rainy season. During our study, WBGT estimates frequently reached values of >30 °C, which is considered critical even under moderate activity [42]. A total of 8 days of heavy rain were experienced by the participants throughout the rainy season. Participants ranked heavy rainfall and flooding as the second most impactful weather event. The number of floods has been increasing over the past few decades in West Africa [22,65] and is a contributing factor to food insecurity and economic loss for subsistence farmers [28]. The rural study population in Nouna, Burkina Faso, is highly exposed to weather extremes in the form of heat and heavy rains, which will likely increase in the future because of climate change [66,67].

Although WBGT was originally introduced as a measure for heat stress for outdoor work in direct sunlight, our model comparison showed that WBGT predicted daily activity better than temperature alone. A reason may be that humidity, a component of WBGT and HI, is an important indicator of a person's heat stress [5]. We calculated WBGT based on weather station measurements, but it should be noted that Lemke and Kjellstrom [68] have found no statistically significant difference between calculated and measured WBGT values. Nevertheless, the assessment of heat exposure using WBGT has its limitations as it, for example, does not include adjustment for clothing [69]. Furthermore, we calculated WBGT based on the equation by Carter et al [39], which was based on cooler climatic conditions than those in our study setting. Therefore, WBGT values should be considered estimates.

Self-Perceived Burden of Heat Exposure

Half (67/137, 48.9%) of the participants reported being negatively affected by heat, and some (30/137, 21.9%) also mentioned heavy rains and flooding as impairing factors in their daily lives. They experienced most disturbances at night, during the hottest times of the day, and during outdoor work. This was also reflected in our objective measurements. The participants already reported many adaptive measures when outdoors and indoors that are also recommended by the World Health Organization [70]. This included resting, seeking shade, working during cooler times of the day, sleeping outdoors or in cooler places, and leaving doors and windows open for ventilation. This adaptive behavior could be seen in the results of their activity journals as well, where they reported most of the outdoor and strenuous work in the morning and more resting during midday and the afternoon, when heat exposure was highest. There is an immediate need for more adaptive and preventive measures for this low-resource population as the future poses even more severe climate conditions.

Limitations

One of the major limitations of our study was incomplete data. Data completeness was low for both accelerometry (daily activity and sleep) and photoplethysmography (HR) data. This could be due to multiple reasons. Environmental factors had a significant impact on reaching the study participants in their homes and synchronizing the wearable data. During the rainy season, for instance, certain villages were inaccessible because of flooding, making many dirt roads impassable for fieldworkers. In addition, the political environment in Burkina Faso hindered the collection of data because of security concerns in some villages (threats of terrorism) and the shutdown of mobile internet in November 2021 and during the coup d'état in January 2022. The limitation of internet connection in most of the study area presented a barrier to the synchronization of participant data at their homes. Another main limitation was damaged wearables. Over the course of the study, >40 devices malfunctioned, of which 20 could be replaced. The most common causes of wearables breaking were water damage (although the devices were waterproof) and the impact of force (eg, during fieldwork). Second, the lack of information regarding the validity of the Withings Pulse HR is a major limitation. The validity of these devices has not been verified in a study setting. Therefore, we can only consider the values as estimates. As Withings only released the processed accelerometry data and not the raw accelerometry data, we lacked clarity regarding the calculations that underpinned these data points for sleep and daily activity. Furthermore, we deemed the pulse rate measured by the photoplethysmography sensor equivalent to HR. The technical aspects of the wearables could have also caused low data completeness and inaccuracy, especially for HR measurements. In previous studies, various noise sources have been identified that impaired the photoplethysmography signals and caused inaccurate measurements [71]. This includes individual, external, and physiological factors such as obesity, skin tone, body site, and motion artifacts. Owing to its relatively narrow photoplethysmography sensor compared with other consumer-grade wearables, the Withings Pulse HR may not have accurately measured participants' HRs while they were in motion or sweating heavily. Furthermore, it has been suggested that melanin disrupts the functionality of photoplethysmography, resulting in poorer performance on darker skin tones, which could explain the low data completeness in our study population [72,73]. To match weather and wearable data, we used weather data from the nearest weather station to each participant's house, not considering when participants were not at home, which might have caused inaccuracy. Regarding the representativeness of our study cohort, we had one major limitation: most of our study population was middle-aged even though most Burkinabé are aged <25 years [74]. It would be important to gain more insights into the sensitivity of children and young adults to heat and heavy rains. Finally, the linear mixed-effects models that we used to explore the relationship between extreme weather exposure and health parameters had some limitations. We tried to address possible overfitting through hierarchical model building and cross-validation. However, the models showed quite low average R^2 values when cross-validated using the LOSO approach. We still deemed our models' performance sufficient for this exploratory analysis as R^2 values for some participants' data as test data were much higher, especially for those participants with good data completeness, which leads us back to our first major limitation.

Conclusions

On the basis of our findings, the rural population in Burkina Faso is exposed to many days of extreme weather in the form of heavy rains, nights with heat stress, and hot days. Heat especially seemed to be associated with shortened sleep duration, which was also confirmed by the subjective perceptions of study participants, who reported that heat had the greatest impact on their daily lives at night. Heat-related sleep disruption is a major issue for the general public's health. During the hottest month (April), daily activity (steps) decreased with increased heat exposure at 30 °C WBGT. Total daily activity (steps) was highest during the rainy season, typically June to September, which had the highest number of heat stress days, and the number of steps even increased as WBGT rose above 30 °C. Most people in rural Burkina Faso are subsistence farmers who depend on their harvests for food and income. This means that most of the agricultural work, especially during the rainy season when most crops are harvested, may coincide with the most extreme weather exposure. Other essential activities such as obtaining water and caring for livestock also potentially expose Burkinabés to weather conditions that may be detrimental to their health. Heavy rainfall was associated with a small increase in sleep duration, slightly decreased average nighttime HR, and decreased daily activity (steps). Participants in the study also recognized the agricultural and economic impacts of heavy rainfall as threats to their daily lives, which should be explored in future research. Study participants reported adverse health impacts of weather and adaptive measures, such as avoiding extreme heat and delaying physical activities until cooler times. This is the first long-term study that, to our knowledge, evaluates the impacts of weather exposure on a rural population in Burkina Faso using objective measures from consumer-grade wearables to better comprehend everyday exposure. On the basis of these findings, new adaptive measures could be implemented in rural, low-resource communities that are highly exposed to climate change to protect people from heat, especially during the night and outdoor work. This could be, for example, in the form of housing cooling methods or personal protective gear.

Acknowledgments

The authors wish to thank the German Research Foundation (*Deutsche Forschungsgemeinschaft*) for supporting this work as part of a *Deutsche Forschungsgemeinschaft*–funded research unit (*Forschungsgruppe* FOR 2936, project: 409670289). The authors acknowledge financial support from the Else Kröner-Fresenius-Stiftung within the Heidelberg Graduate School of Global Health; the German Research Foundation (Deutsche Forschungsgemeinschaft) within the Open Access Publishing funding program by the Baden-Württemberg Ministry of Science, Research, and the Arts and by Ruprecht-Karls-Universität Heidelberg. The funders did not play a role in the design, data collection and analysis, decision to publish, or writing of the manuscript.

Data Availability

The data sets generated during the study and the corresponding R scripts can be made available from the corresponding author in an anonymized form upon individual request.

Authors' Contributions

SB, TB, and AS conceived and designed the study. AS, VB, and GC managed the study in Burkina Faso with the remote guidance of SB, MK, AB, and MAM. MK monitored the data during the study, analyzed the data, and drafted the manuscript in close collaboration with SB. All authors contributed to the critical revision of the draft and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement: guidelines for reporting observational studies.

[PDF File (Adobe PDF File), 127 KB - mhealth_v11i1e46980_app1.pdf]

Multimedia Appendix 2

Activity journal and heat questionnaire.

[PDF File (Adobe PDF File), 284 KB - mhealth_v11i1e46980_app2.pdf]

Multimedia Appendix 3
Full results of the linear mixed-effects models.
[PDF File (Adobe PDF File), 511 KB - mhealth_v11i1e46980_app3.pdf]

References

1. Intergovernmental Panel on Climate Change (IPCC). Human health: impacts, adaptation, and co-benefits. In: Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report. Cambridge, UK: Cambridge University Press; 2014:709-754.
2. Watts N, Amann M, Arnell N, Ayeb-Karlsson S, Beagley J, Belesova K, et al. The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises. Lancet 2021 Jan 09;397(10269):129-170. [doi: 10.1016/S0140-6736(20)32290-X] [Medline: 33278353]
3. Rocque RJ, Beaudoin C, Ndjaboue R, Cameron L, Poirier-Bergeron L, Poulin-Rheault RA, et al. Health effects of climate change: an overview of systematic reviews. BMJ Open 2021 Jun 09;11(6):e046333 [FREE Full text] [doi: 10.1136/bmjopen-2020-046333] [Medline: 34108165]
4. Kenny GP, Notley SR, Flouris AD, Grundstein A. Climate change and heat exposure: impact on health in occupational and general populations. In: Adams W, Jardine J, editors. Exertional Heat Illness. Cham, Switzerland: Springer; 2020. [doi: 10.1007/978-3-030-27805-2_12]
5. Sobolewski A, Młynarczyk M, Konarska M, Bugajska J. The influence of air humidity on human heat stress in a hot environment. Int J Occup Saf Ergon 2021 Mar 25;27(1):226-236. [doi: 10.1080/10803548.2019.1699728] [Medline: 31779525]
6. Minor K, Bjerre-Nielsen A, Jonasdottir SS, Lehmann S, Obradovich N. Rising temperatures erode human sleep globally. One Earth 2022 May;5(5):534-549. [doi: 10.1016/j.oneear.2022.04.008]
7. Rifkin DI, Long MW, Perry MJ. Climate change and sleep: a systematic review of the literature and conceptual framework. Sleep Med Rev 2018 Dec;42:3-9 [FREE Full text] [doi: 10.1016/j.smrv.2018.07.007] [Medline: 30177247]
8. Krause AJ, Simon EB, Mander BA, Greer SM, Saletin JM, Goldstein-Piekarski AN, et al. The sleep-deprived human brain. Nat Rev Neurosci 2017 Jul 18;18(7):404-418 [FREE Full text] [doi: 10.1038/nrn.2017.55] [Medline: 28515433]
9. Irwin MR. Why sleep is important for health: a psychoneuroimmunology perspective. Annu Rev Psychol 2015 Jan 03;66(1):143-172 [FREE Full text] [doi: 10.1146/annurev-psych-010213-115205] [Medline: 25061767]
10. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J 2011 Jun;32(12):1484-1492. [doi: 10.1093/eurheartj/ehr007] [Medline: 21300732]
11. Al-Mohannadi AS, Farooq A, Burnett A, Van Der Walt M, Al-Kuwari MG. Impact of climatic conditions on physical activity: a 2-year cohort study in the Arabian Gulf region. J Phys Act Health 2016 Sep;13(9):929-937. [doi: 10.1123/jpah.2015-0593] [Medline: 27631650]
12. Quante M, Wang R, Weng J, Kaplan ER, Rueschman M, Taveras EM, et al. Seasonal and weather variation of sleep and physical activity in 12-14-year-old children. Behav Sleep Med 2019 Oct 09;17(4):398-410 [FREE Full text] [doi: 10.1080/15402002.2017.1376206] [Medline: 28922020]
13. Flouris AD, Dinas PC, Ioannou LG, Nybo L, Havenith G, Kenny GP, et al. Workers' health and productivity under occupational heat strain: a systematic review and meta-analysis. Lancet Planetary Health 2018 Dec;2(12):e521-e531. [doi: 10.1016/s2542-5196(18)30237-7]
14. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
15. Bellettiere J, LaMonte MJ, Evenson KR, Rillamas-Sun E, Kerr J, Lee I, et al. Sedentary behavior and cardiovascular disease in older women: the Objective Physical Activity and Cardiovascular Health (OPACH) study. Circulation 2019 Feb 19;139(8):1036-1046 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.118.035312] [Medline: 31031411]
16. Kraus WE, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP, 2018 Physical Activity Guidelines Advisory Committee*. Daily step counts for measuring physical activity exposure and its relation to health. Med Sci Sports Exerc 2019 Jun;51(6):1206-1212 [FREE Full text] [doi: 10.1249/MSS.0000000000001932] [Medline: 31095077]
17. Madaniyazi L, Zhou Y, Li S, Williams G, Jaakkola JJ, Liang X, et al. Outdoor temperature, heart rate and blood pressure in Chinese adults: effect modification by individual characteristics. Sci Rep 2016 Feb 15;6(1):21003 [FREE Full text] [doi: 10.1038/srep21003] [Medline: 26876040]
18. De Barros JA, Macartney MJ, Peoples GE, Notley SR, Herry CL, Kenny GP. Effects of sex and wet-bulb globe temperature on heart rate variability during prolonged moderate-intensity exercise: a secondary analysis. Appl Physiol Nutr Metab 2022 Jul 01;47(7):725-736. [doi: 10.1139/apnm-2022-0004] [Medline: 35290752]

19. Boudoulas KD, Borer JS, Boudoulas H. Heart rate, life expectancy and the cardiovascular system: therapeutic considerations. Cardiology 2015 Aug 15;132(4):199-212 [FREE Full text] [doi: 10.1159/000435947] [Medline: 26305771]
20. De Longueville F, Hountondji Y, Kindo I, Gemenne F, Ozer P. Long-term analysis of rainfall and temperature data in Burkina Faso (1950-2013). Int J Climatol 2016 Feb 03;36(13):4393-4405. [doi: 10.1002/joc.4640]
21. Intergovernmental Panel on Climate Change (IPCC). Africa. Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part B: Regional Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report. Cambridge, UK: Cambridge University Press; 2014:1199-1265.
22. Kappelle M. State of the climate in Africa 2019. World Meteorological Organization. 2020 Oct. URL: https://www.researchgate.net/publication/344898712_State_of_the_Climate_in_Africa_2019 [accessed 2022-02-02]
23. Green H, Bailey J, Schwarz L, Vanos J, Ebi K, Benmarhnia T. Impact of heat on mortality and morbidity in low and middle income countries: a review of the epidemiological evidence and considerations for future research. Environ Res 2019 Apr;171:80-91. [doi: 10.1016/j.envres.2019.01.010] [Medline: 30660921]
24. Pasquini L, van Aardenne L, Godsmark CN, Lee J, Jack C. Emerging climate change-related public health challenges in Africa: a case study of the heat-health vulnerability of informal settlement residents in Dar es Salaam, Tanzania. Sci Total Environ 2020 Dec 10;747:141355. [doi: 10.1016/j.scitotenv.2020.141355] [Medline: 32777515]
25. Gilbert R, Cliffe SJ. Public health surveillance. In: Regmi K, Gee I, editors. Public Health Intelligence. Cham, Switzerland: Springer; 2016. [doi: 10.1007/978-3-319-28326-5_5]
26. Nsubuga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public health surveillance: a tool for targeting and monitoring interventions. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, et al, editors. Disease Control Priorities in Developing Countries. 2nd edition. New York, NY: World Bank Publications; 2006.
27. Sié A, Louis VR, Gbangou A, Müller O, Niamba L, Stieglbauer G, et al. The Health and Demographic Surveillance System (HDSS) in Nouna, Burkina Faso, 1993-2007. Glob Health Action 2010 Sep 14;3(1):5284 [FREE Full text] [doi: 10.3402/gha.v3i0.5284] [Medline: 20847837]
28. Mank I, Belesova K, Bliefernicht J, Traoré I, Wilkinson P, Danquah I, et al. The impact of rainfall variability on diets and undernutrition of young children in rural Burkina Faso. Front Public Health 2021 Sep 20;9:693281 [FREE Full text] [doi: 10.3389/fpubh.2021.693281] [Medline: 34616704]
29. Diboulo E, Sié A, Rocklöv J, Niamba L, Yé M, Bagagnan C, et al. Weather and mortality: a 10 year retrospective analysis of the Nouna Health and Demographic Surveillance System, Burkina Faso. Glob Health Action 2012 Nov 23;5(1):6-13 [FREE Full text] [doi: 10.3402/gha.v5i0.19078] [Medline: 23195510]
30. Bunker A, Sewe MO, Sié A, Rocklöv J, Sauerborn R. Excess burden of non-communicable disease years of life lost from heat in rural Burkina Faso: a time series analysis of the years 2000-2010. BMJ Open 2017 Nov 04;7(11):e018068 [FREE Full text] [doi: 10.1136/bmjopen-2017-018068] [Medline: 29102994]
31. Heo S, Bell ML, Lee JT. Comparison of health risks by heat wave definition: applicability of wet-bulb globe temperature for heat wave criteria. Environ Res 2019 Jan;168:158-170. [doi: 10.1016/j.envres.2018.09.032] [Medline: 30316101]
32. Parsons K. Heat stress standard ISO 7243 and its global application. Ind Health 2006 Jul;44(3):368-379 [FREE Full text] [doi: 10.2486/indhealth.44.368] [Medline: 16922180]
33. Sahu KS, Majowicz SE, Dubin JA, Morita PP. NextGen public health surveillance and the Internet of Things (IoT). Front Public Health 2021 Dec 3;9:756675 [FREE Full text] [doi: 10.3389/fpubh.2021.756675] [Medline: 34926381]
34. Huhn S, Matzke I, Koch M, Gunga HC, Maggioni MA, Sié A, et al. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: a case study. Front Public Health 2022 Sep 30;10:972177 [FREE Full text] [doi: 10.3389/fpubh.2022.972177] [Medline: 36249225]
35. Koch M, Matzke I, Huhn S, Gunga HC, Maggioni MA, Munga S, et al. Wearables for measuring health effects of climate change-induced weather extremes: scoping review. JMIR Mhealth Uhealth 2022 Sep 09;10(9):e39532 [FREE Full text] [doi: 10.2196/39532] [Medline: 36083624]
36. Tomalka J, Lange S, Röhrig F, Gornott C. Climate risk profile: Burkina Faso*. Federal Ministry for Economic Cooperation and Development. URL: https://tinyurl.com/35af49fn [accessed 2023-01-04]
37. Barteit S, Boudo V, Ouedraogo A, Zabré P, Ouremi L, Sié A, et al. Feasibility, acceptability and validation of wearable devices for climate change and health research in the low-resource contexts of Burkina Faso and Kenya: study protocol. PLoS One 2021 Sep 30;16(9):e0257170 [FREE Full text] [doi: 10.1371/journal.pone.0257170] [Medline: 34591893]
38. Hijmans RJ, Karney C, Williams E, Vennes C. geosphere: spherical trigonometry, R package version 1.5-18. CRAN R-Project. URL: https://cran.r-project.org/web/packages/geosphere/index.html [accessed 2023-10-23]
39. Carter AW, Zaitchik BF, Gohlke JM, Wang S, Richardson MB. Methods for estimating wet bulb globe temperature from remote and low ‐ cost data: a comparative study in Central Alabama. GeoHealth 2020 May 21;4(5):e2019GH000231. [doi: 10.1029/2019gh000231]
40. Anderson GB, Bell ML, Peng RD. Methods to calculate the heat index as an exposure metric in environmental health research. Environ Health Perspect 2013 Oct;121(10):1111-1119 [FREE Full text] [doi: 10.1289/ehp.1206273] [Medline: 23934704]

41. Perkins SE, Alexander LV. On the measurement of heat waves. J Clim 2013;26(13):4500-4517. [doi: 10.1175/JCLI-D-12-00383.1]
42. Vecellio DJ, Wolf ST, Cottle RM, Kenney WL. Evaluating the 35°C wet-bulb temperature adaptability threshold for young, healthy subjects (PSU HEAT Project). J Appl Physiol (1985) 2022 Feb 01;132(2):340-345 [FREE Full text] [doi: 10.1152/japplphysiol.00738.2021] [Medline: 34913738]
43. Wolf ST, Folkerts MA, Cottle RM, Daanen HA, Kenney WL. Metabolism- and sex-dependent critical WBGT limits at rest and during exercise in the heat. Am J Physiol Regul Integr Comp Physiol 2021 Sep 01;321(3):R295-R302 [FREE Full text] [doi: 10.1152/ajpregu.00101.2021] [Medline: 34259026]
44. What is the heat index? National Weather Service. URL: https://www.weather.gov/ama/heatindex [accessed 2021-12-22]
45. Tan FE, Jolani S, Verbeek H. Guidelines for multiple imputations in repeated measurements with time-dependent covariates: a case study. J Clin Epidemiol 2018 Oct;102:107-114. [doi: 10.1016/j.jclinepi.2018.06.006] [Medline: 29964148]
46. Peters SA, Bots ML, den Ruijter HM, Palmer MK, Grobbee DE, Crouse JR, METEOR study group. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012 Jun;65(6):686-695. [doi: 10.1016/j.jclinepi.2011.11.012] [Medline: 22459429]
47. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
48. Lewis LK, Maher C, Belanger K, Tremblay M, Chaput JP, Olds T. At the mercy of the gods: associations between weather, physical activity, and sedentary time in children. Pediatr Exerc Sci 2016 Feb;28(1):152-163. [doi: 10.1123/pes.2015-0076] [Medline: 26098393]
49. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
50. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156 [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
51. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
52. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
53. Edwards NM, Myer GD, Kalkwarf HJ, Woo JG, Khoury PR, Hewett TE, et al. Outdoor temperature, precipitation, and wind speed affect physical activity levels in children: a longitudinal cohort study. J Phys Act Health 2015 Aug;12(8):1074-1081 [FREE Full text] [doi: 10.1123/jpah.2014-0125] [Medline: 25423667]
54. Burkina Faso acute food insecurity. Burkina Faso | Famine Early Warning Systems Network. URL: https://fews.net/west-africa/burkina-faso [accessed 2022-12-12]
55. Belesova K, Gasparrini A, Sié A, Sauerborn R, Wilkinson P. Household cereal crop harvest and children's nutritional status in rural Burkina Faso. Environ Health 2017 Jun 20;16(1):65 [FREE Full text] [doi: 10.1186/s12940-017-0258-9] [Medline: 28633653]
56. Zheng C, Huang WY, Wong SH. Associations of weather conditions with adolescents' daily physical activity, sedentary time, and sleep duration. Appl Physiol Nutr Metab 2019 Dec;44(12):1339-1344. [doi: 10.1139/apnm-2019-0309] [Medline: 31063697]
57. Klenk J, Büchele G, Rapp K, Franke S, Peter R, ActiFE Study Group. Walking on sunshine: effect of weather conditions on physical activity in older people. J Epidemiol Community Health 2012 May 15;66(5):474-476. [doi: 10.1136/jech.2010.128090] [Medline: 21325149]
58. Bélanger M, Gray-Donald K, O'Loughlin J, Paradis G, Hanley J. Influence of weather conditions and season on physical activity in adolescents. Ann Epidemiol 2009 Mar;19(3):180-186. [doi: 10.1016/j.annepidem.2008.12.008] [Medline: 19217000]
59. Obradovich N, Migliorini R, Mednick SC, Fowler JH. Nighttime temperature and human sleep loss in a changing climate. Sci Adv 2017 May;3(5):e1601555 [FREE Full text] [doi: 10.1126/sciadv.1601555] [Medline: 28560320]
60. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021 Feb 12;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]
61. Buguet A. Sleep under extreme environments: effects of heat and cold exposure, altitude, hyperbaric pressure and microgravity in space. J Neurol Sci 2007 Nov 15;262(1-2):145-152. [doi: 10.1016/j.jns.2007.06.040] [Medline: 17706676]
62. Hargett T. The effect of heat exposure on heart rate and VO2 during time to exhaustion exercise. 2019 Presented at: Kent State University Undergraduate Symposium on Research, Scholarship and Creative Activity; April 17, 2023; Kent, OH URL: https://oaks.kent.edu/ugresearch/2019/effect-heat-exposure-heart-rate-and-vo2-during-time-exhaustion-exercise
63. Ravanelli NM, Hodder SG, Havenith G, Jay O. Heart rate and body temperature responses to extreme heat and humidity with and without electric fans. JAMA 2015 Feb 17;313(7):724-725 [FREE Full text] [doi: 10.1001/jama.2015.153] [Medline: 25688786]

64. Rahimi GR, Albanaqi AL, Van der Touw T, Smart NA. Physiological responses to heat acclimation: a systematic review and meta-analysis of randomized controlled trials. J Sports Sci Med 2019 Jun;18(2):316-326 [FREE Full text] [Medline: 31191102]
65. Kogan F, Guo W, Yang W. Drought and food security prediction from NOAA new generation of operational satellites. Geomat Nat Hazards Risk 2019 Jan 29;10(1):651-666. [doi: 10.1080/19475705.2018.1541257]
66. Hondula DM, Rocklöv J, Sankoh OA. Past, present, and future climate at select INDEPTH member Health and Demographic Surveillance Systems in Africa and Asia. Glob Health Action 2012 Nov 23;5(1):74-86 [FREE Full text] [doi: 10.3402/gha.v5i0.19083] [Medline: 23195511]
67. Schewe J, Levermann A. Sahel rainfall projections constrained by past sensitivity to global warming. Geophysical Research Letters 2022 Sep 20;49(18):8286. [doi: 10.1029/2022GL098286]
68. Lemke B, Kjellstrom T. Calculating workplace WBGT from meteorological data: a tool for climate change assessment. Ind Health 2012;50(4):267-278 [FREE Full text] [doi: 10.2486/indhealth.ms1352] [Medline: 22673363]
69. d'Ambrosio Alfano FR, Malchaire J, Palella BI, Riccio G. WBGT index revisited after 60 years of use. Ann Occup Hyg 2014 Oct;58(8):955-970. [doi: 10.1093/annhyg/meu050] [Medline: 25062982]
70. Heat and health. World Health Organization. 2018 Jun 1. URL: https://www.who.int/news-room/fact-sheets/detail/climate-change-heat-and-health [accessed 2023-02-06]
71. Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126 [FREE Full text] [doi: 10.3390/bios11040126] [Medline: 33923469]
72. Piazena H, Meffert H, Uebelhack R. Spectral remittance and transmittance of visible and infrared-A radiation in human skin-comparison between in vivo measurements and model calculations. Photochem Photobiol 2017 Nov 27;93(6):1449-1461. [doi: 10.1111/php.12785] [Medline: 28471473]
73. Preejith SP, Alex A, Joseph J, Sivaprakasam M. Design, development and clinical validation of a wrist-based optical heart rate monitor. In: Proceedings of the IEEE International Symposium on Medical Measurements and Applications (MeMeA). 2016 Presented at: MeMeA; May 15-18, 2016; Benevento, Italy. [doi: 10.1109/memea.2016.7533786]
74. Burkina Faso - The World Factbook. Central Intelligence Agency. URL: https://www.cia.gov/the-world-factbook/countries/burkina-faso/ [accessed 2023-01-04]

Abbreviations

bpm: beats per minute
HDSS: health and demographic surveillance system
HI: heat index
HR: heart rate
LOSO: leave-one-subject-out
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WBGT: wet-bulb globe temperature

Edited by L Buis; submitted 04.03.23; peer-reviewed by A Gorny, M Khairulbahri; comments to author 03.07.23; revised version received 25.08.23; accepted 15.09.23; published 08.11.23.

Please cite as:
Koch M, Matzke I, Huhn S, Sié A, Boudo V, Compaoré G, Maggioni MA, Bunker A, Bärnighausen T, Dambach P, Barteit S
Assessing the Effect of Extreme Weather on Population Health Using Consumer-Grade Wearables in Rural Burkina Faso: Observational Panel Study
JMIR Mhealth Uhealth 2023;11:e46980
URL: https://mhealth.jmir.org/2023/1/e46980
doi:10.2196/46980
PMID:37938879

Technical Assistance Received by Older Adults to Use Commercially Available Physical Activity Monitors (Ready Steady 3.0 Trial): Ad-Hoc Descriptive Longitudinal Study

Elizabeth A Choma[1,*], DPT; Shannon Hayes[2,*], BSN, PHN; Beth A Lewis[3,*], PhD; Alexander J Rothman[4,*], PhD; Jean F Wyman[2,*], PhD; Weihua Guan[5,*], PhD; Siobhan K McMahon[2,*], PhD

1

2

3

4

5

*all authors contributed equally

Corresponding Author:
Siobhan K McMahon, PhD

Abstract

Background: Despite evidence that regular physical activity (PA) among older adults confers numerous health and functional benefits, PA participation rates are low. Using commercially available wearable PA monitors (PAMs) is one way to augment PA promotion efforts. However, while expert recommendations exist for the specific information needed at the beginning of PAM ownership and the general ongoing need for structures that support as-needed technical troubleshooting, information is lacking about the type, frequency, and modes of assistance needed during initial and long-term ownership.

Objective: This paper describes problems reported and technical assistance received by older adults who used PAMs during the 18 months they participated in a community-based PA trial: Ready Steady 3.0 (RS3).

Methods: This was an ad-hoc longitudinal analysis of process variables representing technical problems reported and assistance received by 113 RS3 study participants in the 18 months after their orientation to PAMs. Variables included date of contact, problem(s) reported, mode of technical assistance, and whether the equipment was replaced. The descriptive analysis included frequencies and incidence rates of distinct contacts, types of problems, and technical assistance modes.

Results: On average, participants were aged 77 (SD 5.2) years. Most identified as female (n=87, 77%), reported experience using smartphones (n=92, 81.4%), and used the PAM between 2 and 18 months. Eighty-two participants (72.6%) reported between 1 to 9 problems with using PAMs, resulting in a total of 150 technical assistance contacts with a mean of 1.3 (SD 1.3) contacts. The incidence rate of new, distinct contacts for technical assistance was 99 per 100 persons per year from 2018 to 2021. The most common problems were wearing the PAM (n=43, 28.7%), reading its display (n=23, 15.3%), logging into its app (n=20, 13.3%), charging it (n=18, 12%), and synchronizing it to the app (n=16, 10.7%). The modalities of technical assistance were in person (n=53, 35.3%), by telephone (n=51, 34%), by email (n=25, 16.7%), and by postal mail (n=21, 14%).

Conclusions: In general, the results of this study show that after receiving orientation to PAMs, problems such as uncomfortable wristbands, difficulty using the PAM or its related app, and obtaining or interpreting relevant personal data were occasionally reported by participants in RS3. Trained staff helped participants troubleshoot and solve these technical problems primarily in person or by phone. Results also underscore the importance of involving older adults in the design, usability testing, and supportive material development processes to prevent technical problems for the initial and ongoing use of PAMs. Clinicians and researchers should further assess technical assistance needed by older adults, accounting for variations in PAM models and wear time, while investigating additional assistance strategies, such as proactive support, short GIF videos, and video calls.

(JMIR Mhealth Uhealth 2023;11:e47891) doi:10.2196/47891

KEYWORDS

wearable device; digital health; physical activity monitor; PAM; older adult; intervention; physical activity; usability; technical assistance; supportive structures; monitoring; promote; community; support

Introduction

Regular physical activity among older adults, including those with chronic conditions and frailty, confers numerous health and functional benefits [1]. Yet less than 16% of people aged over 65 years in the United States meet the recommendations for minimal physical activity among older adults [2]. Technologies such as commercially available wearable physical activity monitors (PAMs) can prompt some behavior change strategies that help promote regular physical activity [3,4]. Although older adults are interested in using PAMs [5,6], most do not use them [7] or stop using them shortly after obtaining them [8]. An important contributor to their lack of PAM use is the limited information received about their purpose, use, and accuracy at the beginning of ownership and limited structures and resources to provide technical assistance during long-term use [9-14]. Although expert recommendations for providing information at the beginning of PAM ownership are specific, recommendations for providing long-term technical support are general. For example, there is consensus that supportive structures to provide on-demand technical assistance are required to facilitate long-term use [10,15,16], yet specific recommendations are lacking about the type, frequency, and mode of support that might be needed. This paper describes problems reported and technical assistance received by older adults who used PAMs during the 18 months they participated in a community-based physical activity trial (Ready Steady 3.0 [RS3]) [17].

Encouraging older adults to use PAMs can augment efforts to promote increased physical activity, particularly when combined with behavior change strategies such as goal setting, prompts, or social support [1,3,18,19]. Results from recent systematic reviews of physical activity interventions that incorporate PAM use have indicated improved levels of physical activity for at least 6 months across a wide range of age groups, including older adults [20,21]. Additionally, prior research suggests that older adults are interested in and willing to use PAMs [5,6]. Moreover, the *Physical Activity Guidelines for Americans* [1] suggest technology such as PAMs should be used to augment efforts to promote increased physical activity and maintain this increase.

Despite evidence supporting PAM use, there are complex and dynamic problems that limit their adoption and long-term use among older adults. Examples include broken or lost PAMs, perceptions that they are too complicated to learn or use and that their data are inaccurate or not helpful, observations that they are uncomfortable or aesthetically unacceptable, and personal experiences that cause lapses in use (eg, illness) [9,13,22]. As a result, in addition to recommendations for improving several facets of PAM designs, experts have provided specific practical recommendations for improving the technical assistance provided to older adults by clinicians and researchers [15,23]. For example, providing introductory information and instructions (verbal and written, with illustrations) to new PAM owners can boost their understanding about the functions, use, and accuracy of PAMs, as well as how to wear them [16]. After this initial orientation, experts recommend continued assistance through structures such as a technology-help phone line [15]. Planning for such supportive structures raises questions about what to expect regarding the frequency, type, and modalities of such PAM technical assistance.

One strategy to improve our understanding of what is needed to provide long-term structured support for older adults' use of PAMs is to examine process data regarding their use in a physical activity trial: RS3. The RS3 study aimed to assess the relative effects of 2 intervention components using distinct behavior change techniques on community-dwelling older adults' physical activity [17]. As part of RS3, participants received and used PAMs (Fitbit Charge 2; Fitbit, Inc) for 18 months [17]. Participants received PAM orientation consistent with published recommendations; they were then encouraged to report PAM problems to the research team and received as-needed technical assistance.

The purpose of this study was to examine the assistance provided to facilitate long-term PAM use among a subsample of RS3 participants. Research questions were as follows:

1. How many participants, after receipt of PAM orientation, reported PAM problems and requested technical assistance during the 18 month follow-up?
    a. How many problems were reported or requests made by each participant?
    b. What was the rate of unique technical assistance contacts per person-year?
2. What types of problems did participants report that required technical assistance?
3. What modes of technical assistance resulted in resolution of PAM problems?

The results provide a detailed description of the types and quantity of assistance to expect when planning supportive structures to facilitate the long-term use of PAMs among older adults.

Methods

Overview

This was an ad hoc, descriptive study using longitudinal process data from 113 of the 309 RS3 participants (Clinical Trials.gov NCT03326141). RS3 was a randomized factorial trial designed to assess the effects of 2 types of behavior change strategies, interpersonal and intrapersonal, in an intervention targeting older adults' physical activity. The full RS3 protocol is described in another paper [17]. We selected 113 consecutive RS3 participants enrolled between February 2018 and July 2021 as the subsample for this ad hoc study. The reasons for not including the remaining 196 RS3 participants were that (1) they enrolled between November 2017 and February 2018, before ad hoc data collection protocols and procedures were fully operational and systematically captured the types and quantity of PAM technical assistance requested and provided (n=152), or (2) they had yet to complete RS3 (n=44).

Ethical Considerations

The University of Minnesota's Institutional Review Board approved and monitored the RS3 trial (1607S90922).

Participants

Community-dwelling adults enrolled in RS3 and included in this ad hoc substudy were aged ≥70 years, had the ability to walk, had at least 1 fall risk factor [24], had levels of physical activity below national and international recommended guidelines [25], had not had lower extremity surgery or injury in the past 6 weeks, had no diagnosis of neurocognitive dysfunction, and did not have a low score (≤4) on the Callahan Cognitive Screener [26]. Participants received a PAM; an 8-week, small-group intervention; and assessments at 4 time points (baseline, immediately postintervention, 6 months postintervention, and 12 months postintervention). Participants were required to use their PAMs for approximately 2 weeks during each assessment time point and encouraged (but not required) to use their PAMs on a daily basis throughout the study [17]. On average, each participant was involved with RS3 for 18 months and was encouraged to keep and use their PAM after study completion.

Settings

Technical assistance was provided to participants in community settings, such as community centers, churches, and public libraries, where the RS3 interventions and assessments occurred in Minneapolis and Saint Paul, Minnesota. We also provided assistance in or near participants' homes, according to participant preferences and COVID-19 restrictions. In addition to community settings, technical assistance was also provided via telephone, postal mail, or email.

Procedures

Training

Part-time research staff were scheduled to cover study implementation, including technical assistance, 9 AM to 5 PM, Monday through Friday. To accomplish this, 12 undergraduate research assistants, 4 graduate research assistants, 3 research professionals, and 3 interventionists were trained to maintain skills and abilities that included facilitating participants' use of PAMs and their apps, in addition to attaining competencies for their primary roles of assessment, coordination, or intervention delivery. In particular, they all acquired competencies to use PAMs through basic training that involved wearing and using all functions of the PAM in RS3 and downloading and using its app on different platforms (ie, smartphone or computer). Research staff also acquired competencies to effectively communicate with research participants who had previous experiences and varied expertise with PAMs. Staff also learned to use teach-back methods during orientation and technical assistance contacts. Finally, staff acquired coordination and documentation competencies to facilitate the team's ability to provide timely, successful assistance and track assistance variables.

Technical Assistance

Participants received 3 types of technical assistance with their PAMs during RS3: orientation, as-needed technical assistance, and pre- or postintervention meetings (Textbox 1). The timing, strategies and materials used for technical assistance were informed by suggestions and feedback from study participants in RS3 and prior research [27,28].

Textbox 1. Technical assistance during the Ready Steady 3.0 study.

Orientation (one-to-one visits, using checklists for staff and illustrated written instructions for participants)

- Visit 1
    1. Distribute PAM (physical activity monitor) with charge cable and block.
    2. Describe the purpose and accuracy of the PAM and how to wear and use it.
    3. Explain how to read the PAM display and charge the PAM using the teach-back method.
    4. Provide contact information and encourage participants to call or write research staff with questions before the second assessment.
    5. Conduct a question and answer session.
- Visit 2 (approximately 1 week after the first assessment visit)
    1. Explore participants' experience with the PAM over the last week and answer questions.
    2. Introduce the PAM app per participants' access and preferences, help participants download it on their phones, and teach them how to synchronize it, using the teach-back method.
    3. Update the PAM display to include steps, distance, active minutes, and battery level.
    4. Provide information about additional functions and provide assistance programming those functions per the participants' preferences (eg, reading sleep data).
    5. Encourage participants to contact the researchers with questions and PAM issues and review the phone number for technical assistance.

As-needed assistance

1. Monitor study telephone line and emails daily for participants' technical assistance requests.
2. Respond to participants' calls and emails immediately or as soon as possible (within 1 to 2 days) using their preferred communication mode.
3. Provide technical assistance over the telephone, by email, and, if necessary, in person.

Orientation to the PAM was standardized and delivered over 2 one-to-one meetings that were guided by the protocol described in Textbox 1, and both were augmented with illustrated guides and the teach-back method [29]. During the first meeting, participants received their PAM and instructions about its purpose, how it works, wearing it, and charging it, with time for questions. During the second meeting, approximately one week after the first meeting, participants were encouraged to ask questions and were given additional information about the PAM display, downloading and using the PAM app, and receiving as-needed technical assistance from the RS3 research team. All orientation meetings were located at the community site of the RS3 study or, when preferred, in participants' homes.

After the initial orientation, participants were encouraged to contact the research team for any PAM-related questions or issues they experienced through a dedicated study telephone number or email. Research staff were responsible for monitoring the telephone and email. They responded to reports of PAM problems within 1 day or up to 2 days if the request was received on a weekend or holiday. If technical assistance over the telephone, via email, or via postal mail did not solve the reported problem, research staff conducted an in-person meeting based on the participant's time and location preferences.

During study intervention meetings, participants were encouraged to regularly use their PAMs. The intervention workbooks included a chapter with information about using the PAM, its displays and functions, and how to get as-needed technical assistance help from research staff. However, assistance was not provided during intervention meetings to avoid interfering with the intervention. Instead, when participants requested technical assistance for PAM problems during an intervention meeting, a separate troubleshooting meeting took place after the intervention meeting or was scheduled for another day.

Data Collection

Real-time data were collected by trained research staff using computers or tablets and REDCap (Research Electronic Data Capture; Vanderbilt University), where it was verified and managed [30]. The REDCap case report forms included a baseline characteristic questionnaire and a repeatable PAM programming and problem-solving tracker to capture information about technical assistance prospectively over the course of 18 months.

Variables

Variables to describe sample characteristics included age at study enrollment, sex (female or male), race (Black/African American, Native American, or White), ethnicity (Hispanic or non-Hispanic), and previous experience with the internet (yes or no), smartphones (yes or no), and tablets (yes or no). Although wear time was measured in the main RS3 study during assessment periods, it was not monitored or measured between those periods due to resource limitations. PAM problems and technical assistance processes included dates of contact and assistance, the problem(s) reported (wearing the PAM, charging the PAM, reading the PAM display, logging into the PAM app, syncing the PAM with the app, downloading the app, using a new function on the app, sensor accuracy, sleep data, or a combination of these), and text notes to describe the problem further if needed.

Modes of assistance provided to resolve PAM problems included documentation that interactions were via telephone, email, in-person meetings (eg, in a participant's home or nearby community center), or postal mail.

Staff also documented when a participant needed a new PAM band (yes or no) and why (broken, uncomfortable, wrong size, or other). Similarly, staff documented when a new device was needed (yes or no) and why (lost, destroyed, malfunctioning, or other).

Statistical Analysis

Deidentified data from aggregated case report forms were exported from REDCap into SPSS Statistics (version 28.0; IBM Corp) and analyzed using descriptive summary statistics, including means, SDs, and frequencies. Additionally, the incidence rate was calculated as the number of new and distinct participant-research staff contacts for technology assistance per 100 persons per year. Several participants made more than one distinct request during their participation in the study, so each of these distinct contacts was considered a new event and included in this analysis. When new distinct requests involved more than one PAM problem (n=15), the primary/initial problem was used in the frequency and incidence rate calculations.

Results

Patient Characteristics

Baseline characteristics of all participants are summarized in Table 1. Data from 113 participants are included in this analysis. A total of 110 participants (97.3%) completed all 18 months of the study. Three participants (2.7%) withdrew due to illness, death, or loss of interest, limiting their follow-up times to 2, 5, and 8 months after enrollment, respectively. On average, participants were aged 76.9 (SD 5.2) years; 87 participants (77%) identified as female, and 92 (81.4%) were experienced smartphone users. As shown in Table 1, participant characteristics were similar across the subsample included in this study and the remaining RS3 sample.

Table . Characteristics of participants.

Variable		Participants who received PAM[a] technical assistance	
		Yes (n=113)	No (n=196)
Age (years), mean (SD)		76.9 (5.2)	77.6 (4.9)
Female, n (%)		87 (77)	153 (78)
Race, n (%)			
	African American/Black	12 (10.6)	36 (18.4)
	Native American	1 (0.9)	0 (0)
	White	100 (88.5)	155 (79)
	Other	0 (0)	5 (2.6)
Ethnicity, n (%)			
	Hispanic/Latino	3 (2.7)	4 (2)
	Non-Hispanic/Latino	100 (88.5)	192 (98)
Living alone, n (%)		51 (45.1)	89 (45.4)
Self-rated health, n (%)			
	Excellent	12 (10.6)	18 (9.2)
	Very good or good	87 (77)	159 (81)
	Fair or poor	14 (12.4)	19 (9.8)
Experience with technology, n (%)			
	Has used internet	106 (93.8)	168 (85.7)
	Has used a smartphone	92 (81.4)	152 (77.6)
	Has used a tablet	54 (47.8)	80 (40.8)

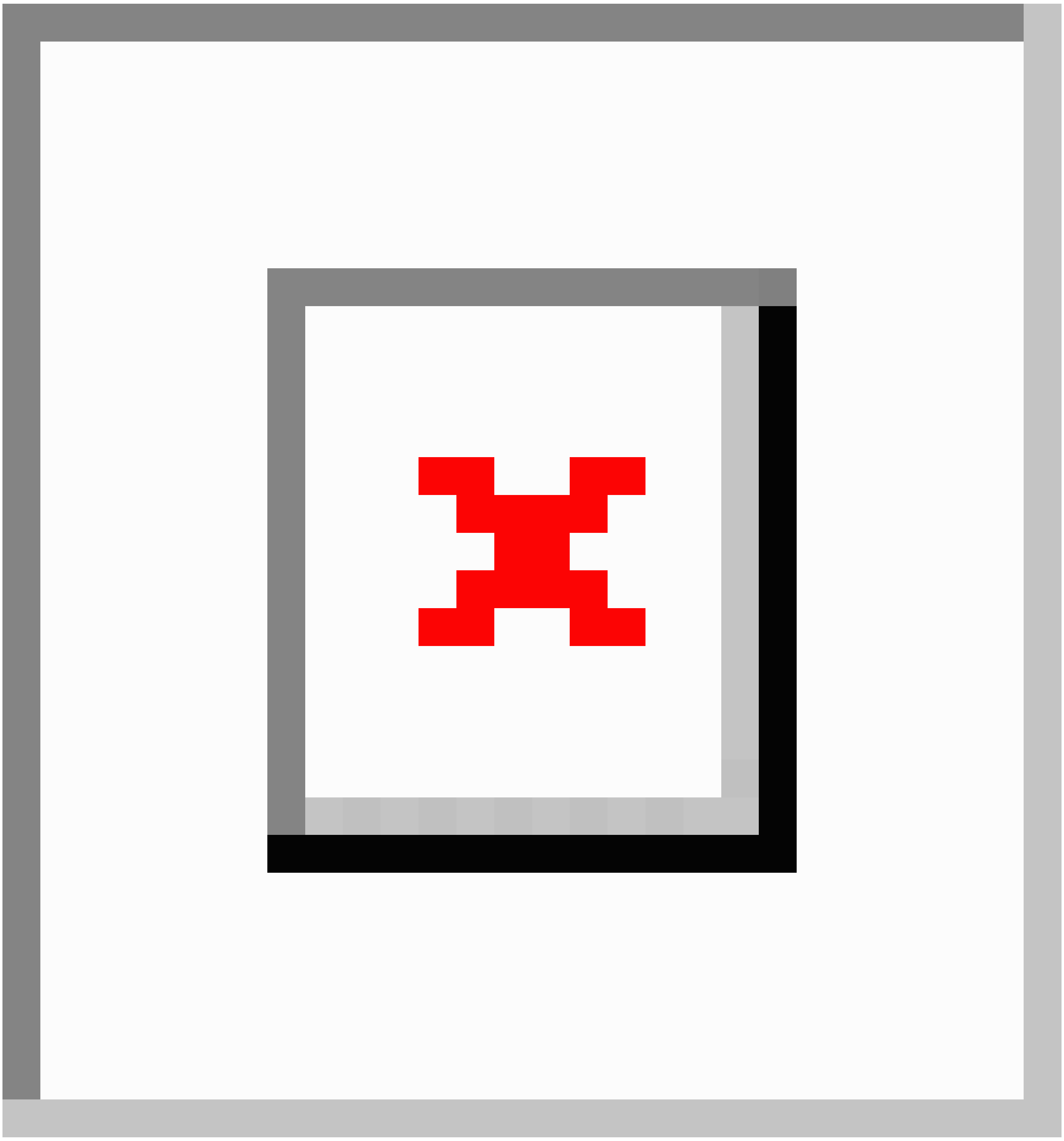
[a]PAM: physical activity monitor.

Participant-Reported PAM Problems Requiring Technical Assistance

Eighty-two participants (72.6%) made between 1 and 9 new and distinct requests for help (150 total) with a total of 169 PAM problems for which they received technical assistance from research staff. Of the 82 participants who made requests, 13 (16%) made 1 request, while 44 (54%), 15 (18%), and 10 (12%) participants made 2, 3, and ≥4 new and distinct requests, respectively. Of the 150 new and distinct requests, 135 (90%) were for 1 PAM-related problem, 11 (7.3%) were for 2 problems, and 4 (2.7%) were for 3 problems. On average, participants in this study made 1.3 (SD 1.3) requests for help with PAM problems. The incidence rate of new, distinct contacts for technical assistance was 99 per 100 persons per year from 2018 to 2021.

Figure 1 illustrates the variation in technical assistance contacts across the 18-month participation time frame. As shown, the frequency of contacts (problems reported and technical assistance provided) peaked during months 1, 10, and 15. These increases corresponded to the study time frames just after initial PAM orientation and the 6- and 12-month research assessments.

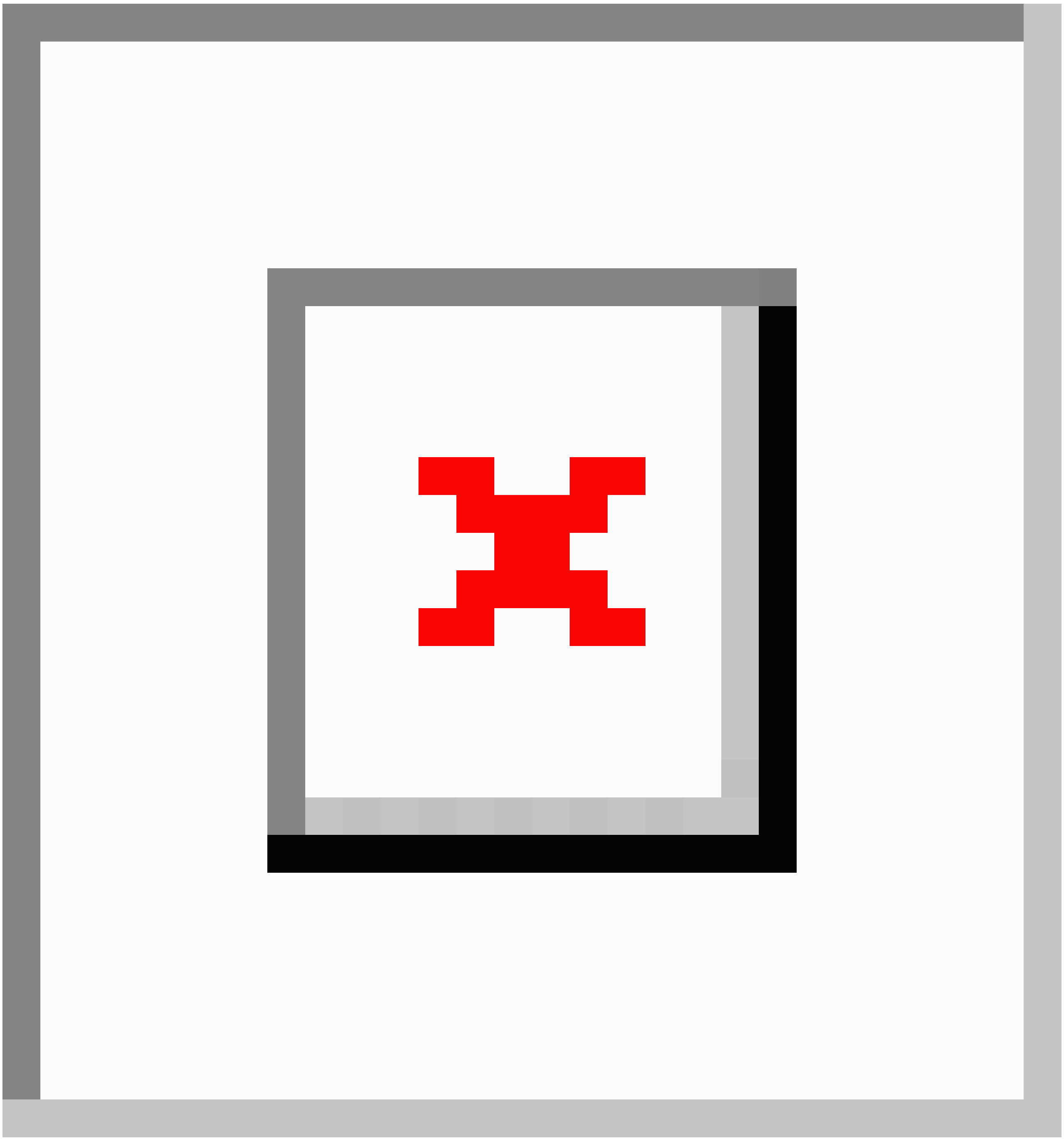
Figure 1. Monthly frequency of physical activity monitor (PAM) problems reported by 82 of 113 study participants requiring technical assistance over 18 months of follow-up after PAM orientation.



Types of PAM Problems and Modes of Technical Assistance

Types, frequencies, and modes of requests for help with PAM problems are illustrated in Figure 2. Among the 150 requests, the most common problem was with (1) wearing the PAM (n=43, 28.7%), followed by (2) reading the PAM display (n=23, 15.3%), (3) logging into the PAM app (n=20, 13.3%), (4) charging the PAM (n=18, 12%), (5) synchronizing the PAM app (n=16, 10.7%), (6) downloading the PAM app to a device (n=13, 8.7%), (7) using or activating a function on the PAM app (n=8, 4.7%), (8) having questions about sensor accuracy (n=7, 4.7%), and (9) reading sleep data (n=2, 1.3%). The modalities used in the 150 contacts to provide solutions for technical assistance included 53 (35.3%) in-person visits at community centers, participants' homes, or locations of their preference in their neighborhoods (eg, a church, library, or coffee shop); 51 (34%) telephone conversations; 25 (16.7%) email conversations; and 21 (14%) notes and instructions sent via postal mail. Twenty-seven of the 150 contacts (18%) comprised multiple visits and modes because initial technical assistance via telephone did not resolve the problem; this was particularly the case for programming or equipment problems.

Figure 2. Physical activity monitor (PAM) problems and types and modes of technical assistance.



In addition to technical assistance, 35 of the 150 (23.3%) PAM problems reported required research staff to distribute new bands, and 5 (3.3%) required new PAM devices. New bands were provided to improve comfort (n=5), improve the ease of fastening the Fitbit (n=16), or improve preferred aesthetics (n=14), such as by using a different color or material. New PAMs were provided due to malfunction (n=2) or loss (n=3).

Discussion

Principal Findings

In this study, we examined the incidence and type of distinct PAM problems reported by older adults participating in an 18-month research trial that required technical assistance to resolve. We also analyzed the mode of technical assistance provided in this study. Our findings suggest that after an initial orientation to PAMs, researchers, clinicians, and others providing technical assistance to facilitate long-term use can expect at least 99 technical assistance contacts per 100 persons per year. Variations across the 18-month follow-up time in this study showed upward trends immediately after PAM orientation and when research staff contacted participants to coordinate study assessments. The problems reported were mostly solvable and related to comfort, using the PAM device and app, and interpreting data. Finally, approximately one-third of the technical assistance contacts were in person, and the remainder were via telephone, email, or postal mail.

To our knowledge, the prior literature has not described the frequency or rate of participants' requests for assistance with PAM technology immediately after receipt or over an extended period. Thus, our findings advance the prior literature, which has elucidated reasons for the limited adoption and use of PAMs [9,13] and recommended improvements to technical assistance provided to older adults who are beginning to use commercially available PAMs, as well as those who use them long term [15,16].

This study supports a new, developing conceptual model [10], which argues that for a person to integrate a PAM into daily life, they initially need help setting it up and learning about its features, and they then need technical assistance over time. It also supports and extends expert recommendations to provide initial orientation and then to maintain supportive structures for as-needed technical assistance. This can facilitate older adults' long-term use of PAMs [15,31] by providing details about the anticipated types and frequency of ongoing assistance. Implications for this finding are that those who encourage older adults to use PAMs (eg, family, friends, clinicians, researchers, public health professionals, and health promotion programs) should plan for providing occasional technical assistance over time. Furthermore, it would be beneficial to combine responsive and proactive strategies for providing technical assistance, given our observation of more technical assistance contacts during months when research staff reached out to participants. That is, in addition to expecting occasional reports of PAM problems that need assistance, it may be helpful to occasionally initiate contacts with PAM users to invite questions and identify problems.

Our findings about the types of problems observed in this study were consistent with previous research that categorized reasons for abandonment as problems related to the PAM, user preferences, or the PAM app [16,32]. The most common problem observed was wearing the out-of-the-box PAM wristband, whose rubber material and clasp were difficult to use, unappealing to some, or broke, which underscores the importance of helping people customize their wearable devices for comfortable everyday use [15,33]. Problems identified by participants, such as reading the PAM display, logging into the PAM app, and synchronizing the PAM to the app, have been identified in prior literature as common usability issues that prevent access to personal physical activity data [16,34]. Our findings support prior recommendations for future design and optimization of wearable technology for older adults [35-37] and demonstrate that many of the problems we observed were solvable with technical assistance.

Lastly, our finding that multiple modes were used to provide PAM-related technical assistance suggests that flexibility is essential. Many technical assistance contacts via phone, email, or postal mail successfully solved the reported problems. Nevertheless, some PAM problems were solved only after in-person interactions within participants' homes or nearby community centers, according to their preference. This finding raises questions about whether additional assistance strategies, such as short GIF videos and video calls, might efficiently augment strategies to provide technical assistance, while also reducing the need for some in-person contacts.

Limitations

This study has several limitations. One is the possibility that we underestimated the frequency and rate of unique technical assistance contacts due to the design of our technical assistance structures in RS3 and our data collection approach. In RS3, we focused on responding to participants' reported PAM problems and requests for assistance. It is possible that a more proactive approach (eg, occasionally reaching out to participants) would uncover additional PAM problems or gradually mitigate problems over time, thereby changing our longitudinal observations. Our data collection procedures focused on interactions between participants and research staff. It is possible that some participants reported PAM problems or requested technical assistance from other people, such as friends, neighbors, family, clinicians, or Fitbit Inc. A second limitation was that we did not collect data about wearing behavior [38] outside the assessment periods. Thus, we cannot analyze the extent to which technical assistance was associated with longitudinal daily PAM-wearing behaviors, and we recommend future studies investigate this. A third limitation is that the RS3 study included only one model of PAMs. As the design of PAMs changes over time and varies between different makes and models, the frequencies and types of technical assistance required by users may also shift.

Conclusions

Encouraging older adults to use PAMs can augment efforts to promote increased physical activity, but adequate technical assistance is needed. Findings from this study show that after receiving orientation to PAMs, problems such as uncomfortable wristbands, difficulty using the PAM or its related app, and obtaining or interpreting relevant personal data are reported. With occasional technical assistance, these problems are solvable over the phone, by email, or by postal mail. However, depending on the problem and an older person's preferences, some PAM problems are best solved in person. The detailed description of PAM problems and contacts provided in this study can inform plans for the quantity, types, and modes of technical assistance required from supportive structures that facilitate the long-term use of commercially available PAMs among older adults. Our results underscore the importance of trying to prevent technical problems by involving older adults in the design of PAMs, from initial design to usability testing and the development of supportive materials and structures for initial and ongoing use. Clinical programs and researchers should further assess the technical assistance needed for PAM problems experienced by older adults, accounting for variations by PAM model and wear time. Future research should also investigate additional assistance strategies, such as proactive support, short GIF videos, and video calls.

Acknowledgments

This work was supported by the National Institutes of Health, National Institutes of Nursing Research (grant 5R01NR016705-05). The authors would like to thank the participants in this study and our community partners.

Conflicts of Interest

None declared.

References

1. Piercy KL, Troiano RP, Ballard RM, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Elgaddal N, Kramarow EA, Reuben C. Physical activity among adults aged 18 and over: United States, 2020. NCHS Data Brief 2022 Aug(443):1-8. [Medline: 36043905]
3. Brickwood KJ, Ahuja KDK, Watson G, O'Brien JA, Williams AD. Effects of activity tracker use with health professional support or telephone counseling on maintenance of physical activity and health outcomes in older adults: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 5;9(1):e18686. [doi: 10.2196/18686] [Medline: 33399541]
4. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
5. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7. [doi: 10.2196/mhealth.4225] [Medline: 26818775]
6. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth 2016 Apr 13;4(2):e35. [doi: 10.2196/mhealth.5120] [Medline: 27076486]
7. Zytnick D, Folta SC, Reid KF, Chomitz VR. Better understanding wearable activity monitor use and non-use among older adults: a qualitative study. J Appl Gerontol 2023 Mar;42(3):447-455. [doi: 10.1177/07334648221137057] [Medline: 36314597]
8. Lee H, Lee Y. A look at wearable abandonment. Presented at: 2017 18th IEEE International Conference on Mobile Data Management (MDM); May 29 to Jun 1, 2017; Daejeon, Korea p. 392-393. [doi: 10.1109/MDM.2017.70]
9. Kononova A, Li L, Kamp K, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e9832. [doi: 10.2196/mhealth.9832] [Medline: 30950807]
10. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 3;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
11. Shin G, Jarrahi MH, Fei Y, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
12. Lazar A, Koehler C, Tanenbaum TJ, Nguyen DH. Why we use and abandon smart devices. Presented at: UbiComp '15: the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan p. 635-646. [doi: 10.1145/2750858.2804288]
13. Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L. No longer wearing: investigating the abandonment of personal health-tracking technologies on Craigslist. Presented at: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan p. 647-658. [doi: 10.1145/2750858.2807554]
14. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164. [doi: 10.2196/mhealth.7311] [Medline: 29084709]
15. Balbim GM, Marques IG, Marquez DX, et al. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. JMIR Mhealth Uhealth 2021 Mar 1;9(3):e25289. [doi: 10.2196/25289] [Medline: 33646135]
16. Steinert A, Haesner M, Steinhagen-Thiessen E. Activity-tracking devices for older adults: comparison and preferences. Univ Access Inf Soc 2018 Jun;17(2):411-419. [doi: 10.1007/s10209-017-0539-7]
17. McMahon SK, Lewis BA, Guan W, Wyman JF, Rothman AJ. Community-based intervention effects on older adults' physical activity and falls: protocol and rationale for a randomized optimization trial. Contemp Clin Trials 2021 Feb;101:106238. [doi: 10.1016/j.cct.2020.106238] [Medline: 33285280]
18. Schorr EN, Gepner AD, Dolansky MA, et al. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103. [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]

19. Zytnick D, Kumar GS, Folta SC, Reid KF, Tybor D, Chomitz VR. Wearable activity monitor use is associated with the aerobic physical activity guidelines and walking among older adults. Am J Health Promot 2021 Jun;35(5):679-687. [doi: 10.1177/0890117120985834] [Medline: 33406889]
20. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
21. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954. [doi: 10.2196/23954] [Medline: 33044175]
22. Ledger D, McCaffrey D. Inside wearables: how the science of human behavior change offers the secret to long-term engagement. Medium. URL: medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2023-11-09]
23. Lee BC, Xie J, Ajisafe T, Kim SH. How are wearable activity trackers adopted in older adults? Comparison between subjective adoption attitudes and physical activity performance. Int J Environ Res Public Health 2020 May 15;17(10):3461. [doi: 10.3390/ijerph17103461] [Medline: 32429258]
24. Stevens JA, Phelan EA. Development of STEADI: a fall prevention resource for health care providers. Health Promot Pract 2013 Sep;14(5):706-714. [doi: 10.1177/1524839912463576] [Medline: 23159993]
25. Topolski TD, LoGerfo J, Patrick DL, Williams B, Walwick J, Patrick MB. The rapid assessment of physical activity (RAPA) among older adults. Prev Chronic Dis 2006 Oct;3(4):A118. [Medline: 16978493]
26. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002 Sep;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]
27. McMahon SK, Wyman JF, Belyea MJ, Shearer N, Hekler EB, Fleury J. Combining motivational and physical intervention components to promote fall-reducing physical activity among community-dwelling older adults: a feasibility study. Am J Health Promot 2016 Nov;30(8):638-644. [doi: 10.4278/ajhp.130522-ARB-265] [Medline: 26389979]
28. McMahon SK, Lewis B, Oakes JM, Wyman JF, Guan W, Rothman AJ. Assessing the effects of interpersonal and intrapersonal behavior change strategies on physical activity in older adults: a factorial experiment. Ann Behav Med 2017 Jun;51(3):376-390. [doi: 10.1007/s12160-016-9863-z] [Medline: 28188585]
29. Ha Dinh TT, Bonner A, Clark R, Ramsbotham J, Hines S. The effectiveness of the teach-back method on adherence and self-management in health education for people with chronic disease: a systematic review. JBI Database System Rev Implement Rep 2016 Jan;14(1):210-247. [doi: 10.11124/jbisrir-2016-2296] [Medline: 26878928]
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
31. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76. [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
32. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Activity tracking: barriers, workarounds and customisation. Presented at: UbiComp ’15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7-11, 2015; Osaka, Japan p. 617-621. [doi: 10.1145/2750858.2805832]
33. Kang J, Binda J, Agarwal P, Saconi B, Choe EK. Fostering user engagement: improving sense of identity through cosmetic customization in wearable trackers. Presented at: PervasiveHealth ’17: The 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2015; Barcelona, Spain p. 11-20. [doi: 10.1145/3154862.3154878]
34. Ahmad N, Rextin A, Kulsoom UE. Perspectives on usability guidelines for smartphone applications: an empirical investigation and systematic literature review. Inf Softw Technol 2018 Feb;94:130-149. [doi: 10.1016/j.infsof.2017.10.005]
35. Khakurel J, Knutas A, Melkas H, Penzenstadler B, Porras J. Crafting usable quantified self-wearable technologies for older adult. Presented at: The AHFE 2018 International Conference on Human Factors and Wearable Technologies, and Human Factors in Game Design and Virtual Environments; Jul 21-25, 2018; Orlando, FL p. 75-87. [doi: 10.1007/978-3-319-94619-1]
36. Klebbe R, Steinert A, Müller-Werdan U. Wearables for older adults: requirements, design, and user experience. In: Perspectives on Wearable Enhanced Learning (WELL): Current Trends, Research, and Practice: Springer; 2019:313-332. [doi: 10.1007/978-3-319-64301-4]
37. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons’ physical activity in daily life: qualitative study of the users’ experiences. JMIR Mhealth Uhealth 2018 Feb 1;6(2):e34. [doi: 10.2196/mhealth.8345] [Medline: 29391342]
38. Jeong H, Kim H, Kim R, Lee U, Jeong Y. Smartwatch wearing behavior analysis: a longitudinal study. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-31. [doi: 10.1145/3131892]

Abbreviations:

PAM: physical activity monitor
REDCap: Research Electronic Data Capture
RS3: Ready Steady 3.0

Edited by L Buis; submitted 04.04.23; peer-reviewed by A Keogh, G Balbim; revised version received 11.08.23; accepted 22.09.23; published 22.11.23.

Please cite as:
Choma EA, Hayes S, Lewis BA, Rothman AJ, Wyman JF, Guan W, McMahon SK
Technical Assistance Received by Older Adults to Use Commercially Available Physical Activity Monitors (Ready Steady 3.0 Trial): Ad-Hoc Descriptive Longitudinal Study
JMIR Mhealth Uhealth 2023;11:e47891
URL: https://mhealth.jmir.org/2023/1/e47891
doi:10.2196/47891

Original Paper

Development and Validation of Multivariable Prediction Algorithms to Estimate Future Walking Behavior in Adults: Retrospective Cohort Study

Junghwan Park[1,2,3,4], BSc; Gregory J Norman[1,5], PhD; Predrag Klasnja[6], PhD; Daniel E Rivera[7], PhD; Eric Hekler[1,2,3], PhD

[1]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, United States

[2]Center for Wireless & Population Health Systems, Calit2's Qualcomm Institute, University of California, San Diego, La Jolla, CA, United States

[3]The Design Lab, University of California, San Diego, La Jolla, CA, United States

[4]Ministry of Health and Welfare, Korean National Government, Sejong, Republic of Korea

[5]Department of Global Access and Evidence, Dexcom Inc., San Diego, CA, United States

[6]School of Information, University of Michigan, Ann Arbor, MI, United States

[7]Control Systems Engineering Laboratory, School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ, United States

Corresponding Author:
Junghwan Park, BSc
Herbert Wertheim School of Public Health and Human Longevity Science
University of California, San Diego
9500 Gilman Dr
La Jolla, CA, 92093
United States
Phone: 1 858 429 9370
Email: jup014@ucsd.edu

Abstract

Background: Physical inactivity is associated with numerous health risks, including cancer, cardiovascular disease, type 2 diabetes, increased health care expenditure, and preventable, premature deaths. The majority of Americans fall short of clinical guideline goals (ie, 8000-10,000 steps per day). Behavior prediction algorithms could enable efficacious interventions to promote physical activity by facilitating delivery of nudges at appropriate times.

Objective: The aim of this paper is to develop and validate algorithms that predict walking (ie, >5 min) within the next 3 hours, predicted from the participants' previous 5 weeks' steps-per-minute data.

Methods: We conducted a retrospective, closed cohort, secondary analysis of a 6-week microrandomized trial of the *HeartSteps* mobile health physical-activity intervention conducted in 2015. The prediction performance of 6 algorithms was evaluated, as follows: logistic regression, radial-basis function support vector machine, eXtreme Gradient Boosting (XGBoost), multilayered perceptron (MLP), decision tree, and random forest. For the MLP, 90 random layer architectures were tested for optimization. Prior 5-week hourly walking data, including missingness, were used for predictors. Whether the participant walked during the next 3 hours was used as the outcome. K-fold cross-validation (K=10) was used for the internal validation. The primary outcome measures are classification accuracy, the Mathew correlation coefficient, sensitivity, and specificity.

Results: The total sample size included 6 weeks of data among 44 participants. Of the 44 participants, 31 (71%) were female, 26 (59%) were White, 36 (82%) had a college degree or more, and 15 (34%) were married. The mean age was 35.9 (SD 14.7) years. Participants (n=3, 7%) who did not have enough data (number of days <10) were excluded, resulting in 41 (93%) participants. MLP with optimized layer architecture showed the best performance in accuracy (82.0%, SD 1.1), whereas XGBoost (76.3%, SD 1.5), random forest (69.5%, SD 1.0), support vector machine (69.3%, SD 1.0), and decision tree (63.6%, SD 1.5) algorithms showed lower performance than logistic regression (77.2%, SD 1.2). MLP also showed superior overall performance to all other tried algorithms in Mathew correlation coefficient (0.643, SD 0.021), sensitivity (86.1%, SD 3.0), and specificity (77.8%, SD 3.3).

Conclusions: Walking behavior prediction models were developed and validated. MLP showed the highest overall performance of all attempted algorithms. A random search for optimal layer structure is a promising approach for prediction engine development. Future studies can test the real-world application of this algorithm in a "smart" intervention for promoting physical activity.

(JMIR Mhealth Uhealth 2023;11:e44296) doi:10.2196/44296

KEYWORDS

mobile health; mHealth; physical activity; walk; prediction; classification; multilayered perceptron; microrandomized trial; MRT; just-in-time adaptive intervention; JITAI; prevention; female; development; validation; application

Introduction

Physical inactivity is associated with numerous chronic diseases, including cancer, cardiovascular disease, type 2 diabetes [1-3], increased health care expenditure [4], and preventable, premature deaths [4]. Insufficient physical activity (PA) cost $53.8 billion worldwide in 2013. Clinical guidelines indicate 8000-10,000 steps per day [5]; nevertheless, the majority of Americans fall short of this goal [6].

In order to increase the level of PA, more than 300 commercial mobile apps have been developed [7]. The recent development of information technologies enabled mobile apps to deliver behavior change support when the users need this the most or when the utility (eg, how much the amount of PA was increased by the in-app notification) is predicted to be high. This new, promising type of intervention is called a just-in-time adaptive intervention (JITAI) [8].

JITAIs are not widely used (eg, 2.2% in 2018 [7]) by commercially available apps. However, it has been shown that JITAIs have the capacity to improve adherence and efficacy [9-11]. In addition, health behavior theories that commonly work as theoretical foundations for JITAIs [9], including social cognitive theory [12] and goal setting theory [13], emphasize the importance of timely feedback and anticipatory intervention [12,14-16]. Adaptation to individual, time-varying needs is theorized to be an effective strategy [14] for implementing time-accurate feedback and anticipatory intervention [16]. Since the opportunity window to intervene depends on the individual's environment, a fully automatic, predictive algorithm that can be run repeatedly is one of the key components of JITAI apps [14]. Thus, developing accurate algorithms to empower JITAIs to promote PA is a central task in overall JITAI development.

Prior JITAI studies used pure randomizations [17], condition-triggered Boolean logic [18,19], a combination of manually designed logics [20], or models that reveal the mathematical relationships between input factors and the behavior (eg, system identification [21]) so that researchers could understand which factors are predictive of the behavior. In this study, the models were evaluated mainly focusing on predictive accuracy rather than explainability [22]. Time series data of walking behavior (ie, steps per minute) measured by a wearable sensor was used to predict future walking behavior. Multiple algorithms were compared using various metrics, including accuracy, Mathew correlation coefficient (MCC), sensitivity, and specificity. If these algorithms can be produced, it would be a critical step toward JITAIs that are cost-efficient and fully autonomous (ie, without human couch interventions), and thus, it could be a valuable part of overall approaches for improving population health. To ensure the model's cost-efficiency and real-time usage feasibility, the training computation time was measured in the standardized computing environment.

Methods

Source of Data

This study used the deidentified Jawbone walking data (ie, steps per minute) from the *HeartSteps* study [23], conducted in the United States from August 2015 to January 2016.

Ethical Considerations

The original study [23] was approved by the University of Michigan Social and Behavioral Sciences Institutional Review Board (HUM00092845) for data collection. As the data in this study were deidentified prior to being provided, the study was deemed as nonhuman subject research by the University of California, San Diego Institutional Review Board. This study adhered to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) statement on reporting development and validation of the multivariate predictive models [24] (Multimedia Appendix 1).

Study Design and Data Processing Protocol

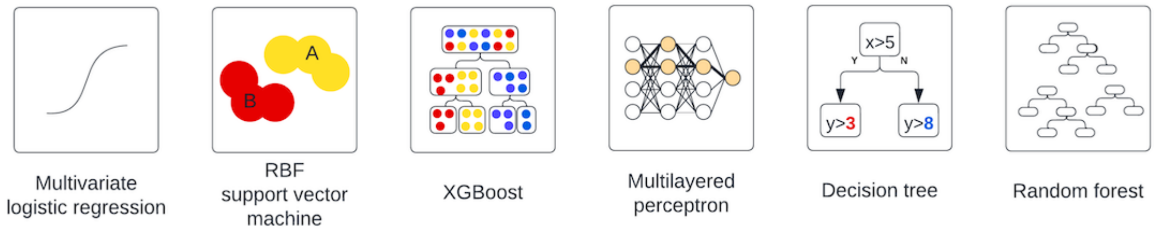
Exclusion and Data Transformation

Minute-by-minute walking data (ie, number of steps per minute) were preprocessed in the following three steps: (1) excluded the participants who have the data of less than 10 days, (2) excluded the data if the participant was inactive (ie, 0 step per minute) or partially active (ie, less than 60 steps per minute) during the minute, and (3) excluded short walks lasted less than 5 minutes. Then, walk data were used to decide whether the participant was active or not during the hour. If there was one or more walks (ie, more than 5 consecutive walking minutes) during the hour, it was marked as an "active hour." Then, the data were transformed to fit the machine learning algorithms (ie, from the time-series DataFrame objects of *Pandas* library to numerical array objects containing vector objects of *NumPy* library).

Training of Machine Learning Algorithms

The hourly walk data of the 5 prior weeks were used to predict the outcome (ie, whether the participant will walk or not during the next 3 hours). The following 6 sets of algorithms were used: logistic regression, radial basis function support vector machine [25], XGBoost [26], multilayered perceptron [27], decision tree, and random forest [28] (Figure 1). We used the implementation of the open-source projects named "scikit-learn" [29], Keras [30], XGBoost [26,31], and "Sci-Keras" [32] for each algorithm.

Figure 1. Brief algorithm descriptions of classification models. RBF: radial basis function.



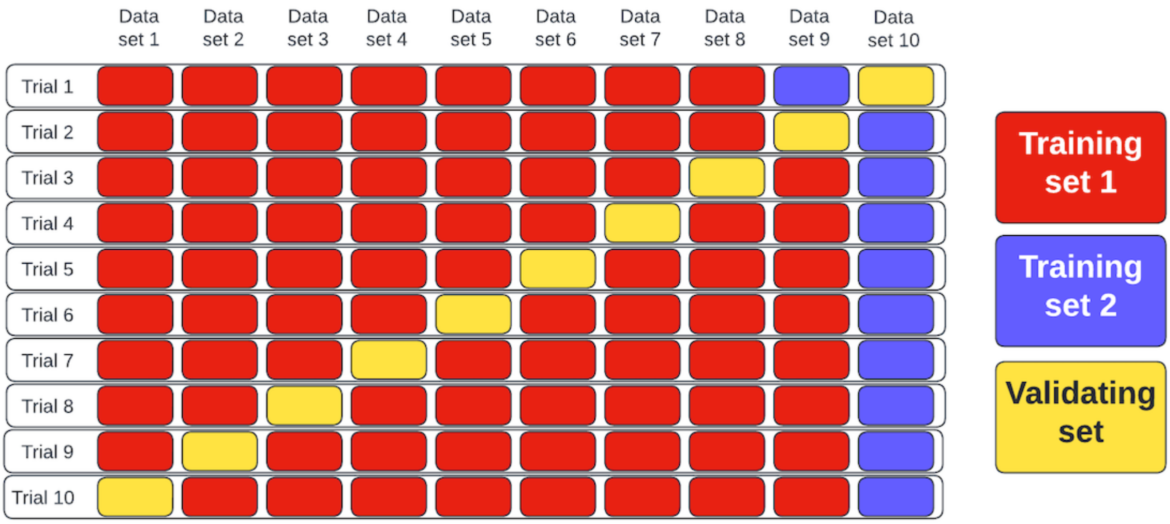
Target Imbalance

Due to sleeping hours and sedentary hours, nonactive hours usually outnumbered active hours. In machine learning algorithms, the phenomena are called "target imbalance" [33,34]. They usually critically reduce the performance of the prediction algorithm. Thus, in this study, we randomly sampled the nonactive hours to attain the same number as that of active hours.

K-fold Validation

After balancing the targets, the data were shuffled to perform K-fold validation [35] (Figure 2). We used K=10 in this study. We divide the shuffled data into 10 parts. Then, 1 part was separated to reduce the risk of overfitting the training data, and 1 part was separated for performance evaluation. In total, 8 out of 10 parts were used for machine learning algorithm training [35]. The process is iterated for 10 times, traversing each part for validation. The method allows us to internally validate the performance of the prediction engine. K (=10) sets of results were compared across the algorithms.

Figure 2. Brief description of K-fold validation method (eg, K=10).



Outcomes

Hourly data were generated during the preprocessing step. For the outcome variable, the activity data for 3 hours were merged. If the participant walked during the 3 hours, the outcome was assigned as "walked."

Predictor Variables

In addition to 5 weeks' hourly walking data, the variables noting the current date and time were used as predictors (Textbox 1). Each variable was encoded by the "One-hot-encoding" method [36]. It was a commonly used method to represent categorical (including ordinal or finite scale) variables in machine learning. The method converts the categorical variables (ie, N possible options) into an N-dimensional vector. Integers such as a current hour or current month were also converted into vectors. Each element of the vector can be ones or zeros. Each position in the vector denotes a particular value of options, and if a certain position was 1, the original value was mapped correspondingly. In a single vector, only one "1" was allowed. Since the encoding method enables the machine learning algorithm to train fast, it was commonly used. The discussion on the impact of the method on prediction performance was inconclusive [36].

Textbox 1. Variables used in classification algorithms.

Predictor variables

- Current hour (24 dichotomous variables, one-hot-encoded)
- Today's day of the week (7 dichotomous variables, one-hot-encoded)
- Current month (12 dichotomous variables, one-hot-encoded)
- Current day of the month (31 dichotomous variables, one-hot-encoded)
- Five Weeks' hourly walking (Yes/No/Missing, 3 dichotomous variables, one-hot-encoded)

Outcome variable

- Whether the individual will walk during the next 3 hours (Yes/No, 1 dichotomous variable)

Random Search for Multilayered Perceptron Model Structure

Unlike other algorithms in this study, the multilayered perceptron (MLP) algorithm uses layer architectures as one of the critical performance factors. Optimization techniques such as evolutionary programming [37] or random search or grid search [38] may be used. A random search was used to minimize the implementation burden while not losing too much performance (Figure 3).

Figure 3. Pseudocode for searching optimal model structure.

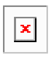
```
K = 10, MAX_LAYER = 10, MIN_N = 10, MAX_N = 1000

db = initialize_db()
For k = 1 to K:                          # experiment K times
For n = 1 to MAX_LAYER:                  # increase number of layers
   model = initialize_model()            # initialize the model
   For i = 1 to n:                       # for each layer
      n_neuron = random(MIN_N, MAX_N)    # decide number of neurons
      model.add_layer(n_neuron)          # add a layer
   model.train(train_data)               # train the model
   metric = model.test(test_data)        # measure the performance
   db.insert(model, metric)              # save the performance metric
```

Validation of the Models

The internal validation was performed by the K-fold validation methods. We used K=10. Individual test results were used to calculate the performance metrics such as accuracy, specificity, sensitivity, or MCCs. Data separation for the K-fold validation was conducted beforehand, which allows us to compare the metrics across the algorithms.

Mathew Correlation Coefficient

MCC [39] was defined as follows:



Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

MCC was sometimes used as an optimization metric. In this study, we measured MCCs as a performance metric, not the optimization metric. Since we have balanced the output (see the Target Imbalance section), accuracy was used as the optimization metric.

Computation Time

To conduct fair comparisons for the computation time, each model was trained in an isolated, standardized computing environment so that the system clock could measure the time elapsed. The system was reset every time a single execution was completed to minimize the fallout of the previous execution to the upcoming execution. Elapsed times were averaged and analyzed per algorithm.

Results

Study Population and Baseline Characteristics

A total of 41 (93%) out of 44 participants were included in the analysis [23]. The population's average age was 35.9 years. Of the 44 study participants, 31 (71%) were female, 26 (59%) were White, and 13 (30%) were Asian, with 36 (82%) having college degree or more. Moreover, 27% (n=12) of the participants had used a fitness app or activity tracker (Table 1).

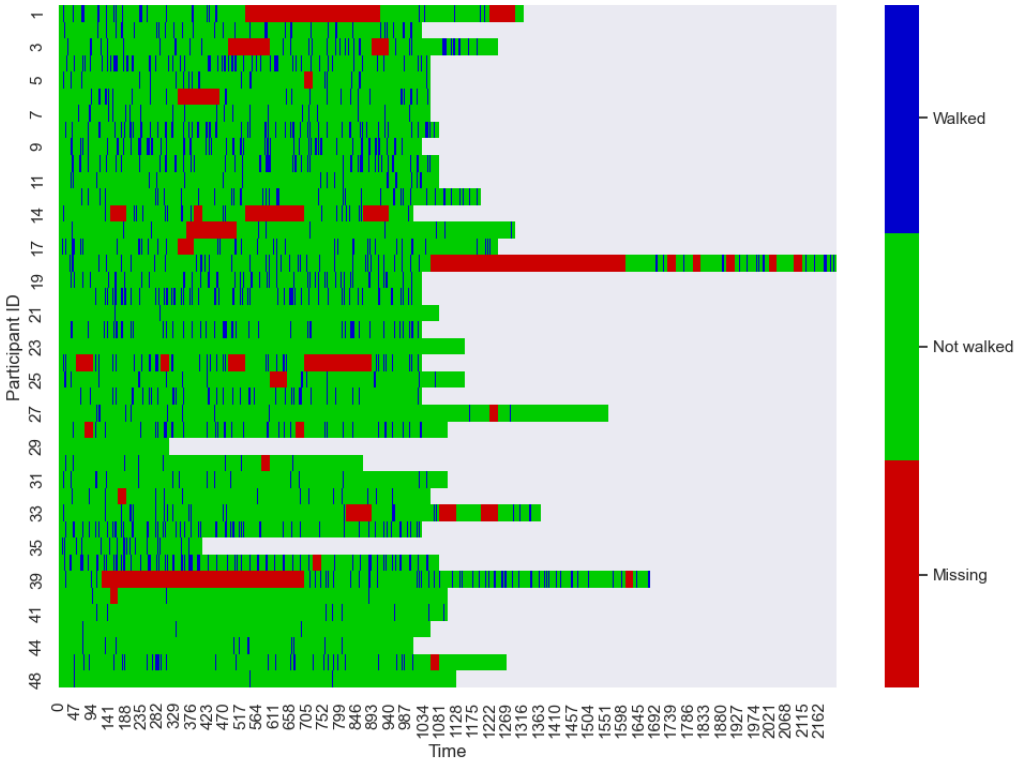
Table 1. Baseline characteristics of participants at study entry.

Variable	Value
Gender, n (%)	
Female	31 (71)
Male	13 (30)
Race, n (%)	
White	26 (59)
Asian	13 (30)
Black or African American	2 (5)
Other	3 (7)
Education, n (%)	
Some college	8 (18)
College degree	13 (30)
Some graduate school or graduate degree	23 (52)
Married or in a domestic partnership, n (%)	15 (34)
Have children, n (%)	16 (36)
Used fitness app before HeartSteps, n (%)	12 (27)
Used activity tracker before HeartSteps, n (%)	10 (22)
Phone used for study app, n (%)	
Used personal phone	21 (48)
Used study-provided phone	23 (52)
Age (years), mean (SD)	35.9 (14.7)

Data Summary for Predictor and Outcome Variables

On average, participants had available walking data for 43.3 (SD 9.1) days and 145.7 (SD 44.6) minutes per day. The average number of walking minutes per participant per day was reduced to 53.3 (SD 26.1) minutes after filtering with the threshold of 60 steps per minute (Methods section). Participants had 2.6 (SD 1.7) walks (ie, 5 or more consecutive walking minutes) every day (Methods section). Average length of each walk was 10.3 (SD 8.0) minutes. In hourly view, the participants had 0.6 (SD 0.1) "walking hours" (ie, the hours in which the participant walked) per day (Figure 4). Missing data were also used as a predictor state (Methods section). There were 18.1 (SD 13.4) missed days on average per participant, equivalent to 36.9% (SD 26.3%) of total days per participant. In the matter of outcome variable, as training and validating data set, 8129 "walking hours" and 37,711 "non-walking hours" (eg, nighttime or sedentary hours) were prepared (Methods section). Across the data, 17.7% of the time included participant activity. Thus, inactive time is 4.64 times more common than active time. The target imbalance was handled by undersampling (Methods section).

Figure 4. Overall distribution of walking data (1 narrow cell=1 hour).



Development of Prediction Algorithms

The calculation time vastly varied (Table 2). The radial basis function support vector machine algorithm and multilayered perceptron algorithm took the longest period to run. Tree-based algorithms such as decision tree and random forests were shorter than others. Random search to discover the optimal layer structure was tried. The optimization process improved the accuracy of the MLP algorithms from 49.8% to 82.1%. The process also improved all other metrics (Figure 5).

Table 2. Performance metrics of tried algorithms.

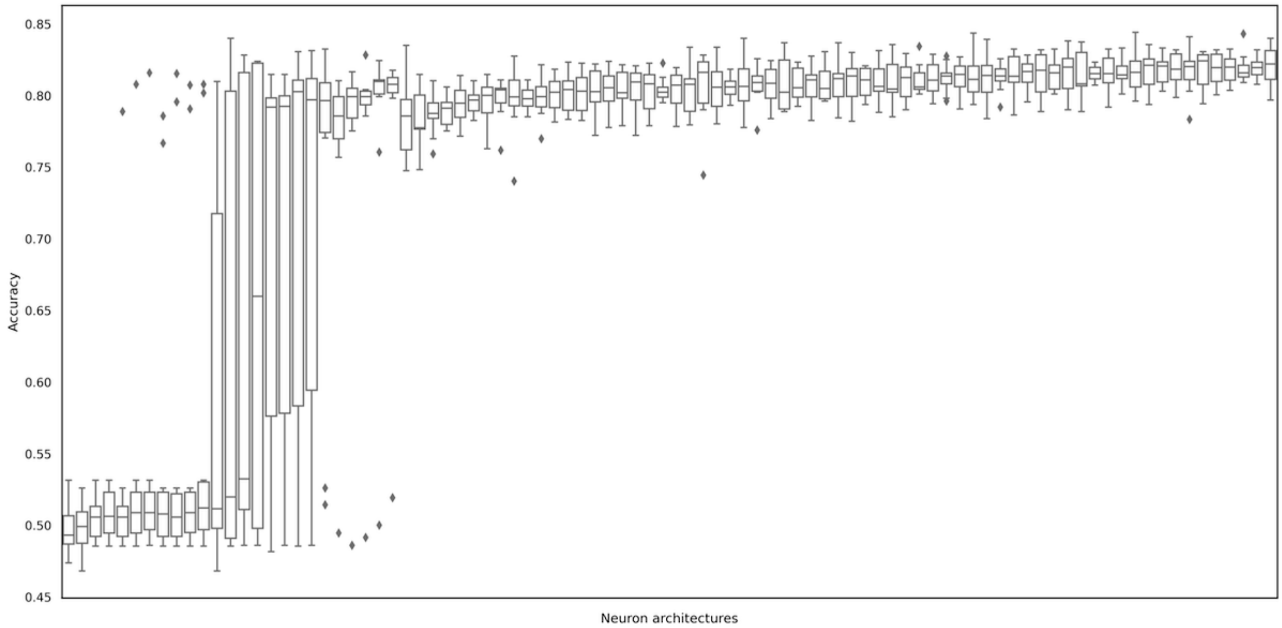
Algorithms	Accuracy, mean (SD)	MCC[a], mean (SD)	Sensitivity, mean (SD)	Specificity, mean (SD)
Logistic regression	0.772 (0.012)	0.545 (0.024)	0.795 (0.015)	0.749 (0.023)
RBF[b] SVM[c]	0.693 (0.010)	0.389 (0.020)	0.746 (0.022)	0.641 (0.017)
XGBoost	0.763 (0.015)	0.530 (0.030)	0.816 (0.010)	0.711 (0.030)
Multilayered perceptron	0.820 (0.011)	0.643 (0.021)	0.861 (0.030)	0.778 (0.033)
Decision tree	0.636 (0.015)	0.281 (0.026)	0.509 (0.075)	0.762 (0.049)
Random forest	0.695 (0.010)	0.396 (0.023)	0.776 (0.019)	0.614 (0.018)

[a]MCC: Mathew correlation coefficient.

[b]RBF: radial basis function.

[c]SVM: support vector machine.

Figure 5. Performance of tried neuron architectures (90 trials).



Validation and Model Performance

The reference algorithm (logistic regression) showed 77.2% (SD 1.2%) accuracy. XGBoost showed 76.3% (SD 1.5%), radial basis function support vector machine showed 69.3% (SD 1.0%), decision tree showed 63.6% (SD 1.5%), and random forest showed 69.5% (SD 1.0%), respectively. MLP performance largely varied from 49.8% (SD 1.7%) to 82.1% (SD 1.3%). Only 3 MLP architectures with the highest accuracies were included (Tables 2 and 3; Figure 6). Sensitivities, specificities, and MCC showed similar patterns to the accuracies. The decision tree algorithm generally showed the lowest performance overall, except on the dimension of specificity. MLP showed the highest performance across metrics (82.0% accuracy, 86.1% sensitivity, and 77.8% specificity).

Table 3. Average confusion matrix of each model of K-fold validation for the validation data set.

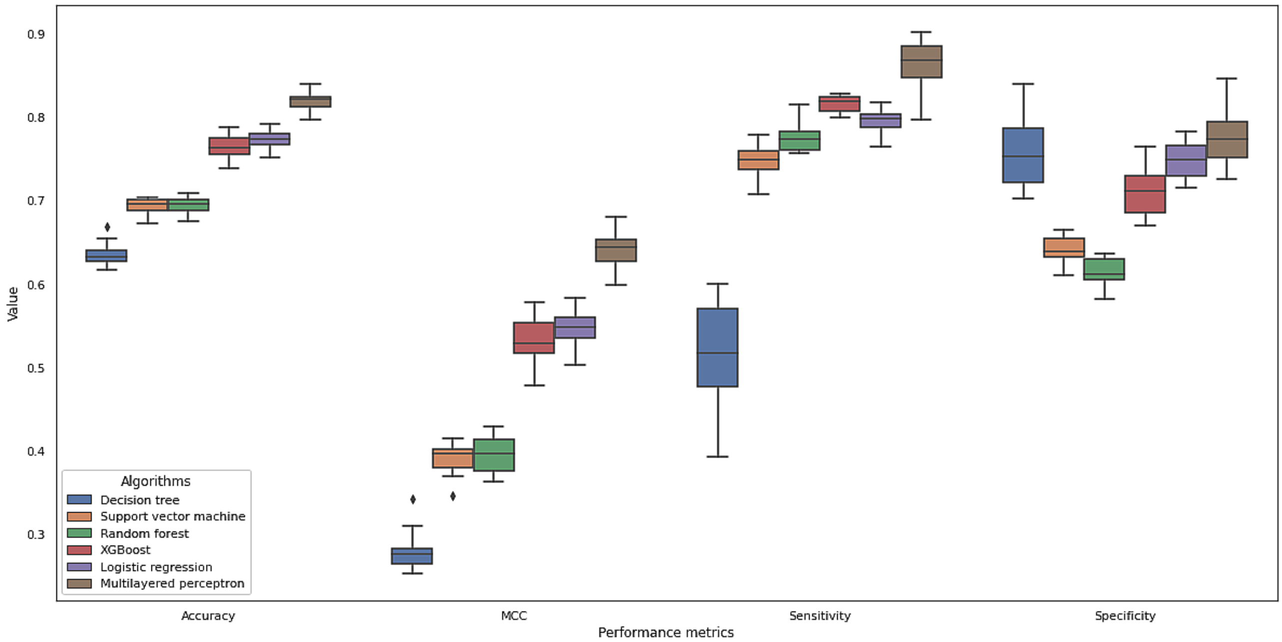
	True positive, mean (SD)	True negative, mean (SD)	False positive, mean (SD)	False negative, mean (SD)
Logistic regression	646.3 (27.3)	609.0 (30.6)	203.5 (18.8)	166.2 (11.7)
RBF[a] SVM[b]	606.3 (25.4)	520.3 (18.3)	292.2 (19.4)	206.2 (19.5)
XGBoost	663.0 (18.3)	577.6 (33.3)	234.9 (24.7)	149.5 (12.3)
MLP[c]	699.9 (35.2)	632.6 (34.7)	180.0 (27.5)	112.6 (24.2)
Decision tree	413.8 (65.4)	619.7 (52.5)	192.8 (39.1)	398.7 (56.5)
Random forest	630.3 (13.6)	499.0 (18.2)	313.5 (20.9)	182.2 (20.7)

[a]RBF: radial basis function.

[b]SVM: support vector machine.

[c]MLP: multilayered perceptron.

Figure 6. Performance metrics of the tried models. The top 3 architectures were chosen among multilayered perceptron engines. MCC: Mathew correlation coefficient.



Computation Time

In all the tested performance indicators, the optimized MLP showed the best performance and showed the second-longest training time of 225 seconds on average (Table 4). If we add up the total training time of all 90 optimization experiments, it took 56 hours. It was feasible to consistently evaluate training speed, accuracy, MCC, sensitivity, and specificity within the standardized performance evaluation framework. Through 90 random experiments, multiple MLP algorithms with optimized performance were obtained. The development, validation, and evaluation protocols can be used for similar prediction or classification problems.

Python 3.7.3, Sci-Kit Learn 1.0.2, Numpy 1.21.6, and Pandas 1.3.5, Tensorflow 2.8.0, xgboost 0.90, keras 2.8.0 were used.

In the matter of computation cost-efficiency (ie, predictive performance vs computation time), each algorithm showed characteristic results. The logistic regression had reasonable prediction performance and relatively low average computation time cost, whereas MLP showed generally higher prediction performance but had the second highest average computation cost (Figure 7).

It was feasible to consistently evaluate training speed, accuracy, MCC, sensitivity, and specificity within the standardized performance evaluation framework. Through 90 random experiments, multiple MLP algorithms with optimized performance were obtained. The development, validation, and evaluation protocols can be used for similar prediction or classification problems (Figure 8).

Table 4. Computation time to reach optimally trained status (seconds[a]).

Algorithms	Minimum	Maximum	Mean (SD)	CI
Logistic regression	20.73	24.89	22.37 (1.50)	19.43-25.31
RBF[b] SVM[c]	413.09	683.62	496.57 (94.58)	311.19-681.96
XGBoost	63.92	73.75	67.79 (4.33)	59.30-76.27
Multilayered perceptron	172.14	300.36	225.35 (38.83)	149.24-301.46
Decision tree	3.30	13.20	5.89 (2.68)	0.65-11.14
Random forest	4.32	13.42	6.63 (2.53)	1.68-11.57

[a]Computation was done in Google Colaboratory Pro+ (High-RAM mode with GPU hardware accelerator); 8 cores of Intel Xeon CPU 2.00 GHz, 53.4GB Memory, Tesla P100-PCIE-16GB.

[b]RBF: radial basis function.

[c]SVM: support vector machine.

Figure 7. The comparisons between algorithms in the matter of mean computation time and mean prediction accuracy. RBF: radial basis function; SVM: support vector machine.

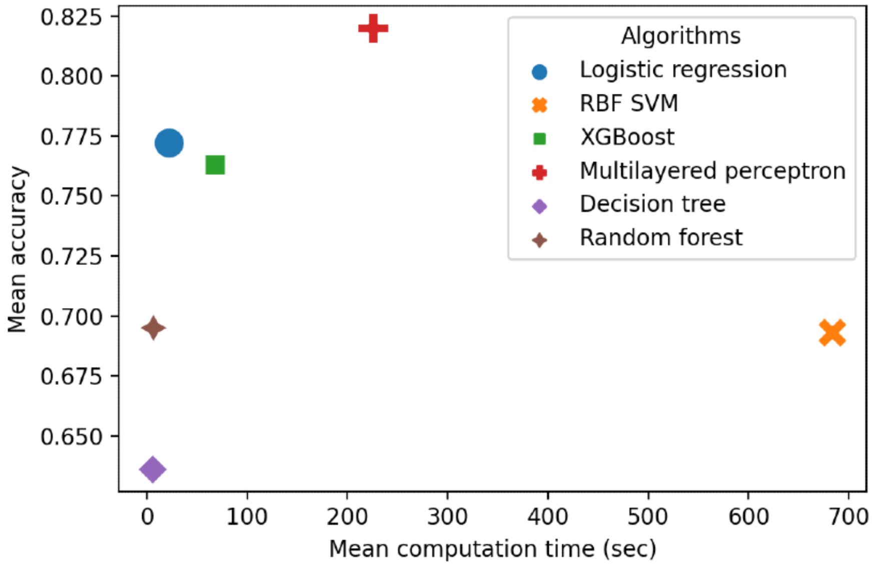
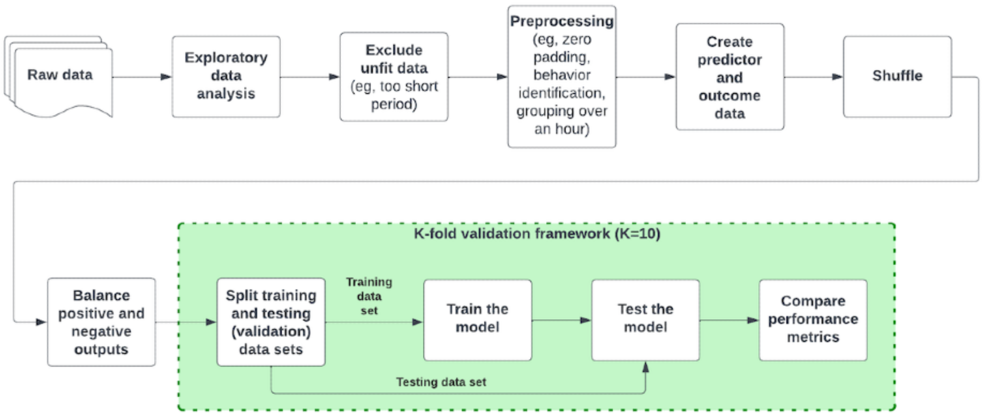


Figure 8. The data processing protocol.



Discussion

Key Implications

The high-level focus of our work is to develop approaches for using data from individuals themselves to create more individualized and adaptive support via digital technologies. In this paper, our goal was to test if predictive models could be generated that would be useful in terms of sensitive and specific probability estimates of the likelihood that someone will walk within an upcoming 3-hour window and that it could be done in a computationally efficient fashion. The latter part is important as computational efficiency is needed to enable the predictive models to be incorporated into future just-in-time adaptive interventions (JITAIs) that could use these predictive models to guide future decision-making. To support robust, automated decision-making within a JITAI to increase walking, our goal was to test if it would be feasible to produce predictive models that are informative for individuals in terms of identifying moments when a person has some chance of walking as opposed to either times when a person will clearly walk and thus does not need support, or times when there was near-zero probability that, in a given 3-hour window, a person will walk. If a predictive model could be produced that would provide this information, it would enable a JITAI that could incorporate these individualized predictions as a signal that could be used for making decisions on whether a given moment would be a *just-in-time* moment to provide a suggestion to go for a walk, with the predictive model used to predict the likelihood that, within the next 3 hours, the person would have the *opportunity* to walk while also having a need for a suggestion (ie, a person would not need a suggestion to walk if they are very likely to walk anyway). Our results, overall, suggest it is possible to generate said models in a scalable fashion, which could then be incorporated into a future JITAI that incorporates these individualized predictive models. Central to this work, the models produced here are definitionally idiographic in nature and thus appropriate for each individual. Thus, the results from the model should not be generalized to other samples. Instead, the key transportable knowledge from this work is the overall approach used for selecting models to guide individualized decision-making in future JITAIs (Figure 8).

Principal Findings

We developed 6 models (one of which was a group of models, and we chose the best 3 model architectures) for predicting future walking behavior within the subsequent 3-hour period using the previous 5 weeks' hourly walking data. MLP algorithm showed the best performance across all 4 metrics within this sample. A random search for MLP architecture produced an optimal model with the best performance. Using predictive engines to decide how to configure JITAIs could enable the mobile physical activity app to deliver more timely, appropriate intervention components such as in-app notifications. To the best of our knowledge, interventions that use predictive models to adjust to participant's behavior are still uncommon. Thus, our study makes a significant contribution by introducing the use of predictive algorithms for optimizing JITAIs.

Methodological Considerations and Comparison With Prior Work

In this study, we designed a protocol to develop and validate a predictive model for walking behavior. While developing the model, we had a few common issues that should be handled as follows.

Small Data Sets and the Potential Risk for Low External Validity

Despite the effort to validate the model with the K-fold cross-validation, since we are using a small number of short time-series data, high levels of external validity are not assumed. However, since the model we developed in this study did not assume any prior knowledge or variability (ie, nonparametric), additional training data are theorized to harness better performance. The model also did not use the pretrained coefficients; we used randomized coefficients. This leaves room for better performance and higher computation efficiency when we use the pretrained model from this study to extend the training. Publicly available lifestyle data, including the All-of-Us project [40] and the ones available on the public data platforms [41], will be a good way to extend the data set.

Target Imbalance

Target imbalance is defined as a significantly unequal distribution between the classes [33]. In numerous clinical [42,43] and behavioral [33] data modeling studies, target imbalance is a common issue. Although a few oversampling methodologies to tackle unbalanced output data have been developed [44], this study used an undersampling approach due to potential concerns of exaggerated accuracy [34]. The separate analysis with oversampling of the same data and methodologies showed 5%-10% increases in the accuracy. It is suspected that the underlying individual behavior patterns in the training samples are partly included in the test and validation samples.

Performance Metrics

Accuracy is the most commonly used performance metric to evaluate classification algorithms. However, the *accuracy* metric is also known to have the inability to distinguish between type 1 and type 2 errors [45]. The metrics of sensitivity and specificity are also commonly used to overcome the limitation of accuracy. The information represented by both metrics is partial (ie, both are addressing either type of error). MCC [46] is used more commonly in recent publications due to its statistical robustness against target imbalance, which is a common issue of clinical and behavioral data. Considering the imbalance of the classification problem of interest, we included MCC as a performance metric.

Limitations of This Study

The original study was designed for the purpose of pilot-testing and demonstrating the potential of microrandomized trials. Thus, these analyses are all secondary in nature. Further, the initial study was a small study, with only a minimum amount of data (n=41) used. Additionally, since the participants were recruited in a homogeneous environment and demographic groups, the external validity of the algorithms may be limited. With that said, the overall approach for formulating predictive models and their selection could feasibly be used in the future and, thus, it is more of our protocol and approach that is likely to be generalizable and generally useful for JITAIs compared to any specific insights from the models we ran. We contend that, for any targeted JITAI, a precondition for this type of approach is the appropriate data available, and that, for any JITAI, it is more valuable to build algorithms that match localized needs and contexts than seek to take insights from some previous samples that are different from a target population and assume they will readily translate. This, of course, can be done with careful tests of transportability using strategies such as directed acyclic graphs to guide the production of estimands [47] that would create formalized hypotheses of transportability. However, this is a much higher bar for transportability that, while valuable, can often be prohibitive for fostering progress in JITAIs. Within our proposed approach, the strategy involves gleaning *good enough* data to enable a localized prediction algorithm appropriate for the targeted population to be produced, with subsequent deployment factoring in strategies and approaches for updating and improving the algorithms as new insights emerge.

Implication and Future Work

The results of our study show that prediction algorithms can be used to predict future walking behavior in a fashion that can be incorporated into a future walking JITAI. In this study, we modeled without contextual information other than the date, time, or day of the week. However, if the machine learning algorithm is trained using the other contextual information such as intervention data (eg, whether the in-app notification message is sent or not, which type of message is sent, and which sentiment is used to draw attention), the prediction engine would be capable of simulating how the intervention components might change the behavior in the multiple hypothetical scenarios. This capability would enable us to use the prediction algorithms uniquely, that is, comparing two or more possible scenarios to decide the optimal intervention mode of a JITAI. We could decide whether to send a message, which message should be sent, or what sentiment we could use to draw attention to our intervention. A pragmatic study that assesses the efficacy of such an approach is necessary.

The search methods for the optimal architectures of MLP could be improved. Evolutionary programming [48] and weight-agnostic neural network [37] are promising approaches. Such improvement could find the MLP architectures' better performance in shorter computation time.

Conclusion

The protocol for developing and validating a prediction engine for health behavior was developed. As a case study, walking behavior classification models were developed and validated. MLP showed the highest overall performance of all tried algorithms, yet it needed relatively higher computation time. A random search for optimal layer structure was a promising approach for prediction engine development.

Acknowledgments

JP conceptualized the research question, analyzed the data, and wrote the manuscript. PK provided the data. DER provided the program code library to assist the analysis. EH provided guidance at each stage of study. All authors contributed to the writing of the manuscript. The National Library of Medicine (R01LM013107) funded JP's stipend.

Conflicts of Interest

JP is an employee of Korean National Government, the Ministry of Health and Welfare. GJN is an employee of Dexcom, Inc.

Multimedia Appendix 1

TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) checklist: prediction model development and validation.

[PDF File (Adobe PDF File), 56 KB - mhealth_v11i1e44296_app1.pdf]

References

1. Thune I, Furberg AS. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S530-50; discussion S609. [doi: 10.1097/00005768-200106001-00025] [Medline: 11427781]
2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. Haskell WL, Blair SN, Hill JO. Physical activity: health outcomes and importance for public health policy. Prev Med 2009 Oct;49(4):280-282. [doi: 10.1016/j.ypmed.2009.05.002] [Medline: 19463850]
4. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. The Lancet 2016 Sep;388(10051):1311-1324. [doi: 10.1016/s0140-6736(16)30383-x]
5. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
6. Blackwell D, Clarke T. State Variation in Meeting the 2008 Federal Guidelines for Both Aerobic and Muscle-strengthening Activities Through Leisure-time Physical Activity Among Adults Aged 18-64: United States, 2010-2015. Natl Health Stat Report 2018 Jun(112):1-22 [FREE Full text] [Medline: 30248007]
7. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of Physical Activity Apps: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2021 Jun 09;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]
8. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
9. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
10. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
11. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
12. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
13. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066X.57.9.705]
14. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

15. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995 May 04;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]
16. Folkman S, Moskowitz JT. Coping: pitfalls and promise. Annu Rev Psychol 2004;55:745-774. [doi: 10.1146/annurev.psych.55.090902.141456] [Medline: 14744233]
17. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
18. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572. [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
19. van Dantzig S, Geleijnse G, van Halteren AT. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246. [doi: 10.1007/s00779-012-0588-0]
20. Mair JL, Hayes LD, Campbell AK, Buchan DS, Easton C, Sculthorpe N. A Personalized Smartphone-Delivered Just-in-time Adaptive Intervention (JitaBug) to Increase Physical Activity in Older Adults: Mixed Methods Feasibility Study. JMIR Form Res 2022 Apr 07;6(4):e34662 [FREE Full text] [doi: 10.2196/34662] [Medline: 35389348]
21. Martin C, Deshpande S, Hekler E, Rivera D. A system identification approach for improving behavioral interventions based on Social Cognitive Theory. 2015 Presented at: American Control Conference (ACC); July 30, 2015; Chicago, IL, USA. [doi: 10.1109/acc.2015.7172261]
22. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. Hastings Cent Rep 2019 Jan 21;49(1):15-21. [doi: 10.1002/hast.973] [Medline: 30790315]
23. Klasnja P, Smith S, Seewald N, Lee A, Hall K, Luers B, et al. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med 2015 May 19;162(10):735-736. [doi: 10.7326/l15-5093-2]
25. Zongben X, Mingwei D, Deyu M. Fast and Efficient Strategies for Model Selection of Gaussian Support Vector Machine. IEEE Trans. Syst., Man, Cybern. B 2009 Oct;39(5):1292-1307. [doi: 10.1109/tsmcb.2009.2015672]
26. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
27. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. JAMA 2018 Sep 18;320(11):1101-1102. [doi: 10.1001/jama.2018.11100] [Medline: 30178065]
28. Ho TK. Random decision forests. 1995 Presented at: Proceedings of 3rd International Conference on Document Analysis and Recognition; August 14-16, 1995; Montreal, QC, Canada p. 278-282.
29. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830.
30. Keras. URL: https://keras.io [accessed 2023-01-12]
31. XGBoost. GitHub. URL: https://github.com/dmlc/xgboost [accessed 2023-01-12]
32. Scikeras. GitHub. URL: https://github.com/adriangb/scikeras [accessed 2023-01-12]
33. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications 2017 May;73:220-239. [doi: 10.1016/j.eswa.2016.12.035]
34. Yap B, Rani K, Rahman H, Fong S, Khairudin Z, Abdullah N. An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. 2013 Presented at: Proceedings of the First International Conference on Advanced Data and Information Engineering; April 8-12, 2013; Brisbane, QLD, Australia p. 13-22. [doi: 10.1007/978-981-4585-18-7_2]
35. Fushiki T. Estimation of prediction error by using K-fold cross-validation. Stat Comput 2009 Oct 10;21(2):137-146. [doi: 10.1007/s11222-009-9153-8]
36. Rodríguez P, Bautista MA, Gonzàlez J, Escalera S. Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing 2018 Jul;75:21-31. [doi: 10.1016/j.imavis.2018.04.004]
37. Gaier A, Ha D. Weight agnostic neural networks. Adv Neural Inf Process Syst 2019;32:1-19.
38. Liashchynskyi P, Liashchynskyi P. Grid search, random search, genetic algorithm: A big comparison for NAS. arXiv Prepr arXiv191206059. Published online 2019:1-11.
39. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS One 2017 Jun 2;12(6):e0177678 [FREE Full text] [doi: 10.1371/journal.pone.0177678] [Medline: 28574989]
40. All of Us Research Program Investigators. The “All of Us” Research Program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/nejmsr1809937]
41. Derry JMJ, Mangravite LM, Suver C, Furia MD, Henderson D, Schildwachter X, et al. Developing predictive molecular maps of human disease through community-based modeling. Nat Genet 2012 Jan 27;44(2):127-130 [FREE Full text] [doi: 10.1038/ng.1089] [Medline: 22281773]

42. Krawczyk B, Schaefer G, Woźniak M. A hybrid cost-sensitive ensemble for imbalanced breast thermogram classification. Artif Intell Med 2015 Nov;65(3):219-227. [doi: 10.1016/j.artmed.2015.07.005] [Medline: 26319694]
43. Jiang J, Liu X, Zhang K, Long E, Wang L, Li W, et al. Automatic diagnosis of imbalanced ophthalmic images using a cost-sensitive deep convolutional neural network. Biomed Eng Online 2017 Nov 21;16(1):132 [FREE Full text] [doi: 10.1186/s12938-017-0420-1] [Medline: 29157240]
44. Quiroz JC, Feng Y, Cheng Z, Rezazadegan D, Chen P, Lin Q, et al. Development and Validation of a Machine Learning Approach for Automated Severity Assessment of COVID-19 Based on Clinical and Imaging Data: Retrospective Study. JMIR Med Inform 2021 Feb 11;9(2):e24572 [FREE Full text] [doi: 10.2196/24572] [Medline: 33534723]
45. Novakovic J, Veljovic A, Ilić S, Papic Ž, Milica T. Evaluation of classification models in machine learning. Theory Appl Math Comput Sci 2017;7(1):39-46.
46. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020 Jan 02;21(1):6 [FREE Full text] [doi: 10.1186/s12864-019-6413-7] [Medline: 31898477]
47. Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. Am Sociol Rev 2021 Jun 04;86(3):532-565. [doi: 10.1177/00031224211004187]
48. Stanley KO, Miikkulainen R. Evolving neural networks through augmenting topologies. Evol Comput 2002 Jun;10(2):99-127. [doi: 10.1162/106365602320169811] [Medline: 12180173]

Abbreviations

JITAI: just-in-time adaptive intervention
MCC: Mathew correlation coefficient
MLP: multilayered perceptron
PA: physical activity
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

Edited by L Buis; submitted 15.11.22; peer-reviewed by Y Wang, D Lewis; comments to author 21.12.22; revised version received 04.01.23; accepted 05.01.23; published 27.01.23.

Please cite as:
Park J, Norman GJ, Klasnja P, Rivera DE, Hekler E
Development and Validation of Multivariable Prediction Algorithms to Estimate Future Walking Behavior in Adults: Retrospective Cohort Study
JMIR Mhealth Uhealth 2023;11:e44296
URL: https://mhealth.jmir.org/2023/1/e44296
doi:10.2196/44296
PMID:36705954

Original Paper

Exploring the Feasibility and Usability of Smartphones for Monitoring Physical Activity in Orthopedic Patients: Prospective Observational Study

Arash Ghaffari[1], MD; Rikke Emilie Kildahl Lauritsen[1], MSc; Michael Christensen[2], MSc; Trine Rolighed Thomsen[3,4], PhD; Harshit Mahapatra[2], MSc; Robert Heck[3], PhD; Søren Kold[1], MD, PhD; Ole Rahbek[1], MD, PhD

[1]Interdisciplinary Orthopaedics, Aalborg University Hospital, Aalborg, Denmark
[2]Alexandra Institute, Aarhus, Denmark
[3]Danish Technological Institute, Aarhus, Denmark
[4]Department of Chemistry and Bioscience, Aalborg Uuniversity, Aalborg, Denmark

Corresponding Author:
Arash Ghaffari, MD
Interdisciplinary Orthopaedics
Aalborg University Hospital
Hobrovej 18 - 22
Aalborg, 9000
Denmark
Phone: 45 91483966
Email: a.ghaffari@rn.dk

Abstract

Background: Smartphones are often equipped with inertial sensors that measure individuals' physical activity (PA). However, their role in remote monitoring of the patients' PAs in telemedicine needs to be adequately explored.

Objective: This study aimed to explore the correlation between a participant's actual daily step counts and the daily step counts reported by their smartphone. In addition, we inquired about the usability of smartphones for collecting PA data.

Methods: This prospective observational study was conducted among patients undergoing lower limb orthopedic surgery and a group of nonpatients as control. The data from the patients were collected from 2 weeks before surgery until 4 weeks after the surgery, whereas the data collection period for the nonpatients was 2 weeks. The participant's daily step count was recorded by PA trackers worn 24/7. In addition, a smartphone app collected the number of daily steps registered by the participants' smartphones. We compared the cross-correlation between the daily steps time series obtained from the smartphones and PA trackers in different groups of participants. We also used mixed modeling to estimate the total number of steps, using smartphone step counts and the characteristics of the patients as independent variables. The System Usability Scale was used to evaluate the participants' experience with the smartphone app and the PA tracker.

Results: Overall, 1067 days of data were collected from 21 patients (n=11, 52% female patients) and 10 nonpatients (n=6, 60% female patients). The median cross-correlation coefficient on the same day was 0.70 (IQR 0.53-0.83). The correlation in the nonpatient group was slightly higher than that in the patient group (median 0.74, IQR 0.60-0.90 and median 0.69, IQR 0.52-0.81, respectively). The likelihood ratio tests on the models fitted by mixed effects methods demonstrated that the smartphone step count was positively correlated with the PA tracker's total number of steps (χ^2_1=34.7, *P*<.001). In addition, the median usability score for the smartphone app was 78 (IQR 73-88) compared with median 73 (IQR 68-80) for the PA tracker.

Conclusions: Considering the ubiquity, convenience, and practicality of smartphones, the high correlation between the smartphones and the total daily step count time series highlights the potential usefulness of smartphones in detecting changes in the number of steps in remote monitoring of a patient's PA.

(JMIR Mhealth Uhealth 2023;11:e44442) doi:10.2196/44442

KEYWORDS

remote monitoring; physical activity; step count; smartphone application; wearable sensors; mixed effects modeling; step count prediction; mobile phone

Introduction

Background

Daily physical activity (PA) is crucial for maintaining physical, mental, and social health [1]. For patients undergoing orthopedic surgery, resuming PA as soon as possible is vital to enhance recovery and prevent complications [2]. In addition, assessing PA after surgery can provide valuable information regarding a patient's health condition, allowing for individualized rehabilitation based on the patient's condition and demands [3-5]. However, some limitations and challenges exist regarding the measurement of PA. Current patient-reported outcome measures (PROMs) such as questionnaires and surveys might seem convenient for evaluating the level of PA; however, they have limitations such as low patient adherence, floor effects, and recall bias and are inefficient in measuring walking as an important PA [6]. In addition, PROMs are often obtained at specific and broad intervals. Therefore, the objective measurement of PA after discharge is of increasing interest [7].

Smartphones and other digital devices are currently equipped with sensors allowing the quantification of an object's motion by converting inertial forces into measurable electrical signals [8]. This makes them valuable tools for remotely monitoring patients' PA during recovery after surgery. Smartphones have also become increasingly prevalent across all age groups and are now ubiquitous [9]. For instance, in Denmark, 90% of the population has access to smartphones [10], making them a widespread technology with the potential for broad societal impact. Using smartphones in remote monitoring the patients also offers the possibility of applying supplementary PROMs. A recent study on patients undergoing hip replacement surgeries demonstrated the patients' interest in using smartphone apps and learning how to use wearable sensors [11]. Collecting activity data and PROMs with a smartphone for this group of patients has proven feasible [12].

Given the increasing prevalence of smartphones in the general population and their growing application in telemedical methods, these devices can play a prominent role in collecting objective PA data. However, their capability has not been fully explored, especially in free-living settings and over extended periods, such as follow-up after surgeries. In addition, some uncertainties have been discussed regarding the validity of the measurements, as the patients usually do not carry their smartphones all the time [13]. Specifically, changing daily life routines during and immediately after surgery may cause the patients not to carry their devices as usual. Accordingly, the amount and the significance of the nonmeasured activity in the perioperative periods are unknown.

Objectives

In this study, we explored the utility of smartphones in measuring daily PA compared with wearable sensors in orthopedic patients during the perioperative period. The PA trackers were used to record step counts during regular continuous walking, sporadic walking, and slow continuous walking. The primary objective of this study was to determine the correlation between the daily step counts obtained from smartphones and the step counts registered by the PA trackers during these different types of walking. In addition, we investigated the ability of smartphones to predict the total number of daily steps taken during each type of walking. The secondary objective was to evaluate the usability of a smartphone app designed to collect health data.

Methods

Study Design and Setting

This prospective observational study was conducted at the Aalborg University Hospital, Denmark, between November 2021 and August 2022. The project was registered at North Jutland Research Database in Denmark (2021-119).

Ethics Approval

This study was approved by the Regional Committee on Health Research Ethics (reference 2021-000438). This study complies with the Strengthening the Reporting of Observational Studies in Epidemiology guidelines [14].

Participants

Overview

We included 2 groups of participants in this study to compare the results of the patients undergoing orthopedic surgeries with those of a control group. First, all participants were informed about the study process and were asked to sign informed consent forms. Subsequently, the participants were instructed to install and use the smartphone app and the PA trackers and transfer the data.

Patients

Patients undergoing lower limb orthopedic surgery were eligible for inclusion if they were smartphone users. No limitation was placed regarding the participant's age or the type of surgery. However, older, frail patients who required a wheelchair for ambulation or who could not walk independently were not included.

Patients' data were collected from at least 2 weeks before surgery until 4 weeks after surgery.

Nonpatients

We also included volunteers without orthopedic problems as the control group. Data regarding the step counts for at least 14 consecutive days were collected in this group.

Data Sources and Measurements

Participants' Characteristics

The patients' basic and demographic information, including their age, sex, BMI, comorbidities (history of medical illness), and previous orthopedic surgery on the lower limbs, were registered in a REDCap (Research Electronic Data Capture; Vanderbilt University) database hosted by the North Jutland Region, Denmark [15].

PA Tracker

SENS sensors (SENS Motion) were used to record the patients' daily number of steps. SENS Motion is a wearable PA sensor worn as a patch on the lateral distal thigh and collects PA data by registering 3D linear acceleration data (Figure 1). Some studies have investigated the reliability and validity of the SENS PA trackers' measurements [16,17] and demonstrated favorable results. As the sensors were attached 24/7 to the patients, we considered their measurements as the total daily step counts. To ensure that the patients wore the sensors for the entire duration, we observed the sensors' relative temperature data in addition to the linear acceleration daily time series.

Figure 1. The photograph demonstrates the SENS Motion physical activity tracker at the lateral side of the distal thigh of 1 of the participants in the study.

The SENS Motion algorithm calculates the number of steps taken during sporadic and continuous walking, as well as training, in three different categories:

1. *Steps-1*: Summarized number of steps during continuous walking and training, based on analysis in the frequency domain.
2. *Steps-2*: Steps taken during sporadic and irregular walking where no continuous frequency can be recognized in the 5-second interval are summarized as 2 steps per 5-second interval.
3. *Steps-3*: Steps taken during slow walking where a continuous frequency can be recognized, but the intensity of the accelerations is lower than that in usual walking.

We calculated the *total PA tracker steps* as the sum of the 3 abovementioned variables.

Smartphone App (OrtoApp)

OrtoApp (Alexandra Institute) is a smartphone app developed to collect step counts and PA data from the Apple HealthKit application programming interface (API) on iOS and the Google Fit API on Android smartphones [18,19]. During the study, the app was installed on patients' smartphones and automatically recorded the steps registered by the Apple HealthKit and the Google Fit APIs. Furthermore, if a person also wears a smartwatch, the Apple HealthKit and Google Fit APIs will collect the data from both devices (the smartwatch and the smartphone), and the step counts will be calculated based on both inputs.

In addition, OrtoApp allows users to record their daily mood and pain levels on an 11-point visual analog scale (0-10). However, we did not use the data regarding the pain and mood scores in this study.

Usability of the Smartphone App and PA Tracker

We used the System Usability Scale (SUS) to evaluate participants' experience with the smartphone app and the PA tracker. The SUS is developed as a survey scale that allows quick and easy assessment of the usability of a given product or service [20]. The original SUS instrument comprises 10 statements scored on a 5-point scale of the strength of agreement. Final SUS scores can range from 0 to 100, with higher scores indicating better usability [21]. In this study, we used the translated and validated Danish version of the SUS [22].

After the data collection period was over, we assessed the usability only in the patient group by distributing the SUS questionnaire via the REDCap web application.

Steps Data Analysis

We generated 5 time series for each participant, including 1 for the daily steps recorded by the smartphones and 4 for the daily PA trackers' measurements (steps-1, steps-2, steps-3, and PA tracker total steps). These time series were then plotted for each participant, and we compared the smartphone data's time series with the different variables of the PA trackers using cross-correlation. Before conducting the cross-correlation analysis, we differentiated the time series data to remove any trends or changes in the mean that may have affected the results. This was done by calculating the difference between consecutive time points (days). Next, we calculated the cross-correlation between the resulting time series using a standard method [23]. We specifically calculated the cross-correlation at 0 days lag (ie, the same day) to assess the immediate relationship between the variables. We used Fisher Z transformation to calculate the 95% CI for the correlation coefficients and to compare the correlation coefficients [24]. The comparisons were performed between various groups based on different criteria, including patient or nonpatient status, preoperative or postoperative status (for patients), age (>60 years or <60 years), comorbidities, history of lower limb surgery, day of data collection (weekday—Monday through Friday—or weekend—Saturday and Sunday), content type of the smartphone used, and the use of a smartwatch.

In addition, we applied mixed effects models to investigate whether the smartphone's step counts could predict the total number of steps. Only the data from the patient group were used for mixed effects modeling. To prepare the data for regression analysis, we applied the moving average method to calculate the average values for the 3 preceding days (trailing moving average with a window of 3 days). In time series data analysis, the moving average method helps discover certain traits by smoothing the variations and reducing the noise [23]. Subsequently, we scaled the data to have a mean 0 and a SD equal to 1.

We used different subjects as random intercepts in the models and by-subject PA tracker–smartphone steps slope variance as random slopes. We included the following variables and all possible interaction effects between the variables to fit the models:

1. Smartphone steps:
   - Scaled 3-days moving average as a continuous variable
2. Participants' characteristics:
   - Age in years as a continuous variable
   - Sex as a categorical variable (male or female)
   - BMI in kg/m^2 as a continuous variable
   - Comorbidity as a categorical variable (yes or no)
   - History of lower limb surgery as a categorical variable (yes or no)
3. Characteristics of data collection day:
   - Preoperative versus postoperative as a categorical variable
4. Smartphone health app:
   - Apple HealthKit versus Google Fit as a categorical variable
5. Smartwatch:
   - Using a smartwatch as a categorical variable (yes or no).

The variables included in the best-performing models were selected by backward elimination, that is, if they did not improve the model, the variables were omitted.

Four models were created for the different variables from the PA tracker (steps-1, steps-2, steps-3, and PA tracker total steps). In the best-fitted models for the steps-2, steps-3, and PA tracker total steps, the selected variables were the period (preoperative or postoperative) and the presence of comorbidities in addition to smartphone steps. However, in the PA steps-1 model, the history of medical disease did not improve the model performance and hence was excluded.

The coefficients for the fixed and random effects variables in the best-fitted models and the performance metrics for the goodness of fit for the models (described in *Statistical Methods* section) were computed. The 95% prediction intervals for the models were created and plotted by bootstrapping techniques.

Statistical Methods

We used the R statistical package (version 4.1.0; R Foundation for Statistical Computing) for the statistical analyses and *lme4* package [25] for the mixed effects models.

Descriptive statistics were used to describe participants' basic information. The counts and percentages were used for the discrete variables, including the number and sex of the participants and the number of days for data collection. Means and SDs were used to describe the participants' age and BMI.

We presented the cross-correlation coefficients between the time series as means and 95% CIs. The SUS values for the smartphone app and PA trackers were provided as median and IQR.

Mixed effects models were created using the restricted maximum likelihood approach. The repeated measures and covariance matrix were modeled as unstructured. No violation of the model assumptions regarding the linearity, homoscedasticity, and normality of residuals was detected. The goodness of fit of the models was assessed by calculating the deviance, Akaike information criterion, Bayesian information criterion [26], intraclass correlation coefficient, and conditional and marginal pseudo-R^2 [27]. Marginal pseudo-R^2 represents the variance explained by the fixed effects, whereas conditional pseudo-R^2 is interpreted as a variance explained by the entire model, that is, both fixed and random effects. The scaled step counts were back transformed into actual values in the plots. We compared the best-fitted models with and without the smartphone step counts by using likelihood ratio tests to calculate P values. The significance level was set at α=.05.

Results

Participants' Characteristics

Overall, 35 participants were included in the study; however, 4 participants were excluded, and data of 31 participants (n=21, 68% patients and n=10, 32% nonpatients) were analyzed. Table 1 presents the characteristics of the participants.

Table 1. Characteristics of the participants in the study.

Variable	Patient (n=21)	Nonpatient (n=10)	Total (n=31)
Age (years), mean (SD)	57.6 (16.4)	49.9 (10.2)	55.1 (14.9)
Sex (female), n (%)	11 (52)	6 (60)	17 (55)
History of lower limb surgery, n (%)	16 (76)	1 (10)	17 (55)
Comorbidities, n (%)	12 (57)	5 (50)	17 (55)
BMI (kg/m^2), mean (SD)	28.7 (5.2)	26.9 (5.3)	28.1 (5.3)
Smartphone health app, n (%)			
Google Fit	17 (81)	8 (80)	25 (81)
Apple HealthKit	4 (19)	2 (20)	6 (19)
Smartwatch, n (%)	4 (19)	2 (20)	6 (19)

Participants were excluded owing to surgery cancellation (2/4, 50%) and technical problems with the sensor (1/4, 25%) or the smartphone app (1/4, 25%). In addition, data from 3 patients only contained preoperative data because one of the patients discontinued collecting data after the surgery, the surgery was postponed in another patient, and the sensor was lost in the operating room in the third patient. The time series from patients who only had preoperative data were used for cross-correlation analysis and comparison, but they were not included in the regression analysis.

In the patient group, the surgical procedures performed included total hip arthroplasty (11/21, 52%), total knee arthroplasty (5/21, 24%), osteosynthesis (3/21, 14%), and high tibial osteotomy (2/21, 10%). The most common symptoms were pain (20/21, 95%), walking problems (18/21, 86%), and joint stiffness (7/21, 33%). In total, 17 participants had comorbidities, and 15 participants took daily medications for high blood pressure (8/21, 38%), heart disease (3/21, 14%), diabetes (2/21, 10%), high cholesterol (2/21, 10%), and other diseases (4/21, 19%). Regarding the history of lower limb surgeries, 7 patients had previous knee surgery, 3 had hip surgery, and 6 had other surgeries. In the nonpatient group, 1 person had previous knee surgery.

We collected 1067 days of data (915 days from the patients and 152 days from the nonpatients). The number of data collection days per patient was between 10 and 16 (mean 14) days in the nonpatient group and between 39 and 69 (mean 49) days in the patient group, except for 3 patients with only preoperative data (with 8-, 10-, and 13-day data).

Step Count Analysis

The median and IQR for the step counts from the PA tracker and the smartphone and the percentages of different step types (steps-1, steps-2, and steps-3) in the total PA tracker step counts in various groups of the participants are provided in Table 2.

Table 2. Median and IQR of step counts measured by smartphone and physical activity (PA) tracker by participant characteristics and distribution of step types (steps-1, steps-2, and steps-3) within PA tracker total steps.

Variables	Days, n	Smartphone steps, median (IQR)	PA tracker total steps, median (IQR)	PA tracker total steps composition, median (IQR)		
				Steps-1[a]	Steps-2[a]	Steps-3[a]
Group						
Patient	915	2000 (700-4800)	6500 (3800-10,800)	48 (35-57)	21 (16-33)	29 (24-32)
Nonpatient	152	4600 (2300-9400)	14,800 (10,600-18,300)	58 (52-65)	17 (14-19)	25 (20-28)
Period						
Preoperative	394	2700 (1000-6300)	9600 (6000-13,600)	53 (47-60)	18 (14-23)	28 (24-31)
Postoperative	521	1400 (400-35,000)	5300 (2800-7700)	42 (22-51)	27 (18-44)	30 (24-33)
Age (years)						
≤60	560	3200 (900-6500)	8000 (4400-13,500)	54 (45-61)	18 (14-24)	27 (23-31)
>60	507	1600 (600-3600)	6800 (4100-11,000)	44 (27-54)	24 (17-39)	29 (23-33)
Sex						
Female	540	2600 (700-6400)	9300 (5000-14,500)	49 (36-58)	20 (16-34)	27 (22-31)
Male	527	2000 (800-4500)	6200 (4000-10,000)	50 (40-59)	19 (14-28)	30 (25-33)
Comorbidity						
Negative	504	2100 (600-5300)	9500 (5200-14,300)	50 (40-59)	20 (16-30)	27 (22-31)
Positive	563	2400 (900-5900)	6200 (3900-10,000)	49 (37-58)	20 (15-30)	29 (25-33)
Previous surgery						
Negative	394	3100 (1200-7300)	7800 (4900-13,800)	55 (47-61)	17 (14-22)	28 (24-32)
Positive	673	1900 (600-4500)	7100 (3900-11,500)	46 (30-55)	22 (17-38)	29 (23-32)
Day of week						
Weekday	755	2200 (800-5600)	7200 (4200-12,000)	50 (38-59)	20 (15-30)	28 (23-32)
Weekend	312	2200 (700-5100)	7900 (4300-12,800)	50 (38-58)	20 (16-29)	28 (24-32)
Smartwatch						
Yes	217	2900 (1500-5900)	6800 (4500-11,100)	42 (27-55)	28 (19-40)	29 (25-32)
No	850	2000 (600-5300)	7600 4200-12,600)	51 (42-59)	19 (15-27)	28 (23-32)
Smartphone health app						
Apple HealthKit	904	2200 (800-5600)	7300 (94,100-12,200)	49 (38-58)	20 (16-29)	29 (24-32)
Google Fit	163	2400 (800-4800)	8400 (94,800-12,400)	54 (39-60)	18 (13-31)	27 (23-31)
All participants	1067	2200 (800-5500)	7400 (4300-12,400)	50 (38-59)	20 (15-30)	28 (23-32)

[a]Correspond to the proportions of total PA tracker steps in percentages.

In Figure 2, the time series data for each patient during the preoperative and postoperative periods and for the nonpatient group are presented for both the smartphone and PA trackers. Table 3 shows the cross-correlation coefficients (*r*) at lag 0 between the smartphone time series and the time series for different PA tracker step counts (steps-1, steps-2, steps-3, and total steps) for each participant in the study.

Table 4 displays the median and IQR of the cross-correlation coefficients between the daily step count time series of smartphones and PA trackers for various variables (steps-1, steps-2, steps-3, and total steps).

Figure 2. The upper panel shows the time series for step counts recorded by the smartphone and physical activity tracker for each patient (P) before and after the surgery, whereas the lower panel displays the same for nonpatient participants (C). Each plot corresponds to 1 participant, and the bold black font indicates their ID, which matches the IDs in Table 3. In the patient group, each gray horizontal gridline represents 5000 steps, and each gray vertical gridline represents 5 days. In the nonpatient group, each gray horizontal gridline represents 5000 steps, and each gray vertical gridline represents 2 days.

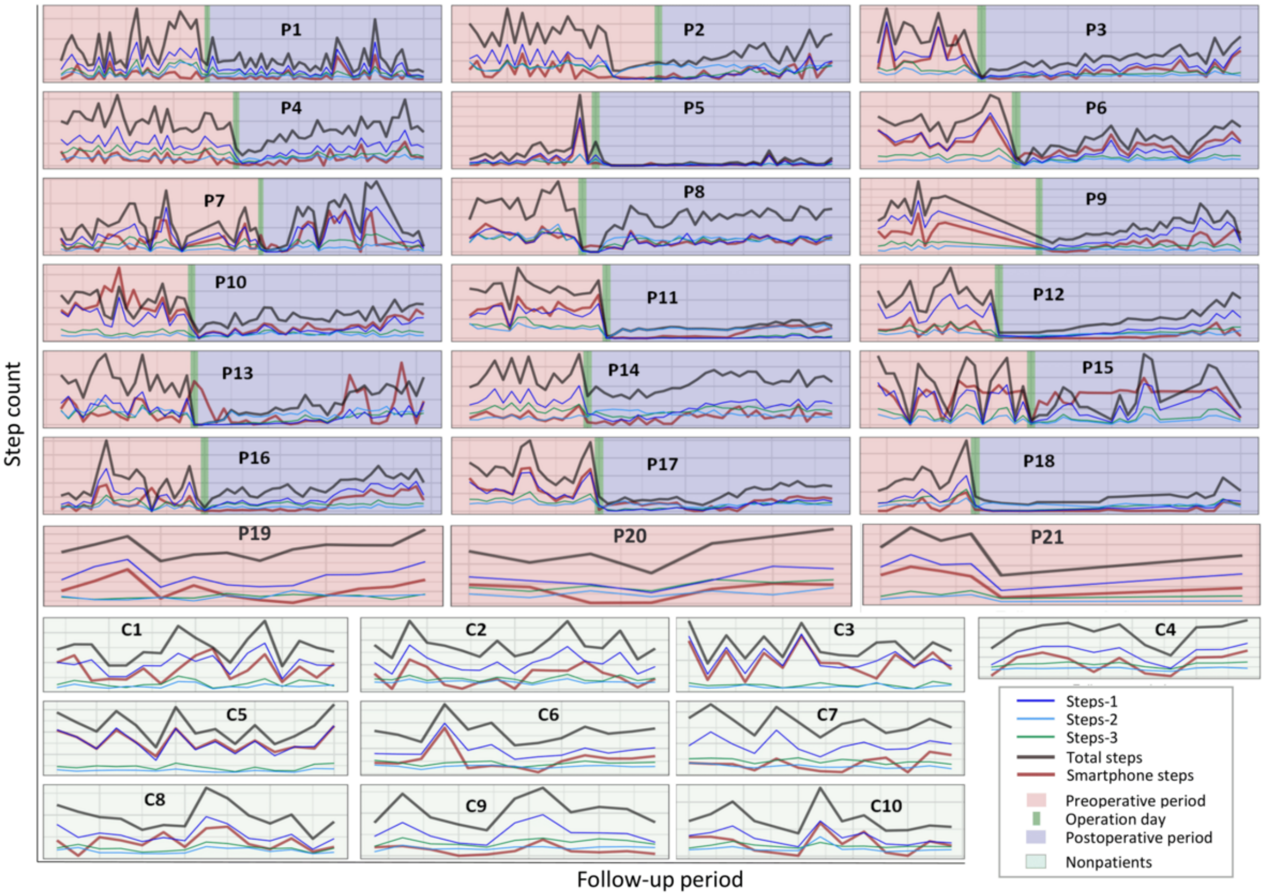


Table 3. Cross-correlation at lag 0 between smartphone and physical activity (PA) tracker step count variables, with correlation coefficients (r) and *P* values for each participant[a].

ID	Group	Steps-1 vs smartphone steps		Steps-2 vs smartphone steps		Steps-3 vs smartphone steps		PA tracker total steps vs smartphone steps	
		r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
P1	Patient	0.70	<.001	0.40	.001	0.47	<.001	0.60	<.001
P2	Patient	0.66	<.001	0.19	.20	0.62	<.001	0.65	<.001
P3	Patient	0.92	<.001	0.26	.009	0.41	<.001	0.84	<.001
P4	Patient	0.78	<.001	0.57	<.001	0.72	<.001	0.73	<.001
P5	Patient	0.97	<.001	0.77	<.001	0.78	<.001	0.93	<.001
P6	Patient	0.94	<.001	0.45	.003	0.74	<.001	0.91	<.001
P7	Patient	0.72	<.001	0.48	<.001	0.57	<.001	0.68	<.001
P8	Patient	0.88	<.001	0.48	.001	0.63	<.001	0.75	<.001
P9	Patient	0.86	<.001	0.44	.005	0.72	<.001	0.84	<.001
P10	Patient	0.78	<.001	0.34	.01	0.27	.052	0.67	<.001
P11	Patient	0.95	<.001	0.32	.03	0.93	<.001	0.96	<.001
P12	Patient	0.78	<.001	0.63	<.001	0.79	<.001	0.80	<.001
P13	Patient	0.30	.03	0.13	.40	0.11	.40	0.16	.30
P14	Patient	0.74	<.001	0.28	.06	0.66	<.001	0.74	<.001
P15	Patient	0.55	<.001	0.50	.002	0.51	.001	0.55	<.001
P16	Patient	0.82	<.001	0.27	.047	0.26	.07	0.70	<.001
P17	Patient	0.96	<.001	0.78	<.001	0.89	<.001	0.96	<.001
P18	Patient	0.92	<.001	0.64	<.001	0.82	<.001	0.90	<.001
P19	Patient (only preoperative)	0.92	.01	0.44	.001	0.30	.08	0.72	.10
P20	Patient (only preoperative)	0.83	.08	0.20	.02	0.46	.02	0.66	.30
P21	Patient (only preoperative)	0.73	.02	0.49	.02	0.45	.06	0.53	.08
C1	Nonpatient	0.66	.005	0.29	.20	0.26	.30	0.61	.01
C2	Nonpatient	0.90	<.001	0.15	.50	0.17	.50	0.82	<.001
C3	Nonpatient	0.84	.01	0.06	.80	0.00	.90	0.68	.01
C4	Nonpatient	0.91	.002	0.69	.02	0.88	.002	0.90	.002
C5	Nonpatient	0.78	.007	0.15	.50	0.11	.60	0.63	.007
C6	Nonpatient	0.78	.01	0.26	.30	0.39	.10	0.66	.01
C7	Nonpatient	0.86	.003	0.76	.006	0.51	.007	0.82	.003
C8	Nonpatient	0.99	<.001	0.52	.04	0.62	.01	0.96	<.001
C9	Nonpatient	0.55	.10	0.09	.70	0.24	.40	0.38	.10
C10	Nonpatient	0.82	.01	0.24	.40	0.54	.06	0.73	.01

[a]The corresponding time series are demonstrated in Figure 2.

Table 4. Cross-correlation coefficients and corresponding 95% CIs between daily smartphone step counts and different step counts from the physical activity (PA) tracker (steps-1, steps-2, steps-3, and total steps)[a].

	Steps-1 vs smartphone steps, mean (95% CI)	*P* value	Steps-2 vs smartphone steps, mean (95% CI)	*P* value	Steps-3 vs smartphone steps, mean (95% CI)	*P* value	PA tracker total steps vs smartphone steps, mean (95% CI)	*P* value
Group		.02		.80		.50		.07
Patient	0.79 (0.72-0.85)		0.35 (0.23-0.47)		0.48 (0.38-0.58)		0.70 (0.62-0.77)	
Nonpatient	0.88 (0.74-0.94)		0.33 (0.09-0.53)		0.41 (0.20-0.59)		0.77 (0.66-0.88)	
Period		.02		.06		.06		.02
Preoperative	0.82 (0.72-0.89)		0.41 (0.21-0.59)		0.53 (0.37-0.67)		0.74 (0.62-0.83)	
Postoperative	0.75 (0.62-0.84)		0.28 (0.16-0.39)		0.42 (0.31-0.52)		0.64 (0.53-0.74)	
Age (years)		.09		.10		.10		.10
≤60	0.83 (0.73-0.89)		0.39 (0.27-0.53)		0.46 (0.34-0.56)		0.75 (0.65-0.82)	
>60	0.79 (0.70-0.86)		0.30 (0.15-0.44)		0.48 (0.33-0.61)		0.70 (0.58-0.79)	
Sex		.40		.60		.30		.40
Female	0.80 (0.69-0.88)		0.34 (0.18-0.48)		0.45 (0.30-0.57)		0.71 (0.58-0.81)	
Male	0.82 (0.75-0.87)		0.36 (0.21-0.50)		0.50 (0.39-0.59)		0.74 (0.67-0.79)	
Smartwatch		.30		.60		.40		.97
Yes	0.84 (0.61-0.94)		0.31 (0.01-0.55)		0.40 (0.15-0.60)		0.73 (0.46-0.87)	
No	0.80 (0.74-0.85)		0.36 (0.24-0.47)		0.49 (0.39-0.58)		0.72 (0.65-0.78)	
Smartphone health app		.003		.40		.30		.01
Apple HealthKit	0.84 (0.77-0.90)		0.37 (0.24-0.48)		0.49 (0.39-0.58)		0.75 (0.68-0.81)	
Google Fit	0.63 (0.49-0.73)		0.25 (0.06-0.42)		0.35 (0.26-0.44)		0.53 (0.42-0.62)	
All participants	0.82 (0.64-0.90)	—[b]	0.30 (0.10-0.56)	—	0.45 (0.23-0.66)	—	0.70 (0.53-0.83)	—

[a]*P* values for group comparisons are provided.

[b]Not available.

Regression Models for Step Counts

Tables 5-7 present the coefficients for the fixed and random effects for the regression models that best fit the data for steps-1, steps-2, steps-3, and the total steps recorded by the PA trackers, along with the goodness-of-fit metrics.

Table 5. Fixed effects for the best-fitted models estimating daily step counts using smartphone step counts[a].

Variables[a]	Models							
	Steps-1	*P* value	Steps-2	*P* value	Steps-3	*P* value	PA tracker total steps	*P* value
Intercept, α (95% CI)	0.33 (0.07 to 0.59)	<.001	0.68 (0.14 to 1.23)	.01	0.76 (0.32 to 1.2)	.002	0.67 (0.37 to 0.96)	<.001
Slope, β (95% CI)								
Smartphone steps	0.82 (0.66 to 0.99)	<.001	0.34 (0.15 to 0.52)	.001	0.85 (0.56 to 1.14)	<.0001	0.85 (0.67 to 1.04)	<.001
Period (postoperative)	−0.44 (−0.61 to −0.26)	<.001	−0.30 (−0.53 to −0.06)	.02	−0.52 (−0.72 to −0.32)	<.001	−0.47 (−0.65 to −0.30)	<.001
Positive medical history	—[b]	—	−0.84 (−1.55 to −0.13)	.02	−0.53 (−1.09 to 0.03)	.06	−0.53 (−0.89 to −0.18)	.007

[a]The values in this table regard the scaled step counts.

[b]Not available.

Table 6. Random effects variances for the best-fitted models estimating daily step counts using smartphone step counts.

Random effects	Models			
	Steps-1	Steps-2	Steps-3	PA tracker total steps
The variance between individuals' intercepts	0.22	0.47	0.32	0.15
The variance of PA tracker—smartphone steps slope between individuals	0.08	0.02	0.25	0.03
The variance of the residuals	0.07	0.20	0.18	0.10

Table 7. Goodness-of-fit metrics for the best-fitted models estimating daily step counts using smartphone step counts.

Model metrics	Models			
	Steps-1	Steps-2	Steps-3	PA tracker total steps
AIC[a]	402	1211	1111	663
BIC[b]	449	1263	1163	714
Deviance	38	1189	1089	641
ICC[c]	0.83	0.78	0.80	0.75
Conditional pseudo-R^2	0.94	0.83	0.90	0.92
Marginal pseudo-R^2	0.65	0.25	0.51	0.68

[a]AIC: Akaike information criterion.

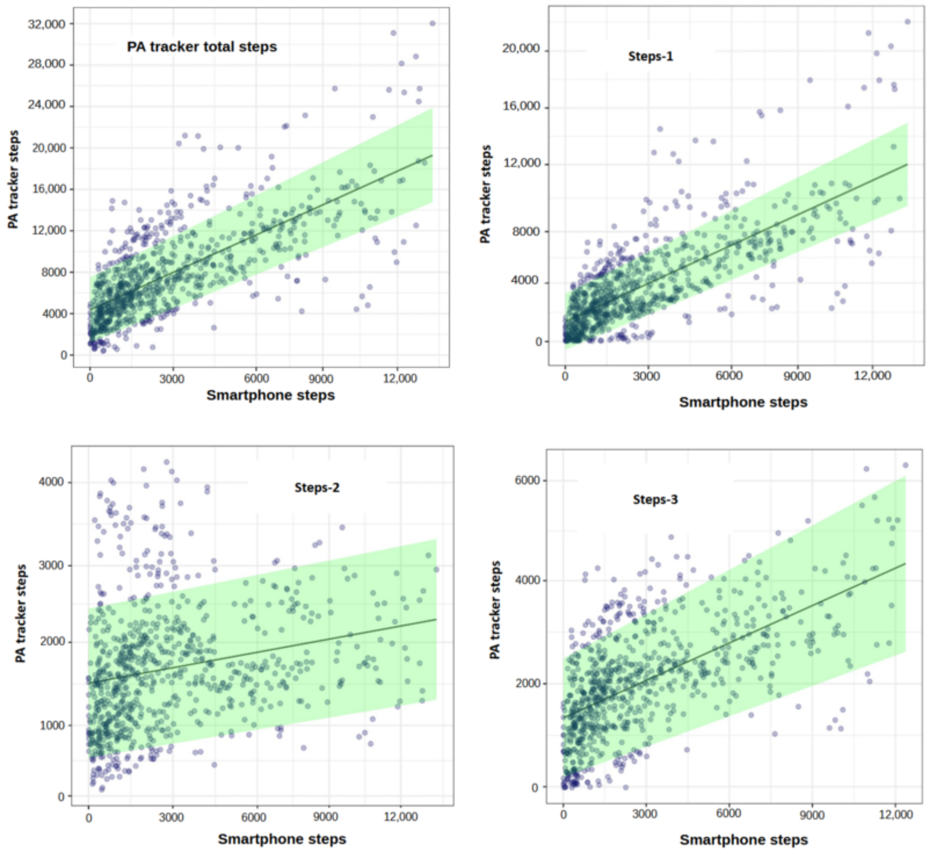
[b]BIC: Bayesian information criterion.

[c]ICC: intraclass correlation coefficient.

Figure 3 displays the outcomes of various models, along with the 95% prediction intervals for all patients.

The models with the smartphone steps provided a better fit for the total step counts than the models without this variable. The likelihood ratio tests for comparing the selected models with and without smartphone steps demonstrated that the smartphone steps were positively correlated with PA tracker total steps (χ^2_1=34.7, *P*<.001), steps-1 (χ^2_1=36.8, *P*<.001), steps-2 (χ^2_1=11.4, *P*<.001), and steps-3 (χ^2_1=22.1, *P*<.001).

Figure 3. Results of the different models for estimating the daily step counts of the physical activity (PA) tracker, including the total steps, steps-1, steps-2, and steps-3. The mean values are depicted by solid lines, whereas the 95% prediction intervals are shown as light green shaded areas for each model.



Questionnaires and SUS Scores

Overall, 94% (17/18) of the patients filled out the questionnaires regarding SUS. The median scores were 78 (IQR 73-88) for the smartphone app and 73 (IQR 68-80) for the PA tracker, respectively. The scores were higher in female patients and in those aged <60 years (Table 8).

Table 8. The median and IQR of the System Usability Scale (SUS) scores for the smartphone app and the physical activity (PA) tracker for different age and sex groups.

Variables	SUS, median (IQR)	
	Smartphone app	PA tracker
Age (years)		
≤60	93 (83-96)	88 (75-95)
>60	73 (65-76)	70 (66-73)
Sex		
Male	73 (65-80)	69 (65-76)
Female	83 (73-95)	78 (70-80)
Total	78 (73-88)	73 (68-80)

Discussion

Principal Findings

In this study, we explored the feasibility of using smartphones for remote monitoring of orthopedic patients' PA. To achieve this, we analyzed the correlation between the step counts recorded by a smartphone and a 24/7 PA tracker. Our results indicated a high correlation (r=0.70) between the time series of daily smartphone steps and daily PA tracker total steps. In addition, we found that the number of steps recorded by the smartphone was a strong predictor of changes in total daily steps. However, the absolute number of daily steps predicted using smartphone data was neither precise nor reliable.

The role of smartphones in remote monitoring of patients' PA has not yet been clearly defined because of 2 main reasons. First, concerns persist regarding the validity and reliability of PA data collected by smartphones, as conflicting results have been reported in the literature [28]. For example, in a systematic review, the difference between smartphone measurements and a gold standard in a laboratory setting varied from 0.1% to 79.3%, and the reliability of smartphone measurements ranged from poor to excellent (intraclass correlation coefficient between 0.02 and 0.99) [13]. Second, the relationship between smartphone PA data and total PA data in different individuals is not fully understood and depends on various factors. The smartphone only records a variable proportion of the total daily PA, which is the time the person carries the device. Ignoring this point can lead to conflicting results, especially in studies with free-living settings. In this study, we investigated the relationship between the 2 variables and found that, despite considerable variability, a high correlation exists between smartphone step counts and total daily step counts.

The correlation between smartphone and total daily steps can vary significantly in a free-living setting, both between and within individuals. Several studies have found inferior results regarding the validity and reliability of smartphone measurements in free-living measurements compared with laboratory settings [29-31]. The variations may be even higher in orthopedic patients owing to pain and mobility issues during the early postoperative period, which could affect smartphone use and measurements. In a recent pilot study, Vorrink et al [32] found a mean correlation of 0.88 between smartphone and PA tracker measurements in a group of nonorthopedic patients, which was higher than the correlation we found in this study. However, we calculated the correlation between the time series after differentiating and detrending. Our analysis of different step count variables from the PA tracker revealed that the correlation with smartphone steps was the highest for steps-1 and the lowest for steps-2. We also found that the correlation between PA tracker's steps-1 and PA data collected by smartphone was higher in the nonpatient group than in the patient group and during the preoperative period compared with the postoperative period. However, the correlation remained relatively high even during the postoperative period (r=0.64 and r=0.75 for total steps and steps-1, respectively). This discrepancy in the correlation could be attributed to the possibility that patients do not carry their smartphones as frequently during the postoperative period as they would under normal circumstances, or it could be because of the lower measurement accuracy in lower walking velocities, which has been demonstrated in previous studies [33,34]. Regarding the PA tracker's steps-2 and steps-3, we could not find a significant difference in the correlations between subject groups with different characteristics (the P values were between .06 and .80).

Most participants (>80%) in our study used iOS smartphones, and we observed a stronger correlation in PA tracker's steps-1 and total steps with smartphones equipped with Apple HealthKit APIs. However, we were unable to compare different smartphone types owing to the small sample size of participants with Google Fit API in our study. Several studies have investigated the impact of smartphone type on the accuracy and precision of PA measurements [35-38]. For instance, Höchsmann et al [38] found lower accuracy in an Android smartphone during low-velocity gait when compared with other smartphones and PA trackers. Moreover, we did not observe a high correlation between smartwatch users and the total PA tracker steps. This finding can be attributed to the lower proportion of steps-1 in the total steps composition among smartwatch users (ie, smartwatch users took fewer continuous regular walking steps [steps-1] in this study), as shown in Table 2. As the highest correlation between the smartphone and PA tracker step counts was observed for steps-1, we would not expect an increase in the correlation between the smartphone and the total PA tracker steps.

We applied mixed effects modeling to predict different step types (continuous regular walking, sporadic walking, and slow continuous walking) by using the smartphone step counts. Mixed effects models are a type of regression analysis and are especially useful in longitudinal studies with repeated measurements or when the measurements are made on cluster units [39]. Although we could fit mixed effects models with relatively high-performance metrics, the bootstrapping methods demonstrated wide prediction intervals. Therefore, estimating the daily number of steps by using the smartphone step counts without further precalibration would be imprecise and inaccurate. The best-fitted model was achieved for continuous regular walking (steps-1), which is consistent with the observation of the highest correlation between smartphone step counts and continuous regular walking (steps-1). On the basis of the models' coefficients, we found that the postoperative period and a positive medical history were negatively associated with the total daily steps. The mixed effects models could also describe the variance in data between and within different individuals. We found that the variation between individuals in both the intercept and the slope of the PA tracker–smartphone steps was higher for sporadic walking (steps-2) and slow continuous walking (steps-3), which makes estimating these variables more difficult. In all 4 fitted models, the variance of the random effects intercept between individuals was more pronounced than that of the random effects slopes.

In this study, the PA tracker and the smartphone app obtained SUS score higher than the acceptable value, which was assumed to be 70 [21]. However, the SUS score cannot independently make absolute judgments about the *goodness* of a product. Factors such as success rate and the nature of the observed failures should play a prominent role in product usability [40]. During this study, we observed 1 smartphone app failure, which led to participant exclusion. This participant unintentionally removed the app from her smartphone and could not reinstall it because of technical issues. Furthermore, we found higher usability scores in patients aged <60 years and female patients. The effects of age and sex were analyzed in SUS applied for different products. A significant but not strong negative correlation has been demonstrated between SUS scores and age; however, no significant difference has been found in the mean SUS scores between female participants and male participants [21]. Some studies have also shown that the young adults and female participants were associated with higher PA tracker use [41,42].

Strengths and Weaknesses of the Study

This longitudinal study is the first of its kind to evaluate the correlation between the daily steps recorded by a smartphone with the total number of steps in patients undergoing orthopedic surgeries for several weeks before and after surgery and in a nonpatient group. We also analyzed different walking types (regular continuous, sporadic, and slow continuous walking) and demonstrated that smartphones are more competent in capturing the steps during regular continuous walking. Detecting different gait patterns by smartphones and PA trackers has recently received considerable attention [43-45]. Indisputably, we must acknowledge the limitation that the validity of the 3 categories of steps measured by the PA tracker in this study has not yet been fully explored and must be scrutinized. Furthermore, our study had other limitations, such as the inability to obtain information regarding the smartphone use habits of the participants, including how and where the user carries the smartphone. Nevertheless, we used mixed effects modeling and random effects variables to account for individual differences to increase the generalizability of the findings. Another limitation of the study was that owing to the setting of the study, we could not use direct observation as the gold standard for counting the steps. However, to reduce data collection bias, we used a previously validated PA tracker that measured PA continuously 24/7.

Implications and Future Research

In this study, we found a high correlation between the number of steps recorded by smartphones and the total number of daily steps. However, owing to the limitations and impact of participant dropouts and missing data, we recommend interpreting the findings with caution and conducting further investigations with larger sample sizes and more robust data collection methods. In addition, further investigations with larger sample sizes and more robust data collection methods are necessary to explore determining factors in the predictability of smartphone measurements and their role in remote patient monitoring. The study also demonstrated the predictive value of the postoperative period and positive medical history in estimating the total daily steps, but more homogenous samples may increase the precision of these prediction models. In future research, it would be valuable to compare the measurements of other well-known PA trackers with varying characteristics to smartphone measurements [46].

Conclusions

This study highlights the potential of smartphones for monitoring changes in PA, showing a strong correlation between daily steps recorded by smartphones and total daily steps, especially during continuous walking. This finding suggests that smartphones could be a valuable tool for remote patient activity monitoring. However, accurately predicting the precise daily step counts from smartphone data still requires further investigation, as our results suggest that the current methods may lack the necessary precision and accuracy.

Acknowledgments

The development of the OrtoApp was supported by the Danish Ministry of Higher Education and Science through a performance contract titled “Digital Health and Welfare,” awarded to the Alexandra Institute and the Danish Technological Institute. The funders were not involved in the study’s design, data collection, analysis, interpretation, manuscript writing, or the decision to publish the results.

Conflicts of Interest

None declared.

References

1. World Health Organization. Global Action Plan on Physical Activity 2018-2030 More Active People for a Healthier World. Geneva: World Health Organization; 2019.
2. Stevens M, Reininga IH, Bulstra SK, Wagenmakers R, van den Akker-Scheek I. Physical activity participation among patients after total hip and knee arthroplasty. Clin Geriatr Med 2012 Aug;28(3):509-520. [doi: 10.1016/j.cger.2012.05.003] [Medline: 22840311]
3. Almeida GJ, Khoja SS, Piva SR. Physical activity after total joint arthroplasty: a narrative review. Open Access J Sports Med 2018 Mar;Volume 9:55-68. [doi: 10.2147/oajsm.s124439]
4. Atallah L, Jones GG, Ali R, Leong JJ, Lo B, Yang GZ. Observing recovery from knee-replacement surgery by using wearable sensors. In: Proceedings of the 2011 International Conference on Body Sensor Networks. 2011 Presented at: 2011 International Conference on Body Sensor Networks; May 23-25, 2011; Dallas, TX, USA. [doi: 10.1109/bsn.2011.10]
5. Paxton RJ, Melanson EL, Stevens-Lapsley JE, Christiansen CL. Physical activity after total knee arthroplasty: a critical review. World J Orthop 2015 Sep 18;6(8):614-622 [FREE Full text] [doi: 10.5312/wjo.v6.i8.614] [Medline: 26396937]
6. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. Sports Med 2001 Feb;31(2):91-100. [doi: 10.2165/00007256-200131020-00002] [Medline: 11227981]
7. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc 2014 Jan;46(1):99-106. [doi: 10.1249/MSS.0b013e3182a0595f] [Medline: 23793232]

8. Maenaka K. MEMS inertial sensors and their applications. In: Proceedings of the 5th International Conference on Networked Sensing Systems. 2008 Presented at: 5th International Conference on Networked Sensing Systems; Jun 17-19, 2008; IKanazawa, Japan. [doi: 10.1109/inss.2008.4610859]
9. Berenguer A, Goncalves J, Hosio S, Ferreira D, Anagnostopoulos T, Kostakos V. Are smartphones ubiquitous?: an in-depth survey of smartphone adoption by seniors. IEEE Consumer Electron Mag 2017 Jan;6(1):104-110. [doi: 10.1109/mce.2016.2614524]
10. Elektronik i hjemmet. Statistics Denmark. URL: https://www.dst.dk/da/Statistik/emner/oekonomi/forbrug/elektronik-i-hjemmet [accessed 2022-06-03]
11. Kurtz SM, Higgs GB, Chen Z, Koshut WJ, Tarazi JM, Sherman AE, et al. Patient perceptions of wearable and smartphone technologies for remote outcome monitoring in patients who have hip osteoarthritis or arthroplasties. J Arthroplasty 2022 Jul;37(7S):S488-92.e2. [doi: 10.1016/j.arth.2022.02.026] [Medline: 35277311]
12. Lyman S, Hidaka C, Fields K, Islam W, Mayman D. Monitoring patient recovery after THA or TKA using mobile technology. HSS J 2020 Dec 12;16(Suppl 2):358-365 [FREE Full text] [doi: 10.1007/s11420-019-09746-3] [Medline: 33380968]
13. Silva AG, Simões P, Queirós A, Rodrigues M, Rocha NP. Mobile apps to quantify aspects of physical activity: a systematic review on its reliability and validity. J Med Syst 2020 Jan 08;44(2):51. [doi: 10.1007/s10916-019-1506-z] [Medline: 31915935]
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
15. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
16. Ghaffari A, Rahbek O, Lauritsen RE, Kappel A, Kold S, Rasmussen J. Criterion validity of linear accelerations measured with low-sampling-frequency accelerometers during overground walking in elderly patients with knee osteoarthritis. Sensors 2022 Jul 15;22(14):5289. [doi: 10.3390/s22145289]
17. Pedersen BS, Kristensen MT, Josefsen CO, Lykkegaard KL, Jønsson LR, Pedersen MM. Validation of two activity monitors in slow and fast walking hospitalized patients. Rehabil Res Pract 2022 May 16;2022:9230081 [FREE Full text] [doi: 10.1155/2022/9230081] [Medline: 35615755]
18. HealthKit. Apple Developer Documentation. URL: https://developer.apple.com/documentation/healthkit [accessed 2022-10-27]
19. Google Fit homepage. Google Fit. URL: https://developers.google.com/fit [accessed 2022-10-27]
20. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
21. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
22. Hvidt JC, Christensen LF, Sibbersen C, Helweg-Jørgensen S, Hansen JP, Lichtenstein MB. Translation and validation of the system usability scale in a Danish mental health setting using digital technologies in treatment interventions. Int J Hum Comput Interact 2020;36(8):709-716 [FREE Full text] [doi: 10.1080/10447318.2019.1680922]
23. Shumway RH, Stoffer DS. Time Series Analysis and Its Applications With R Examples. Cham: Springer; 2017.
24. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 1915 May;10(4):507-521. [doi: 10.2307/2331838]
25. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
26. Merlo J, Yang M, Chaix B, Lynch J, Råstam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. J Epidemiol Community Health 2005 Sep 01;59(9):729-736 [FREE Full text] [doi: 10.1136/jech.2004.023929] [Medline: 16100308]
27. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 2012 Dec 03;4(2):133-142 [FREE Full text] [doi: 10.1111/j.2041-210x.2012.00261.x]
28. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
29. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci 2018 Aug;36(15):1695-1704. [doi: 10.1080/02640414.2017.1409855] [Medline: 29179653]
30. Leong JY, Wong JE. Accuracy of three Android-based pedometer applications in laboratory and free-living settings. J Sports Sci 2017 Jan 07;35(1):14-21. [doi: 10.1080/02640414.2016.1154592] [Medline: 26950687]

31. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
32. Vorrink SN, Kort HS, Troosters T, Lammers JW. A mobile phone app to stimulate daily physical activity in patients with chronic obstructive pulmonary disease: development, feasibility, and pilot studies. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e11 [FREE Full text] [doi: 10.2196/mhealth.4741] [Medline: 26813682]
33. Fokkema T, Kooiman TJ, Krijnen WP, VAN DER Schans CP, DE Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
34. Dybus A, Paul L, Wyke S, Brewster S, Gill JM, Ramsay A, et al. Validation of smartphone step count algorithm used in STARFISH smartphone application. Technol Health Care 2017 Dec 04;25(6):1157-1162. [doi: 10.3233/THC-170970] [Medline: 28946599]
35. Amagasa S, Kamada M, Sasai H, Fukushima N, Kikuchi H, Lee IM, et al. How well iPhones measure steps in free-living conditions: cross-sectional validation study. JMIR Mhealth Uhealth 2019 Jan 09;7(1):e10418 [FREE Full text] [doi: 10.2196/10418] [Medline: 30626569]
36. Adamakis M. Criterion validity of iOS and android applications to measure steps and distance in adults. Technologies 2021 Jul 29;9(3):55. [doi: 10.3390/technologies9030055]
37. Beltrán-Carrillo VJ, Jiménez-Loaisa A, Alarcón-López M, Elvira JL. Validity of the "Samsung Health" application to measure steps: a study with two different Samsung smartphones. J Sports Sci 2019 Apr 17;37(7):788-794. [doi: 10.1080/02640414.2018.1527199] [Medline: 30332917]
38. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
39. Gomes DG. Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? PeerJ 2022;10:e12794 [FREE Full text] [doi: 10.7717/peerj.12794] [Medline: 35116198]
40. ISO 9241-11:1998 ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: guidance on usability. ISO. 1998 Mar. URL: https://www.iso.org/standard/16883.html [accessed 2022-11-12]
41. Macridis S, Johnston N, Johnson S, Vallance JK. Consumer physical activity tracking device ownership and use among a population-based sample of adults. PLoS One 2018;13(1):e0189298 [FREE Full text] [doi: 10.1371/journal.pone.0189298] [Medline: 29293532]
42. Voskuil VR, Stroup S, Leyden M. Acceptability and usability of a wearable activity tracker and application among inactive adolescent girls. Phys Act Health 2020 Jun 3;4(1):52-61. [doi: 10.5334/paah.51]
43. Hong JH, Han KS, Kim K. Walking pattern analysis using an acceleration sensor device. J Electrical Eng Technol 2017 Jan 02;12(1):396-401. [doi: 10.5370/JEET.2017.12.1.396]
44. Bayat A, Bayat AH, Ghasemi AS. Classifying human walking patterns using accelerometer data from smartphone. Int J Comput Sci Mob Computing 2017 Dec;3:1-6 [FREE Full text]
45. Alizadeh V. Gait pattern recognition using accelerometers. arXiv 2017 Mar 11 [FREE Full text]
46. Mandigout S, Lacroix J, Perrochon A, Svoboda Z, Aubourg T, Vuillerme N. Comparison of step count assessed using wrist- and hip-worn Actigraph GT3X in free-living conditions in young and older adults. Front Med (Lausanne) 2019 Nov 5;6:252 [FREE Full text] [doi: 10.3389/fmed.2019.00252] [Medline: 31828072]

Abbreviations

API: application programming interface
PA: physical activity
PROM: patient-reported outcome measure
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by L Buis; submitted 19.11.22; peer-reviewed by T Aubourg, W Lin; comments to author 13.04.23; revised version received 26.04.23; accepted 05.06.23; published 04.07.23.

Please cite as:
Ghaffari A, Kildahl Lauritsen RE, Christensen M, Rolighed Thomsen T, Mahapatra H, Heck R, Kold S, Rahbek O
Exploring the Feasibility and Usability of Smartphones for Monitoring Physical Activity in Orthopedic Patients: Prospective Observational Study
JMIR Mhealth Uhealth 2023;11:e44442
URL: https://mhealth.jmir.org/2023/1/e44442
doi:10.2196/44442
PMID:37283228

Original Paper

Exploring Variations in Sleep Perception: Comparative Study of Chatbot Sleep Logs and Fitbit Sleep Data

Hyunchul Jang[1], PhD; Siwoo Lee[1], MD, PhD; Yunhee Son[1], PhD; Sumin Seo[1], MSc; Younghwa Baek[1], PhD; Sujeong Mun[1], MD, PhD; Hoseok Kim[1], MSc; Icktae Kim[1], MSc; Junho Kim[1], MSc

KM Data Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea

Corresponding Author:
Siwoo Lee, MD, PhD
KM Data Division
Korea Institute of Oriental Medicine
1672 Yuseong-daero, Yuseong-gu
Daejeon, 34054
Republic of Korea
Phone: 82 42 868 9555
Email: bfree@kiom.re.kr

Abstract

Background: Patient-generated health data are important in the management of several diseases. Although there are limitations, information can be obtained using a wearable device and time-related information such as exercise time or sleep time can also be obtained. Fitbits can be used to acquire sleep onset, sleep offset, total sleep time (TST), and wakefulness after sleep onset (WASO) data, although there are limitations regarding the depth of sleep and satisfaction; therefore, the patient's subjective response is still important information that cannot be replaced by wearable devices.

Objective: To effectively use patient-generated health data related to time such as sleep, it is first necessary to understand the characteristics of the time response recorded by the user. Therefore, the aim of this study was to analyze the characteristics of individuals' time perception in comparison with wearable data.

Methods: Sleep data were acquired for 2 weeks using a Fitbit. Participants' sleep records were collected daily through chatbot conversations while wearing the Fitbit, and the two sets of data were statistically compared.

Results: In total, 736 people aged 30-59 years were recruited for this study, and the sleep data of 543 people who wore a Fitbit and responded to the chatbot for more than 7 days on the same day were analyzed. Research participants tended to respond to sleep-related times on the hour or in 30-minute increments, and each participant responded within the range of 60-90 minutes from the value measured by the Fitbit. On average for all participants, the chat responses and the Fitbit data were similar within a difference of approximately 15 minutes. Regarding sleep onset, the participant response was 8 minutes and 39 seconds (SD 58 minutes) later than that of the Fitbit data, whereas with respect to sleep offset, the response was 5 minutes and 38 seconds (SD 57 minutes) earlier. The participants' actual sleep time (AST) indicated in the chat was similar to that obtained by subtracting the WASO from the TST measured by the Fitbit. The AST was 13 minutes and 39 seconds (SD 87 minutes) longer than the time WASO was subtracted from the Fitbit TST. On days when the participants reported good sleep, they responded 19 (SD 90) minutes longer on the AST than the Fitbit data. However, for each sleep event, the probability that the participant's AST was within ±30 and ±60 minutes of the Fitbit TST-WASO was 50.7% and 74.3%, respectively.

Conclusions: The chatbot sleep response and Fitbit measured time were similar on average and the study participants had a slight tendency to perceive a relatively long sleep time if the quality of sleep was self-reported as good. However, on a participant-by-participant basis, it was difficult to predict participants' sleep duration responses with Fitbit data. Individual variations in sleep time perception significantly affect patient responses related to sleep, revealing the limitations of objective measures obtained through wearable devices.

(JMIR Mhealth Uhealth 2023;11:e49144) doi:10.2196/49144

KEYWORDS

sleep; sleep time; chat; self-report; sleep log; sleep diary; wearables; Fitbit; patient-generated health data; PGHD

Introduction

Patient-generated health data play an important role in the management of many diseases. Various types of health-related information can be obtained from patients, ranging from subjective feelings or pain to objectively measurable steps and sitting times. Although some information remains unobtainable and there are various restrictions, it is possible to obtain diverse types of information from patients using wearable devices. In relation to sleep, time-related information such as the time an individual falls asleep and time they wake up can be obtained, as well as sleep quality and sleep environment information [1,2].

The current standard for clinical sleep evaluation is polysomnography (PSG) [3,4]. However, because PSG is usually performed in a hospital, actigraphy is used as an alternative in outpatient environments [5-7]. Actigraphy is less accurate than PSG, but is generally considered to be more accurate than sleep diaries. As PSG is performed in a sleep laboratory, many studies have used actigraphy to measure bedtime or wake-up time in everyday life and to study sleep-related diseases [8-13].

To track the state of sleep, an app installed on a smartphone or a sensor installed on a mattress or around it is used, although the accuracy of such devices is lower than that of a device worn directly [14,15]. In addition, the sleep state can be obtained through a sleep diary or questionnaire, which is less accurate but nevertheless useful in that the subjective sleep information of the user can be obtained [15]. As wearables still have limitations in assessing the depth or quality of sleep, there is a need to utilize the user's perceived sleep, and user feedback is required until an objective diagnostic test technique can exclude the user's subjective feelings. Each of these sleep measurement methods has advantages and disadvantages in terms of accuracy, convenience, and cost, and information can only be obtained through subjective methods. There is a need to use two or more methods together to create a synergistic effect between objective and subjective methods and to compensate for each of their disadvantages [15].

Commercially available wearables for actigraphy include various smart watches and fitness trackers. Although wearables are less reliable, they provide acceptable levels of sleep monitoring and are promising monitoring tools [16]. One representative wearable device is the fitness tracker Fitbit [17-25]. In addition to movement, Fitbits measure heart rate and other characteristics to provide sleep values [26-28]. Fitbit was reported to calculates total sleep time (TST) by 9 minutes more and sleep onset latency (SOL) by 4 minutes less compared with PSG, and a correlation between sleep onset, sleep offset, TST, and wakefulness after sleep onset (WASO) compared with PSG was reported [29]. Fitbit has been shown to be accurate to some extent for measuring sleep time, although there are limitations with sleep depth; however, there is no device that accurately measures sleep stages [30-34]. Therefore, a more personalized model is required to determine sleep stages or sleep quality using wearables [35].

Balancing user acceptance and monitoring performance is the biggest challenge in sleep-monitoring system research in terms of cost and efficiency [36]. A separate process such as charging and wearing may be required to wear a wearable device for a long period of time [37]. Wearables have several advantages, although they also have well-known disadvantages. There are many difficulties such as not wearing them, not wearing them properly, and the devices not accurately identifying the wearer. When using various devices, there are problems related to differences in operation methods or algorithms [38]; even when using a single device, the measurement process or results may change because of changes in firmware or algorithms. Therefore, further research on standardized performance evaluation systems for sleep-tracking technology is required [39].

In the United States, women sleep more on average than men [40]. Women also have better objective sleep quality, sleep duration, and sleep efficiency than men; however, they report poor sleep [41]. One study reported that subjective sleep quality was low in women [42]. In Australia, men stated that they think that their quality of sleep is better than that of women [43], and a report in China based on a Pittsburgh Sleep Quality Index (PSQI) survey suggested that women have worse quality of sleep than men [44]. Although many studies have addressed gender differences in sleep, few have addressed the differences between healthy men and women. In general, adult men and women require approximately 7 hours of sleep [45], and many websites do not distinguish the appropriate sleep times for adults by age. The difference in sleep time between the ages of 30 and 50 years is not large [46]. The role of BMI can vary depending on age, although it is considered that the higher the BMI, the shorter the sleep time and the lower the BMI, the longer the sleep time [45]. People with a high BMI of 30 kg/m^2 have a slightly shorter than average sleep time [47]. People with obesity complain of insomnia or sleep disorders more often than those without obesity, and an association between obesity and increased daytime sleepiness or fatigue has been reported [48,49].

The difference between the amount of sleep measured by a Fitbit and how much sleep users feel they had is not well known. It is also not known how sleep time differs from day to day, other than rough information obtained through questionnaires. It is very important to understand how the perception of average sleep time, which reflects the quality of sleep for a certain period, differs from the daily recorded sleep time. The user's recognition can be obtained through a sleep diary or survey, which also has limitations. Conversation apps offer a potential solution in this respect, which have been widely used recently and can be used to obtain periodic and immediate feedback. Therefore, it is necessary to compare the data obtained on the same day through chatbot conversations and Fitbit data to reveal more accurate user perception differences. However, to obtain daily information, wearables and daily user feedback are required, user convenience needs to be considered, and the user response must be minimized. Accordingly, the aim of this study was to analyze the characteristics of users' time responses to sleep by comparing data obtained through Fitbit and chatbot conversations on the same day.

Methods

Recruitment

The Korean Medicine Daejeon Citizen Cohort study is being conducted over a 9-year period between 2017 and 2025, including 2000 adults living in Daejeon [50]. The cohort inclusion criteria are as follows: (1) men and women aged 30-55 years, (2) residents of Daejeon, and (3) individuals who provided informed consent. However, individuals are excluded if they (1) have been diagnosed with a malignant tumor or cardiovascular disease (myocardial infarction, angina, stroke/apoplexy); (2) are deemed to have difficulty following study instructions, such as having difficulty completing and understanding the questionnaire; or (3) determined by the researcher to be inappropriate to participate in this study. This study was conducted among the cohort participants who agreed to wear a Fitbit. For approximately 2 years, from October 10, 2020, to November 9, 2022, participants who agreed to participate in the PSQI survey, wear a Fitbit device, and have chatbot conversations were recruited, and sleep information was obtained. The participants were adults without special health problems who were in their 30s to 50s that agreed to wear a Fitbit and installed the Telegram-based chatbot app on their smartphone. The PSQI survey was conducted on the day of the hospital visit with those who wished to participate, and they were asked to wear a Fitbit device for approximately 2 weeks and to log a sleep diary through chatbot conversations.

Ethical Approval

This study was approved by the Institutional Review Board (DJDSKH-17-BM-12) of Daejeon Korean Medicine Hospital of Daejeon University and written informed consent was obtained from all participants.

PSQI Survey

The Korean version of the PSQI was used to measure sleep quality [51]. The PSQI consists of 18 questions divided into 7 subfactors to subjectively evaluate sleep in the past month. The PSQI survey inquired about the time going to bed and how long it took to fall asleep. The higher the PSQI total score, the poorer the sleep condition (range 0 to 21 points).

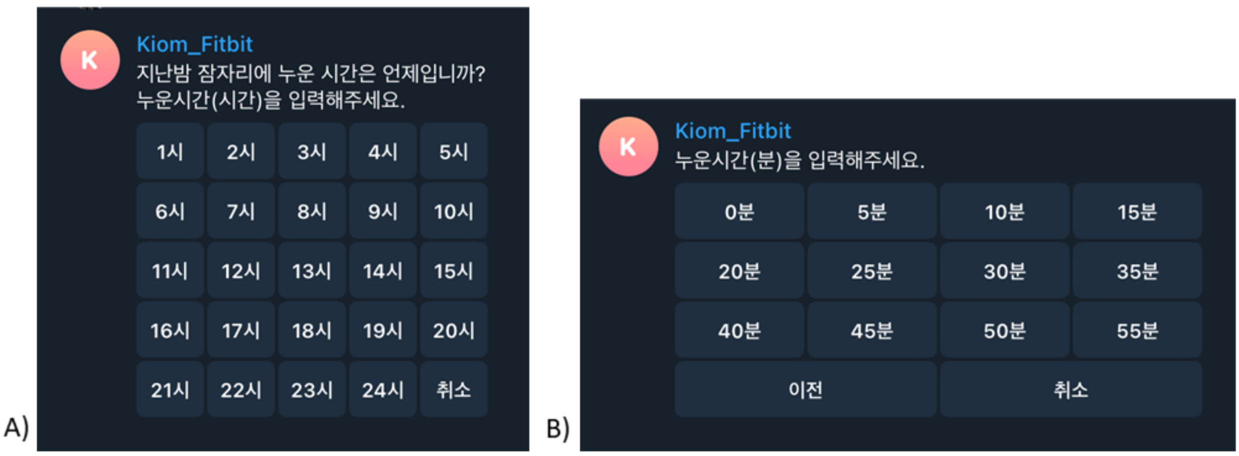
Fitbit Inspire 2 Recordings

A Fitbit Inspire 2 (Fitbit Inc, San Francisco, CA, USA) was used to obtain the sleep life log data. Participants were instructed to wear the Fitbit Inspire 2 for 24 hours a day for 2 weeks to measure the amount of activity and sleep efficiency during the day; the Fitbit could be worn on either the right or left wrist according to the participant's preference. However, the participants were instructed to take off the Fitbit when in the water for a long time, such as showering and swimming. The participants were instructed to sync their Fitbit app every morning after waking up. The data stored on the Fitbit server were collected using the Fitbit web application programming interface. The sleep information provided and collected by Fitbit included sleep variables such as the time the user fell asleep, woke up, TST, times of waking up during sleep, and sleep stages (wake, rapid eye movement, light, and deep). The administrator checked the participants' Fitbit data after 14 days, and if the Fitbit-wearing duration was less than 10 days, they were instructed to add 7 or 14 days.

Chatbot Conversation Recordings

While wearing the Fitbit, the sleep diary data of the study participants were obtained using a Telegram-based chatbot. The participants installed Telegram, added and registered a chatbot channel, and were requested to conduct conversations for 2 weeks. The participants received questions from the chatbot at 9 AM and logged sleep diaries by responding to these questions. Through chatbot conversations, the participants were asked about the time they went to bed, when they fell asleep, when they awoke, how many times they woke up during sleep, how long they actually slept, the quality of the sleep, whether there was any strenuous physical activity during the day, and how long they spent sitting. Opportunities for correction were provided, with the function of returning to the previous step during the answer and reviewing the content of the answer after the end of the conversation. The chatbot was implemented using the Python-Telegram-Bot (version 20.3) [52] (Figure 1).

Figure 1. Screenshots of the Kiom_Fitbit Telegram-based chatbot app. Question-and-answer screen about the (A) hour and (B) minute of going to bed the night before.



For answers related to time, hours and minutes were divided and entered by clicking a button. In the case of hours, 24 buttons were presented from “1:00” to “24:00,” and in the case of minutes, 12 buttons were presented from “0 min” to “55 min.” The participants were asked whether their quality of sleep was “Very good,” “Quite good,” “Quite poor,” and “Very poor,” whereas “Yes” and “No” buttons were presented for the presence or absence of strenuous physical activity. Regarding the number of awakenings, 26 buttons were presented ranging from “1” to “more than 26.” When participants were asked to wear the Fitbit, they were instructed to continue the chatbot conversations for the same period.

Statistical Analysis

The *startTime*, *endTime*, *endTime–startTime*, and *minutesAwake* values were used as the variables representing sleep onset, sleep offset, TST, and WASO from the Fitbit data [53]. The time of falling asleep, waking up, and actual sleep time (AST) in the chatbot response were compared to the Fitbit data. The calculated sleep time, obtained by subtracting the time of falling asleep from the time of waking up, was used as the TST, and the AST in the chatbot response was compared with the time obtained by subtracting the Fitbit WASO from the Fitbit TST.

For each participant, only the sleep information on the day when both the Fitbit data and chat responses were obtained was used and the mean value was used for each participant. To calculate the mean of the time values, the time information was converted into seconds; if necessary, 24 hours was added to prevent errors and later subtracted. The mean difference (SD) was used to compare the Fitbit data and chatbot responses, and box plots and Bland-Altman plots were used for visualization.

To observe the response characteristics of the participants, the values from 0 to 59 minutes were calculated in 5-minute increments from the response time values of the participants. Responses to the PSQI survey and Fitbit data were collected in close proximity in units of 5 minutes. Although the number of sleep days obtained by each participant differed, all the frequencies of the participants were cumulatively collected.

All statistical analyses and data processing were performed in Python (version 3.9) [54]. The PSQI survey results and chatbot responses were exported to Microsoft Excel files and read using the *pandas* tool library (version 1.5.3). Data imported from the Fitbit server were stored in an Oracle database, separated by a delimiter, exported as a CSV file, and read using *pandas* [55]. The *pandas* and *NumPy* packages (version 1.24.1) were used for data processing [56]. Box plots were drawn using the *matplotlib* library (version 3.6.3), Bland-Altman plots were drawn using the *statsmodelsPython* module (version 0.13.5), and *P* values were calculated using *SciPy* (version 1.10.0) [57-59].

Results

Participant Selection Conditions and Demographic Characteristics

Participants were recruited for approximately 2 years, from October 10, 2020, to November 9, 2022. A total of 736 participants participated in this study and agreed to wear the Fitbit device for 2 weeks. Among them, 731 (99.3%) participants acquired the Fitbit data and collected the main sleep data defined by the Fitbit. During the first 14-day wearing request, 589 (80.0%) participants collected sleep data for 10 or more days. By requesting 1 or 2 weeks of additional wear, 63 (8.6%) participants collected sleep data for 10 or more days. As a result, 652 (88.6%) participants acquired main sleep data for 10 or more days, while 79 (10.7%) participants obtained less than 10 days of sleep data.

Of the 652 participants, 150 provided Fitbit data for 10-14 days and 502 provided data for 15 days or more. For participants whose data were collected for more than 14 days, only data up to 14 days were used for the analysis (Table 1).

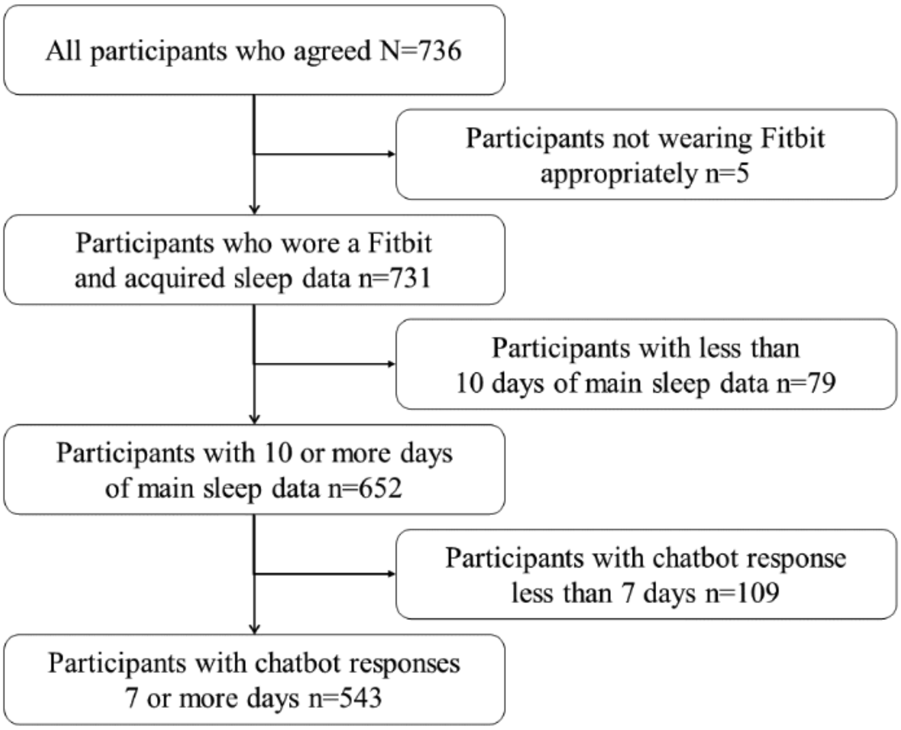
Table 1. Number of days of Fitbit sleep data provided by the study participants (N=652).

Days of sleep data	Participants, n (%)
0 (Fitbit not worn)	5 (0.7)
Less wear than requested	79 (10.7)
1-3	27 (0.4)
4-6	25 (0.4)
7-9	27 (0.4)
Worn as requested	150 (20.4)
10	13 (0.2)
11	13 (0.2)
12	20 (0.3)
13	36 (0.6)
14	68 (0.7)
Worn more than requested[a]	502 (68.2)
15	75 (11.5)
16-18	130 (19.9)
19-21	92 (14.1)
22	205 (31.4)

[a]Only data for the first 14 days were included in the analysis.

For chatbot responses, the time to go to bed, fall asleep, and wake up should be in the order of time; however, if the response value broke this order, it was considered an input error and excluded. In addition, the answerable button presented by the chatbot was set to respond to 1 of the 24 buttons from “1:00” to “24:00”; thus, responses that were considered to be wrong with respect to AM and PM were also excluded as input errors after comparison with the Fitbit data. Responses with a difference of more than 9 hours were excluded. Participants whose chatbot responses were collected for 7 or more days on the day the Fitbit main sleep data were collected were set as participants who did log chatbot responses normally. Finally, 543 (73.8%) participants’ data were analyzed, excluding 109 participants whose chatbot responses were collected over less than 7 days (Figure 2).

Figure 2. Flow of the study participants in the final analysis.



For the chat responses, participants responded to the sleep question sent at 9 AM in an average of 5 hours and 11 minutes, 288 (53.0%) responded within an average of 3 hours, and 505 (93.0%) responded within an average of 15 hours; 38 (7.0%) participants responded after an average of 15 hours (Table 2).

Table 2. Average chat response times by the study participants (N=543).

Variable	Within 3 h	Within 6 h	Within 9 h	Within 12 h	Within 15 h	After 15 h
Number of participants (%)	288 (53.0)	105 (19.3)	63 (11.6)	38 (7.0)	11 (2.0)	38 (7.0)
Cumulative number of participants (%)	288 (53.0)	393 (72.4)	456 (84.0)	494 (91.0)	505 (93.0)	543 (100.0)

Based on the PSQI total score, the 543 participants were divided into a good sleep group (5 points or less) and poor sleep group (more than 5 points). There were 318 (58.6%) participants in the good sleep group and 215 (39.6%) in the poor sleep group; this classification could not be made for 10 participants or the participants did not respond correctly to the questions. The breakdown of participants classified in each sleep group according to demographic characteristics is shown in Table 3. The majority of the participants were women; in terms of age, the greatest proportion were in their 40s, followed by 50s and 30s. According to BMI, most of the participants were in the normal group, preobese group, or obesity class I group; the BMI classification followed the Korean Society for the Study of Obesity Guidelines [60].

Table 3. Demographics of the analyzed participants and distribution of the Pittsburgh Sleep Quality Index sleep groups.

Characteristics	Total, n (%)	Good sleep group, n (%)	Poor sleep group, n (%)	Not classified, n (%)
All participants	543 (100.0)	318 (58.6)	215 (39.6)	10 (1.8)
Gender				
Men	155 (28.5)	101 (65.2)	53 (34.2)	1 (0.6)
Women	388 (71.5)	217 (55.9)	162 (41.8)	9 (2.3)
Age (decade)				
30s	115 (21.2)	71 (61.7)	43 (37.4)	1 (0.9)
40s	259 (47.7)	158 (61.0)	97 (37.5)	4 (1.5)
50s	169 (31.1)	89 (52.7)	75 (44.4)	5 (3.0)
BMI				
Underweight, <18.5	7 (1.3)	4 (57.1)	3 (42.9)	0 (0.0)
Normal, 18.5-22.9	205 (37.8)	119 (58.0)	84 (41.0)	2 (1.0)
Preobese, 23-24.9	129 (23.8)	75 (58.1)	52 (40.0)	2 (1.6)
Obesity class I, 25.0-29.9	162 (29.8)	94 (58.0)	62 (38.3)	6 (3.7)
Obesity class II, ≥30	40 (7.4)	26 (65.0)	14 (35.0)	0 (0.0)

Time Response Characteristics for the PSQI Survey and Chatbot Responses

The frequency was calculated for the minute values of the time data of the PSQI survey, chatbot conversations, and Fitbit data. The time data represent the time participants went to bed, time they fell asleep, time they woke up, and AST. However, the time taken to fall asleep in the PSQI survey was calculated by adding the time taken to fall asleep to the time of going to bed, and the AST of the Fitbit was obtained by subtracting WASO from the TST. For the PSQI responses, 533 cases were analyzed once per participant and 6276 cases, representing all sleeps of the 543 participants, were analyzed for chatbot responses and Fitbit data.

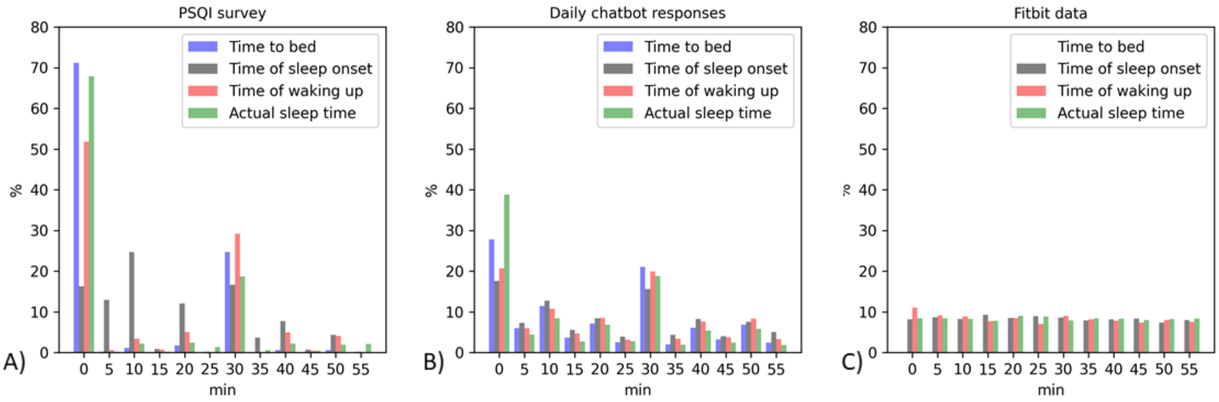
The proportion of respondents answering the questions related to sleep variables in the PSQI survey, chatbot, and recorded by Fitbit on the hour or in 30-minute intervals are presented in Table 4. The response distribution broken down per 5 minutes of participant data is shown in Figure 3. In the PSQI survey, most of the participants provided answers for the time they went to bed, followed by the actual sleep time and time they woke up, whereas only slightly more than one-third of participants provided the time they fell asleep. The percentages of participants responding to these questions in conversations with the chatbot were all much lower than those given on the PSQI survey, ranging from 33.2% for the time they fell asleep to 57.6% for the actual sleep time. Most respondents provided answers in 60- or 30-minute intervals; therefore, the lower response rate for falling asleep might be due to the fact that the PSQI adds the time taken to fall asleep to the time participants went to bed. The Fitbit data excluded the time when the participants went to bed; unlike the participants' responses, similar levels of data were collected for each time period.

Table 4. Sleep data in 30- and 60-minute increments for Pittsburg Sleep Quality Index (PSQI) surveys, conversations with the chatbot, and acquired with the Fitbit.

Variable	PSQI survey (N=533), n (%)	Chatbot (N=6276), n (%)[a]	Fitbit (N=6276), n (%)
Time went to bed	513 (95.9)	3192 (50.9)	N/A[b]
Time fell asleep	176 (32.9)	2078 (33.1)	1045 (16.7)
Time woke up	433 (81.0)	2551 (40.6)	1254 (20.0)
Actual sleep time	463 (86.6)	3616 (57.6)	1022 (16.3)

[a]These values represent the predictable aspects of chatbot design. If the sliding interface is difficult to use, the user is also likely to leave it at 0 minutes.
[b]N/A: not applicable.

Figure 3. Response distribution per 5 minutes of participant data in (A) the Pittsburg Sleep Quality Index (PSQI) survey, (B) conversations with the chatbot, and (C) Fitbit data.



Differences Between Chatbot Responses and Fitbit Data

The response distribution of participants for chatbot conversations and Fitbit data is shown in Figure 4 and Table 5. The average time the participants fell asleep by processing the chatbot conversation was 12:28:21 AM with an SD-1.96 of the difference of 65 minutes. The average time they woke up was 7:22:01 AM and the SD-1.96 was 62 minutes. The average TST calculated from the two times was 6 hours and 53 minutes (SD 49 minutes). The average time the participants fell asleep obtained by processing the Fitbit data was 12:19:42 AM (SD 70 minutes), the average time they woke up was 7:27:40 AM (SD 68 minutes), and the average TST calculated from the two times was 7 hours and 8 minutes (SD 53 minutes).

Figure 4. Participants' sleep distribution obtained by chatbot conversations and Fitbit data for (A) sleep onset, (B) sleep offset, and (C) chat AST and Fitbit TST-WASO. AST: actual sleep time; TST-WASO: total sleep time-wakefulness after sleep onset.

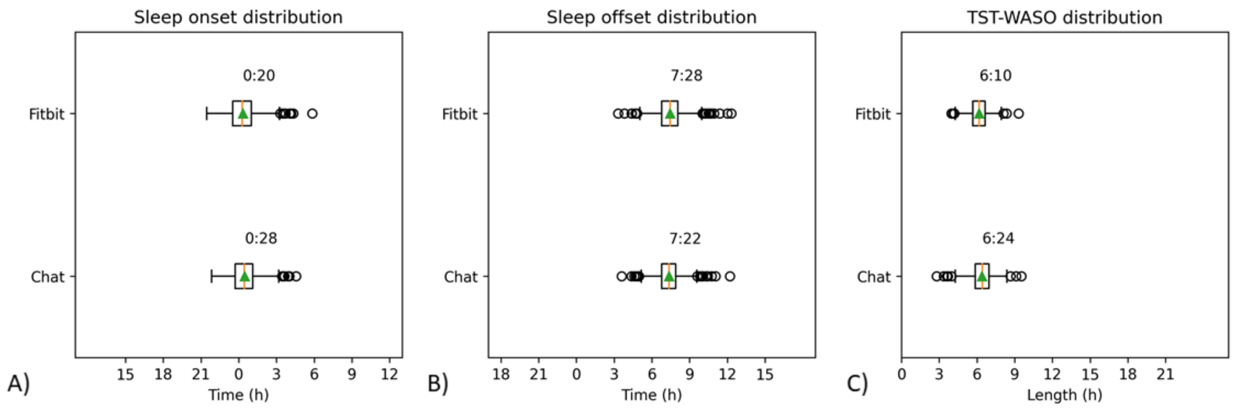


Table 5. Time (hour:minute:second) of falling asleep, waking up, and sleep time obtained from chatbot conversations compared to Fitbit data.

Variable	Chatbot, mean (SD)	Fitbit, mean (SD)	Difference, mean (SD 1.96)
Sleep onset	12:28:21 AM (65.1 min)	12:19:42 AM (69.7 min)	–0:08:39 (58.1)
Sleep off	7:22:01 AM (62.3 min)	7:27:40 AM (68.4 min)	0:05:38 (57.1)
TST[a]	6:53:41 AM (48.9 min)	7:07:58 AM (52.8 min)	0:14:17 (78.1)
AST[b]/TST-WASO[c]	6:23:45 (51.5 min)	6:10:06 AM (45.5 min)	–0:13:39 (87.0)

[a]TST: total sleep time (sleep off–sleep on).

[b]AST: chatbot actual sleep time.

[c]TST-WASO: Fitbit total sleep time-wakefulness after sleep onset.

The average AST answered by participants in the chatbot conversations was 6 hours and 24 minutes, which was 30 minutes shorter than the chat TST calculated by subtracting sleep onset from sleep offset and 44 minutes shorter than the Fitbit TST. Compared with the time minus WASO, the response time was 14 minutes longer. The mean difference between chatbot TST and AST was approximately 30 minutes and the average Fitbit WASO was approximately 58 minutes. The mean difference is a comparison between the Fitbit data and chat responses. In the case of sleep onset or offset, a negative value indicates a chatbot response in time later than the Fitbit and a positive value indicates a chatbot response in time earlier than the Fitbit. In the case of the chatbot TST or AST, a negative value indicated a longer time than the Fitbit and a positive value indicated a shorter time than the Fitbit (Table 5).

Bland-Altman statistics and plots comparing the Fitbit and chatbot responses are shown in Table 6 and Figure 5, respectively.

Table 6. Bland-Altman statistics for Fitbit and chatbot responses.

Variable	Fitbit–chatbot, mean	Lower LoA[a]	Upper LoA	*P* value[b]
Sleep onset	–8.6	–66.8	49.5	.04
Sleep off	5.6	–51.4	62.7	.16
TST-WASO[c] (AST[d])	–13.6	–100.6	73.3	<.001

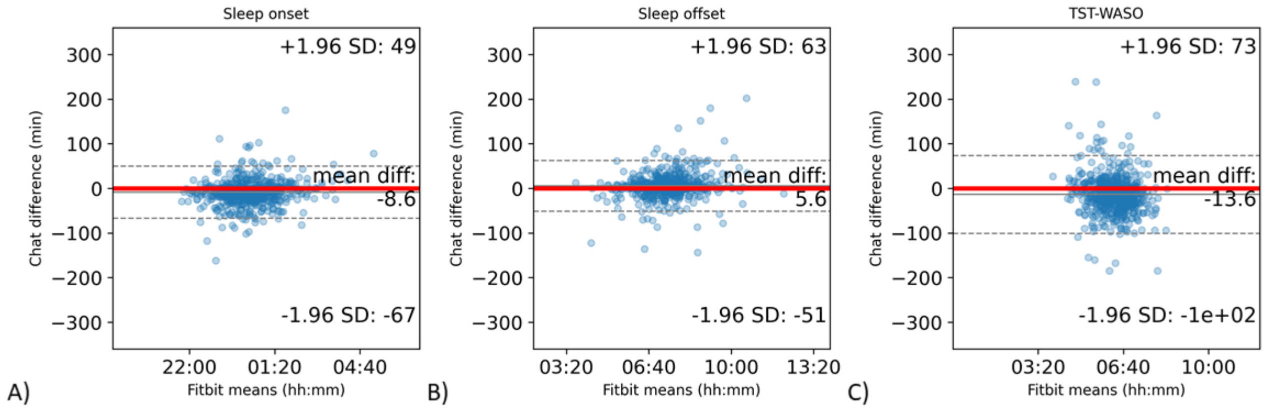
[a]LoA: limit of agreement (SD 1.96).

[b]*P* values calculated from paired *t* tests.

[c]TST-WASO: Fitbit total sleep time (sleep off–sleep onset)-wakefulness after sleep onset.

[d]AST: Chatbot actual sleep time.

Figure 5. Bland-Altman plots for the time of falling asleep, waking up, and actual sleep time. The x-axis displays the Fitbit variables and the y-axis denotes the chatbot response differences based on Fitbit data. (A) Sleep onset, (B) sleep offset, and (C) chatbot AST compared to Fitbit TST-WASO. AST: actual sleep time; TST-WASO: total sleep time-wakefulness after sleep onset.



Difference Between Chat AST and Fitbit TST-WASO According to the PSQI Survey and Demographic Information

According to the results of the PSQI survey analysis and the demographic classification, the mean values of chatbot TST and AST of participants by group (good sleep and poor sleep) were compared with the mean values of Fitbit TST and TST-WASO (Table 7). The results of the PSQI survey showed that the Fitbit measurement TST-WASO of the two groups was similar; however, the good sleep group responded to the AST for a relatively longer time than the poor sleep group. The AST levels in chats were similar for men and women, although the Fitbit TST or Fitbit TST-WASO was longer for women. By age, the TST-WASO measured by Fitbit was similar, although participants in their 30s and 40s indicated a longer AST than those in their 50s. There was no significant difference according to BMI, although the normal BMI group measured and responded to the AST longer, whereas the obese class I group measured and responded to the AST for relatively shorter periods. The number of participants in the underweight and obese class II groups was too small for comparison. The mean Fitbit WASO was 56 to 60 minutes (SD 21-24) in all groups.

Table 7. Mean differences in sleep variables determined by the chatbot and Fitbit according to sleep groups and demographic characteristics.

Variables	Chatbot			Fitbit		Difference	
	TST[a] (h:min), mean (SD 1.96)	AST[b] (h:min), mean (SD 1.96)	Difference (min), mean (SD)	TST (h:min), mean (1.96 SD)	TST-WASO[c] (h:min), mean (1.96 SD)	Fitbit–chatbot TST (min), mean (SD)	TST-WASO–AST (min), mean (SD)
PSQI[d]							
Good sleep (n=318)	6:59 (47)	6:34 (45)	–25 (61)	7:10 (51)	6:12 (43)	11 (68)	–21 (66)
Poor sleep (n=215)	6:46 (52)	6:09 (58)	–37 (77)	7:04 (56)	6:0 (49)	18 (91)	–2 (108)
Gender							
Men (n=155)	6:53 (44)	6:23 (45)	–30 (51)	6:56 (49)	5:58 (42)	3 (70)	–26 (68)
Women (n=388)	6:54 (51)	6:24 (54)	–30 (75)	7:13 (54)	6:1 (46)	19 (79)	–9 (92)
Age (decade)							
30s (n=115)	7:03 (49)	6:27 (49)	–35 (54)	7:07 (57)	6:11 (50)	4 (73)	–16 (71)
40s (n=259)	6:52 (48)	6:26 (48)	–26 (71)	7:08 (52)	6:10 (45)	16 (73)	–15 (84)
50s (n=169)	6:50 (50)	6:18 (58)	–32 (73)	7:09 (51)	6:09 (43)	19 (86)	–9 (100)
BMI (kg/m^2)							
Underweight, <18.5 (n=7)	7:14 (37)	6:39 (31)	–35 (41)	7:59 (77)	6:55 (62)	45 (159)	16 (124)
Normal, 18.5-22.9 (n=205)	7:01 (49)	6:32 (53)	–29 (66)	7:21 (51)	6:22 (44)	20 (76)	–10 (90)
Preobese, 23-24.9 (n=129)	6:50 (51)	6:22 (48)	–28 (80)	7:03 (50)	6:06 (42)	13 (64)	–16 (74)
Obesity class I, 25.0-29.9 (n=162)	6:49 (47)	6:16 (54)	–33 (66)	6:58 (52)	6:00 (45)	9 (82)	–16 (93)
Obesity class II, ≥30 (n=40)	6:45 (46)	6:15 (43)	–30 (55)	6:51 (48)	5:54 (42)	6 (81)	–22 (65)

[a]TST: total sleep time.

[b]AST: chatbot actual sleep time.

[c]TST-WASO: total sleep time-wakefulness after sleep onset.

[d]PSQI: Pittsburgh Sleep Quality Index.

The Bland-Altman statistics and plots for these comparisons are presented in Table 8 and Figure 6, respectively. The poor sleep group responded with chatbot responses that statistically matched Fitbit data, while the good sleep group reported sleeping longer than the Fitbit data.

Table 8. Bland-Altman statistics in sleep variables determined by the chatbot and Fitbit according to sleep groups and demographic characteristics.

Variables	TST-WASO[a]–AST[b] (minutes), mean difference	Lower LoA[c] (minutes)	Upper LoA (minutes)	*P* value[d]
PSQI[e]				
Good sleep (n=318)	–21.4	–87.2	44.7	<.001
Poor sleep (n=215)	–2.3	–110.5	105.9	.66
Gender				
Men (n=155)	–25.8	–93.9	42.3	<.001
Women (n=388)	–8.8	–100.6	83.0	.02
Age (decade)				
30s (n=115)	–16.1	–87.0	54.8	.01
40s (n=259)	–15.5	–99.7	68.8	<.001
50s (n=169)	–9.2	–108.7	90.3	.10
BMI (kg/m^2)				
Underweight, <18.5 (n=7)	15.8	–107.8	139.4	.59
Normal, 18.5-22.9 (n=205)	–10.2	–100.0	79.7	.04
Preobese, 23-24.9 (n=129)	–15.9	–90.3	58.6	.005
Obesity class I, 25.0-29.9 (n=162)	–15.6	–109.0	77.8	.005
Obesity class II, ≥30 (n=40)	–21.5	–86.8	43.8	.03

[a]TST-WASO: Fitbit total sleep time-wakefulness after sleep onset.

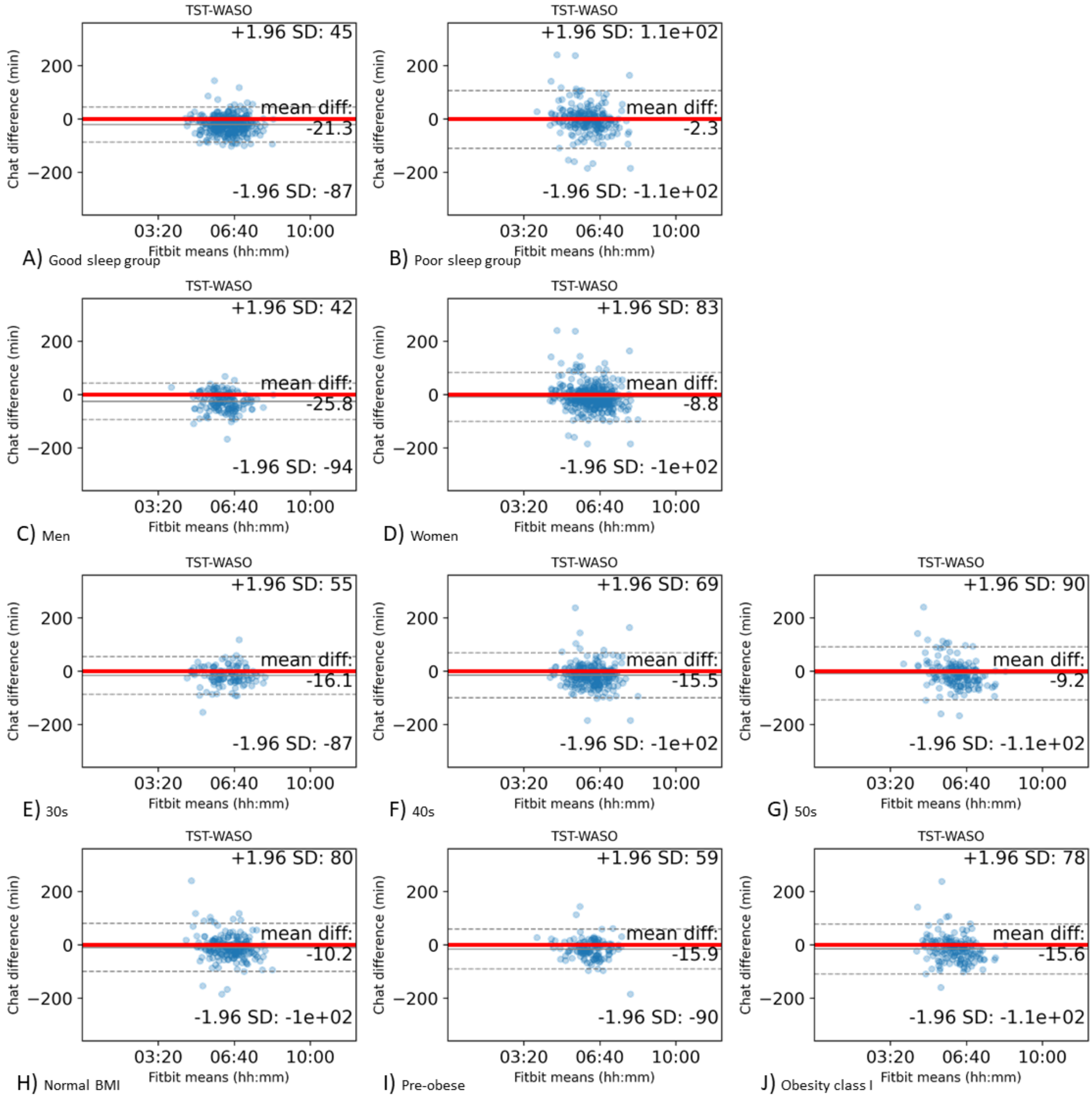
[b]AST: chatbot actual sleep time.

[c]LoA: limit of agreement; 1.96 times the SD around the bias.

[d]*P* values calculated by paired *t* tests.

[e]PSQI: Pittsburgh Sleep Quality Index.

Figure 6. Bland-Altman plots for the Fitbit TST-WASO and the chatbot AST. The x-axis displays the Fitbit TST-WASO and the y-axis denotes the chatbot response differences based on Fitbit data. (A) Good sleep group, (B) poor sleep group; (C) men, (D) women; (E) 30s age group, (F) 40s age group, (G) 50s age group; (H) normal BMI, (I) preobese, (J) obesity class I. TST-WASO: Fitbit total sleep time-wakefulness after sleep onset; AST: chatbot actual sleep time.



Differences According to Chatbot Responses for Sleep Quality

For the chatbot responses, the days on which the participants responded “Very good” or “Quite good” for the quality of sleep the previous night were considered good sleep and the days on which they responded “Very poor” or “Quite poor” were considered poor sleep. The mean AST for good sleep responses was longer than that of the Fitbit data, and both the AST response and TST-WASO measured by the Fitbit were longer than those of the chatbot. Both the mean value of AST chatbot responses and Fitbit-measured TST-WASO of the poor sleep group were relatively short. The Fitbit WASO was 58 (SD 24) minutes for the good sleep group and was 57 (SD 35) minutes for the poor sleep group (Table 9).

Table 9. Mean differences between chatbot and Fitbit sleep data according to chatbot sleep quality responses.

Sleep quality	Chatbot			Fitbit		Difference	
	TST[a] (h:min), mean (SD)	AST[b] (h:min), mean (SD)	Difference, mean (SD 1.96)	TST (h:min), mean (SD)	TST-WASO[c] (h:min), mean (SD)	Fitbit TST–chatbot TST, mean (SD 1.96)	TST-WASO–AST, mean (SD 1.96)
Good sleeps (n=524, 4380 nights)	7:09 (52)	6:43 (55)	–26 (88)	7:22 (58)	6:24 (50)	13 (84)	–19 (90)
Poor sleeps (n=432, 1896 nights)	6:23 (78)	5:40 (76)	–43 (121)	6:39 (88)	5:42 (75)	17 (152)	3 (151)

[a]TST: total sleep time.

[b]AST: chatbot actual sleep time.

[c]TST-WASO: total sleep time-wakefulness after sleep onset.

Table 10 and Figure 7 show the Bland-Altman statistics and plots, respectively. Participants who responded with poor sleeps received chatbot responses that matched the Fitbit data, whereas those who responded with good sleeps reported sleeping longer than recorded in the Fitbit data.

Table 10. Bland-Altman statistics comparing sleep time according to sleep quality responses in the chatbot.

Sleep quality	Fitbit TST-WASO[a]–chatbot AST[b] (minutes)	Lower LoA[c] (minutes)	Upper LoA (minutes)	*P* value[d]
Good sleeps (n=524, 4380 nights)	–19.2	–108.9	70.5	<.001
Poor sleeps (n=432, 1896 nights)	2.5	–148.8	153.9	.63

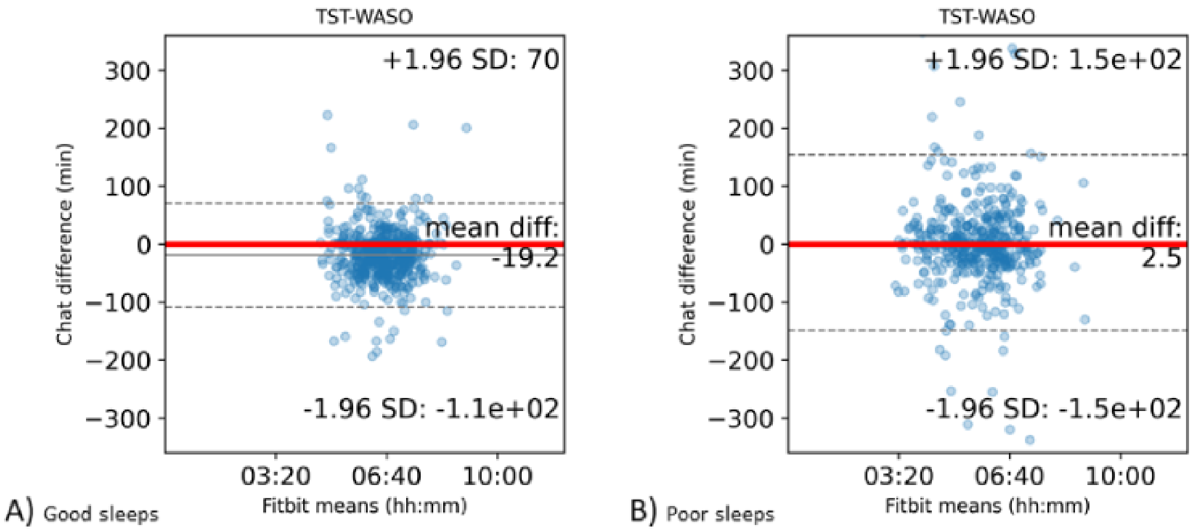
[a]TST-WASO: Fitbit total sleep time-wakefulness after sleep onset.

[b]AST: chatbot actual sleep time.

[c]LoA: limit of agreement; 1.96 SD around the bias.

[d]*P* values are based on paired *t* tests.

Figure 7. Bland-Altman plots for the Fitbit TST-WASO and the chatbot AST. The x-axis displays the Fitbit TST-WASO and the y-axis denotes the chatbot response differences based on Fitbit data. (A) good sleeps, (B) poor sleeps. AST: chatbot actual sleep time; TST-WASO: Fitbit total sleep time-wakefulness after sleep onset.



Comparison of Fitbit TST and TST-WASO to the Participants' Responses

To determine the probability of representing user responses using Fitbit data, we counted the number of sleeps in which the difference between the chatbot and Fitbit TST values was less than 30 minutes. In only 3483 (55.5%) of 6276 sleeps, the participants' chat and Fitbit TST values were within 30 minutes, and only 59.0% was covered even when modified with the mean difference and quality of sleep information revealed in the previous statistics. For chat AST and Fitbit TST-WASO, less sleep was within 30 min. Even if we expanded the period to less than 60 minutes, only approximately 74% of sleeps were applicable (Table 11).

Table 11. Comparison of Fitbit and chatbot sleep times.

Variable	Chat TST[a] and Fitbit TST			Chat AST[b] and Fitbit TST-WASO[c]		
Used variables	None	Mean difference	Mean difference by QoS[d]	None	Mean difference	Mean difference by QoS
Number of sleeps within ±30 min (%)	3483 (55.5)	3,701 (59.0)	3681 (58.7)	2960 (47.2)	3174 (50.6)	3180 (50.7)
Number of sleeps within ±60 min (%)	4711 (75.1)	4816 (76.7)	4808 (76.6)	4514 (71.9)	4658 (74.2)	4661 (74.3)

[a]TST: total sleep time.

[b]AST: chatbot actual sleep time.

[c]WASO: wakefulness after sleep onset.

[d]QoS: quality of sleep.

Discussion

Principal Results

Whether the Fitbit was worn well for the requested 14 days was based on whether sleep data were collected for more than 10 days. Of the 736 participants who wished to participate in the study, 589 (80.0%) provided their data when first requested and 63 (8.6%) provided their data in response to the second request. The data of 543 (73.8%) participants were analyzed by limiting the number of participants who responded to chats and wore the Fitbit together for more than 7 days. In the case of chatbot responses, on average, participants responded within 5 hours and 11 minutes after the chat was delivered, with 53.0% of the 543 participants responding within 3 hours and a cumulative 91.0% responding within 12 hours.

In the Fitbit data, the distribution of data in 5-minute increments was uniform, but in the case of responses to the PSQI survey or chatbot conversations, participants had a high tendency to respond on the hour and in 30-minute increments.

The mean difference between the participants' responses on sleep onset and sleep offset and their Fitbit data was within 10 minutes, indicating that each participant responded earlier or later within the range of up to 60 minutes compared to the Fitbit data. The mean difference in the TST was approximately 14 minutes longer for the Fitbit data. Considering that Fitbit calculates the TST by 9 minutes more and SOL by 4 minutes more, the TST of the participants, Fitbit data, and sleep diary collected through chatbot conversations were found to be quite consistent on average. However, there was a deviation of up to 80 minutes depending on the participant because of the tendency to respond in units of 30 minutes, perception of time according to the participant, and accuracy of the response accordingly.

The AST of the participants' responses was similar to the time obtained by subtracting WASO from the TST of the Fitbit data. On average, the AST was answered 14 minutes longer than the TST-WASO of the Fitbit data and each participant showed a maximum deviation of approximately 90 minutes.

In the PSQI survey, participants were asked to describe their sleep status the month before the chatbot conversation started. When participants were divided into the good and poor sleep groups, the sleep time measured by the Fitbit was similar, although the AST of the good sleep group was longer than that of the poor sleep group. By gender, men and women responded similarly to the AST in chatbot responses, although the sleep time measured by the Fitbit was longer in women. In addition, the TST-WASO measured by Fitbit was similar according to age group; however, the ASTs for participants in their 30s and 40s were longer than those of participants in their 50s. There was no significant difference according to BMI, although the chat AST and the Fitbit TST-WASO of the normal BMI group were longer than those of the obesity class I group.

For the chatbot responses, when we compared the sleep data answered as good sleep to those answered as poor sleep, both the AST and the TST-WASO values for the good sleep group were longer than those for the poor sleep group. The AST for responses corresponding to poor sleep was almost the same as the TST-WASO.

On average for all participants, the chat response and the Fitbit data seemed to match; however, the Fitbit data could not represent the participants' responses due to the individual differences of the participants. When tested within 30 minutes and within 60 minutes, the probability that the participant's response was close to the Fitbit recorded data was in the 50% and 70% range, respectively.

Limitations

Changes in Fitbit's algorithm were not considered in this study. A Fitbit product of the same name was used; however, possible changes to the hardware or software were not considered. In addition, we cannot guarantee that the study participants wore their Fitbits and responded to the chatbot themselves, and a confirmation process for this was not included in the analysis. Sleep determined by the Fitbit was targeted as the main sleep source, and differences in sleep due to naps or occupational characteristics were not considered. We also did not take into account whether the Fitbit was worn on the participant's dominant wrist, which could affect accuracy.

As a result of obtaining the mean difference between the Fitbit data and chatbot responses by sequentially increasing the number of days from the 1st to the 14th, the change in the mean difference according to the period was sufficiently small after approximately 7 days. Using data from at least 7 days was considered appropriate for analysis of the mean difference and SD. Considering the fatigue from continuous Fitbit wear and repeated chatting, only data from up to 14 days were used in this analysis.

In the chatbot response, if the participants fell asleep earlier than the time they went to bed or the time they woke up was earlier than the time they went to bed or fell asleep, it was regarded as an input error and was excluded. The difference between the Fitbit-measured time and reply time was very large, at approximately 12 hours, and data that could be seen as AM and PM input errors were also excluded as errors. Therefore, it is necessary to supplement the user interface to prevent user input errors. Since we did not implement a slider-like interface that allows minute input with a single touch, we did not receive every minute input. Depending on the chatbot interface, the results may vary to some extent.

The PSQI survey was conducted on sleep status during the month before the Fitbit and chatbot conversations. Under the premise that the sleep information obtained through the PSQI survey did not change rapidly, it was expected that the sleep state, based on analysis of the PSQI survey, would be applied for the next month; thus, sleep states that could not be reflected based on this assumption were not considered.

However, previous studies have shown high test-retest reliability of the PSQI. One study found that within-class correlations ranged from 0.709 to 0.813 in a retest with 30 health care workers after 2 weeks, when reliability was considered acceptable if within-class correlations were greater than 0.70 [61]. Various studies have also demonstrated high test-retest reliability of the PSQI score after 2 days or 2 to 4 weeks [62-66].

Conclusions

There was a greater tendency to respond in 30-minute increments in the PSQI survey asking about the status of the past month than in the chatbot conversation asking about daily status. This tendency can be large when asking about the average value of past periods, which are difficult to specify, and small when asking about daily values. In addition, because this tendency is relatively small at the time of falling asleep or waking up, it can be expected to be smaller when asking about the easy-to-remember value for each day. This tendency may be greater in questions about situations that are difficult to remember or specify, such as when you went to bed and for how long you were actually asleep.

The results did not change when only the sleep data for which the chat response time was answered within 6 and 12 hours were used to determine whether the time taken to respond to the chatbot was related to the correct answer. There was no significant difference between the previous day's sleep information answered in the morning after waking up and sleep information answered in the afternoon or evening. To reduce the causes of large interindividual variation, it is necessary to include methods that can help the process by requiring clearer queries and more accurate answers.

When the chatbot responses were compared with the sleep-related times obtained from the study participants' Fitbit data, the mean difference for all participants was approximately 10 minutes. Considering the response rate at 30-minute intervals in chatbot responses, it can be considered that participants' responses, on average, represented sleep time information similar to that recorded by Fitbit. Considering the distribution of PSQI sleep quality and the demographic characteristics of the study participants, the AST subjectively assessed by the participants was relatively longer than the Fitbit TST-WASO in the group with good sleep quality than in the group with poor sleep quality.

Depending on the participant, there was a deviation of up to 60-90 minutes, and it was difficult to predict whether the individual response time was earlier or later than the Fitbit data or whether the response time was short or long. This deviation may occur because each user's sleep characteristics and response tendencies in chatbot conversations are different. It was also difficult to predict whether these differences were related to the perception of waking time during sleep, depth of sleep, or quality of sleep. It may be meaningful to provide this information or to clarify the difference between people who sleep for short periods but feel that they had good-quality sleep and people who sleep for long periods but feel that they had poor-quality sleep. It would be essential to analyze whether individuals with deep sleep patterns tend to report shorter sleep durations and whether those with shallow sleep patterns tend to report longer sleep durations to achieve similar levels of satisfaction.

If an individual's perceived sleep time is important, their report will still be meaningful, and if their cooperation is possible, daily diary reporting will be effective. In addition, it is expected that conversations through chatbots will be able to obtain this information efficiently. To provide a clearer conclusion on the difference in user perception, it is necessary to improve the quality of sleep or depth recognition performance of wearables and to establish appropriate methods to reduce the deviation in user responses.

Acknowledgments

This study was supported by the "Development of Korean Medicine Original Technology for Preventive Treatment based on Integrative Big Data" grant for the Korea Institute of Oriental Medicine (KSN1731121). The authors thank all participants for their cooperation and also thank Bok-Nam Seo and Kyoungsik Jeong for assisting with data collection and data processing.

Authors' Contributions

HJ, YB, SM, and SL conceived the study. HJ, YS, SS, and IK contributed to implementation and data acquisition. HJ, YS, SS, YB, SM, HK, JK, and SL contributed to analysis and interpretation. HJ, YS, SS, and SL wrote the manuscript. All authors contributed to manuscript revisions. All authors approved the submitted version.

Conflicts of Interest

None declared.

References

1. Umetani T, Ishii M, Tamura Y, Saiwaki N, Yokoyama K. Change detection of sleeping conditions based on multipoint ambient sensing of comforter on bed. 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 17-21, 2018; Honolulu, Hawaii p. 4997-5001. [doi: 10.1109/embc.2018.8513477]
2. Feustel C, Shu N, Clifford G, Anderson D, Zimring C. Practical high-fidelity sensing of the sleep environment in the home. 2022 Presented at: PETRA '22: 15th International Conference on PErvasive Technologies Related to Assistive Environments; June 29-July 1, 2022; Corfu, Greece p. 155-158. [doi: 10.1145/3529190.3529219]
3. Hirshkowitz M. Polysomnography and beyond. In: Kryger MH, Roth T, Dement WC, editors. Principles and practice of sleep medicine. 6th ed. Amsterdam: Elsevier; 2017:1564-1566.
4. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001 Sep;2(5):389-396. [doi: 10.1016/s1389-9457(00)00098-8] [Medline: 14592388]
5. Martin JL, Hakim AD. Wrist actigraphy. Chest 2011 Jun;139(6):1514-1527 [FREE Full text] [doi: 10.1378/chest.10-1872] [Medline: 21652563]
6. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557 [FREE Full text] [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
7. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med 2018 Jul 15;14(7):1209-1230 [FREE Full text] [doi: 10.5664/jcsm.7228] [Medline: 29991438]
8. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 01;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
9. Sadeh A, Hauri PJ, Kripke DF, Lavie P. The role of actigraphy in the evaluation of sleep disorders. Sleep 1995 May;18(4):288-302. [doi: 10.1093/sleep/18.4.288] [Medline: 7618029]
10. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011 Aug;15(4):259-267. [doi: 10.1016/j.smrv.2010.10.001] [Medline: 21237680]
11. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
12. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med 2018 Jul 15;14(7):1231-1237 [FREE Full text] [doi: 10.5664/jcsm.7230] [Medline: 29991437]
13. Tsanas A, Woodward E, Ehlers A. Objective characterization of activity, sleep, and circadian rhythm patterns using a wrist-worn actigraphy sensor: insights into posttraumatic stress disorder. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e14306 [FREE Full text] [doi: 10.2196/14306] [Medline: 32310142]
14. Reilly ED, Robinson SA, Petrakis BA, Gardner MM, Wiener RS, Castaneda-Sceppa C, et al. Mobile intervention to improve sleep and functional health of Veterans with insomnia: randomized controlled trial. JMIR Form Res 2021 Dec 09;5(12):e29573 [FREE Full text] [doi: 10.2196/29573] [Medline: 34889746]
15. Ibáñez V, Silva J, Cauli O. A survey on sleep assessment methods. PeerJ 2018;6:e4849 [FREE Full text] [doi: 10.7717/peerj.4849] [Medline: 29844990]
16. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
17. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
18. Bian J, Guo Y, Xie M, Parish AE, Wardlaw I, Brown R, et al. Exploring the association between self-reported asthma impact and Fitbit-derived sleep quality and physical activity measures in adolescents. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e105 [FREE Full text] [doi: 10.2196/mhealth.7346] [Medline: 28743679]
19. Weatherall J, Paprocki Y, Meyer TM, Kudel I, Witt EA. Sleep tracking and exercise in patients with type 2 diabetes mellitus (Step-D): pilot study to determine correlations between Fitbit data and patient-reported outcomes. JMIR Mhealth Uhealth 2018 Jun 05;6(6):e131 [FREE Full text] [doi: 10.2196/mhealth.8122] [Medline: 29871856]

20. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
21. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit Activity Trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
22. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-Analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
23. Katzan I, Schuster A, Kinzy T. Physical activity monitoring using a Fitbit device in ischemic stroke patients: prospective cohort feasibility study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e14494 [FREE Full text] [doi: 10.2196/14494] [Medline: 33464213]
24. Balbim GM, Marques IG, Marquez DX, Patel D, Sharp LK, Kitsiou S, et al. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. JMIR Mhealth Uhealth 2021 Mar 01;9(3):e25289 [FREE Full text] [doi: 10.2196/25289] [Medline: 33646135]
25. Rayward AT, Vandelanotte C, Van Itallie A, Duncan MJ. The association between logging steps using a website, app, or Fitbit and engaging with the 10,000 steps physical activity program: observational study. J Med Internet Res 2021 Jun 18;23(6):e22151 [FREE Full text] [doi: 10.2196/22151] [Medline: 34142966]
26. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465 [FREE Full text] [doi: 10.2196/20465] [Medline: 33038869]
27. Lee XK, Chee NIYN, Ong JL, Teo TB, van Rijn E, Lo JC, et al. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. J Clin Sleep Med 2019 Sep 15;15(9):1337-1346 [FREE Full text] [doi: 10.5664/jcsm.7932] [Medline: 31538605]
28. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. Sleep 2020 Jul 13;43(7):zsaa045 [FREE Full text] [doi: 10.1093/sleep/zsaa045] [Medline: 32215550]
29. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
30. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019 Jun 19;7(6):e13084 [FREE Full text] [doi: 10.2196/13084] [Medline: 31219048]
31. Thota D. Evaluating the relationship between Fitbit sleep data and self-reported mood, sleep, and environmental contextual factors in healthy adults: pilot observational cohort study. JMIR Form Res 2020 Sep 29;4(9):e18086 [FREE Full text] [doi: 10.2196/18086] [Medline: 32990631]
32. Stucky B, Clark I, Azza Y, Karlen W, Achermann P, Kleim B, et al. Validation of Fitbit Charge 2 sleep and heart rate estimates against polysomnographic measures in shift workers: naturalistic study. J Med Internet Res 2021 Oct 05;23(10):e26476 [FREE Full text] [doi: 10.2196/26476] [Medline: 34609317]
33. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e13384 [FREE Full text] [doi: 10.2196/13384] [Medline: 31172956]
34. Stone JD, Rentz LE, Forsey J, Ramadan J, Markwald RR, Finomore VS, et al. Evaluations of commercial sleep technologies for objective monitoring during routine sleeping conditions. Nat Sci Sleep 2020;12:821-842 [FREE Full text] [doi: 10.2147/NSS.S270705] [Medline: 33149712]
35. Gashi S, Alecci L, Lascio ED, Debus ME, Gasparini F, Santini S. The role of model personalization for sleep stage and sleep quality recognition using wearables. IEEE Pervasive Comput 2022 Apr 1;21(2):69-77. [doi: 10.1109/mprv.2022.3164334]
36. Pan Q, Brulin D, Campo E. Current status and future challenges of sleep monitoring systems: systematic review. JMIR Biomed Eng 2020 Aug 26;5(1):e20921 [FREE Full text] [doi: 10.2196/20921]
37. Fortunato M, Adusumalli S, Chokshi N, Harrison J, Rareshide C, Patel M. Usability of wearable devices to remotely monitor sleep patterns among patients with ischemic heart disease: observational study. JMIR Form Res 2020 Apr 07;4(4):e14508 [FREE Full text] [doi: 10.2196/14508] [Medline: 32254044]
38. Tobin SY, Williams PG, Baron KG, Halliday TM, Depner CM. Challenges and opportunities for applying wearable technology to sleep. Sleep Med Clin 2021 Dec;16(4):607-618. [doi: 10.1016/j.jsmc.2021.07.002] [Medline: 34711385]
39. Benedetti D, Menghini L, Vallat R, Mallett R, Kiss O, Faraguna U, et al. Call to action: an open-source pipeline for standardized performance evaluation of sleep-tracking technology. Sleep 2023 Feb 08;46(2):zsac304. [doi: 10.1093/sleep/zsac304] [Medline: 36611112]
40. Burgard SA, Ailshire JA. Gender and time for sleep among U.S. adults. Am Sociol Rev 2013 Feb;78(1):51-69 [FREE Full text] [doi: 10.1177/0003122412472048] [Medline: 25237206]

41. Krishnan V, Collop NA. Gender differences in sleep disorders. Curr Opin Pulm Med 2006 Nov;12(6):383-389. [doi: 10.1097/01.mcp.0000245705.69440.6a] [Medline: 17053485]
42. Ruggiero AR, Peach HD, Zendels P, Gaultney JF. Sleep attitudes as a predictor of sleep outcomes: a secondary data analysis. Health Psychol Behav Med 2020 Dec 08;8(1):623-635 [FREE Full text] [doi: 10.1080/21642850.2020.1852939] [Medline: 34040889]
43. Tribl GG, Schmeiser-Rieder A, Rosenberger A, Saletu B, Bolitschek J, Kapfhammer G, et al. Sleeping habits in the Austrian population. Sleep Med 2002 Jan;3(1):21-28. [doi: 10.1016/s1389-9457(01)00117-4] [Medline: 14592249]
44. Tang J, Liao Y, Kelly BC, Xie L, Xiang Y, Qi C, et al. Gender and regional differences in sleep quality and insomnia: a general population-based study in Hunan Province of China. Sci Rep 2017 Mar 06;7:43690. [doi: 10.1038/srep43690] [Medline: 28262807]
45. Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, et al. Recommended amount of sleep for a healthy adult: a Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. Sleep 2015 Jun 01;38(6):843-844 [FREE Full text] [doi: 10.5665/sleep.4716] [Medline: 26039963]
46. Grandner MA, Schopfer EA, Sands-Lincoln M, Jackson N, Malhotra A. Relationship between sleep duration and body mass index depends on age. Obesity 2015 Dec;23(12):2491-2498 [FREE Full text] [doi: 10.1002/oby.21247] [Medline: 26727118]
47. Jaiswal SJ, Quer G, Galarnyk M, Steinhubl SR, Topol EJ, Owens RL. Association of sleep duration and variability with body mass index: sleep measurements in a large US population of wearable sensor users. JAMA Intern Med 2020 Dec 01;180(12):1694-1696 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2834] [Medline: 32926073]
48. Pearson NJ, Johnson LL, Nahin RL. Insomnia, trouble sleeping, and complementary and alternative medicine: analysis of the 2002 national health interview survey data. Arch Intern Med 2006 Sep 18;166(16):1775-1782. [doi: 10.1001/archinte.166.16.1775] [Medline: 16983058]
49. Vgontzas AN, Bixler EO, Tan T, Kantner D, Martin LF, Kales A. Obesity without sleep apnea is associated with daytime sleepiness. Arch Intern Med 1998 Jun 22;158(12):1333-1337. [doi: 10.1001/archinte.158.12.1333] [Medline: 9645828]
50. Baek Y, Seo B, Jeong K, Yoo H, Lee S. Lifestyle, genomic types and non-communicable diseases in Korea: a protocol for the Korean Medicine Daejeon Citizen Cohort study (KDCC). BMJ Open 2020 Apr 09;10(4):e034499 [FREE Full text] [doi: 10.1136/bmjopen-2019-034499] [Medline: 32276954]
51. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. Sleep Breath 2012 Sep;16(3):803-812. [doi: 10.1007/s11325-011-0579-9] [Medline: 21901299]
52. Python Telegram Bot. URL: https://python-telegram-bot.org [accessed 2023-10-22]
53. Get sleep log by date. FitBit developer. URL: https://dev.fitbit.com/build/reference/web-api/sleep/get-sleep-log-by-date/ [accessed 2023-10-22]
54. Python. URL: https://www.python.org [accessed 2023-10-22]
55. Pandas. URL: https://pandas.pydata.org [accessed 2023-10-22]
56. NumPy. URL: https://numpy.org [accessed 2023-10-22]
57. Matplotlib. URL: https://matplotlib.org [accessed 2023-10-22]
58. Statsmodels. URL: https://www.statsmodels.org [accessed 2023-10-22]
59. SciPy. URL: https://scipy.org [accessed 2023-10-22]
60. Seo MH, Lee W, Kim SS, Kang J, Kang J, Kim KK, Committee of Clinical Practice Guidelines, Korean Society for the Study of Obesity (KSSO). 2018 Korean Society for the Study of Obesity Guideline for the Management of Obesity in Korea. J Obes Metab Syndr 2019 Mar;28(1):40-45 [FREE Full text] [doi: 10.7570/jomes.2019.28.1.40] [Medline: 31089578]
61. Wang L, Wu Y, Lin Y, Wang L, Zeng Z, Xie X, et al. Reliability and validity of the Pittsburgh Sleep Quality Index among frontline COVID-19 health care workers using classical test theory and item response theory. J Clin Sleep Med 2022 Feb 01;18(2):541-551 [FREE Full text] [doi: 10.5664/jcsm.9658] [Medline: 34534069]
62. Tzeng JI, Fu Y, Lin C. Validity and reliability of the Taiwanese version of the Pittsburgh Sleep Quality Index in cancer patients. Int J Nurs Stud 2012 Jan;49(1):102-108. [doi: 10.1016/j.ijnurstu.2011.08.004] [Medline: 21924421]
63. Kotronoulas GC, Papadopoulou CN, Papapetrou A, Patiraki E. Psychometric evaluation and feasibility of the Greek Pittsburgh Sleep Quality Index (GR-PSQI) in patients with cancer receiving chemotherapy. Support Care Cancer 2011 Nov 23;19(11):1831-1840. [doi: 10.1007/s00520-010-1025-4] [Medline: 20972588]
64. Passos MHP, Silva HA, Pitangui ACR, Oliveira VMA, Lima AS, Araújo RC. Reliability and validity of the Brazilian version of the Pittsburgh Sleep Quality Index in adolescents. J Pediatr 2017;93(2):200-206 [FREE Full text] [doi: 10.1016/j.jped.2016.06.006] [Medline: 27520731]
65. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
66. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. J Psychosom Res 2002 Sep;53(3):737-740. [doi: 10.1016/s0022-3999(02)00330-6] [Medline: 12217446]

Abbreviations

AST: actual sleep time
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
SOL: sleep onset latency
TST: total sleep time
WASO: wakefulness after sleep onset

Edited by L Buis; submitted 24.05.23; peer-reviewed by Z Liang, Z Zhang; comments to author 12.08.23; revised version received 11.09.23; accepted 18.10.23; published 21.11.23.

Please cite as:
Jang H, Lee S, Son Y, Seo S, Baek Y, Mun S, Kim H, Kim I, Kim J
Exploring Variations in Sleep Perception: Comparative Study of Chatbot Sleep Logs and Fitbit Sleep Data
JMIR Mhealth Uhealth 2023;11:e49144
URL: https://mhealth.jmir.org/2023/1/e49144
doi:10.2196/49144
PMID:37988148

Review

Augmented Reality in Real-time Telemedicine and Telementoring: Scoping Review

Alana Dinh[1], BA; Andrew Lukas Yin[2], MBA, MD; Deborah Estrin[3], PhD; Peter Greenwald[4], MS, MD; Alexander Fortenko[4], MPH, MD

[1]Medical College, Weill Cornell Medicine, New York, NY, United States

[2]Department of Internal Medicine, Weill Cornell Medicine, New York, NY, United States

[3]Department of Computer Science, Cornell Tech, New York, NY, United States

[4]Emergency Medicine, NewYork-Presyterian Hospital, New York, NY, United States

Corresponding Author:
Alana Dinh, BA
Medical College
Weill Cornell Medicine
1300 York Ave
New York, NY, 10021
United States
Phone: 1 212 746 1067
Email: ald4006@med.cornell.edu

Abstract

Background: Over the last decade, augmented reality (AR) has emerged in health care as a tool for visualizing data and enhancing simulation learning. AR, which has largely been explored for communication and collaboration in nonhealth contexts, could play a role in shaping future remote medical services and training. This review summarized existing studies implementing AR in real-time telemedicine and telementoring to create a foundation for health care providers and technology developers to understand future opportunities in remote care and education.

Objective: This review described devices and platforms that use AR for real-time telemedicine and telementoring, the tasks for which AR was implemented, and the ways in which these implementations were evaluated to identify gaps in research that provide opportunities for further study.

Methods: We searched PubMed, Scopus, Embase, and MEDLINE to identify English-language studies published between January 1, 2012, and October 18, 2022, implementing AR technology in a real-time interaction related to telemedicine or telementoring. The search terms were “augmented reality” OR “AR” AND “remote” OR “telemedicine” OR “telehealth” OR “telementoring.” Systematic reviews, meta-analyses, and discussion-based articles were excluded from analysis.

Results: A total of 39 articles met the inclusion criteria and were categorized into themes of patient evaluation, medical intervention, and education. In total, 20 devices and platforms using AR were identified, with common features being the ability for remote users to annotate, display graphics, and display their hands or tools in the local user’s view. Common themes across the studies included consultation and procedural education, with surgery, emergency, and hospital medicine being the most represented specialties. Outcomes were most often measured using feedback surveys and interviews. The most common objective measures were time to task completion and performance. Long-term outcome and resource cost measurements were rare. Across the studies, user feedback was consistently positive for perceived efficacy, feasibility, and acceptability. Comparative trials demonstrated that AR-assisted conditions had noninferior reliability and performance and did not consistently extend procedure times compared with in-person controls.

Conclusions: Studies implementing AR in telemedicine and telementoring demonstrated the technology’s ability to enhance access to information and facilitate guidance in multiple health care settings. However, AR’s role as an alternative to current telecommunication platforms or even in-person interactions remains to be validated, with many disciplines and provider-to-nonprovider uses still lacking robust investigation. Additional studies comparing existing methods may offer more insight into this intersection, but the early stage of technical development and the lack of standardized tools and adoption have hindered the conduct of larger longitudinal and randomized controlled trials. Overall, AR has the potential to complement and advance the capabilities of remote medical care and learning, creating unique opportunities for innovator, provider, and patient involvement.

(JMIR Mhealth Uhealth 2023;11:e45464) doi:10.2196/45464

KEYWORDS

augmented reality; telemedicine; telehealth; telementoring; teleguidance; telecommunication; teleconsultation; telecollaboration; scoping review; mobile phone

Introduction

Background

Augmented reality (AR) is an emerging technology that can enhance how the real world is experienced by the user. Compared with virtual reality (VR), in which the user is immersed in a completely synthesized world, AR combines both the virtual and real by overlaying the external world with computer-generated sensory data such as audio, video, and graphics. AR technology, often accessed through head-mounted devices (HMDs) or software on personal devices, can be used to display information and virtual objects that facilitate learning and navigation through tasks in the real world [1,2].

In medicine, VR and AR have been explored in educational, diagnostic, and treatment settings, with an increasing number of publications in the last decade [3-6]. A 2012 to 2017 review of 338 original studies using AR in medicine, most related to surgery and simulation learning, estimated the technology readiness level of AR to be at the stage of a prototype that has yet to be completed and tested in its intended environment [7]. AR has since appeared in the literature across many specialized fields, with reviews since 2019 describing AR technology in emergency medicine (EM) [8], dermatology [9], radiology [10,11], orthopedics [12], nursing [13], and many more.

A promising use of AR technology is in remote collaboration, an application seen in many industry- and engineering-related tasks over the last 2 decades [14,15]. As COVID-19 pushed health care to explore remote health solutions, providers, caregivers, and students have become increasingly aware of technology's role in enabling access to care [16-18]. A cohort study of 36.5 million individuals in the United States found that 23.6% of ambulatory visits in 2020 were billed as telehealth visits compared with 0.3% in 2019 [19]. Although videoconferencing programs allow health care workers to remotely connect with each other, trainees, and patients, current systems limit the extent of care that can be delivered during such interactions [20]. Innovations using AR technology offer an opportunity to expand real-time remote health services such as consultation and telesurgery [21]. Recent literature on AR includes studies on caregiver perspectives on the technology [22], frameworks for AR-assisted remote medical communication [23], and remote health care delivery devices incorporating AR [24].

Objectives

Remote medical communication is a defining feature of *telemedicine*, a concept that first arose with the use of telephones to share information across hospital systems [25]. Since the conception of the internet and personal devices, remote health care visits and interventions have become possible, with the scope of care expanding as technological advances are introduced. Currently, there is no literature summarizing the applications of AR in *synchronous telemedicine*, defined as the use of electronic devices for real-time communication in health care services related to patient encounters, treatment, and consultation [25,26]. *Telementoring*, a subcategory of telemedicine, is the real-time remote guidance of health care procedures or skills. Telementoring also plays a role in granting remote access to care and expertise, enabling the sharing of specialized knowledge and education [27]. Similar to telemedicine, it is also a topic scarcely studied in relation to AR and has the potential to evolve with technological innovation [28]. By exploring how AR is used for real-time telemedicine and telementoring, this review could better inform health care providers and developers of AR's future potential and current limitations.

Methods

Scoping Method

Scoping reviews entail a systematic selection of literature with the purpose of examining the extent and nature of an area of interest [29,30]. Compared with systematic reviews, scoping studies allow for the integration of a range of study designs, especially in fields with emerging evidence that may lack randomized controlled trials (RCTs). By mapping the existing evidence of AR in telemedicine, the scoping approach allows for the identification of gaps that may inform future studies and innovations.

Research Questions

Which devices and platforms using AR have been studied in the published literature in the context of real-time telemedicine and remote education? In which areas of medicine have these been integrated and for what purposes? How are outcomes evaluated and what variables have yet to be measured? What are the overall findings of existing studies?

Identifying Studies in the Literature

The literature was reviewed using PubMed, Scopus, Embase, and MEDLINE for articles or trials published from January 1, 2012, to October 18, 2022, with search queries submitted and articles accessed on October 18, 2022. The search terms were "augmented reality" OR "AR" AND "remote" OR "telemedicine" OR "telehealth" OR "telementoring." The PubMed search was performed using article titles and abstracts. The Scopus search was performed using article titles, abstracts, and keywords, with articles and conference papers in the areas of "medicine," "health professions," and "nursing" included.

The Embase search was performed using article titles, abstracts, and keywords, with articles and conference papers included. The MEDLINE search was also performed using article titles, abstracts, and keywords. Only articles available in English were included.

Article Selection

Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, the titles and abstracts of all articles were independently reviewed by 2 researchers (AD and AF) for relevance to AR and real-time communication between separated individuals. Articles were excluded if they were unrelated to medicine or were reviews or discussions. Any articles that were included by one reviewer but not the other were included in the full-text screening, which was also performed independently by the 2 researchers. As this review focused on the use of AR rather than the development of related equipment or software, articles were excluded if the technological design, rather than the implementation, was the focus or if the technology was not intended for remote interaction as described by the inclusion criteria. Articles that included both technological design and implementation data were included, with the review focusing on the latter. Articles that described mixed reality devices capable of both AR and VR were included if the implementation primarily used and studied AR features. Reviews, perspectives, discussion-based articles, and study proposals without results were excluded. Correspondence was sent to the authors of articles for which the full text was not available; a lack of response resulted in the exclusion of these articles. Any disagreements regarding an article's inclusion were resolved through discussion between the 2 reviewers.

Data Charting

The articles were reviewed based on the context in which the AR was implemented. Articles describing mixed reality devices were analyzed for data relevant to AR use and not VR. Unique devices and platforms using AR across the articles were identified. The articles were later grouped into one of the 3 identified areas: patient evaluation, medical intervention, and education. "Patient evaluation" included articles that described the examination of patients and processes that obtained information for clinical decision-making. "Medical intervention" included articles that described procedures related to the initiation or provision of therapy. "Education" included articles that described the mentorship or training of a less experienced individual for a task or procedure. Subgroups for surgical versus nonsurgical tasks in the latter 2 groups were created. Articles that fell into more than one of the 3 identified areas were organized in the Results section of this paper based on which area was the primary focus. Finally, common objective and subjective end variables discussed in the articles were identified.

Collation and Summary

For this scoping review, we first provide an overview of the devices and platforms that use AR and the types of tasks in which they appear. We then summarize the implementation and measurement of these AR-capable tools within 3 areas: patient evaluation, medical intervention, and education. Finally, we review the common methodologies and end variables observed across the studies.

Results

Overview

The PubMed, Scopus, Embase, and MEDLINE searches yielded 298, 195, 187, and 274 articles, respectively. This totaled 954 articles, 558 (58.5%) of which were identified as duplicates. The abstracts and titles of the remaining 396 articles were reviewed, with 62 (15.7%) identified as meeting the inclusion criteria. In total, 5% (3/62) of the articles, for which the full text was not available, were excluded after no response was received from the original authors. A total of 39 articles were included following full-text screening. The selection process is depicted in Figure 1. The publication years of the selected articles spanned 2014 to 2022, with 64% (25/39) published in the last 3 years.

From the included articles, 20 unique devices and platforms with AR features were identified, with 10 (50%) being commercial HMDs in which AR features were projected before the wearer's eyes and another 4 (20%) being "virtual presence"–type platforms in which a remote viewer can superimpose video of their hand or tools over the live stream of a local site or procedure; the hybrid video is accessed by both local and remote users via a smartphone, tablet, computer, or monitor. The remaining 30% (6/20) involved systems with no commercial HMDs or virtual presence—these included a smartphone app and systems built specifically for tele-ultrasonography, physical rehabilitation, and surgical telementoring. Overall, all the identified devices and platforms (20/20, 100%) allowed remote individuals to view the perspective or environment of the local user. Common AR features of these devices included annotation (12/20, 60%) and graphical overlay over the local user's view, specifically 2D or 3D images (9/20, 45%) and the remote viewer's hands or tools (8/20, 40%). The identified devices and platforms are listed in Table 1, with similar device models included in the same row.

Figure 1. Selection process for the articles. AR: augmented reality.

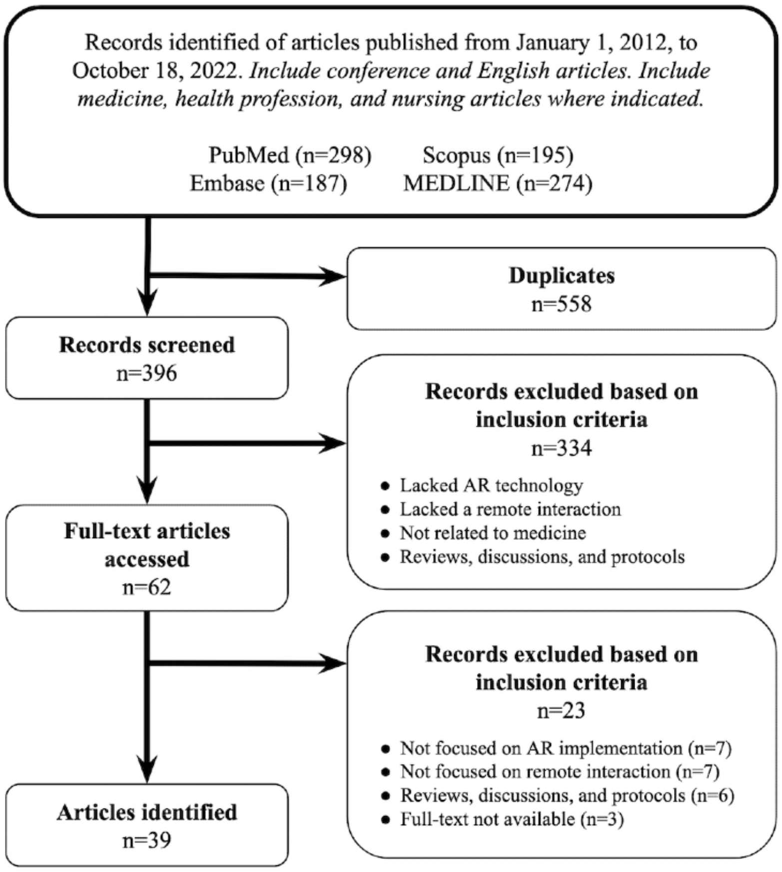


Table 1. Overview of devices and features.

Tool (year)	Communication features	Visual features	Relevant studies
Commercial HMDs[a]			
Vuzix Wrap 920AR (2010)	• Internet transmission of video data	• Camera-captured video feed of mentor's hand gestures is overlaid on mentee's HMD, and vice versa	• Chinthammit et al [31]
Vuzix Wrap 1200DX (2013)	• 2-way audio-video communication over Wi-Fi	• Camera-captured images of mentor's hand gestures are transmitted to mentee's HMD	• Mather et al [32]
Recon Jet (2013)	• 2-way audio communication over phone or Wi-Fi • Integration with custom software for 1-way video feed to computer	• Integration with custom Android-based triage app	• Follmann et al [33]
Google Glass (2013-2017)	• Compatible with Google Hangouts for audio-video streaming on remote devices and 2-way audio communication	• Projects SMS text messages from remote viewers into local user's view • Remote viewer moves a mouse cursor in local user's view • Remote viewer's webcam captures image of their hands or tools that superimpose onto local user's view	• Broach et al [34] • Ponce et al [35] • Armstrong et al [36]
Microsoft HoloLens (2016)	• Compatible with Skype or other Windows applications for audio-video streaming on remote devices and 2-way audio communication	• Displays instructions, patient data, and images • Remote viewers annotate local user's view and create 3D surgical trajectories on virtual limbs • Displays hand gestures captured by sensor device used by remote mentor	• Kaylor et al [37] • Cofano et al [38] • Liu et al [39] • Hanna et al [40] • Wang et al [41]
Moverio BT-300 (2016) and Moverio BT-350 (2017)	• Compatible with TeamViewer app for audio-video streaming on remote devices and 2-way audio communication	• Remote viewers directly annotate local user's view • Remote viewers create and annotate screenshots to be displayed in local user's view	• Cofano et al [38]
Vuzix Blade (2018)	• Compatible with TeamViewer app for audio-video streaming on remote devices and 2-way audio communication	• Remote viewers directly annotate local user's view • Remote viewers create and annotate screenshots to be displayed in the local user's view	• Cofano et al [38]
Magic Leap One (2018)	• 2-way audio communication over Wi-Fi	• Displays holographic patients and monitors that can be modulated by remote viewers	• Hess et al [42]
Microsoft HoloLens 2 (2019)	• Compatible with Microsoft Teams for audio-video streaming on remote devices and 2-way audio communication	• SMS texts messages from remote viewers projected onto local user's view • Display patient data and images • Remote viewers directly annotate and blend pictures or videos into local user's view • 3D object annotation	• Hill [43] • Rigamonti et al [44] • Martin et al [45] • Van der Putten et al [46] • Rafi et al [47] • Bala et al [48] • Mill et al [49]
Virtual presence tools			
Original augmented reality telementoring platform (2014)	• Wired connection allows for the sharing of video feeds and audio communication	• Image of mentor's laparoscopic instruments is superimposed onto mentee's monitor; hybrid video seen at both sites	• Vera et al [50]

Tool (year)	Communication features	Visual features	Relevant studies
Virtual interactive presence and augmented reality platform (2013)	• 2-way video streaming via internet • Can combine with Skype or other teleconferencing software	• Remote viewer's hand or instrument is superimposed over local video feed • Remote viewer can freeze screen or 2D annotate image using pen tool	• Ponce et al [51] • Ponce et al [35] • Vyas et al [52] • Davis et al [53]
Help Lightning mobile app (2016)	• 2-way internet transmission of audio and video between phones	• Foreground of physician's video (eg, physician's hand) is superimposed over patient's video for live gesturing • Physician annotates over patient's live video	• Ponce et al [54]
Proximie (2016)	• Audio-video streaming from local site can be accessed by remote computers via internet • Viewers from different remote sites can access the same live stream and talk with each other and the local site	• Virtual hand pointer or pen to mark video feed from local site • Image of remote viewer's hand is superimposed over video from local site • Overlaying video with 2D images and 3D models • Computer vision algorithm allows for the anchoring of annotations	• Hassan et al [55] • El-Asmar et al [56] • Greenfield et al [57] • Patel et al [58]
Other original systems			
System for Telementoring with Augmented Reality platform (2015)	• 2-way internet transmission of audio and video between tablets or to HMD	• Remote viewer annotates over local site's video • Remote viewer places scalable instrument icons or labels over local site's video • Local user can access prerecorded video clips to guide procedure in the event that the remote connection is unstable [59]	• Rojas-Muñoz et al [60-62] • Andersen et al [59,63]
Vuforia Chalk mobile app (2017)	• 2-way internet transmission of audio between devices • Internet transmission of camera feed from local user to remote expert	• Remote viewer annotates over local user's live video • Annotations remain anchored to objects in the video even if local camera moves	• Ramsingh et al [64]
Original tele-ultrasound system (2018)	• Open-source software for communication between remote viewer and local user • Internet transmission of ultrasound video and live video of local user's environment sent to remote viewer's laptop	• Remote viewer draws or writes directly on ultrasound images being streamed by local user; hybrid video seen at both sites	• Carbone et al [65]
Augmented Reality–based Telerehabilitation System with Haptics (2019)	• 2-way internet transmission of audio and visual data via computers • Haptic devices relay force feedback and motion to each other through networked computer	• Camera data used to generate image of remote and local users sitting across from each other in a virtual space seen on 3D televisions	• Borresen et al [66,67]
Telestration with coaxial projective imaging (2022)	• 2-way audio-video streaming over the internet	• Obtains images from local field that remote viewer can annotate • System projects annotations onto local field directly	• Zhang et al [68]
Original remote training platform (2022)	• 2-way audio-video streaming over the internet	• HMD optics has see-through transparency and enables display of instructional information and video	• Stone et al [69]

[a]HMD: head-mounted device.

Of the 39 studies, 22 (56%) were related to surgery, 6 (15%) were related to EM, and 4 (10%) were related to hospital medicine. When looking at the setting and structure of the tasks from each study, 51% (20/39) involved an operation or technique used in the operating room, where a mentor figure used AR to remotely interact with a task performer. The remaining 19 studies involved nonoperative tasks, 11 (58%) of which also involved a mentor remotely interacting with a task performer, whereas 8 (42%) involved the mentor performing the task instead, with AR enhancing what their remote spectators saw.

The articles were divided into 3 sections based on the AR-assisted task performed: patient evaluation, medical intervention, and education (Multimedia Appendix 1-3). Medical intervention and education are further subdivided based on whether AR supported a nonsurgical or surgical task. Notably, some articles discussed in the medical intervention section were also relevant to education.

AR in Remote Patient Evaluation

The 26% (10/39) of articles included in this section (Multimedia Appendix 1 [33,34,37,43-45,54,65-67]) described implementations of AR in remote triage, wound assessment, musculoskeletal examination, sonography, and hospital rounding.

The potential of AR to perform fast-paced triage assessments remotely was explored in 20% (2/10) of the studies. Broach et al [34] investigated the use of Google Glass to relay what paramedics saw to other remote providers, with the intention of allowing the remote EM physicians to perform secondary triage before the patient's arrival at a hospital. The remote physicians accessed the perspective of the paramedics at a simulated disaster scene and could send instructional SMS text messages projected onto the paramedics' Glass. When comparing the remote assessments of physicians with the in-person assessments by different EM physicians, the study found high interrater agreement (0.923), which was not significantly different from the interrater agreement within the same assessment condition (0.976; *P*=.41) [34]. Follmann et al [33] investigated how AR implementation could affect triage time and accuracy. The performance of non–AR-assisted first responders was compared with that of 2 groups using AR-capable glasses, one that displayed an interactive triage algorithm and the other that streamed footage to a remote EM physician who could verbally guide the on-site individual. The results revealed that AR assistance increased accuracy at the cost of time, with the accuracy of the 3 groups being 58%, 92% (*P*=.04), and 90% (*P*=.01), whereas the duration of triage was 16.6, 37.0 (*P*=.001), and 35.0 (*P*=.01) seconds, respectively [33].

Other studies focused on AR's potential to enhance remote wound assessments. Ponce et al [54] demonstrated the use of AR with mobile devices to allow orthopedic surgeons and neurosurgeons to conduct remote postoperative visits. Using a virtual interactive presence (VIP) smartphone app, surgeons overlaid live camera footage of their hands over the camera footage of the patient's postoperative wound. Surveys from users revealed that both patients and surgeons found utility (27 of 28 and 26 of 29 with positive responses) in the virtual experience [54]. Kaylor et al [37] and Hill [43] focused on the use of AR for consultations during inpatient wound assessments. Using Microsoft HoloLens, bedside nurses in the former study could send live video of a patient's wound and communicate with a remote wound, ostomy, and continence (WOC) nurse. The remote WOC nurse could provide annotations or images to guide the local nurse during the assessment. When comparing remote assessments with in-person assessments performed by a different WOC nurse, the study found the interrater agreement of treatment plans to be 100% across 21 cases [37]. The bedside nurses in the study by Hill [43] used the Microsoft HoloLens 2 to consult for complications at night or over weekends for patients undergoing negative pressure wound therapy. Compared with a control group of previous cases that did not use AR, the study group underwent fewer unplanned surgical revisions (*P*=.002) and admissions related to wound infection (*P*=.004) [43].

Borresen et al [66,67] introduced the AR-based Telerehabilitation System with Haptics (ARTESH) to perform remote strength and range-of-motion examinations. Both the local and remote sites were equipped with a haptic device, a Kinect camera, and a 3D-capable television that allowed the physician and patient to view each other seated together at a virtual table. The 7-point Likert scale surveys from the pilot study showed positive ratings from 5 physicians on the ability to evaluate arm strength and visualize limb movement (6/7 and 5.87/7) [66]. A follow-up study measured interrater agreement between in-person examinations and remote evaluations using ARTESH for different components of the upper extremity examination. The highest levels of agreement were observed in strength testing of elbow flexion, shoulder abduction, and protraction (κ=0.63, 95% CI 0-1.0), with the percentage of interrater agreement across evaluations for all 15 patients ranging from 30 to 100 [67].

Rigamonti et al [44] and Carbone et al [65] conducted studies that integrated AR technology to allow for supervision and consultation while performing ultrasound examinations. Rigamonti et al [44] interviewed engineering and sports science professionals across 6 countries after they accessed live video footage from a Microsoft HoloLens 2 worn by an ultrasound operator in Germany. Users were able to annotate the stream, overlay pictures and videos onto the display, and communicate in real time as they watched the examination from the operator's perspective. Interview responses from the spectators revealed AR's potential in both education and enhancing remote examinations with the supplementation of live data and feedback [44]. Carbone et al [65] developed a tele-ultrasound system that allowed rural hospital clinicians to contact a consultant. The remote consultant, who viewed live video of the user's environment and ultrasound sequences, could provide audio feedback and overlay annotations or a cursor on the ultrasound imaging seen at the local site. Although the connected parties primarily used the platform to discuss the diagnosis, in 5 of 12 cases, the consultant directed the user on the device's probe position [65].

The study by Martin et al [45] used an AR-capable Microsoft HoloLens 2 in multiple COVID-19 wards. The senior member of a clinical team would wear the headset and personal protective equipment (PPE) while examining patients, allowing the remaining members of the team to watch and interact remotely. Viewers could remotely annotate objects as well as overlay patient imaging and data from the electronic health record onto the view of the headset user and the live video feed. In comparison with teams without the device, the AR-assisted teams saw less COVID-19 exposure time by 51.5% and decreased PPE use by 83.1% over a week (*P*=.002 and *P*=.02) [45].

AR in Remote Medical Intervention

The 41% (16/39) of articles included in this section (Multimedia Appendix 2 [31,35,38,39,46,52,53,55-57,59-61,63,64,68]) described AR in remote nonsurgical and surgical contexts. The latter included studies that focused on surgical efficiency, long-distance consultation, and differences between telesurgical systems.

Nonsurgical

Chinthammit et al [31] introduced the “Ghostman” system, in which a patient and a remote physical therapist are connected via AR-capable headsets (Vuzix Wrap 920AR). The hands and tools of the therapist were overlaid onto the patient’s headset, allowing the patient to obtain real-time feedback during the session. To test the system’s potential for telerehabilitation, an RCT was designed with 2 groups of volunteers receiving training on how to use chopsticks, one group with AR assistance and the other with face-to-face mentoring. Assessments performed immediately after, 1 day after, and 7 days after training found no significant difference in total skill errors or time to task completion between the 2 groups [31].

Ramsingh et al [64] described the use of a commercial smartphone app (Vuforia Chalk) to allow experts in Loma Linda, California, United States, to support an ultrasound-guided popliteal nerve block performed in Port-au-Prince, Haiti. The remote expert viewed the patient and ultrasound monitor through the local smartphone’s camera and created annotations that appeared on the local smartphone screen to guide the procedure. Both local and remote users rated the quality of the video communication as 5/5, whereas the local user rated the clarity of the AR annotations in guiding probe placement as 5/5 and the identification of relevant anatomy in ultrasound imaging as 4/5 [64].

Surgical

The efficacy of various AR-capable tools in the operative setting was tested in 25% (4/16) of the studies. Rojas-Muñoz et al [60] designed an RCT comparing audio-only telementoring against the System for Telementoring with AR (STAR) combined with an HMD (HMD-STAR) in the setting of emergency cricothyroidotomies by first responders. HMD-STAR use, in which a remote mentor placed helpful annotations and icons in the responder’s line of sight, increased performance scores when considering all-experience groups (*P*=.01) and those with low first-responder experience (*P*=.01) and low procedure experience (*P*=.03) [60]. Cofano et al [38] surveyed orthopedic surgeons who used various AR-capable HMDs for telementoring and visualization of 3D anatomical reconstructions. The surgeons reported positive feedback on the ergonomy of the headsets and perceived them as a beneficial tool that would shorten procedures and reduce postoperative complications [38]. Hassan et al [55] and El-Asmar et al [56] described neurointerventional and urological case series in which the telesurgery platform Proximie was used. Proximie allows a remote surgeon to overlay live video of their hands and tools onto a live stream of the operating field, thereby giving the consulting surgeon, who sees the hybrid video on a monitor in the operating room, real-time guidance. Hassan et al [55] observed no complications after 10 neurovascular procedures and noted no significant difference in contrast dye use or fluoroscopy times compared with similar on-site procedures (*P*=.38 and *P*=.85) [55]. El-Asmar et al [56] compared 21 AR-proctored aquablation procedures with 38 on-site guided cases and found no significant difference in length of stay, hospitalization, and 3-month adverse events [56].

Other studies (6/16, 38%) used AR in long-distance consultation and global surgery. Greenfield et al [57] described a case report in which a surgeon in Gaza, Palestine, connected with a specialist in Beirut, Lebanon, via Proximie for a hand reconstruction procedure. Ponce et al [35] described a case report that used VIP and AR (VIPAAR) to connect a remote consultant in Atlanta, Georgia, with a Google Glass–wearing orthopedic surgeon in Birmingham, Alabama, for a successful shoulder replacement. Similar to Proximie, VIPAAR allows the remote user, equipped with a computer and web camera, to create annotations and superimpose video of their hands or instruments over the local video of the operating field. However, instead of a hybrid video appearing on a monitor at the local site, the manipulations of the remote viewer would appear on the Glass worn by the local surgeon. Vyas et al [52] and Davis et al [53] focused on VIPAAR for telesurgery across continents. Vyas et al [52] described surgeons in Peru performing pediatric cleft lip repairs while using VIPAAR to connect with expert surgeons in California, United States. Davis et al [53] reported a pediatric neurosurgery case (endoscopic third ventriculostomy) in Ho Chi Minh City, Vietnam, with consultation from a specialist in Birmingham, Alabama, United States. Liu et al [39] described a case of cross-continental telesurgery while also introducing a 3D point-tracking module compatible with the Microsoft HoloLens to accurately track a scalpel’s location. During a skin grafting and fasciotomy of a rabbit model, a surgical trainee wearing the headset in Anhui, China, was able to visualize surgical trajectories drawn by a surgeon in Columbus, Ohio, United States [39]. Van der Putten et al [46] featured the unplanned use of the Microsoft HoloLens 2 to allow a product manager to remotely guide a surgeon through the installation of an implant that would address a complication encountered during total knee arthroplasty. This study uniquely involved a nonsurgeon as the mentoring individual and described the minimal learning curve required for the remote consultant to instruct through the HMD [46].

Andersen et al [59,63], Rojas-Muñoz et al [61], and Zhang et al [68] developed trials to compare the procedural efficiencies of different AR tool setups during telesurgery. Andersen et al [63] compared conventional telestration, in which the hybrid video with the expert’s annotations is displayed on a separate monitor outside the surgical field, with the STAR platform, in which the hybrid video is shown on a tablet directly above the field and the surgeon's hands. Premedical and medical students were guided through a placement task for would-be laparoscopic ports, followed by abdominal incisions on a model. The STAR-assisted group saw lower placement errors ($P<.01$) and focus shifts away from the operating field ($P<.001$), with time to task completion being slower but not statistically significant ($P=.17$) [63]. A subsequent study by Andersen et al [59] focused on the addition of offline references during a cricothyroidotomy with network limitations. A control group using conventional telestration was compared with a STAR group with access to offline video references showing future steps. Less idle time ($P<.001$) and higher performance scores were observed ($P<.05$ for both raters) in the STAR group [59]. Rojas-Muñoz et al [61] later compared STAR with HMD-STAR, in which the remote viewer's modifications were displayed on a headset instead of a tablet. The RCT used 2 groups of medical students performing a similar marking and incision task. Across the 2 tasks, the HMD-STAR group had fewer placement errors ($P<.001$ and $P=.01$) and focus shifts ($P<.001$ and $P<.004$) but took more time ($P<.001$ and $P<.02$) [61]. Zhang et al [68] implemented a system (coaxial projective imaging) that allowed a remote mentor's annotation to directly project onto the local operating field. The study compared the performance of trainees using this system with that of a control group using conventional telestration during a skin cancer surgery simulation. The experimental group demonstrated higher accuracy, shorter operating times, and fewer focus shifts away from the operating field ($P<.05$ each) [68].

AR in Remote Education

The articles discussed in this section (Multimedia Appendix 3 [32,36,40-42,47-51,58,62,69]) are divided based on nonsurgical contexts, which include clinical skills, autopsy, and sonography, and surgical contexts, which include procedure observation, tool-specific training, simulations, and intraoperative learning. Of note, one-third (6/19, 32%) of the articles that involved education have been described in the previous section and will be briefly mentioned in this section.

Nonsurgical

In total, 16% (3/19) of the articles described proof-of-concept studies in which a clinician wore a Microsoft HoloLens 2 to live stream footage of patient examinations during general medicine rounds to remote students. The Microsoft HoloLens 2 allowed students to not only see through the clinician's eyes using their personal devices but also simultaneously review overlaid 2D patient data and imaging. Rafi et al [47] featured a cardiovascular examination that was remotely spectated by final-year medical students. Using the Microsoft Teams application compatible with the HoloLens, the study measured student engagement in the form of written "chat" comments from students during the session [47]. Bala et al [48] conducted a similar study with remote fourth-year medical students observing a 1-hour session with a patient interview followed by a data interpretation and management planning session. Survey responses from the students found unanimous positive ratings regarding the tool's impact on accessibility of education, whereas free-response feedback from students, staff, and patients revealed that the technology was a feasible and acceptable method for providing clinical education [48]. Mill et al [49] accommodated 53 fourth-year students split across 3 sessions that each featured a case discussion, bedside review, and debriefing. Survey responses from students and instructors found favorable feedback regarding the teaching quality of the sessions despite one-third of respondents reporting issues with audio and video quality [49].

A few studies (3/19, 16%) integrated AR tools with nonsurgical telementoring. Hanna et al [40] described the use of the Microsoft HoloLens by pathology staff to communicate with an attending pathologist during an autopsy. The headset allowed trainees to view holograms of tissue specimens and web-based procedure manuals, whereas the attending could provide guidance remotely during the procedure [40]. Wang et al [41] introduced the HoloLens to sonography training in a trial comparing an AR-assisted group with an audio-assisted group. Undergraduate and paramedic students were equipped with an AR headset to view the remote mentor's hand gestures while being instructed on how to perform the right upper quadrant portion of the Focused Assessment with Sonography in Trauma examination. A control group of similar low-experience students wore headphones instead of the headset, with the mentor being able to view their progress through on-site cameras. The results revealed that performance scores were not significantly different between the 2 groups ($P=.53$), although the task completion time was longer for the AR group ($P=.008$) [41]. Mather et al [32] implemented 2 AR-capable HMDs (Vuzix Wrap 1200DX) to create an educational system called "Helping Hands." A remote instructor's HMD could capture their hand movements to be overlaid onto the display of a student's HMD while the student's hands could be visualized by the instructor. The pilot study involved guiding students through handwashing, with free-text survey responses from students showing favorable impressions of the system [32].

Hess et al [42] described a remote advanced cardiovascular life support simulation using Magic Leap One headsets distributed to second-year medical and physician assistant students. The HMDs used AR to display the holographic simulation apparatus (eg, patients, beds, and monitors) modulated by instructors, thereby allowing students to attend the simulation from their homes. Postsession interviews yielded positive feedback regarding experiential satisfaction and value in practicing communication skills [42].

Surgical

AR-capable HMDs also allowed surgical trainees to remotely observe a procedure from the operating surgeon's perspective. Cofano et al [38], discussed in the previous section, described 2 spine surgeries in which surgical interns and medical students received live commentary from the operating surgeon while also viewing reference models and images.

Vera et al [50] and Patel et al [58] implemented AR for training on the use of surgical tools. Vera et al [50] conducted an RCT using portable laparoscopic training boxes with AR telementoring features to train medical students. An experimental group of students used the AR telementoring platform, in which the instructor's laparoscopic instruments were superimposed in real time on the student's monitor during a training session. Compared with a control group that was mentored in person, the AR telementoring group had significantly faster skill acquisition (*P*<.001) and more completed attempts during a posttraining suturing task (*P*=.02) [50]. Patel et al [58] used Proximie to remotely teach medical students how to use robotic surgery tools (da Vinci Skills Simulator), with postexperience Likert surveys showing positive ratings for ease of use and quality of the audio-video feed.

Other articles (4/19, 21%) discussed surgical education performed through simulated procedures. Andersen et al [63] and Rojas-Muñoz et al [61] both featured trials using medical students for abdominal incision tasks and have been discussed in the previous section. Another study by Rojas-Muñoz et al [62] compared HMD-STAR with textbook review for leg fasciotomies performed by medical students and surgical residents. The HMD-STAR group could receive verbal guidance and annotations from a remote mentor, whereas the control group performed the procedure with independent review of the procedure beforehand. When comparing both groups, the HMD-STAR group had a 10% higher weighted individual performance score (*P*=.03) with 67% fewer errors (*P*=.04) and no significant difference between task completion times [62]. Stone et al [69] introduced a novel training system that allowed for the remote instruction of a transperineal prostate biopsy and rectal spacer placement on an anatomical model. The system involved a pair of HMDs that allowed users to view ultrasound imaging and the procedural field simultaneously. The students observed the mentor perform the procedure before practicing on the model with remote guidance. Both learners and instructors reported that the displayed images were adequate for the procedures and that the HMDs did not affect performance negatively [69].

Education in the form of intraoperative telementoring was featured in 21% (4/19) of the studies. A study by Armstrong et al [36] entailed a case report in which a junior resident wore Google Glass while performing a delayed primary closure of a plantar defect under the supervision of a remote attending surgeon. Accessing the audio-video feed through Google Hangouts, the attending could provide verbal feedback and use a mouse on their computer to point to items seen by the Glass [36]. Ponce et al [51] implemented a VIP platform for a pilot study in which surgical residents performed arthroscopic shoulder procedures under the remote mentorship of an attending surgeon. The platform was similar to the VIPAAR platform described previously: video of the operating site was viewed by a remote mentor, allowing the mentor to create annotations or superimpose their own hand and instruments over the video feed. A monitor at the operating site showed the hybrid video to the residents to allow for real-time feedback. Survey responses from those involved showed positive ratings for ease and utility of the tool in anatomical learning, with all in agreement that the system did not compromise safety [51]. Vyas et al [52] and Davis et al [53] were described previously in the context of global surgery, but both demonstrated examples of long-distance surgical training. Mentoring surgeons in the overseas curriculum described by Vyas et al [52] evaluated the local mentees on various aspects of cleft lip repair following both in-person and telementored surgeries. In-person procedures preferentially improved intraoperative decision-making (*P*<.001) and repair principles (*P*<.001), whereas remote sessions preferentially improved understanding of anatomy (*P*<.01) and increased procedural efficiency (*P*<.001) [52].

Evaluation of AR-Assisted Remote Tasks

Of the 39 studies, 21 (54%) had a comparative design, with 6 (15%) being RCTs. Most studies (23/39, 59%) gathered data through user feedback surveys or interviews, 70% (16/23) of which used numerical Likert scales. Tables 2 and 3 summarize the variables measured and discussed across the articles in nonsurgical and surgical tasks, respectively.

Table 2. Range of variables studied in augmented reality–assisted nonsurgical tasks.

Variable and subcategory	Study, year
Objective	
Performance score	• Wang et al [41], 2017 • Rigamonti et al [44], 2021
Time for task	• Chinthammit et al [31], 2014 • Wang et al [41], 2017
Accuracy	• Chinthammit et al [31], 2014 • Follmann et al [33], 2019
Reliability across task performers	• Broach et al [34], 2018 • Kaylor et al [37], 2019 • Borresen et al [67], 2022
Patient complications	• Hill [43], 2022
Subjective efficacy	
Usefulness	• Chinthammit et al [31], 2014 • Ponce et al [54], 2016 • Wang et al [41], 2017 • Broach et al [34], 2018 • Carbone et al [65], 2018 • Mather et al [32], 2018 • Follmann et al [33], 2019 • Ramsingh et al [64], 2022 • Mill et al [49], 2021 • Rigamonti et al [44], 2021 • Hess et al [42], 2022
Improved patient care	• Martin et al [45], 2020
Improved performance	• Chinthammit et al [31], 2014 • Wang et al [41], 2017
Efficacy in communication	• Wang et al [41], 2017 • Borresen et al [66], 2019 • Ramsingh et al [64], 2019 • Martin et al [45], 2020 • Bala et al [48], 2021 • Mill et al [49], 2021 • Hess et al [42], 2022
Superiority to other communication methods	• Ponce et al [54], 2016 • Wang et al [41], 2017
Feasibility	
Ease of use	• Wang et al [41], 2017 • Broach et al [34], 2018 • Carbone et al [65], 2018 • Mather et al [32], 2018 • Borresen et al [66], 2019 • Martin et al [45], 2020 • Mill et al [49], 2021 • Rigamonti et al [44], 2021 • Hess et al [42], 2022
Interference in task	• Broach et al [34], 2018 • Carbone et al [65], 2018 • Mill et al [49], 2021

Variable and subcategory	Study, year
Comfort of device	• Mather et al [32], 2018 • Martin et al [45], 2020 • Rigamonti et al [44], 2021
Acceptability	
Satisfaction	• Ponce et al [54], 2016 • Borresen et al [66], 2019 • Bala et al [48], 2021 • Mill et al [49], 2021 • Hess et al [42], 2022
Favorability and interest	• Mather et al [32], 2018 • Rigamonti et al [44], 2021 • Mill et al [49], 2021 • Rafi et al [47], 2021 • Hess et al [42], 2022
Trust in use	• Carbone et al [65], 2018
Graphics quality	• Carbone et al [65], 2018 • Borresen et al [66], 2019 • Ramsingh et al [64], 2019 • Bala et al [48], 2021 • Mill et al [49], 2021
Other measures	
Efficacy of device training and instructions	• Chinthammit et al [31], 2014 • Broach et al [34], 2018 • Borresen et al [66], 2019 • Hess et al [42], 2022
User safety and PPE[a] use	• Martin et al [45], 2020

[a]PPE: personal protective equipment.

The most common objective variable measured was time to task completion, with other common variables being task performance and procedure-related complications. Several studies (7/39, 18%) on the topics of telerehabilitation and telesurgery found no difference in task time compared with non–AR-assisted conditions, whereas Vera et al [50] found that less time was needed when using AR for laparoscopic training. In contrast, Follmann et al [33], Wang et al [41], and Ponce et al [35] found that AR increased the time needed compared with non-AR conditions for triage assessment, ultrasound examination, and shoulder replacement, respectively.

Performance, measured using scoring systems or frequency of errors, was seen to improve with AR use in telesurgery tasks described by Vera et al [50], Rojas-Muñoz et al [60,62], and Andersen et al [63], whereas Wang et al [41] and Chinthammit et al [31] found no difference when comparing with non-AR groups in tele-ultrasonography and a telerehabilitation-related task, respectively. When comparing AR-assisted tools with each other, as done by Andersen et al [63] and Rojas-Muñoz et al [61], improved performance was observed as AR-enhanced displays were brought closer to the eyes of the user. AR's impact on resource use was different based on context, with Martin et al [45] finding that AR-assisted rounding during COVID-19 saved PPE, whereas El-Asmar et al [56] and Hassan et al [55] found no significant increases in general anesthesia use in aquablation procedures or contrast use for neuroradiological interventions, respectively.

Subjective measures included in the studies were related to device efficacy, feasibility, and acceptability. All studies that examined feedback-related efficacy, such as the perceived "usefulness" of AR assistance, reported a majority of positive ratings. However, when comparing the ratings of AR-assisted conditions with those of non-AR conditions, Wang et al [41] and Chinthammit et al [31] notably found no differences. Studies that measured ratings for feasibility, such as ease or comfort of use, also found a majority of positive responses. The lowest percentage of positive Likert-scale ratings for ease was observed in first responders in the study by Broach et al [34] in the context of using Google Glass for triage assessments, with 64% (9/14) of users giving a rating of 4 out of 5 or higher. Studies that examined acceptability ratings, such as experience satisfaction or interest, similarly found a consistent majority of positive ratings from AR users. Further study into potential differences between patient and provider ratings may be warranted as Ponce et al [54] found that postoperative patients using a virtual presence–type mobile app were more likely to be satisfied with the overall experience (average rating of 4.6 vs 4.2 out of 5; $P<.05$) and view the virtual interaction as superior to email and SMS text messaging (4.7 vs 4.4 out of 5; $P<.05$) than surgeons.

Table 3. Range of variables studied in augmented reality–assisted surgical tasks.

Variable and subcategory	Study, year
Objective	
Performance	• Vera et al [50], 2014 • Andersen et al [63], 2017 • Vyas et al [52], 2020 • Andersen et al [59], 2019 • Rojas-Muñoz et al [61], 2019 • Rojas-Muñoz et al [60], 2020 • Rojas-Muñoz et al [62], 2020 • Zhang et al [68], 2022
Procedural efficiency	• Andersen et al [63], 2017 • Andersen et al [59], 2019 • Rojas-Muñoz et al [61], 2019 • Zhang et al [68], 2022
Procedure time	• Ponce et al [35], 2014 • Ponce et al [51], 2014 • Vera et al [50], 2014 • Andersen et al [63], 2017 • Rojas-Muñoz et al [61], 2019 • Rojas-Muñoz et al [60], 2020 • Rojas-Muñoz et al [62], 2020 • Hassan et al [55], 2021 • El-Asmar et al [56], 2021 • Zhang et al [68], 2022
OR[a] resource use	• Hassan et al [55], 2021 • El-Asmar et al [56], 2021
Complications	• Ponce et al [35], 2014 • Vyas et al [52], 2020 • Hassan et al [55], 2021 • El-Asmar et al [56], 2021 • Van der Putten et al [46], 2022
Patient length of stay and rehospitalization rates	• El-Asmar et al [56], 2021
Skill acquisition	• Vera et al [50], 2014
Accuracy of device	• Andersen et al [63], 2017 • Liu et al [39], 2021
Subjective efficacy	
Usefulness	• Ponce et al [51], 2014 • Vera et al [50], 2014 • Davis et al [53], 2017 • Patel et al [58], 2021
Efficiency	• Rojas-Muñoz et al [61], 2019 • Rojas-Muñoz et al [60], 2020
User improvement	• Vera et al [50], 2014 • Vyas et al [52], 2020
Procedural safety	• Ponce et al [51], 2014 • Vyas et al [52], 2020
Feasibility	
Ease of use	• Ponce et al [51], 2014 • Rojas-Muñoz et al [61], 2019 • Cofano et al [38], 2021 • Patel et al [58], 2021

Variable and subcategory	Study, year
Frustration in use	• Rojas-Muñoz et al [61], 2019 • Rojas-Muñoz et al [60], 2020
Interference in task	• Ponce et al [35], 2014 • Andersen et al [63], 2017
Comfort of device	• Rojas-Muñoz et al [61], 2019 • Cofano et al [38], 2021 • Stone et al [69], 2022
Acceptability	
Satisfaction	• Ponce et al [35], 2014
Favorability and interest	• Vera et al [50], 2014 • Rojas-Muñoz et al [61], 2019 • Patel et al [58], 2021 • Cofano et al [38], 2021
Graphics quality	• Ponce et al [35], 2014 • Ponce et al [51], 2014 • Patel et al [58], 2021 • Stone et al [69], 2022
Other measures	
Device latency	• Ponce et al [35], 2014 • Ponce et al [51], 2014 • Davis et al [53], 2017 • Stone et al [69], 2022
Financial cost	• Davis et al [53], 2017

[a]OR: operating room.

Although most survey- or interview-based studies measured user feedback about AR's usefulness (15/23, 65%), far fewer studies (2/23, 9%) surveyed AR's ability to substitute in-person methods in telemedicine. Borresen et al [66], who used the ARTESH system for motor and strength examinations, found an average Likert rating (4/7) from physicians when they were asked whether in-person examinations would have similar results. Surveyed patients similarly had a lower percentage of positive responses (9/15, 60%) compared with other feasibility and acceptability questions when asked about the device's potential to substitute an in-person examination [66]. In telementoring, the literature shows that AR may have more potential. Patel et al [58], who used Proximie to remotely mentor students in using robotic surgery tools, observed average ratings above 4/5 for utility as an alternative to in-person mentoring.

Across all 39 studies, there were gaps in the longitudinal measurements related to patient outcomes, such as costs, hospitalization course, and quality of life. Only the study by Davis et al [53] analyzed the finances of implementing an AR system, estimating a cost of US $14,900 per calendar year for a VIPAAR system used for telementoring pediatric neurosurgery in Vietnam. Meanwhile, only El-Asmar et al [56] measured postoperative length of stay and 3-month adverse events to find no significant difference between AR and non-AR conditions. Furthermore, there were few studies (2/39, 5%) that focused on the long-term benefits of AR-enabled remote education, such as retention of learned material and performance over an extended period. Chinthammit et al [31] measured the performance of trainees multiple times over a period of a week after training; however, studies that examined nonsurgical skills over longer periods are absent in this review. Vyas et al [52] described a 13-month overseas course in pediatric cleft lip repair; however, the curriculum combined both in-person and remote intraoperative learning sessions rather than comparing the methods.

Discussion

Principal Findings

In this scoping review, we discussed 39 studies that used AR in real-time telemedicine and telementoring. From these studies, 20 unique devices and platforms, most of which involved an HMD, were identified and found to have common features such as annotation, graphical references, and the ability to overlay a remote viewer's hands or tools onto a local user's screen.

AR builds on the remote examinations of current audio- and videoconferencing tools by enhancing the remote acquisition and exchange of information. AR technology can supply users with visual aids, such as electronic health record data and guidelines, or facilitate communication with specialists trained to identify specific conditions or manipulate diagnostic equipment. Studies on AR-assisted remote patient evaluations took place in both outpatient and inpatient settings, including emergency triage, wound evaluation, and hospital rounding. Although the devices vary across settings, the comparative studies in this review found noninferior accuracy and reliability of AR-assisted remote conditions compared with in-person controls. Devices dedicated to remote musculoskeletal and sonographic examinations were also developed, with positive user ratings. As technology improves and becomes more accessible, the types of examinations that can be performed remotely will expand and allow for greater access to care, especially by patients and providers in low-resource areas.

The research on AR in the remote delivery of medical interventions predominantly focused on surgery, with other treatment modalities such as physical therapy being less studied [70]. Telesurgery implemented AR features such as virtual presence and annotation to facilitate procedural guidance, mentorship, and consultation. The distance of such connections was tested, with several case reports describing AR in the telementoring of operations across continents. Various types of surgical procedures were telementored using AR, including minimally invasive surgery, orthopedic surgery, and pediatric neurosurgery. The literature surrounding AR and telesurgery focused primarily on the technology's potential to improve procedural efficiency and communication with experts, with the differences across systems for surgical telementoring studies being investigated in more recent articles. From the data available so far, the performance of AR-assisted procedures appears equivalent to other remotely mentored conditions, with limited research suggesting AR's potential to substitute in-person mentorship without sacrificing resources or short-term outcomes. Although such findings require further validation, AR plays a promising role in enabling less experienced providers to perform more specialized treatments and avoid the unnecessary transfer of patients across hospital systems. Research on AR-assisted telerehabilitation and psychotherapy could offer insight into whether these nonsurgical treatments could also be effectively performed as remote visits.

Medical education emerged as a common theme underlying AR's integration into telemedicine and telementoring. In addition to surgery, remote training using AR was studied for clinical skills that are important to inpatient wards and procedures related to pathology and EM. Both surgical and nonsurgical studies implemented AR tools in similar ways, including observational learning, real-time audiovisual feedback, and teaching by demonstration. When in-person mentorship was not possible or inconvenient, the studies showed how AR-enhanced communication allowed distant learners to not just observe but also engage with clinicians and surgeons. Distance learning of surgical and sonographic equipment could also be achieved with AR systems, with few controlled trials thus far indicating noninferior performance results compared with in-person mentoring and superior results compared with unassisted conditions, particularly for less experienced trainees. As the equipment for telecollaboration in medicine becomes readily available, AR technology is expected to advance the sharing of knowledge related to clinical skills, health care devices, and procedural techniques, which directly and indirectly improves patient care outcomes.

Given the early stage of platform development, many studies included in this review (25/39, 64%) used small cohorts of ≤15 patients or participants per study group; although approximately half (21/39, 54%) of the studies included comparative data, far fewer were RCTs. This finding is expected as AR is an emerging technology in medicine, and its use in telemedicine and telementoring is still developing. Existing comparative trials that measured time consumption and performance have so far found that AR-assisted conditions mostly have noninferior results compared with other remote and in-person methods. However, as these studies are heterogeneous in methodology and equipment, dedicated RCTs with standardized designs are needed to understand whether certain AR systems for specific tasks can effectively substitute current remote alternatives or in-person methods. Feedback surveys and interviews were the most common form of data collection observed in this review; subjective variables such as perceived efficacy and ease ratings were commonly measured with consistently positive findings despite the novelty of the technology, suggesting interest across user groups, including providers, trainees, and patients.

Future Directions

Overall, interventions and learning that require active patient or family caregiver participation (eg, physical therapy, psychotherapy, preventative medicine, chemotherapy, and dialysis) have yet to achieve a similar level of investigation in the space of AR and telemedicine or telementoring as surgery. From the nonsurgical studies included in this review, there is an interest in AR for tele-sonography, EM, and hospital medicine, with many nonsurgical fields yet to be represented. In the realm of patient evaluation, AR's role is still being separately investigated for electrograms [71], diagnostic procedures related to endoscopy [72], biopsies [73], and urology [74], as well as specialty-specific examinations pertinent to dentistry [75], ophthalmology [76], and dermatology [9], to name a few.

Considering the diversity of medical fields and levels of experience across users, AR-enabled remote interactions are likely to appear in certain settings or users sooner than in others. In the interventional and educational space, studies so far have primarily implemented AR tools for remote consultation between current or future medical professionals. Few trials using untrained individuals for AR-assisted procedural, nonsurgical tasks exist; it would be reasonable to anticipate future research with AR facilitating remote provider-to-patient or provider-to-caregiver interactions in the therapeutic context [77]. To support innovations focused on remote interactions with home caregivers and patients, perspectives from and qualitative studies on these particular end users, rather than solely clinician users, are necessary [22,78].

Although more dedicated RCTs would be needed to assess the efficacy of AR-assisted communication in the treatment setting, AR's impact on variables related to costs beyond procedure time and outcome measures beyond complication rates remains relatively unexplored. These include hospital-based measurements such as length of stay, equipment costs, and intra- and postoperative pain medication use but also patient-centered variables such as treatment cost, posttreatment functional status, and quality of life. The literature in this review focused mainly on short-term variables, which reveals a lack of longitudinal research that could provide insight into both the long-term outcomes of patients and system-wide effects on productivity and sustainability of AR use. Longitudinal studies with dedicated comparisons, which are likely to increase as devices improve in wearability, usability, and affordability, are also needed to fully understand whether AR could enhance knowledge and skill retention in the remote learning environment.

Limitations

This review was limited to studies from the last 10 years in 4 medical research databases. Many AR developments in telemedicine and telementoring that exist in the private sector have not been described in published articles and so cannot be systematically evaluated. Notably, a 2019 systematic review of AR in telementoring has also explored additional databases [28]; however, our review included "remote," "telemedicine," and "telehealth" as search terms to locate a wider variety of health care tasks that may not rely on equipment or procedures typically associated with "telementoring." Furthermore, this scoping review placed more emphasis on the diversity of systems using AR and the nonsurgical specialties incorporating them. The small cohorts and predominant collection of subjective data were expected given the novel nature of this intersection and the technology.

Although the potential benefits of AR in telemedicine are promising, the challenges facing this technology are the early stage of research and prototype development for these application contexts and a lack of standardized devices. Outside the original and surgery-specific platforms, most of the hardware observed in this review is available to consumers but at costs that limit widespread use [79]. Furthermore, the adjunct programs and applications used with the hardware greatly varied across the studies. The diversity of tools and their availability could limit the design and generalizability of future trials, especially if the technology is custom-made and difficult to reproduce. Combined with low awareness and a lack of guidelines on how to evaluate AR technology, innovators face difficulty in developing appropriate tools and introducing them into current or even unexplored health care spaces. Future research focusing on the utility and feasibility of AR compared with current technology in the medical setting is paramount, but studies looking into the costs of implementation, user readiness, and user-friendly design are also necessary for successful adoption.

Conclusions

This scoping review discussed studies that combined AR with real-time telemedicine and telementoring, including patient evaluation, medical intervention, and education. Commonly explored applications for this novel intersection include consultation and procedural guidance, particularly in telesurgery. AR-assisted telecommunication was studied to complement or even improve the capability of remote visits, treatments, and training, but more RCTs are needed to validate task-specific benefits as well as understand the long-term effects for all users. As technology evolves and use at the consumer and industry levels becomes more widespread, research on AR in health care is expected to see larger cohorts, standardized equipment, and more rigorous methods of evaluation. Developing AR tools in medicine must balance user-friendly design with limited research and uptake; such challenges create an opportunity for institutional involvement and a need for perspectives from all those involved in health care, including but not limited to clinicians, caregivers, and patients.

Acknowledgments

The authors would like to acknowledge DE and AF as the senior coauthors of this paper. DE was funded by a grant from the National Science Foundation (2026577) and research funding from OptumLabs. The authors also give special thanks to the XR Collaboratory at Cornell Tech; the Center for Virtual Care at Weill Cornell Medicine; and those who supported the conception, discussion, and revision of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of studies using augmented reality in remote patient evaluation.

[DOCX File , 21 KB - mhealth_v11i1e45464_app1.docx]

Multimedia Appendix 2

Overview of studies using augmented reality in remote medical intervention. Articles denoted with an asterisk (*) also involved educational contexts but are not listed in Multimedia Appendix 3.

[DOCX File , 23 KB - mhealth_v11i1e45464_app2.docx]

Multimedia Appendix 3

Overview of studies using augmented reality in remote medical education. Multimedia Appendix 2 includes additional articles in educational contexts that are also discussed in the Results section.

[DOCX File , 21 KB - mhealth_v11i1e45464_app3.docx]

References

1. Berryman DR. Augmented reality: a review. Med Ref Serv Q 2012;31(2):212-218. [doi: 10.1080/02763869.2012.670604] [Medline: 22559183]
2. Brigham TJ. Reality check: basics of augmented, virtual, and mixed reality. Med Ref Serv Q 2017;36(2):171-178. [doi: 10.1080/02763869.2017.1293987] [Medline: 28453428]
3. Yeung AW, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
4. Kamphuis C, Barsom E, Schijven M, Christoph N. Augmented reality in medical education? Perspect Med Educ 2014 Sep 25;3(4):300-311 [FREE Full text] [doi: 10.1007/s40037-013-0107-7] [Medline: 24464832]
5. Hu H, Feng X, Shao Z, Xie M, Xu S, Wu X, et al. Application and prospect of mixed reality technology in medical field. Curr Med Sci 2019 Feb;39(1):1-6. [doi: 10.1007/s11596-019-1992-8] [Medline: 30868484]
6. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020 Jan 18;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]
7. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
8. Munzer BW, Khan MM, Shipman B, Mahajan P. Augmented reality in emergency medicine: a scoping review. J Med Internet Res 2019 Apr 17;21(4):e12368 [FREE Full text] [doi: 10.2196/12368] [Medline: 30994463]
9. Sharma P, Vleugels RA, Nambudiri VE. Augmented reality in dermatology: are we ready for AR? J Am Acad Dermatol 2019 Nov;81(5):1216-1222. [doi: 10.1016/j.jaad.2019.07.008] [Medline: 31302186]
10. Uppot RN, Laguna B, McCarthy CJ, De Novi G, Phelps A, Siegel E, et al. Implementing virtual and augmented reality tools for radiology education and training, communication, and clinical care. Radiology 2019 Jun;291(3):570-580. [doi: 10.1148/radiol.2019182210] [Medline: 30990383]
11. Elsayed M, Kadom N, Ghobadi C, Strauss B, Al Dandan O, Aggarwal A, et al. Virtual and augmented reality: potential applications in radiology. Acta Radiol 2020 Sep 13;61(9):1258-1265. [doi: 10.1177/0284185119897362] [Medline: 31928346]
12. Laverdière C, Corban J, Khoury J, Ge SM, Schupbach J, Harvey EJ, et al. Augmented reality in orthopaedics. Bone Joint J 2019 Dec;101-B(12):1479-1488. [doi: 10.1302/0301-620x.101b12.bjj-2019-0315.r1]
13. Kim SK, Lee Y, Yoon H, Choi J. Adaptation of extended reality smart glasses for core nursing skill training among undergraduate nursing students: usability and feasibility study. J Med Internet Res 2021 Mar 02;23(3):e24313 [FREE Full text] [doi: 10.2196/24313] [Medline: 33650975]
14. Wang P, Bai X, Billinghurst M, Zhang S, Zhang X, Wang S, et al. AR/MR remote collaboration on physical tasks: a review. Robot Comput Integr Manuf 2021 Dec;72:102071. [doi: 10.1016/j.rcim.2020.102071]
15. Wang J, Qi Y. A multi-user collaborative AR system for industrial applications. Sensors (Basel) 2022 Feb 09;22(4):1319 [FREE Full text] [doi: 10.3390/s22041319] [Medline: 35214221]
16. Kong SS, Azarfar A, Ashour A, Atkins C, Bhanusali N. Awareness and attitudes towards telemedicine among medical students in the United States. Cureus 2020 Nov 19;12(11):e11574 [FREE Full text] [doi: 10.7759/cureus.11574] [Medline: 33364099]
17. Garfan S, Alamoodi A, Zaidan B, Al-Zobbi M, Hamid RA, Alwan JK, et al. Telehealth utilization during the Covid-19 pandemic: a systematic review. Comput Biol Med 2021 Nov;138:104878 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
18. Egan KJ, Clark P, Deen Z, Paputa Dutu C, Wilson G, McCann L, et al. Understanding current needs and future expectations of informal caregivers for technology to support health and well-being: national survey study. JMIR Aging 2022 Jan 27;5(1):e15413 [FREE Full text] [doi: 10.2196/15413] [Medline: 35084339]
19. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021 Mar 01;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]
20. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
21. Kenoui M. Telemedicine meets augmented reality: healthcare services delivery and distance training. In: Proceedings of the 2020 4th International Symposium on Informatics and its Applications (ISIA). 2020 Presented at: 4th International Symposium on Informatics and its Applications (ISIA); Dec 15-16, 2020; M'sila, Algeria. [doi: 10.1109/isia51297.2020.9416532]
22. Janßen M, Prilla M. Usability and utility of head mounted displays in care according to caregivers. In: Proceedings of the 7th International Conference on Infrastructures for Healthcare. 2019 Presented at: InfraHealth 2019: Proceedings of the 7th International Conference on Infrastructures for Healthcare; May 30-31, 2019; Vienna, Austria. [doi: 10.18420/ihc2019_007]

23. Anton D, Kurillo G, Yang A, Bajcsy R. Augmented telemedicine platform for real-time remote medical consultation. In: MultiMedia Modeling. Cham: Springer; 2017.
24. Gasques D, Johnson J, Sharkey T, Feng Y, Wang R, Xu Z, et al. ARTEMIS: a collaborative mixed-reality system for immersive surgical telementoring. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445576]
25. Teoli D, Aeddula N. Telemedicine. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023.
26. Shulkin D. Telemedicine: 5 questions asked and answered. VA News. 2016 Jun 13. URL: https://blogs.va.gov/VAntage/28397/telemedicine/ [accessed 2022-04-15]
27. Ayoub CH, El-Asmar JM, Abdulfattah S, El-Hajj A. Telemedicine and telementoring in urology: a glimpse of the past and a leap into the future. Front Surg 2022 Feb 22;9:811749 [FREE Full text] [doi: 10.3389/fsurg.2022.811749] [Medline: 35273996]
28. Bui DT, Barnett T, Hoang HT, Chinthammit W. Tele-mentoring using augmented reality technology in healthcare: a systematic review. Australasian J Educational Technol 2021;37(4):68-88. [doi: 10.14742/ajet.6243]
29. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
30. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
31. Chinthammit W, Merritt T, Pedersen S, Williams A, Visentin D, Rowe R, et al. Ghostman: augmented reality application for telerehabilitation and remote instruction of a novel motor skill. Biomed Res Int 2014;2014:646347 [FREE Full text] [doi: 10.1155/2014/646347] [Medline: 24829910]
32. Mather C, Barnett T, Broucek V, Saunders A, Grattidge D, Huang W. Helping hands: using augmented reality to provide remote guidance to health professionals. Stud Health Technol Inform 2017;241:57-62. [Medline: 28809183]
33. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical support by smart glasses during a mass casualty incident: a randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. J Med Internet Res 2019 Jan 03;21(1):e11939 [FREE Full text] [doi: 10.2196/11939] [Medline: 30609988]
34. Broach J, Hart A, Griswold M, Lai J, Boyer EW, Skolnik AB, et al. Usability and reliability of smart glasses for secondary triage during mass casualty incidents. Proc Annu Hawaii Int Conf Syst Sci 2018 Jan 03;2018:1416-1422 [FREE Full text] [Medline: 29398976]
35. Ponce BA, Menendez ME, Oladeji LO, Fryberger CT, Dantuluri PK. Emerging technology in surgical education: combining real-time augmented reality and wearable computing devices. Orthopedics 2014 Nov;37(11):751-757. [doi: 10.3928/01477447-20141023-05] [Medline: 25361359]
36. Armstrong DG, Rankin TM, Giovinco NA, Mills JL, Matsuoka Y. A heads-up display for diabetic limb salvage surgery: a view through the google looking glass. J Diabetes Sci Technol 2014 Sep;8(5):951-956 [FREE Full text] [doi: 10.1177/1932296814535561] [Medline: 24876445]
37. Kaylor J, Hooper V, Wilson A, Burkert R, Lyda M, Fletcher K, et al. Reliability testing of augmented reality glasses technology: establishing the evidence base for telewound care. J Wound Ostomy Continence Nurs 2019;46(6):485-490. [doi: 10.1097/WON.0000000000000585] [Medline: 31633610]
38. Cofano F, Di Perna G, Bozzaro M, Longo A, Marengo N, Zenga F, et al. Augmented reality in medical practice: from spine surgery to remote assistance. Front Surg 2021 Mar 30;8:657901 [FREE Full text] [doi: 10.3389/fsurg.2021.657901] [Medline: 33859995]
39. Liu P, Li C, Xiao C, Zhang Z, Ma J, Gao J, et al. A wearable augmented reality navigation system for surgical telementoring based on microsoft hololens. Ann Biomed Eng 2021 Jan 05;49(1):287-298. [doi: 10.1007/s10439-020-02538-5] [Medline: 32504141]
40. Hanna MG, Ahmed I, Nine J, Prajapati S, Pantanowitz L. Augmented reality technology using Microsoft Hololens in anatomic pathology. Arch Pathol Lab Med 2018 May;142(5):638-644. [doi: 10.5858/arpa.2017-0189-OA] [Medline: 29384690]
41. Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, et al. Augmented reality as a telemedicine platform for remote procedural training. Sensors (Basel) 2017 Oct 10;17(10):2294 [FREE Full text] [doi: 10.3390/s17102294] [Medline: 28994720]
42. Hess O, Qian J, Bruce J, Wang E, Rodriguez S, Haber N, et al. Communication skills training using remote augmented reality medical simulation: a feasibility and acceptability qualitative study. Med Sci Educ 2022 Oct 04;32(5):1005-1014 [FREE Full text] [doi: 10.1007/s40670-022-01598-7] [Medline: 35966166]
43. Hill R. Using augmented reality to improve patient outcomes with negative pressure wound therapy. Wounds 2022 Feb;33(2):47-50 [FREE Full text] [doi: 10.25270/wnds/2022.4750] [Medline: 35108668]
44. Rigamonti L, Secchi M, Lawrence JB, Labianca L, Wolfarth B, Peters H, et al. An augmented reality device for remote supervision of ultrasound examinations in international exercise science projects: usability study. J Med Internet Res 2021 Oct 05;23(10):e28767 [FREE Full text] [doi: 10.2196/28767] [Medline: 34609312]

45. Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, PanSurg Collaborative. Use of the HoloLens2 mixed reality headset for protecting health care workers during the COVID-19 pandemic: prospective, observational evaluation. J Med Internet Res 2020 Aug 14;22(8):e21486 [FREE Full text] [doi: 10.2196/21486] [Medline: 32730222]
46. van der Putten K, Anderson MB, van Geenen RC. Looking through the lens: the reality of telesurgical support with interactive technology using Microsoft's Hololens 2. Case Rep Orthop 2022;2022:5766340 [FREE Full text] [doi: 10.1155/2022/5766340] [Medline: 35273816]
47. Rafi D, Stackhouse AA, Walls R, Dani M, Cowell A, Hughes E, et al. A new reality: bedside geriatric teaching in an age of remote learning. Future Healthc J 2021 Nov 12;8(3):e714-e716 [FREE Full text] [doi: 10.7861/fhj.2021-0108] [Medline: 34888473]
48. Bala L, Kinross J, Martin G, Koizia LJ, Kooner AS, Shimshon GJ, et al. A remote access mixed reality teaching ward round. Clin Teach 2021 Aug 30;18(4):386-390. [doi: 10.1111/tct.13338] [Medline: 33786988]
49. Mill T, Parikh S, Allen A, Dart G, Lee D, Richardson C, et al. Live streaming ward rounds using wearable technology to teach medical students: a pilot study. BMJ Simul Technol Enhanc Learn 2021 May 25;7(6):494-500 [FREE Full text] [doi: 10.1136/bmjstel-2021-000864] [Medline: 35520979]
50. Vera AM, Russo M, Mohsin A, Tsuda S. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. Surg Endosc 2014 Dec;28(12):3467-3472. [doi: 10.1007/s00464-014-3625-4] [Medline: 24962856]
51. Ponce BA, Jennings JK, Clay TB, May MB, Huisingh C, Sheppard ED. Telementoring: use of augmented reality in orthopaedic education: AAOS exhibit selection. J Bone Joint Surg Am 2014 May 21;96(10):e84. [doi: 10.2106/JBJS.M.00928] [Medline: 24875036]
52. Vyas RM, Sayadi LR, Bendit D, Hamdan US. Using virtual augmented reality to remotely proctor overseas surgical outreach: building long-term international capacity and sustainability. Plast Reconstr Surg 2020 Nov;146(5):622e-629e. [doi: 10.1097/PRS.0000000000007293] [Medline: 33136961]
53. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
54. Ponce BA, Brabston EW, Zu S, Watson SL, Baker D, Winn D, et al. Telemedicine with mobile devices and augmented reality for early postoperative care. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:4411-4414. [doi: 10.1109/EMBC.2016.7591705] [Medline: 28269256]
55. Hassan AE, Desai SK, Georgiadis AL, Tekle WG. Augmented reality enhanced tele-proctoring platform to intraoperatively support a neuro-endovascular surgery fellow. Interv Neuroradiol 2022 Jun 04;28(3):277-282. [doi: 10.1177/15910199211035304] [Medline: 34346826]
56. El-Asmar JM, Labban M, El-Hajj A. Integration of aquablation through telemetry: an alternative to onsite proctoring? World J Urol 2021 Sep 06;39(9):3473-3479 [FREE Full text] [doi: 10.1007/s00345-021-03603-x] [Medline: 33547926]
57. Greenfield MJ, Luck J, Billingsley ML, Heyes R, Smith OJ, Mosahebi A, et al. Demonstration of the effectiveness of augmented reality telesurgery in complex hand reconstruction in Gaza. Plastic Reconstructive Surgery Global Open 2018;6(3):e1708. [doi: 10.1097/gox.0000000000001708]
58. Patel E, Mascarenhas A, Ahmed S, Stirt D, Brady I, Perera R, et al. Evaluating the ability of students to learn and utilize a novel telepresence platform, Proximie. J Robot Surg 2022 Aug 06;16(4):973-979 [FREE Full text] [doi: 10.1007/s11701-021-01330-4] [Medline: 34741714]
59. Andersen DS, Cabrera ME, Rojas-Muñoz EJ, Popescu VS, Gonzalez GT, Mullis B, et al. Augmented reality future step visualization for robust surgical telementoring. Simul Healthc 2019 Feb;14(1):59-66. [doi: 10.1097/SIH.0000000000000334] [Medline: 30395078]
60. Rojas-Muñoz E, Lin C, Sanchez-Tamayo N, Cabrera ME, Andersen D, Popescu V, et al. Evaluation of an augmented reality platform for austere surgical telementoring: a randomized controlled crossover study in cricothyroidotomies. NPJ Digit Med 2020 May 21;3(1):75 [FREE Full text] [doi: 10.1038/s41746-020-0284-9] [Medline: 32509972]
61. Rojas-Muñoz E, Cabrera ME, Andersen D, Popescu V, Marley S, Mullis B, et al. Surgical telementoring without encumbrance: a comparative study of see-through augmented reality-based approaches. Ann Surg 2019 Aug;270(2):384-389. [doi: 10.1097/SLA.0000000000002764] [Medline: 29672404]
62. Rojas-Muñoz E, Cabrera ME, Lin C, Andersen D, Popescu V, Anderson K, et al. The system for telementoring with augmented reality (STAR): a head-mounted display to improve surgical coaching and confidence in remote areas. Surgery 2020 Apr;167(4):724-731. [doi: 10.1016/j.surg.2019.11.008] [Medline: 31916990]
63. Andersen D, Popescu V, Cabrera ME, Shanghavi A, Mullis B, Marley S, et al. An augmented reality-based approach for surgical telementoring in austere environments. Military Med 2017 Mar;182(S1):310-315. [doi: 10.7205/milmed-d-16-00051]
64. Ramsingh D, Van Gorkom C, Holsclaw M, Nelson S, De La Huerta M, Hinson J, et al. Use of a smartphone-based augmented reality video conference app to remotely guide a point of care ultrasound examination. Diagnostics (Basel) 2019 Oct 24;9(4):159 [FREE Full text] [doi: 10.3390/diagnostics9040159] [Medline: 31652998]

65. Carbone M, Ferrari V, Marconi M, Piazza R, Del Corso A, Adami D, et al. A tele-ultrasonographic platform to collect specialist second opinion in less specialized hospitals. Updates Surg 2018 Sep 18;70(3):407-413. [doi: 10.1007/s13304-018-0582-9] [Medline: 30121846]
66. Borresen A, Wolfe C, Lin C, Tian Y, Raghuraman S, Nahrstedt K, et al. Usability of an immersive augmented reality based telerehabilitation system with haptics (ARTESH) for synchronous remote musculoskeletal examination. Int J Telerehabil 2019 Jun 12;11(1):23-32 [FREE Full text] [doi: 10.5195/ijt.2019.6275] [Medline: 31341544]
67. Borresen A, Chakka K, Wu R, Lin C, Wolfe C, Prabhakaran B, et al. Comparison of in-person and synchronous remote musculoskeletal exam using augmented reality and haptics: A pilot study. PM R 2022 Aug 03. [doi: 10.1002/pmrj.12883] [Medline: 36197806]
68. Zhang F, Contreras CM, Shao P, Zhao L, Wu B, Li C, et al. Co-axial projective imaging for augmented reality telementoring in skin cancer surgery. Ann Biomed Eng 2022 Dec 05;50(12):1846-1856 [FREE Full text] [doi: 10.1007/s10439-022-03000-4] [Medline: 35788468]
69. Stone NN, Wilson MP, Griffith SH, Immerzeel J, Debruyne F, Gorin MA, et al. Remote surgical education using synthetic models combined with an augmented reality headset. Surg Open Sci 2022 Oct;10:27-33 [FREE Full text] [doi: 10.1016/j.sopen.2022.06.004] [Medline: 35866070]
70. Zhang J, Yu N, Wang B, Lv X. Trends in the use of augmented reality, virtual reality, and mixed reality in surgical research: a global bibliometric and visualized analysis. Indian J Surg 2022 Feb 24;84(Suppl 1):1-18 [FREE Full text] [doi: 10.1007/s12262-021-03243-w] [Medline: 35228782]
71. Bifulco P, Narducci F, Vertucci R, Ambruosi P, Cesarelli M, Romano M. Telemedicine supported by Augmented Reality: an interactive guide for untrained people in performing an ECG test. Biomed Eng Online 2014 Nov 21;13:153 [FREE Full text] [doi: 10.1186/1475-925X-13-153] [Medline: 25413448]
72. Mahmud N, Cohen J, Tsourides K, Berzin TM. Computer vision and augmented reality in gastrointestinal endoscopy. Gastroenterol Rep (Oxf) 2015 Aug;3(3):179-184 [FREE Full text] [doi: 10.1093/gastro/gov027] [Medline: 26133175]
73. Kashiwagi S, Asano Y, Goto W, Morisaki T, Shibutani M, Tanaka H, et al. Optical see-through head-mounted display (OST-HMD)-assisted needle biopsy for breast tumor: a technical innovation. In Vivo 2022 Mar 02;36(2):848-852 [FREE Full text] [doi: 10.21873/invivo.12772] [Medline: 35241541]
74. Schoeb DS, Schwarz J, Hein S, Schlager D, Pohlmann PF, Frankenschmidt A, et al. Mixed reality for teaching catheter placement to medical students: a randomized single-blinded, prospective trial. BMC Med Educ 2020 Dec 16;20(1):510 [FREE Full text] [doi: 10.1186/s12909-020-02450-5] [Medline: 33327963]
75. Joda T, Gallucci G, Wismeijer D, Zitzmann N. Augmented and virtual reality in dental medicine: a systematic review. Comput Biol Med 2019 May;108:93-100. [doi: 10.1016/j.compbiomed.2019.03.012] [Medline: 31003184]
76. Leitritz MA, Ziemssen F, Suesskind D, Partsch M, Voykov B, Bartz-Schmidt KU, et al. Critical evaluation of the usability of augmented reality ophthalmoscopy for the training of inexperienced examiners. Retina 2014 Apr;34(4):785-791. [doi: 10.1097/IAE.0b013e3182a2e75d] [Medline: 24670999]
77. Lindeman DA, Kim KK, Gladstone C, Apesoa-Varano EC. Technology and caregiving: emerging interventions and directions for research. Gerontologist 2020 Feb 14;60(Suppl 1):S41-S49 [FREE Full text] [doi: 10.1093/geront/gnz178] [Medline: 32057082]
78. Dorner SC, Wint AJ, Brenner PS, Keefe B, Palmisano J, Iezzoni LI. Patient perceptions of in-home urgent care via mobile integrated health. Am J Manag Care 2022 Apr 8;28(4):152-158 [FREE Full text] [doi: 10.37765/ajmc.2022.88859] [Medline: 35420743]
79. Murali S, Paul KD, McGwin G, Ponce BA. Updates to the current landscape of augmented reality in medicine. Cureus 2021 May 16;13(5):e15054 [FREE Full text] [doi: 10.7759/cureus.15054] [Medline: 34150401]

Abbreviations

AR: augmented reality
ARTESH: Augmented Reality–based Telerehabilitation System with Haptics
EM: emergency medicine
HMD: head-mounted device
HMD-STAR: head-mounted device with System for Telementoring with Augmented Reality
PPE: personal protective equipment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
STAR: System for Telementoring with Augmented Reality
VIP: virtual interactive presence
VIPAAR: virtual interactive presence and augmented reality
VR: virtual reality
WOC: wound, ostomy, and continence

Edited by L Buis; submitted 04.01.23; peer-reviewed by V Nambudiri, E Presseller; comments to author 27.02.23; revised version received 01.03.23; accepted 15.03.23; published 18.04.23.

Please cite as:
Dinh A, Yin AL, Estrin D, Greenwald P, Fortenko A
Augmented Reality in Real-time Telemedicine and Telementoring: Scoping Review
JMIR Mhealth Uhealth 2023;11:e45464
URL: https://mhealth.jmir.org/2023/1/e45464
doi:10.2196/45464
PMID:37071458

Original Paper

Willingness to Use and Pay for Digital Health Care Services According to 4 Scenarios: Results from a National Survey

Junbok Lee[1,2], MA; Yumi Oh[3], PhD; Meelim Kim[1,4], PhD; Belong Cho[2,5], MD, PhD; Jaeyong Shin[6,7,8], MD, MPH, PhD

[1]Health-IT Center, Yonsei University Health System, Seoul, Republic of Korea
[2]Department of Human Systems Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
[3]Korea Health Promotion Institute, Seoul, Republic of Korea
[4]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States
[5]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[6]Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[7]Institute of Health Services Research, Yonsei University, Seoul, Republic of Korea
[8]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Jaeyong Shin, MD, MPH, PhD
Department of Preventive Medicine
Yonsei University College of Medicine
50-1 Yonsei-Ro
Seodaemun-Gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 1881
Email: DRSHIN@yuhs.ac

Abstract

Background: Smartphones and their associated technology have evolved to an extent where these devices can be used to provide digital health interventions. However, few studies have been conducted on the willingness to use (WTU) and willingness to pay (WTP) for digital health interventions.

Objective: The purpose of this study was to investigate how previous service experience, the content of the services, and individuals' health status affect WTU and WTP.

Methods: We conducted a nationwide web-based survey in 3 groups: nonusers (n=506), public service users (n=368), and private service users (n=266). Participants read scenarios about an imagined health status (such as having a chronic illness) and the use of digital health intervention models (self-management, expert management, and medical management). They were then asked to respond to questions on WTU and WTP.

Results: Public service users had a greater intention to use digital health intervention services than nonusers and private service users: scenario A (health-risk situation and self-management), nonusers=odd ratio [OR] .239 (SE .076; *P*<.001) and private service users=OR .138 (SE .044; *P*<.001); scenario B (health-risk situation and expert management), nonusers=OR .175 (SE .040; *P*<.001) and private service users=OR .219 (SE .053; *P*<.001); scenario C (chronic disease situation and expert management), nonusers=OR .413 (SE .094; *P*<.001) and private service users=OR .401 (SE .098; *P*<.001); and scenario D (chronic disease situation and medical management), nonusers=OR .480 (SE .120; *P*=.003) and private service users=OR .345 (SE .089; *P*<.001). In terms of WTP, in scenarios A and B, those who used the public and private services had a higher WTP than those who did not (scenario A: β=–.397, SE .091; *P*<.001; scenario B: β=–.486, SE .098; *P*<.001). In scenario C, private service users had greater WTP than public service users (β=.264, SE .114; *P*=.02), whereas public service users had greater WTP than nonusers (β=–.336, SE .096; *P*<.001). In scenario D, private service users were more WTP for the service than nonusers (β=–.286, SE .092; *P*=.002).

Conclusions: We confirmed that the WTU and WTP for digital health interventions differed based on individuals' prior experience with health care services, health status, and demographics. Recently, many discussions have been made to expand digital health care beyond the early adapters and fully into people's daily lives. Thus, more understanding of people's awareness and acceptance of digital health care is needed.

(JMIR Mhealth Uhealth 2023;11:e40834) doi:10.2196/40834

KEYWORDS

digital health intervention; service experience; willingness to pay; willingness to use; digital health; health technology

Introduction

Since their introduction, smartphones and their associated technology have evolved to an extent where these devices can be used to provide personal health care services through various mobile apps [1-4]. The number of such apps continues to increase [5,6].

Through such digital health interventions, people can manage their health anytime and anywhere [7,8]. Additionally, digital health interventions have the advantage of enabling health management through features that use objective, numerical, and health-related data, such as the step counter and heart rate tracker [9]. Digital health interventions aid in continuous health management, strengthening the potential to prevent chronic diseases by promoting constant individual health monitoring and to reduce medical expenses [10,11]. Interest in these interventions has further increased because of the COVID-19 pandemic and the consequent restrictions on outdoor activities and movements [12-15]. Given the growing preference for services that do not require physical contact, digital health interventions will only become more prevalent [16]. Digital health interventions are being developed for various medical conditions to complement traditional medical care and improve patient experience [17,18]. The US Food and Drug Administration approved several digital health apps, such as those for the management of diabetes (BlueStar) or the treatment of substance use disorder (reSET) [19]. In Germany, digital health apps approved by BfArM (*Bundesamt für Arzneimittel und Medizinprodukte*; the German Federal Institute for Drugs and Medical Devices) could be included in the DiGA (*digitale Gesundheitsanwendung*; digital health applications) directory for reimbursement [20].

Even with the convenience, usefulness, and potential for future development of digital health interventions, only some people manage their health using digital health care tools [21,22]. Additionally, the retention rate of digital health intervention services is low [23-25]. It is necessary to provide opportunities for more people to experience the service and make them use the service continuously. Thus, it is vital to identify factors that affect people's willingness to use (WTU) and willingness to pay (WTP) for these services.

Only a few studies have been conducted on the WTU and WTP for digital health interventions [26-30]. Previous studies have identified demographic and health-related factors that affect the WTP for digital health interventions [26]. Research showed that the absolute WTP of those in the UK-representative cohort was £196 (US $258) and the marginal WTP was £160 (US $211), whereas those who availed the national digital health program had an absolute WTP of £162 (US $214) and a marginal WTP of £151 (US $199). Another study conducted an experimental vignette to identify factors affecting people's use of and payment for mobile health care apps in the context of 4 different business models [27]. It showed that doctors' recommendations helped increase both the WTU and WTP in Germany and the Netherlands.

This study intended to investigate how individuals' previous service experience, the content of the services, and health status influence the WTU and WTP of digital health interventions. Referring to previous research [26], we surveyed not only those who availed public digital health intervention services but also those who had experience with private services in South Korea. We subdivided digital health interventions into self-management, expert (nonmedical) management, and medical personnel management to identify differences by service type.

Methods

Digital Health Interventions: Public and Private Service

Mobile Healthcare at public health centers is a free health care service program provided by the South Korean government. The service team at public health centers helps individuals manage their daily lives by setting health goals and counseling them via smartphones with activity trackers. As part of the program, they visit the public health centers for counseling and examinations and revisit after 3 and 6 months for check-ups.

Company N's digital health intervention is a mobile-based app service whose users aim to lose weight and prevent diabetes through lifestyle changes. Based on behavioral science and psychology, health care coaches communicate with users to set health care goals and provide nutrition and exercise feedback, which help them achieve those goals. This service is used in the Centers for Disease Control and Prevention's Diabetes Prevention Program in the United States.

Participants

For this study, a nationwide web-based and mobile survey of people aged 19 to 59 years was conducted. We recruited participants from 3 groups: nonusers (n=506), public service users (n=368), and private service users (n=266). Public service users were participants who took part in the Mobile Healthcare program at public health centers. Private service users were people who experienced Company N's digital health intervention. In the case of public and private service users, recruitment notices were posted on the notice board of the mobile apps. People who expressed their intention to participate in the survey received a survey link. Nonusers were recruited via emails to a large-scale web-based panel of a research company. Based on the Mobile Healthcare project promoted by the Korean Ministry of Health and Welfare, samples of public service users were recruited using a proportional allocation of gender, age, and residence in 2020. Nonservice users were sampled using proportional rates based on gender, age, and residence as of 2020 in South Korea for national representation.

All participants responded through a web page developed by the research company, and the data were stored in the research company's database. The participants received ₩1500 (South Korean won; US $1=₩1100) as a reward for the survey. Table 1 shows the demographic distribution of the study participants.

Table 1. Demographic distribution of the participants.

	Nonuser (n=506)	Public service user (n=368)	Private service user (n=266)
Demographics			
Age range (years), n (%)			
19-29	121 (23.9)	23 (6.3)	64 (24.1)
30-39	113 (22.3)	92 (25)	78 (29.3)
40-49	136 (26.9)	160 (43.5)	72 (27.1)
≥50	136 (26.9)	93 (25.3)	52 (19.5)
Gender, n (%)			
Men	258 (51)	112 (30.4)	122 (45.9)
Women	248 (49)	256 (69.6)	144 (54.1)
Residence, n (%)			
Seoul Capital Area	268 (53)	76 (20.7)	80 (30.1)
Others	238 (47)	292 (79.3)	186 (69.9)
Health status			
Medication, n (%)			
Yes	160 (31.6)	45 (12.2)	57 (21.4)
No	346 (68.4)	323 (87.8)	209 (78.6)
Hypertension or diabetes, n (%)			
Yes	119 (23.5)	56 (15.2)	41 (15.4)
No	387 (76.5)	312 (84.8)	225 (84.6)

Design and Procedure

People's WTU and WTP may change according to the contents of digital health interventions, and several factors can influence them. Therefore, in this study, we created 3 digital health intervention models referring to the Evidence Standards Framework developed by The National Institute for Health and Care Excellence in the United Kingdom [31]. According to the functional aspect and potential risks, this framework classifies the level of digital health technology (DHT) into tier 1, tier 2, tier 3a, and tier 3b. Tier 1 includes DHT that provides systemic benefits but no direct benefits to the patient. Tier 2 includes DHT that cannot evaluate a patient's health outcomes but can help them live a healthy life by providing information and offering simple monitoring services based on the patient's health-related data. Tier 3a is a service for preventive behavioral modifications and management (which allows users to record and selectively exchange data with specialists) designed to modify health behaviors and eating habits using DHT. Tier 3b refers to DHT with measurable improvements, such as treatment and diagnosis devices. These include devices that provide treatment for diagnosed diseases, provide automated information records and data to experts, and perform calculations that affect clinical decisions. Tier 1 was excluded from this study because it does not directly serve patients.

All the study participants read the scenarios (Textbox 1) and responded to the questions (Table 2). They first read situation scenarios about their health status to imagine themselves as part of a high-risk group. Subsequently, they read the content of digital health interventions for self-management and then indicated their WTU and WTP. Thereafter, they read the scenario for digital health interventions administered by a health care professional (nonmedical person) and responded with their WTU and WTP in the same way.

Next, the participants read the chronic disease patient scenario to imagine themselves as patients with a chronic disease. They read the content of digital health interventions offered by a health care professional (nonmedical person; same as the previous service scenario) and responded with their WTU and WTP. Lastly, they responded with their WTU and WTP for the mobile app that verified the treatment effect and was managed by a doctor.

Textbox 1. Text of the scenarios.

Health-risk situation

- Imagine that a routine medical check-up reveals that you are at a high risk of becoming diabetic. The doctor advises you to come back for a check-up after three months of regular exercising and eating a healthy diet instead of prescribing medication.

Self-management

- A service that allows people to enter and monitor health-related data weekly or monthly, such as their food intake, steps walked, weight, blood pressure level, pulse rate, blood sugar level, and so on.

Expert management

- A service wherein a healthcare expert (non-medical person) sets up an exercise and diet plan based on the health-related information (diet, weight, etc.) provided, and sends messages via the application on a regular basis for counselling, or to share educational information and advice.

Chronic disease situation

- Imagine that you started taking diabetes medicine because your fasting blood sugar level did not drop, and the doctor recommended also utilizing the suggested service.

Expert management

- A service wherein a healthcare expert (non-medical person) sets up an exercise and diet plan based on the health-related information (diet, weight…) provided, and sends messages via the application on a regular basis for counselling, or to share educational information and advice.

Medical management

- A mobile application-based service that proves the effectiveness of diabetes treatment and a doctor checks the medical data entered by the patient undergoing treatment as well as provides a customized exercise and diet plan.

Table 2. Scenario design.

	Self -management (tier 2)	Expert management (tier 3a)	Medical management (tier 3b)
Health-risk situation	Scenario A	Scenario B	N/A[a]
Chronic disease situation	N/A	Scenario C	Scenario D

[a]N/A: not applicable.

Covariates

Sociodemographic and Health Status

Before reading the scenarios, the participants provided information about their age, gender, and residence. After the survey ended, they provided information about their income and occupation. They also indicated their subjective health status, diagnosed disease (high blood pressure, diabetes, etc), and whether they had taken medication for 3 months or longer within the last year for their disease.

Dependent Variables, WTU, and WTP Questions

The participants read 4 scenarios (A to D) and indicated their WTU the health care service app in each scenario on a 4-point scale (1=not at all, 2=not very much, 3=somewhat willing, and 4=highly willing). Participants who responded with "somewhat" and "a lot" were asked how much they were WTP per month for the service.

Statistical Analysis

In this study, logistic regression was used to identify factors affecting the WTU digital health interventions with completed questionnaires. Multiple linear regression analysis was conducted to confirm the WTP. Regarding the WTP, the analysis was performed by log transformation. We also conducted an ANOVA to find out the difference between participants' WTU and WTP according to their service experience. STATA (version 16; StataCorp) software was used for the analysis.

Ethics Approval

This study was approved by the institutional review board of the Korea Health Promotion Institute (120160811107AN01-2020-HR-049-02). All participants agreed to participate in the study after reading the explanation page, including the purpose of this study, the number of participants, and the data storage period.

Results

WTP

In scenario A, the average WTP was ₩21,909 (SD ₩22,418) for private service users, ₩17,020 (SD ₩15,877) for public service users, and ₩11,913 (SD ₩11,090) for nonusers. In scenario B, the average WTP was ₩17,636 (SD ₩15,356) for private service users, ₩15,392 (SD ₩15,278) for public service users, and ₩10,279 (SD ₩10,428) for nonusers. In scenario C, the average WTP was ₩21,322 (SD ₩21,836) for private service users, ₩14,516 (SD ₩14,977) for public service users, and ₩11,906 (SD ₩12,268) for nonusers. In scenario D, the average WTP was ₩23,520 (SD ₩25,277) for private service users, ₩15,500 (SD ₩16,035) for public service users, and ₩13,084 (SD ₩13,288) for nonusers. Table 3 shows the summary statistics of the WTP.

Table 3. Willingness to pay (₩; US $1=₩1100).

	Nonuser (n=506), ₩ (US $)	Public service user (n=368), ₩ (US $)	Private service user (n=266), ₩ (US $)
Scenario A			
Mean	11,913 (10.8)	17,020 (15.5)	21,909 (19.9)
SD	11,090 (10.1)	15,877 (14.4)	22,418 (20.4)
Median	10,000 (9.1)	10,000 (9.1)	14,000 (12.7)
Range	0-65,000 (0-59.1)	0-100,000 (0-90.9)	0-109,000 (0-99.1)
Scenario B			
Mean	10,279 (9.3)	15,392 (14.0)	17,636 (16.0)
SD	10,428 (9.5)	15,278 (13.9)	15,356 (14.0)
Median	7000 (6.4)	10,000 (9.1)	10,000 (9.1)
Range	0-55,000 (0-50)	0-90,000 (0-81.8)	0-80,000 (72.7)
Scenario C			
Mean	11,906 (10.8)	14,516 (13.2)	21,322 (19.4)
SD	12,268 (11.2)	14,977 (13.6)	21,836 (19.9)
Median	8000 (7.3)	10,000 (9.1)	11,000 (10.0)
Range	0-70,000 (0-63.6)	0-90,000 (0-81.8)	0-100,000 (0-90.9)
Scenario D			
Mean	13,084 (11.9)	15,500 (14.1)	23,520 (21.4)
SD	13,288 (12.1)	16,035 (14.6)	25,277 (23.0)
Median	10,000 (9.1)	10,000 (9.1)	15,000 (13.6)
Range	0-80,000 (0-72.7)	0-100,000 (0-90.9)	0-150,000 (0-136.4)

WTU and WTP in Scenario A (Health-Risk Situation and Self-management)

We explored the WTU and WTP of those in the health-risk situation on the self-management service (see Multimedia Appendix 1). Linear regression analyses were conducted to identify the factors influencing the WTU digital health interventions. In scenario A, age, gender, income, type of service experience, medication use, and residence affected the WTU. Specifically, younger people (odds ratio [OR] .961, SE .010; *P*<.001), women (OR .470, SE .099; *P*<.001), people with high income (OR 1.364, SE .099; *P*<.001), and people who lived in the metropolitan area (OR .629, SE .144; *P*=.04) were more WTU the self-management service. In the case of service experience, public service users had a greater intention to use the self-management service than nonusers and private service users (nonusers: OR .239, SE .076; *P*<.001; private service users: OR .138, SE .044; *P*<.001). People who were on medication for 1 year were more likely to use the self-management service (OR 3.171, SE 1.082; *P*=.001). The explanatory power of the model with all predictors was 13.1%. We also conducted logistic regression analyses to examine the WTP in scenario A. Those who used the public and private services had a higher WTP than those who did not (β=–.397, SE .091; *P*<.001). The explanatory power of WTP was 5.2%.

WTU and WTP in Scenario B (Health-Risk Situation and Expert Management)

We investigated the WTU and WTP of those in the health-risk situation on the expert management service (see Multimedia Appendix 2). The type of service experience and medication use affected the WTU in scenario B. Specifically, public service users showed greater intention to use the service than private service users and nonusers (nonusers: OR .175, SE .040; *P*<.001; private service users: OR .219, SE .053; *P*<.001). People who were on medication for 1 year were more WTU the expert management service (OR 1.773, SE .365; *P*=.005). The explanatory power of the model with all predictors was 7.2%. We also performed logistic regression analyses to examine the WTP. Private and public service users had higher WTP than nonusers (β=–.486, SE .098; *P*<.001). In terms of demographics, women had a greater WTP for the service than men (β=.233, SE .082; *P*=.005). The explanatory power of WTP was 5.9%.

WTU and WTP in Scenario C (Chronic Patient Situation and Expert Management)

We conducted linear regression analyses to explore the WTU in scenario C (see Multimedia Appendix 3). Several factors influenced the WTU: age, gender, type of service experience, and having high blood pressure or diabetes. To be specific, younger people (β=.982, SE .008; *P*=.03) and women (β=.705, SE .119; P=.04) were more WTU the service. In the case of service experience, public service users showed greater WTU the expert management service than nonusers and private service users (nonusers: OR .413, SE .094; P<.001; private service users: OR .401, SE .098; P<.001). Regarding health status, those with high blood pressure or diabetes showed more WTU the service (OR 1.751, SE .487; P=.04). The explanatory power of the model with all predictors was 3.4%. The result of logistic analyses of the WTP showed that private service users had greater intention to pay than public service users (β=.264, SE .114; P=.02), whereas public service users had greater WTP than nonusers (β=–.336, SE .096; P<.001). Women were more WTU for the service than men (β=.250, SE .080; P=.002), similar to scenario B. The explanatory power of the WTP was 5.4%.

WTU and WTP in Scenario D (Chronic Patient Situation and Medical Management)

We performed linear regression analyses to investigate the WTU of those in the chronic disease situation on the medical management service (see Multimedia Appendix 4). Age, gender, type of service experience, and having high blood pressure or diabetes influenced the WTU in scenario D. Specifically, younger people (OR .967, SE .009; P<.001) and women (OR .569, SE .106; P=.002) were more WTU the service. Similar to the other scenarios, public service users showed greater WTU the medical management service (nonusers: OR .480, SE .120; P=.003; private service users: OR .345, SE .089; P<.001). Those with high blood pressure or diabetes had higher intention to use the service (OR 1.894, SE .596; P=.04). The explanatory power of the model with all predictors was 5.2%. The result of logistic analyses of the WTP revealed that private and public service users were more WTP for the service than nonusers (β=–.286, SE .092; P=.002). Those who experienced private services had a marginally higher WTP than those who used public services (β=.193, SE .111; P=.08). Younger people (β=–.010, SE .004; P=.007) and women (β=.177, SE .779; P=.02) had a higher likelihood of paying for the medical management service. The explanatory power of the model for WTP was 4.4%.

Discussion

Principal Findings

We conducted a web-based survey to investigate the WTU and WTP according to the type of digital health intervention, wherein the respondents were divided into 3 groups (nonusers, public service users, and private service users). We also aimed to identify the factors that affect the WTU and WTP for digital health interventions.

Participants' WTU and WTP for digital health interventions differed significantly based on their prior experience with health care services. Public service users tended to use digital health intervention more than nonusers and private service users, whereas private service users were more WTP for digital health interventions than the others. This trend was true for all 4 scenarios. Private service users had an average WTP that is 1.5 to 2 times higher than nonusers. However, in this study, it is not clear whether those with high WTP used private services or whether their WTP increased because of positive experiences with the services. Public service users were 1.2 to 1.5 times more WTP than nonusers. In the health-risk situation, there was no difference in the WTP between public and private service users, but the WTP of private service users was much higher than that of public service users in the chronic disease situation.

At first, we expected that private service users would have a higher intention to use digital health interventions. Contrary to our expectations, public services users showed higher WTU such interventions. This might imply higher motivation and interest in terms of health care among public service users. They must have visited the public health center at least thrice to avail the service for additional examination and counseling and receive activity trackers. As the public service is linked to the national health examination, perhaps they realized how severe their health condition was and felt the need for health care. Hence, they showed high intention to use digital health interventions in our study.

An individual's health status is one of the most critical aspects of the WTP for digital health interventions. Even when the same service was to be provided by a nonmedical person (scenarios B and C), participants were WTP more after reading scenario C (chronic disease situation) than in scenario B (health-risk situation). Additionally, in the health-risk situation scenarios (A and B), having high blood pressure and diabetes did not affect their WTU digital health interventions. Rather, participants with such ailments were more WTU digital health interventions than the others in the chronic disease situation scenarios (C and D).

The content of the services provided is a factor that affects people's WTP and WTU. This scenario was developed based on The National Institute for Health and Care Excellence's Evidence Standards Framework. Scenario A is a self-care service that allows people to manage their health. Services in scenarios B and C help people manage exercise and diet with health care experts. In contrast, scenario D is a service in which medical professionals manage diseases with services verified to be effective. Participants wanted to use and pay more for the service in scenario D than in scenarios B and C. This result shows that the more professional and advanced the service, the more willing people are to use and pay for the service.

In scenario D, the importance of validating effectiveness in digital health care services was confirmed. Compared to other scenarios, people were WTU and WTP more for the service that validated their effectiveness. The previous study showed that the effectiveness of digital health care services is an essential factor for British health professionals [32]. Our result also indicated that service users also recognized the importance of effectiveness by clinical evidence.

Another factor highlighted in this study is the percentage of people who are WTP for services. Similar previous studies, the proportion of people WTP for services was at the level of 50% to 60%, and the rest were unwilling to pay [26]. This result indicates that it is necessary to work on the maturation of DHT.

Like previous research [26,27,29], this study showed that age, gender, and income affected the WTP for digital health interventions. Younger individuals and men were more WTP for digital health intervention services. However, the results regarding the WTU somewhat differed from previous studies. Women were more WTU digital health interventions than men in scenarios A, C, and D.

Limitation

First, the participants responded to hypothetical scenarios, which means that they answered based on what they imagined about a given situation. Their response to a similar real-life situation may distinctly differ. Their reactions to similar situations in practice may be different because of the service's various features that determine the price, such as governmental regulatory approval with significant evidence or the existence of a physician guide. Additionally, this study cited diabetes as an example of a chronic disease. Patients with other serious diseases may want to pay more for services. Despite this limitation, this study might help provide a lot of insight into developing user-centered services. Second, different countries have different health insurance systems, so the WTU and WTP in other nations may differ from the result of this study. The WTP presented in this study is the general price average for hypothetical scenarios, and attention is needed to interpret it directly. Third, the WTP in scenario A was higher than those in other scenarios. This may be attributed to the question order bias. To reduce this bias, the order of scenarios A to D could have been presented at random. However, we did not do so because we had to consider the presence or absence of disease and services. Fourth, the explanatory power of the regression analysis was not high. However, even in a previous study [27], the explanatory power of the WTP was 3% to 8%, close to the WTP explanatory power of 4.4% to 5.9% for scenarios A to D in this study.

Conclusion

Digital health care technology has continued to develop and is expected to grow further. More people are WTU their smartphones to manage their health, creating various health care innovations. Recently, there have been many discussions about expanding digital health care for general people, not only early adopters. However, studies have yet to be conducted on the WTU and WTP for digital health care. It is necessary to develop a deeper understanding of people's awareness and acceptance of digital health care. Digital health care companies should develop their product based on this understanding. Since digital health care needs to work within the health care system, it is essential to evaluate the effectiveness of the services with clinical evidence.

Acknowledgments

This research was supported by the National Health Promotion Fund, funded by the Ministry of Health and Welfare, Republic of Korea, and by the Technology Innovation Program (20018246, Promotion of Digital Therapeutics Industry for Global Expansion), funded by the Ministry of Trade, Industry & Energy (MOTIE; Republic of Korea). We thank the Korea Health Promotion Institute and Noom for helping to recruit participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Willingness to use (WTU) and willingness to pay (WTP) in scenario A (health-risk situation and self-management).
[DOCX File , 31 KB - mhealth_v11i1e40834_app1.docx]

Multimedia Appendix 2
Willingness to use (WTU) and willingness to pay (WTP) in scenario B (health-risk situation and expert management).
[DOCX File , 31 KB - mhealth_v11i1e40834_app2.docx]

Multimedia Appendix 3
Willingness to use (WTU) and willingness to pay (WTP) in scenario C (chronic disease situation and expert management).
[DOCX File , 32 KB - mhealth_v11i1e40834_app3.docx]

Multimedia Appendix 4
Willingness to use (WTU) and willingness to pay (WTP) in scenario D (chronic disease situation and medical management).
[DOCX File , 32 KB - mhealth_v11i1e40834_app4.docx]

References

1. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
2. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: a review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]

3. Walker DM, Diana ML. Hospital adoption of health information technology to support public health infrastructure. J Public Health Manag Pract 2016;22(2):175-181. [doi: 10.1097/PHH.0000000000000198] [Medline: 26811967]
4. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
5. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR mHealth uHealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
6. Aitken M. Digital Health Trends. Danbury, CT: IQVIA Institute; 2021.
7. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014 Feb 5;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
8. Lundgren AS, Lindberg J, Carlsson E. "Within the hour" and "wherever you are": exploring the promises of digital healthcare apps. Journal of Digital Social Research 2021 Oct 26;3(3):32-59. [doi: 10.33621/jdsr.v3i3.77]
9. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013 Feb 13;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
10. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR mHealth uHealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
11. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
12. Keesara S, Jonas A, Schulman K. COVID-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
13. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
14. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
15. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
16. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19: cross-sectional analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
17. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep;3:38-38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
18. Győrffy Z, Radó N, Mesko B. Digitally engaged physicians about the digital health transition. PLoS One 2020 Sep 28;15(9):e0238658 [FREE Full text] [doi: 10.1371/journal.pone.0238658] [Medline: 32986733]
19. Patel NA, Butte AJ. Characteristics and challenges of the clinical pipeline of digital therapeutics. NPJ Digit Med 2020 Dec 11;3(1):159 [FREE Full text] [doi: 10.1038/s41746-020-00370-8] [Medline: 33311567]
20. The fast-track process for digital health applications (DiGA) according to section 139e SGB V. Federal Institute for Drugs and Medical Devices (BfArM). URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf;jsessionid=2C5B67183372C3B0DAAD7DC886953C1C.1_cid354?__blob=publicationFile&v=2%20[accessed%202023-02-02] [accessed 2023-02-02]
21. Dehzad F, Hilhorst C, de Bie C, Claassen E. Adopting health apps, what's hindering doctors and patients? Health 2014 Sep;06(16):2204-2217. [doi: 10.4236/health.2014.616256]
22. Ha Dinh TT, Bonner A, Clark R, Ramsbotham J, Hines S. The effectiveness of the teach-back method on adherence and self-management in health education for people with chronic disease: a systematic review. JBI Database System Rev Implement Rep 2016 Jan;14(1):210-247. [doi: 10.11124/jbisrir-2016-2296] [Medline: 26878928]
23. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. There's an app for that, but nobody's using it: insights on improving patient access and adherence to digital therapeutics in Germany. Digit Health 2022 Jul 03;8:20552076221104672 [FREE Full text] [doi: 10.1177/20552076221104672] [Medline: 35811758]
24. Pfammatter AF, Mitsos A, Wang S, Hood SH, Spring B. Evaluating and improving recruitment and retention in an mHealth clinical trial: an example of iterating methods during a trial. Mhealth 2017 Nov 1;3:49 [FREE Full text] [doi: 10.21037/mhealth.2017.09.02] [Medline: 29184901]
25. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]

26. Somers C, Grieve E, Lennon M, Bouamrane M, Mair FS, McIntosh E. Valuing mobile health: an open-ended contingent valuation survey of a national digital health program. JMIR mHealth uHealth 2019 Jan 17;7(1):e3 [FREE Full text] [doi: 10.2196/mhealth.9990] [Medline: 30664488]
27. Lupiáñez-Villanueva F, Folkvord F, Vanden Abeele M. Influence of the business revenue, recommendation, and provider models on mobile health app adoption: three-country experimental vignette study. JMIR mHealth uHealth 2020 Jun 04;8(6):e17272 [FREE Full text] [doi: 10.2196/17272] [Medline: 32496204]
28. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
29. Shariful Islam SM, Lechner A, Ferrari U, Seissler J, Holle R, Niessen LW. Mobile phone use and willingness to pay for SMS for diabetes in Bangladesh. J Public Health (Oxf) 2016 Mar;38(1):163-169. [doi: 10.1093/pubmed/fdv009] [Medline: 25687131]
30. Callan A, O'Shea E. Willingness to pay for telecare programmes to support independent living: results from a contingent valuation study. Soc Sci Med 2015 Jan;124:94-102. [doi: 10.1016/j.socscimed.2014.11.002] [Medline: 25461866]
31. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence (NICE). 2019 Mar. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2022-07-07]
32. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR mHealth uHealth 2020 Jul 06;8(7):e17704 [FREE Full text] [doi: 10.2196/17704] [Medline: 32628118]

Abbreviations

BfArM: Bundesamt für Arzneimittel und Medizinprodukte (German Federal Institute for Drugs and Medical Devices)
DiGA: digitale Gesundheitsanwendung (digital health applications)
DHT: digital health technology
OR: odds ratio
WTP: willingness to pay
WTU: willingness to use

Edited by L Buis; submitted 25.07.22; peer-reviewed by N Maglaveras, J Shull; comments to author 06.01.23; revised version received 01.02.23; accepted 06.03.23; published 29.03.23.

Please cite as:
Lee J, Oh Y, Kim M, Cho B, Shin J
Willingness to Use and Pay for Digital Health Care Services According to 4 Scenarios: Results from a National Survey
JMIR Mhealth Uhealth 2023;11:e40834
URL: https://mhealth.jmir.org/2023/1/e40834
doi:10.2196/40834
PMID:36989025

Original Paper

Economic Evaluation of Digital Therapeutic Care Apps for Unsupervised Treatment of Low Back Pain: Monte Carlo Simulation

Daniel Lewkowicz[1], MSc; Erwin Bottinger[1], Prof Dr; Martin Siegel[2], Prof Dr

[1]Digital Health Center, Hasso Plattner Insitute, University of Potsdam, Potsdam, Germany

[2]Department of Empirical Health Economics, Technische Universität Berlin, Berlin, Germany

Corresponding Author:
Daniel Lewkowicz, MSc
Digital Health Center
Hasso Plattner Insitute
University of Potsdam
Prof.-Dr.-Helmert-Straße 2-3
Potsdam, 14482
Germany
Phone: 49 331 55094
Email: daniel.lewkowicz@hpi.de

Abstract

Background: Digital therapeutic care (DTC) programs are unsupervised app-based treatments that provide video exercises and educational material to patients with nonspecific low back pain during episodes of pain and functional disability. German statutory health insurance can reimburse DTC programs since 2019, but evidence on efficacy and reasonable pricing remains scarce. This paper presents a probabilistic sensitivity analysis (PSA) to evaluate the efficacy and cost-utility of a DTC app against treatment as usual (TAU) in Germany.

Objective: The aim of this study was to perform a PSA in the form of a Monte Carlo simulation based on the deterministic base case analysis to account for model assumptions and parameter uncertainty. We also intend to explore to what extent the results in this probabilistic analysis differ from the results in the base case analysis and to what extent a shortage of outcome data concerning quality-of-life (QoL) metrics impacts the overall results.

Methods: The PSA builds upon a state-transition Markov chain with a 4-week cycle length over a model time horizon of 3 years from a recently published deterministic cost-utility analysis. A Monte Carlo simulation with 10,000 iterations and a cohort size of 10,000 was employed to evaluate the cost-utility from a societal perspective. Quality-adjusted life years (QALYs) were derived from Veterans RAND 6-Dimension (VR-6D) and Short-Form 6-Dimension (SF-6D) single utility scores. Finally, we also simulated reducing the price for a 3-month app prescription to analyze at which price threshold DTC would result in being the dominant strategy over TAU in Germany.

Results: The Monte Carlo simulation yielded on average a €135.97 (a currency exchange rate of EUR €1=US $1.069 is applicable) incremental cost and 0.004 incremental QALYs per person and year for the unsupervised DTC app strategy compared to in-person physiotherapy in Germany. The corresponding incremental cost-utility ratio (ICUR) amounts to an additional €34,315.19 per additional QALY. DTC yielded more QALYs in 54.96% of the iterations. DTC dominates TAU in 24.04% of the iterations for QALYs. Reducing the app price in the simulation from currently €239.96 to €164.61 for a 3-month prescription could yield a negative ICUR and thus make DTC the dominant strategy, even though the estimated probability of DTC being more effective than TAU is only 54.96%.

Conclusions: Decision-makers should be cautious when considering the reimbursement of DTC apps since no significant treatment effect was found, and the probability of cost-effectiveness remains below 60% even for an infinite willingness-to-pay threshold. More app-based studies involving the utilization of QoL outcome parameters are urgently needed to account for the low and limited precision of the available QoL input parameters, which are crucial to making profound recommendations concerning the cost-utility of novel apps.

(JMIR Mhealth Uhealth 2023;11:e44585) doi:10.2196/44585

KEYWORDS

cost-utility analysis; cost; probabilistic sensitivity analysis; Monte Carlo simulation; low back pain; pain; economic; cost-effectiveness; Markov model; digital therapy; digital health app; mHealth, mobile health; health app; mobile app; orthopedic; QUALY; DALY; quality-adjusted life years; disability-adjusted life years; time horizon; veteran; statistics

Introduction

Background

Low back pain (LBP) poses a tremendous health burden for patients and health care systems worldwide, with a lifetime prevalence of up to 85% [1,2]. For patients with nonspecific and nonacute LBP, current clinical guidelines recommend conservative treatment with physiotherapy at regular intervals and increased physical activity [3,4]. Smartphone or web-based digital therapeutic care (DTC) apps offer a novel unsupervised treatment modality for patients with nonspecific LBP [5]. Although DTC apps are now offered by numerous providers, they all follow the same treatment approach, in that video-based exercises aim to replace face-to-face physiotherapy and the provided educational material aims to reinforce patients' coping abilities for everyday life [5]. A major strength of DTC apps lies in their potential inclusion of decision support interventions, which include tailored push notifications and personalized exercise recommendations that guide subscribed patients through the treatment program [5-7]. These decision support interventions may stimulate persistent engagement and thereby enhance coping abilities and support long-term treatment compliance [8,9].

In Germany, the *Digital Health Care Act* allows statutory health insurance providers to reimburse DTC apps since December 2019, if sound scientific evidence indicates that they are an effective treatment alternative [10]. At present, there are 2 companies, namely ViViRa and HelloBetter, which have developed apps that can provide digital therapeutic via the smartphone or PC and that are now listed in the Digital Health Applications (DiGA) directory to be prescribed for patients with LBP via International Classification of Diseases-10 (ICD-10) code M54 [10]. This paper explores potential trade-offs between higher chances of achieving better long-term health outcomes through lasting behavioral changes, as well as the risk of reimbursing the cost without any benefit for the patients because of higher attrition rates for unsupervised DTC programs as compared to the treatment as usual (TAU; ie, physiotherapy and medication for temporary pain relief [11]).

Objectives

We applied a probabilistic sensitivity analysis (PSA) to address uncertainties in the transition probabilities, attrition rates, cost components, and health-related quality of life (QoL) scores, which were beyond the scope of the deterministic analysis recently published by Lewkowicz et al [11]. Amending the recently published deterministic analysis offers a relevant contribution to the literature because decision-making based on Markov chains, or other at least moderately complex or nonlinear models, should not be based solely on deterministic models but should include parameter uncertainty as well [12]. Moreover, we intended to explore to what extent the results in this probabilistic analysis differ from the results in the base case analysis and to what extent a shortage of outcome data concerning QOL metrics impacts the overall results. Hence, this underlying PSA intends to reveal the incapacity of a deterministic sensitivity analysis to overcome the challenges of a small patient cohort to simulate the long-term uncertain utility of an intervention. Accordingly, this study aims to inform researchers and decision-makers equally—both to underline the importance of a large data set of QoL data gathered from a large patient cohort and for future approvals of DTC apps for LBP regarding a potential price range, for which such apps may be expected to be a cost-effective alternative to the TAU.

Methods

Ethical Considerations

Because this was a simulation study with no human participants, ethics approval was not sought.

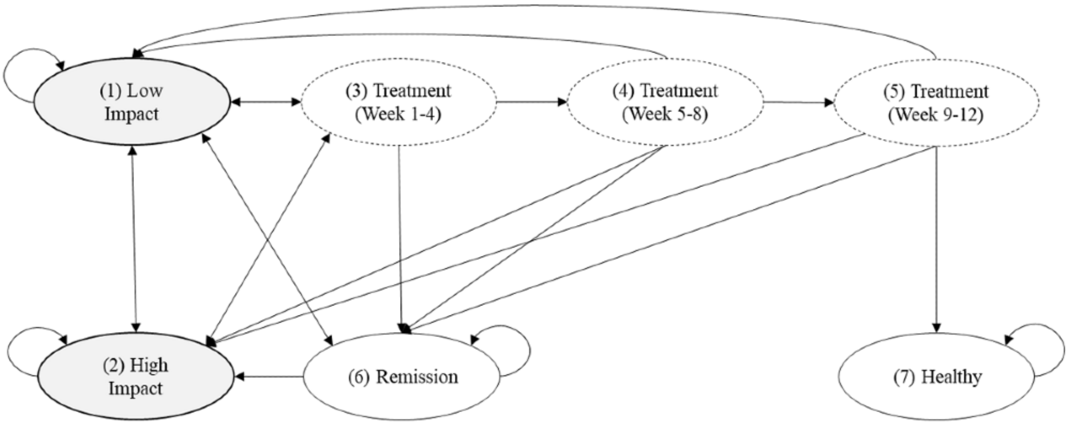
Model Framework

This paper builds on a recent analysis of the cost-utility of a DTC program for patients with nonacute LBP in Germany from a societal perspective [11]. The adopted state-transition model in Figure 1 comprises seven distinct health states: (1) low impact of LBP, (2) high impact of LBP, (3) treatment weeks 1 to 4, (4) treatment weeks 5 to 8, (5) treatment weeks 9 to 12, (6) remission, and (7) healthy. States 3, 4, and 5 represent different phases of the treatment progress. State 6 is a state of only temporary improvement, which allows for reoccurring phases with higher or lower pain intensities in the simulation, and state 7 is the final healthy state where no recrudescence can occur.

Like Lewkowicz et al [11], we covered a model time horizon of 3 years and used a cycle length of 4 weeks to allow the inclusion of different treatment states and for patients to drop out before finishing the 3-month course of treatment. Since no published evaluation studies for the ViViRa or HelloBetter DTC apps were available, Lewkowicz et al [11] employed outcome data from an evaluation of the Kaia Health app against 6 face-to-face physiotherapy sessions over a period of 12 weeks [13], arguing that the Kaia Health app is sufficiently similar to the 2 apps currently listed in the DiGA directory.

The transition probabilities for states 3, 4, and 5 were derived from the attrition rates reported in the Kaia Health app study [13]. Patients undergoing app-based treatment continued the program with a chance of 87.5% after each month. In the TAU group, 93.5% of the patients continued the recommended treatment program after the first month, and 95.7% continued after the second month. A recent systematic review on the effects of DTC apps for patients with LBP confirmed this pattern and found that attrition rates can even peak up to 80% in noncontrolled retrospective studies [5].

Figure 1. Discrete health state-transition Markov chain with 7 health states (adapted and reprinted from Lewkowicz et al [11]).



Lewkowicz et al [11] incorporated several assumptions in their model to be able to specify transition probabilities for their Markov chain. First, the probability of LBP patients visiting a general practitioner, and thus entering treatment, was set to 75% for low-impact LBP and to 80% for high-impact LBP. Second, 50% of the dropout patients were assumed to experience health improvements and thus move to the temporary remission state (state 6). The other 50% of the dropout patients were assumed to have stopped because of coping issues, lack of motivation, or time constraints. Of these, 82.5% fell back into the low-impact LBP state (state 1) and 17.8% fell back into the high-impact LBP state (state 2). Finally, the decision support interventions integrated into the DTC app were assumed to yield a 5% higher chance to transfer to the healthy state (state 7) [8,9,11] than in the TAU strategy [13]. We use the same figures here and display the resulting transition matrices for DTC and TAU in Table 1.

Lewkowicz et al [11] utilized the Veterans RAND 6-Dimension (VR-6D) preference single-utility index [14] derived from the Kaia Health study data [13] for QoLin states 1, 3, 4, and 5. For the remaining states, utility scores based on the Short-Form 6-Dimension (SF-6D) scale were retrieved from other lower back pain (LBP studies [15,16]). The cost components taken from [11] include direct costs for general practitioner and orthopedic consultations, diagnostic procedures, medication, and indirect costs through nonproductive time due to LBP. The price for the DTC app is the current reimbursement price of the ViViRa app of €239.96 for a 3-month prescription (a currency exchange rate of EUR €1=US $1.069 is applicable throughout this paper) [17]. The cost of face-to-face physiotherapy was set to €21.11 per session according to the binding German medical fee schedule [18]. The included utility scores and cost data were discounted with a discount factor of 3% [11].

Table 1. Transition matrix of the Markov chain.

To and from	Low impact (state 1)	High impact (state 2)	Treatment weeks 1-4 (state 3)	Treatment weeks 5-8 (state 4)	Treatment weeks 9-12 (state 5)	Remission (state 6)	Healthy (state 7)
Low impact (state 1)							
DTC[a]	0.2	0.0125	0.75[b]	0	0	0.0375	0
TAU[c]	0.2	0.0125	0.75[b]	0	0	0.0375	0
High impact (state 2)							
DTC	0.042	0.158	0.8[b]	0	0	0	0
TAU	0.042	0.158	0.8[b]	0	0	0	0
Treatment weeks 1-4 (state 3)							
DTC	0.0513	0.0111	0	0.87 [b]	0	0.0625	0
TAU	0.0267	0.0057	0	0.935[b]	0	0.0325	0
Treatment weeks 5-8 (state 4)							
DTC	0.0513	0.0111	0	0	0.875[b]	0.0625	0
TAU	0.0176	0.0038	0	0	0.957[b]	0.0215	0
Treatment weeks 9-12 (state 5)							
DTC	0.2350	0.0509	0	0	0	0.614[b]	0.1[b]
TAU	0.2761	0.0598	0	0	0	0.614[b]	0.05[b]
Remission (state 6)							
DTC	0.5047	0.1092	0	0	0	0.386	0
TAU	0.5047	0.1092	0	0	0	0.386	0
Healthy (state 7)							
DTC	0	0	0	0	0	0	1
TAU	0	0	0	0	0	0	1

[a]DTC: digital therapeutic care.

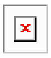
[b]Transition probabilities taken from the literature. All other transition probabilities in the respective rows are calculated from conditional probabilities given the respective event based on [11].

[c]TAU: treatment as usual.

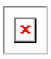
PSA Measure

For the PSA, which is a robust method to evaluate the impact of parameter uncertainties [12], we employed the aforementioned model and performed a Monte Carlo simulation with 10,000 iterations. In each iteration, the input parameters were randomly drawn from a priori–defined probability distributions for an entire cohort of 10,000 hypothetical patients. The model time horizon was 3 years with a state length of 4 weeks. We employed a beta distribution to simulate transition probabilities and QoL parameters and a gamma distribution to simulate costs.

We considered the input parameters for transition probabilities and QoL outcomes from the literature as “most likely” values and applied the Program Evaluation and Review Technique (PERT) approximation [19-21] to transform them into estimates for our mean and SD calculations [22] (Multimedia Appendix 1A). We then obtained the shape parameters α and β for the beta distribution through the method of moments [18,21]:



We applied the gamma distribution for all cost components, which requires the mean and SD of the cost components as input parameters. We used the results for direct and indirect cost components of chronic LBP over a 6-month period reported in a large German cost-of-illness study [23] to obtain cost estimates for health states 1, 2, 3, 4, and 5. We assumed costs to be distributed evenly over time and rescale the reported mean costs and the upper and lower limit of the 95% CIs to monthly costs. We derive the SD from the rescaled 95% CIs by dividing the range between the upper and lower limit by twice the 97.5% quantile of the normal distribution [24]:



where n=51 [23].

We deviated from the assumption in [11] that all physiotherapy costs occur in the first treatment cycle and allocated costs for weekly physiotherapy sessions to states 3 and 4 because they can only be paid if patients continue their treatment. Costs for 4 of the 6 physiotherapy sessions were allocated to state 3, and the remainder was allocated to state 4. The adapted input parameters, including the corresponding distribution parameters, are shown in Tables 2, 3, and 4. Multimedia Appendix 1B contains a full list of all parameters and probability density functions, and Multimedia Appendices 2-5 contain histograms of the parameters and matrices.

We derived cost-effectiveness acceptability curves (CEACs) to illustrate the probability of DTC apps being a cost-effective measure given a certain willingness-to-pay (WTP) threshold. The CEAC indicated the fraction of iterations considered to be cost-effective given a specific WTP. Graphically, the WTP threshold was a line through the origin with a slope equal to the respective WTP, and the outcome of an iteration in the Monte Carlo simulation was considered to be cost-effective if it lies below the WTP threshold in the cost-utility plane [22].

Some health care systems may only adopt novel technologies which are more effective than TAU, (ie, if its incremental effect is nonnegative). We derived an additional CEAC where we included only outcomes that lay in the southeast quadrant or in the northeast quadrant under the WTP threshold in the cost-utility plane to account for this constraint. Moreover, we computed the number of iterations where DTC strictly dominates TAU (ie, where $cost_{\_DTC}<cost_{\_TAU}$ and $effect_{\_DTC}>effect_{\_TAU}$, and vice-versa).

Table 2. Transition probabilities and beta parameters for simulation after PERT[q] transformation.

Transition probability	Expected value[b]	SD[c]	α[d]	β[d]
Temporary health states				
Low impact (state 1) to				
Low impact (state 1)	0.30000	0.16667	1.96800	4.59200
High impact (state 2)	0.02439	0.01234	3.78404	151.36171
Treatment weeks 1-4 (state 3)	0.66667	0.16667	4.66667	2.33333
Remission (state 6)	0.06977	0.03612	3.40097	45.34629
High impact (state 2) to				
Low Impact (state 1)	0.07749	0.04027	3.33817	39.74008
High Impact (state 2)	0.27200	0.16667	1.66697	4.46160
Treatment weeks 1-4 (state 3)	0.70000	0.16667	4.59200	1.96800
Remission (state 6) to				
Low impact (state 1)	0.50314	0.16667	4.02493	3.97471
High impact (state 2)	0.17938	0.09793	2.57361	11.77401
Remission (state 6)	0.42400	0.16667	3.30384	4.48823
DTC[e]				
Treatment weeks 1-4 (state 3) to				
Low impact (state 1)	0.09318	0.04880	3.21274	31.26755
High impact (state 2)	0.02177	0.01100	3.80694	171.09847
Treatment weeks 5-8 (state 4)	0.80000	0.11024	9.73257	2.43314
Remission (state 6)	0.11111	0.05871	3.07279	24.58235
Treatment weeks 5-8 (state 4) to				
Low impact (state 1)	0.09318	0.04880	3.21274	31.26755
High impact (state 2)	0.02177	0.01100	3.80694	171.09847
Treatment weeks 9-12 (state 5)	0.80000	0.11024	9.73257	2.43314
Remission (state 6)	0.11111	0.05871	3.07279	24.58235
Treatment weeks 9-12 (state 5) to				
Low impact (state 1)	0.32339	0.16667	2.22404	4.65314
High impact (state 2)	0.09241	0.04838	3.21882	31.61410
Remission (state 6)	0.57600	0.16667	4.48823	3.30384
Healthy (state 7)	0.16667	0.09045	2.66255	13.31276
TAU[f]				
Treatment weeks 1-4 (state 3) to				
Low impact (state 1)	0.05072	0.02601	3.55883	66.60730
High impact (state 2)	0.01144	0.00575	3.89784	336.89235
Treatment weeks 5-8	0.88496	0.06090	23.40459	3.04260
Remission	0.06103	0.03146	3.47283	53.42822
Treatment weeks 5-8 (state 4) to				
Low impact (state 1)	0.03414	0.01736	3.69970	104.67097
High impact (state 2)	0.00760	0.00381	3.93198	513.71610
Treatment weeks 9-12	0.92081	0.04119	38.65633	3.32444
Remission (state 6)	0.04123	0.02104	3.63908	84.62987

Transition probability	Expected value[b]	SD[c]	α[d]	β[d]
Treatment weeks 9-12 (state 5) to				
Low impact (state 1)	0.35079	0.16667	2.52522	4.67334
High impact (state 2)	0.10684	0.05633	3.10581	25.96486
Remission (state 6)	0.57600	0.16667	4.48823	3.30384
Healthy (state 7)	0.09091	0.04756	3.23069	32.30693

[a]PERT: Program Evaluation and Review Technique.
[b]First moment: "Most likely" (expected) value taken from [11].
[c]SD for calculation of the second moment taken from [11].
[d]Shape parameters α and β for beta distribution were calculated using the method of moments.
[e]DTC: digital therapeutic care.
[f]TAU: treatment as usual.

Table 3. Cost components.

Cost components (health state)	Mean[a] (SD[b])	α[c]	β[c]
Low impact (state 1)	441.74 (476.74)	0.8584	514.5364
High impact (state 2)	588.96 (476.74)	1.5261	385.9023
Treatment weeks 1-4 (state 3)			
GP[d] consultation	20.47 (43.93)	0.2171	94.2767
Medication	16.81 (35.36)	0.226	74.3801
Diagnostic procedure	29.24 (53.72)	0.2962	98.6948
Indirect cost	147.74 (476.74)	0.0953	1543.6092
App price (only DTC[e])	239.96 (N/A[f])	N/A	N/A
4 × physiotherapy (only TAU[g])	102.88 (44.4266)	5.363	19.184
Treatment weeks 5-8 (state 4)			
Medication	16.81 (35.36)	0.226	74.3801
2 × physiotherapy (only TAU)	46.44 (22.2133)	4.3711	10.6249
Treatment weeks 9-12 (state 5)			
Medication	16.81 (35.36)	0.226	74.3801

[a]Mean values taken from [11].
[b]SD calculated from 95% CIs reported in [23].
[c]Parameters α and β for Gamma distribution calculated from mean and SD values.
[d]GP: general practitioner.
[e]DTC: digital therapeutic care.
[f]N/A: not applicable.
[g]TAU: treatment as usual.

Table 4. Health-related QoL[a] utility scores after PERT[b] transformation.

Health-related QoL (QALY[c] weight)	Expected value[d]	SD	α[e]	β[e]
Health states				
Low impact (state 1)	0.655	0.0743[f]	26.1445	13.7708
High impact (state 2)	0.61	0.1248[g]	8.7032	5.5643
Remission (state 6)	0.806	0.0713[g]	23.9639	5.7679
Healthy (state 7)	0.806	0.0713[g]	23.9639	5.7679
DTC[h]				
Treatment weeks 1-4 (state 3)	0.655	0.0766[f]	24.5159	12.9129
Treatment weeks 5-8 (4)	0.699	0.0695[f]	29.6712	12.7768
Treatment weeks 9-12 (5)	0.748	0.0699[f]	28.1058	9.4687
TAU[i]				
Treatment weeks 1-4 (3)	0.655	0.0691[f]	30.2894	15.954
Treatment weeks 5-8 (4)	0.717	0.0834[f]	20.1705	7.9613
Treatment weeks 9-12 (5)	0.729	0.0862[f]	18.6139	6.9196

[a]QoL: quality of life.
[b]PERT: Program Evaluation and Review Technique.
[c]QALY: quality-adjusted life year.
[d]First moment: “most likely” (expected) value taken from [11].
[e]Shape parameters α and β for beta distribution were calculated using the method of moments.
[f]SD calculated from [13].
[g]SD calculated from [15].
[h]DTC: digital therapeutic care.
[i]TAU: treatment as usual.

Results

The 10,000 iterations of the Monte Carlo simulation yielded average costs of €2263.96 with an average of 0.6941 QALYs per person and year for DTC and an average cost of €2127.99 with an average of 0.6902 QALYs per person and year for TAU. Thus, the mean incremental cost is €135.97, and the mean incremental QALYs are 0.004 per person and year for the DTC app. The corresponding incremental cost-utility ratio (ICUR) amounts to an additional €34,315.19 per additional QALY. Table 5 shows the summary statistics of the relevant cost and effectiveness outcomes.

Figure 2 shows the simulation results per person and year in the cost-utility plane, where each of the dots reflects 1 outcome of one of the 10,000 Monte Carlo simulations. The histograms on the axes confirm the numbers from the table, which indicate that the mean and median incremental effect, as well as the mean and median incremental cost, are positive. The diagonal line visualizes the estimated average ICUR of 34,315.19.

Table 5. Summary statistics of the relevant cost and effectiveness outcomes[a].

Parameter	Mean (SD)	Median	Min[b]	Max[c]
DTC[d] cost (€)	2263.96 (1467.69)	1853.92	413.53	22108.91
DTC cost (€) (hypothetical if app price is € 0)	1830.94 (1456.95)	1420.13	99.77	21544.24
DTC QALYs[e]	0.6941 (0.0321)	0.6944	0.5608	0.8223
TAU[f] cost (€)	2127.99 (1459.20)	1736.76	251.55	21033.72
TAU QALYs	0.6902 (0.0309)	0.6909	0.5711	0.7997
Incremental cost (€)	135.97 (484.54)	149.88	−4748.22	4551.22
Incremental cost (€) (hypothetical, if app price is 0)	1830.9 (1456.9)	1420.1	99.7704	21544.2
Incremental QALYs	0.0040 (0.0296)	0.0038	−0.0950	0.1484

[a]Table shows summary statistics of the simulation results for the regular app price of €239 (a currency exchange rate of EUR €1=US $1.069 is applicable) and the hypothetical scenario with an app price of €0 per person and year.

[b]Min: minimum.

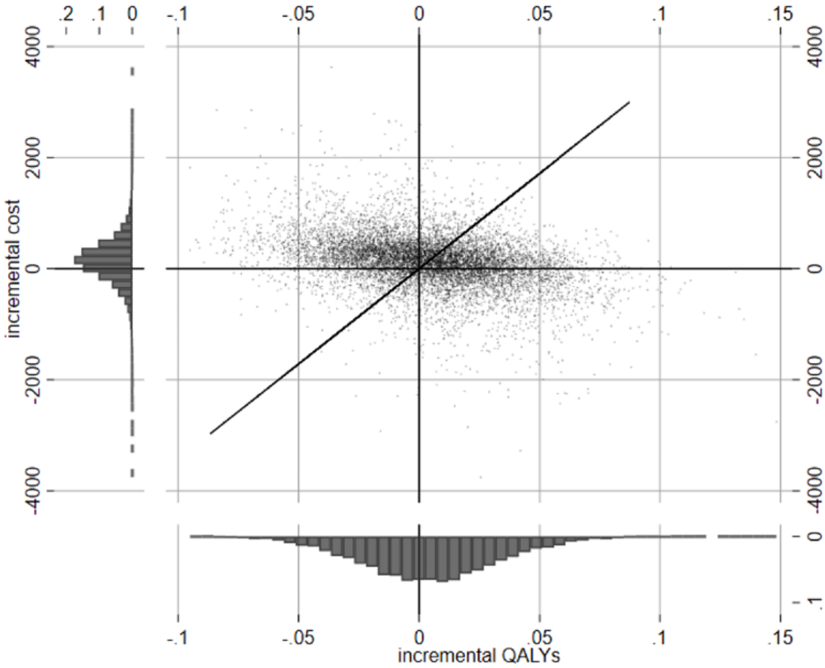
[c]Max: maximum.

[d]DTC: digital therapeutic care.

[e]QALY: quality-adjusted life year.

[f]TAU: treatment as usual.

Figure 2. Monte Carlo simulation results per person and year in the cost-utility plane. Each dot represents incremental quality-adjusted life years (QALYs) and incremental costs for one simulated outcome in the cost-utility-plane. Histograms on axes visualize the marginal distributions of incremental costs and incremental QALYs.



DTC was costlier than TAU in 66.53% of the iterations but also yielded more QALYs in 54.96% of the iterations. DTC dominated TAU in 24.04% of the iterations, whereas TAU dominated DTC in 35.61% of the iterations. Table 6 gives an overview of the number of iterations, which indicate the different findings.

The CEAC in Figure 3 illustrates the probability of cost-effectiveness for given WTP thresholds. The solid black line depicts the probability of the DTC strategy being cost-effective given a certain WTP when taking all potential health outcomes into account. The dashed line indicates the probability of DTC being cost-effective at a given WTP under the additional condition that DTC is only acceptable if it produces better health outcomes than TAU. The solid gray line at 54.96% indicates the highest probability of cost-effectiveness at an infinite WTP. Since only 54.96% of the iterations yielded a positive incremental effect and negative incremental effects are unacceptable at an infinite WTP even without the additional condition, both CEACs approximate this threshold.

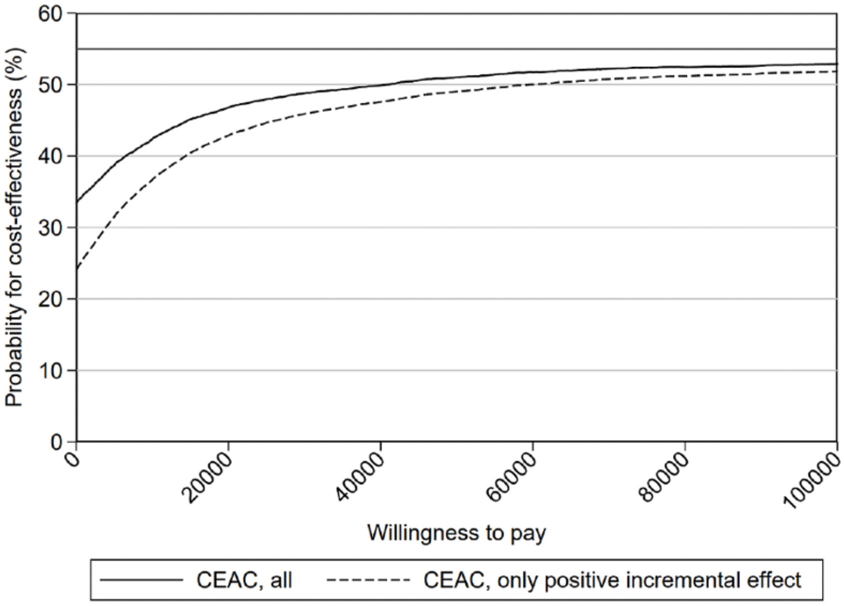
Table 6. Overview of the numbers of iterations, which indicate the different outcomes.

Parameter	Current app price (€239), %	Hypothetical app price (€0), %
Positive incremental treatment outcome	54.96	54.96
Negative incremental treatment outcome	45.04	45.04
Positive incremental cost	66.53	17.64
Negative incremental cost	33.47	82.36
DTC[a] dominant	24.04	48.85
TAU[b] dominant	35.61	11.53

[a]DTC: digital therapeutic care.

[b]TAU: treatment as usual.

Figure 3. Cost-effectiveness acceptability curve (CEAC) for quality-adjusted life years (QALYs).



When including iterations with negative incremental effects, the minimum probability of DTC being effective was 33.47%, corresponding to the fraction of iterations with negative incremental costs. The CEAC reached 50% at a WTP of approximately €41,000, flattened at a WTP of around €80,000, and approximated the maximum possible probability of cost-effectiveness of 54.96% when the WTP tended to infinity. When excluding outcomes with negative incremental effects, DTC was only considered to be cost-effective with a probability of 24.04% for a WTP of €0, corresponding to the fraction of iterations in which DTC strictly dominated TAU. The restricted CEAC reached a probability of cost-effectiveness of 50% only at a WTP of approximately €60,000. Like the unrestricted CEAC, the restricted CEAC flattened around a WTP of €80,000 and approximated the maximum possible probability of cost-effectiveness of 54.96% when WTP tended to infinity.

We reran the Monte Carlo simulation using the same aforementioned figures but with the app cost set to €0 to assess the cost-effectiveness of DTC if the app was available free of charge. Decreasing the app price to €0 yielded a decrease in the incremental cost to €−297.04 and thus a decrease in the ICUR to €−74,964.87. Note that using the same random seed in both simulations assured that the effects and simulated courses of treatment and compliance remained unchanged. Comparing the ICUR with app prices of €239 and €0 allowed us to determine the association between app price and ICUR, which amounts to an increase in the ICUR of €455.41 for each additional Euro charged for a 3-month period. Although the ICUR would be negative for an app price below €164,61, the estimated probability of DTC being more effective than TAU was only 54.96%.

Discussion

Principal Findings

This paper presents a PSA to evaluate the potential benefits of an app-based DTC program for patients with LBP in comparison to the TAU in Germany. We found the resulting ICUR to be substantially higher compared to the ICUR in the deterministic base case analysis, indicating that DTC apps are not clearly cost-effective at the current app price of €239 compared to TAU in Germany. The PSA yielded incremental costs of €135.97 and 0.004 incremental QALYs per patient and year for the DTC app. The resulting ICUR was €34,315.19 per QALY gained, as compared to €5,486 reported in [11]. The highest probability of cost-effectiveness for DTC in the PSA was 54.96% at an infinite WTP. Reducing the app price in the simulation from €239.96 to €164.61 for a 3-month prescription could yield a negative ICUR and thus make DTC the dominant strategy, even though the estimated probability of DTC being more effective than TAU is only 54.96%.

The large difference between the ICUR of 34,315.19 found in the PSA and the ICUR of 5,486 reported in [11] can be attributed to the differences in the incremental effects: DTC yielded 0.6941 QALYs per patient and year in the PSA, whereas Lewkowicz et al [11] found 0.697 QALYs per year for DTC. The PSA yielded 0.6902 QALYs per year for TAU, which is similar to the 0.689 QALYs per year reported for TAU in [9].

Overall, the stark difference between the outcome from the PSA and from [11] may be explained by the infinitesimally small incremental effect, indicating that DTC and TAU were similarly effective both in the PSA and in [11]. Since the incremental QALYs appear in the denominator and are close to 0, even small differences may produce drastically different ICURs. With this outcome, a high measurement precision would be required to allow reliable inference from the results, but the available QoL estimate is a single-study outcome derived from 42 participants of the Kaia Health App trial [13], which found no significant difference between DTC and TAU. By including additional states for temporary (state 6) and lasting (state 7) health improvements and simulating a 3-year period, our PSA goes beyond the information available in [13] but still produces similar findings in terms of QoL.

Although a recent review found 12 studies on 6 different DTC apps with implemented decision support interventions, the control groups in those studies received no specific treatment [5]. To the best of our knowledge, the only existing relevant study comparing a DTC app with physiotherapy for our evaluation is the Kaia Health App trial [13], which offered only imprecise estimates for the treatment effect. The limited precision of the available QoL input parameters is reflected in the rather flat histogram of incremental QALYs in Figure 2, which clearly calls for further studies to explore the effects of DTC and decision support interventions on compliance and QoL outcomes for patients with LBP. Particularly, considering that the underlying randomized controlled trial (RCT) [13] only involved a small patient cohort in the app-based intervention group, studies with greater patient cohorts are needed to achieve more precise estimates and to outweigh potential outliers.

The incremental costs of €135.97 found in the PSA are fairly similar to the €121.59 reported in [11]. The primary cost driver in the DTC strategy is the fixed app prescription cost, which occurs every time a patient starts a new treatment program, entering state 3 in the model. These high initial fees may backfire for such highly scalable and easily available app programs, especially if patients' compliance is unobservable, and there is a high risk for early discontinuation of the DTC. In our simulation, we allowed that the DTC could be prescribed multiple times for 1 patient, which we considered realistic. The higher attrition rate in DTC than in TAU reinforces this major cost driver since the cost of DTC in health state 3 is €239.96 and thus substantially higher than the cost of 4 physiotherapy sessions of €102.88 in the first month. However, it is unclear how often a physician will prescribe the DTC app for the same patient in real life if that patient repeatedly aborts treatment.

Our scenario analysis focused on the effects of the app cost and investigated how the reimbursement price could be updated to render app-based treatment as a cost-effective alternative. The results suggest that an adjusted app reimbursement price less than €164.61, which would be slightly higher than the presumed costs for physiotherapy in the TAU, could lead to negative incremental costs, thus yielding a negative ICUR for the DTC app. Therefore, according to our model, a reimbursement price below €54.87 per month could make DTC somewhat less costly than face-to-face physiotherapy, while the health outcomes cannot be considered to differ significantly between TAU and DTC.

Different DTC programs with different app components and divergently progressed decision support interventions are associated with different overall cost-utility outcomes. While the core components and the core method of health care delivery are similar among these apps, further implementations such as virtual reality guidance during exercises or personalized feedback interventions through push notifications may improve the efficacy of DTC programs and generate increased effects on the QoL of LBP patients. Extended capabilities of decision support interventions may have a significantly positive impact on the long-term outcome [5,9].

To the best of our knowledge, along with [11], this is the first cost-effectiveness analysis for a DTC app based on a RCT for patients with LBP. While we found no clear evidence for a positive incremental effect on health-related QoL but a noticeable increase in cost for the DTC app for LBP, recent studies found DTC apps to be a cost-effective and promising approach for the treatment of unipolar depression [25] and essential hypertension [24].

Limitations

The shortage of data may involve potential biases in the parameters of the distributions. We applied the PERT approach to derive probability density functions for the transition probabilities and considered the base-case values from [11] as "most likely" values. However, even though most of the probabilities represent reasonable scenarios in the treatment of LBP, not all parameter values could be derived from clinical findings.

For the gamma distribution, the input values for the standard deviation parameter were derived from a German cost-of-illness study and adopted for the cost components in the PSA. Since we found no information in the literature on potential correlations between different cost components, we sampled each cost component independently in the PSA. The cost outcome may thus be biased either upward or downward, depending on whether higher costs in 1 component increase (eg, if more physician visits trigger more prescriptions) or decrease (eg, if seeing the physician more often avoids costs in other components) the costs in other components. However, since indirect costs make up the largest part of total cost and all cost parameters except for the app reimbursement price and cost of face-to-face physiotherapy are equally included in both strategies, we argue that the missing correlations may have only a relatively small impact on our overall findings.

Our model focused on the direct comparison between the cost of unsupervised DTC and personal physiotherapy, and we excluded inpatient and rehabilitation care, as well as minor ambulatory treatment modalities. Overall, only 81% of total LBP-related health care expenditures were considered in our simulation [23]. It remains unclear what effect an increased use of DTC would have on the utilization of, for instance, injection therapy or surgery. However, we argue that the exclusion of such treatment options does not influence the incremental cost outcome, especially since injection therapy and surgery are usually applied in acute and highly severe cases.

Finally, measuring QoL through 2 different metrics (ie, the SF-6D and VR-6D) is another potential limitation. We acknowledge that using different outcome metrics for 1 simulation is not recommended but argue that SF-6D and VR-6D tend to be highly correlated and yield comparable outcomes, so they may be used interchangeably [14]. Since for both strategies each metric was used similarly for a respective health state, we argue that this methodological choice does not have an impact on the overall results. In addition, probing the results by rerunning the simulation as a cost-effectiveness analysis with pain reduction on a numerical rating scale yielded a similar distribution of the incremental treatment effect (results are available from the authors on request).

Conclusion

Allowing for parameter uncertainty yielded a significantly higher ICUR than the previously published deterministic approach. The CEACs indicate that the DTC approach is not very likely to be cost-effective, as the probability of cost-effectiveness remains below 55% even for an infinite WTP. One reason for the inconclusive result for QoL may be the high uncertainty, especially in health outcomes. At present, decision-makers should be cautious when considering the reimbursement of DTC apps, since no significant incremental effect on health was found. However, future developments of DTC apps may involve further decision support interventions, which may improve compliance, decrease attrition, and eventually yield better health outcomes. Future evaluations of DTC programs should strive to improve the precision of QoL outcome data and preferably aim to evaluate DTC apps with decision support interventions in a real-life environment.

Acknowledgments

This work has received funding from the European Union’s Horizon 2020 research and innovation program Smart4Health (grant 826117). It was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 491466077

Data Availability

All data generated or analyzed during this study are included in this published paper and its Multimedia Appendix files.

Authors' Contributions

DL and MS conceptualized the study. DL and MS were in charge of the Monte Carlo simulation and analyses, interpretation of the results, and writing of the manuscript. DL, MS, and EB contributed to refining all sections and critically editing the manuscript. All authors approved the submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A. Applying Program Evaluation and Review Technique (PERT) and Method of Moments (MoM) methods. B. Summary of probabilistic sensitivity analysis (PSA) input parameters and probability density functions.

[PDF File (Adobe PDF File), 905 KB - mhealth_v11i1e44585_app1.pdf]

Multimedia Appendix 2

Histogram: cost parameter.

[PDF File (Adobe PDF File), 322 KB - mhealth_v11i1e44585_app2.pdf]

Multimedia Appendix 3

Histogram: quality of life (QoL) parameter.

[PDF File (Adobe PDF File), 391 KB - mhealth_v11i1e44585_app3.pdf]

Multimedia Appendix 4

Histogram: digital therapeutic care (DTC) transition matrix.

[PDF File (Adobe PDF File), 406 KB - mhealth_v11i1e44585_app4.pdf]

Multimedia Appendix 5
Histogram: treatment-as-usual (TAU) transition matrix.
[PDF File (Adobe PDF File), 425 KB - mhealth_v11i1e44585_app5.pdf]

References

1. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jun 24;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428] [Medline: 24665116]
2. Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. Lancet 2012 Feb;379(9814):482-491. [doi: 10.1016/s0140-6736(11)60610-7]
3. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot J, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018 Nov;27(11):2791-2803. [doi: 10.1007/s00586-018-5673-2] [Medline: 29971708]
4. Chenot J, Greitemann B, Kladny B, Petzke F, Pfingsten M, Schorr SG. Non-specific low back pain. Dtsch Arztebl Int 2017 Dec 25;114(51-52):883-890 [FREE Full text] [doi: 10.3238/arztebl.2017.0883] [Medline: 29321099]
5. Lewkowicz D, Slosarek T, Wernicke S, Winne A, Wohlbrandt AM, Bottinger E. Digital therapeutic care and decision support interventions for people with low back pain: systematic review. JMIR Rehabil Assist Technol 2021 Nov 19;8(4):e26612 [FREE Full text] [doi: 10.2196/26612] [Medline: 34807837]
6. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
7. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
8. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
9. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020 Sep 09;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
10. The fast-track process for digital health applications according to section 139e SGB V: a guide for manufacturers, service providers and users. Federal Institute for Drugs and Medical Devices (BfArM). 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf?__blob=publicationFile [accessed 2022-06-02]
11. Lewkowicz D, Wohlbrandt AM, Bottinger E. Digital therapeutic care apps with decision-support interventions for people with low back pain in Germany: cost-effectiveness analysis. JMIR Mhealth Uhealth 2022 Feb 07;10(2):e35042 [FREE Full text] [doi: 10.2196/35042] [Medline: 35129454]
12. Thom H. Deterministic and probabilistic analysis of a simple Markov model: how different could they be? Appl Health Econ Health Policy 2022 May 20;20(3):447-449 [FREE Full text] [doi: 10.1007/s40258-021-00700-1] [Medline: 35048285]
13. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
14. Selim A, Rogers W, Qian S, Brazier J, Kazis LE. A preference-based measure of health: the VR-6D derived from the veterans RAND 12-Item Health Survey. Qual Life Res 2011 Oct;20(8):1337-1347. [doi: 10.1007/s11136-011-9866-y] [Medline: 21336657]
15. Cassidy JD, Côté P, Carroll LJ, Kristman V. Incidence and course of low back pain episodes in the general population. Spine (Phila Pa 1976) 2005 Dec 15;30(24):2817-2823. [doi: 10.1097/01.brs.0000190448.69091.53] [Medline: 16371911]
16. Herman P, Lavelle T, Sorbero M, Hurwitz E, Coulter I. Are nonpharmacologic interventions for chronic low back pain more cost effective than usual care? Proof of concept results from a Markov model. Spine (Phila Pa 1976) 2019;44(20):1456-1464. [doi: 10.1097/brs.0000000000003097]
17. DiGA-Verzeichnis. Federal Institute for Drugs and Medical Devices. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2022-02-14]
18. Bundeseinheitliches positionsnummernverzeichnis für heilmittelleistungen. GKV-Spitzenverband. URL: https://www.gkv-datenaustausch.de/media/dokumente/leistungserbringer_1/sonstige_leistungserbringer/ [accessed 2021-06-02]
19. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford, UK: Oxford University Press; 2006.
20. Farnum NR, Stanton LW. Some results concerning the estimation of beta distribution parameters in PERT. J Oper Res Soc 2017 Dec 20;38(3):287-290. [doi: 10.1057/jors.1987.45]

21. Malcolm DG, Roseboom JH, Clark CE, Fazar W. Application of a technique for research and development program evaluation. Operations Research 1959 Oct;7(5):646-669. [doi: 10.1287/opre.7.5.646]
22. O'Brien BJ, Briggs AH. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. Stat Methods Med Res 2002 Dec 02;11(6):455-468. [doi: 10.1191/0962280202sm304ra] [Medline: 12516984]
23. Becker A, Held H, Redaelli M, Strauch K, Chenot JF, Leonhardt C, et al. Low back pain in primary care: costs of care and prediction of future health care utilization. Spine (Phila Pa 1976) 2010 Aug 15;35(18):1714-1720. [doi: 10.1097/brs.0b013e3181cd656f] [Medline: 21374895]
24. Nomura A, Tanigawa T, Kario K, Igarashi A. Cost-effectiveness of digital therapeutics for essential hypertension. Hypertens Res 2022 Oct 20;45(10):1538-1548 [FREE Full text] [doi: 10.1038/s41440-022-00952-x] [Medline: 35726085]
25. Baumann M, Stargardt T, Frey S. Cost-utility of internet-based cognitive behavioral therapy in unipolar depression: a Markov model simulation. Appl Health Econ Health Policy 2020 Aug 15;18(4):567-578 [FREE Full text] [doi: 10.1007/s40258-019-00551-x] [Medline: 32060822]

Abbreviations

CEAC: cost-effectiveness acceptability curve
DiGA: Digital Health Applications
DTC: digital therapeutic care
ICD-10: International Classification of Diseases 10th revision
ICUR: incremental cost-utility ratio
LBP: low back pain
PERT: Program Evaluation and Review Technique
PSA: probabilistic sensitivity analysis
QALY: quality-adjusted life year
QoL: quality of life
RCT: randomized controlled trial
SF-6D: Short-Form 6-Dimension
TAU: treatment as usual
VR-12D: Veterans RAND 12-Item Health Survey
VR-6D: Veterans RAND 6-Dimension
WTP: willingness to pay

Edited by L Buis; submitted 25.11.22; peer-reviewed by A Kotlo, M Ghazisaeedi, M Ali, M Zamandi; comments to author 14.04.23; revised version received 26.04.23; accepted 23.05.23; published 29.06.23.

Please cite as:
Lewkowicz D, Bottinger E, Siegel M
Economic Evaluation of Digital Therapeutic Care Apps for Unsupervised Treatment of Low Back Pain: Monte Carlo Simulation
JMIR Mhealth Uhealth 2023;11:e44585
URL: https://mhealth.jmir.org/2023/1/e44585
doi:10.2196/44585
PMID:37384379

Original Paper

The Impact of a Digital Weight Loss Intervention on Health Care Resource Utilization and Costs Compared Between Users and Nonusers With Overweight and Obesity: Retrospective Analysis Study

Ellen Siobhan Mitchell[1], PhD; Alexander Fabry[1], MPhil; Annabell Suh Ho[1], PhD; Christine N May[1], PhD; Matthew Baldwin[1], PhD; Paige Blanco[1], MPH; Kyle Smith[1], BSBA; Andreas Michaelides[1], PhD; Mostafa Shokoohi[2], PhD; Michael West[3], MSc; Kim Gotera[2], MSc; Omnya El Massad[2], MSc; Anna Zhou[2], MSc

[1]Academic Research, Noom, Inc, New York City, NY, United States
[2]Eversana, Burlington, ON, Canada
[3]Eversana, Sydney, NS, Canada

Corresponding Author:
Christine N May, PhD
Academic Research
Noom, Inc
450 W 33rd St
New York City, NY, 10001
United States
Phone: 1 8882665071
Email: christinem@noom.com

Abstract

Background: The Noom Weight program is a smartphone-based weight management program that uses cognitive behavioral therapy techniques to motivate users to achieve weight loss through a comprehensive lifestyle intervention.

Objective: This retrospective database analysis aimed to evaluate the impact of Noom Weight use on health care resource utilization (HRU) and health care costs among individuals with overweight and obesity.

Methods: Electronic health record data, insurance claims data, and Noom Weight program data were used to conduct the analysis. The study included 43,047 Noom Weight users and 14,555 non–Noom Weight users aged between 18 and 80 years with a BMI of ≥25 kg/m² and residing in the United States. The index date was defined as the first day of a 3-month treatment window during which Noom Weight was used at least once per week on average. Inverse probability treatment weighting was used to balance sociodemographic covariates between the 2 cohorts. HRU and costs for inpatient visits, outpatient visits, telehealth visits, surgeries, and prescriptions were analyzed.

Results: Within 12 months after the index date, Noom Weight users had less inpatient costs (mean difference [MD] −US $20.10, 95% CI −US $30.08 to −US $10.12), less outpatient costs (MD −US $124.33, 95% CI −US $159.76 to −US $88.89), less overall prescription costs (MD −US $313.82, 95% CI −US $565.42 to −US $62.21), and less overall health care costs (MD −US $450.39, 95% CI −US $706.28 to −US $194.50) per user than non–Noom Weight users. In terms of HRU, Noom Weight users had fewer inpatient visits (MD −0.03, 95% CI −0.04 to −0.03), fewer outpatient visits (MD −0.78, 95% CI −0.93 to −0.62), fewer surgeries (MD −0.01, 95% CI −0.01 to 0.00), and fewer prescriptions (MD −1.39, 95% CI −1.76 to −1.03) per user than non–Noom Weight users. Among a subset of individuals with 24-month follow-up data, Noom Weight users incurred lower overall prescription costs (MD −US $1139.52, 95% CI −US $1972.21 to −US $306.83) and lower overall health care costs (MD −US $1219.06, 95% CI −US $2061.56 to −US $376.55) per user than non–Noom Weight users. The key differences were associated with reduced prescription use.

Conclusions: Noom Weight use is associated with lower HRU and costs than non–Noom Weight use, with potential cost savings of up to US $1219.06 per user at 24 months after the index date. These findings suggest that Noom Weight could be a cost-effective weight management program for individuals with overweight and obesity. This study provides valuable evidence for health care providers and payers in evaluating the potential benefits of digital weight loss interventions such as Noom Weight.

(JMIR Mhealth Uhealth 2023;11:e47473) doi:10.2196/47473

KEYWORDS

mobile health; mHealth; obesity; overweight; Noom Weight; digital weight loss intervention; health care resource utilization; costs; electronic health record; EHR; insurance claims; inverse probability of treatment weighting; IPTW; mobile phone

Introduction

Background

Rising rates of obesity globally [1] have led to substantial increases in related health care expenditures. From 2000 to 2018, the age-adjusted rate of obesity in the United States increased from 30.5% to 42.4%, with 76.5% of the adult population classified as either overweight or obese in 2018 [2]. Between 1998 and 2016, obesity-related health care spending in the United States increased from US $111.7 to US $170.3 billion [3] to US $182 to US $288 billion (2021 US dollars) [4,5]. Obesity is associated with an increase in direct annual medical costs ranging from US $1961 [5] to US $3423 [4] per individual (2021 US dollars). Furthermore, obesity-related comorbidities are among the greatest contributors to total US annual medical expenditures, including US $126 billion for diabetes, US $101.2 billion for ischemic heart disease, US $89.5 billion for hypertension, and US $29.9 billion for hyperlipidemia (2021 US dollars) [6].

There is a pressing need for effective strategies to address these rising costs and the growing prevalence of obesity [7]. Standard dietary interventions that maintain an energy deficit typically produce an average maximal weight loss of 4 to 12 kg after 6 months, with smaller sustained losses of 4 to 10 kg after 1 year and only 3 to 4 kg after 2 years [8]. Lifestyle changes are typically required to sustain weight losses, and, as such, a lifestyle intervention is an effective approach [9-11].

In-person interventions, although effective, can be time-consuming, expensive, and thus unappealing to many potential participants [12,13]. In addition to these barriers, limited program availability and potential lack of reimbursement [14,15] further limit widespread participation. Remote interventions using mobile health (mHealth) technologies such as telephone calls, SMS text messages, and smartphone apps have been effective in the treatment of obesity [16-19] and can address many of the limitations associated with in-person treatment [20]. By maintaining regular interaction with health care providers and directly receiving educational content, support, and motivation in a widely accessible, convenient, and affordable format, patient engagement and adherence are improved [17,21]. Although the evidence for clinical effectiveness continues to grow, the literature lacks sufficient data on health care cost savings from mHealth programs [22].

Noom (Noom Inc) is an mHealth program that delivers a comprehensive lifestyle intervention through educational articles, coaching, support groups, diet and exercise tracking, and techniques based on the principles of cognitive behavioral therapy. Noom has 2 health programs: Noom Weight for weight management and Noom Mood for stress management. Previous work has shown that 56% of Noom Weight starters achieve weight loss of ≥5% of initial body weight after 6 months [23], a threshold shown to produce clinically meaningful improvements in health by improving lipid profiles and reducing the risks of developing diabetes and hypertension [8,24]. A retrospective analysis of >11,000 users who opened the program at least once after week 8 showed that the majority achieved ≥5% weight loss at 32 weeks (79%) and 52 weeks (82%), with the proportion of users losing ≥10% body weight increasing from 30% to 40% over the same period [25]. The degree of weight loss achieved has also been demonstrated to be strongly associated with user engagement levels [16,25-29]. Although the clinical benefit of Noom Weight has been established, its economic impact, such as that on health care resource utilization (HRU) and associated costs, has not been thoroughly evaluated or reported in the literature to date.

Objectives

We conducted a retrospective study using real-world data from the Noom Weight user database, electronic health records (EHRs), and a commercial insurance claims database to assess the impact of Noom Weight on HRU and health care costs for Noom Weight users in the United States. Using propensity score analyses, HRU and costs among Noom Weight users with overweight and obesity were compared with those of demographically similar non–Noom Weight users at 12 and 24 months of follow-up. We hypothesized that Noom Weight users would demonstrate lower HRU and health care costs than individuals who did not use Noom Weight. A secondary aim was to explore the potential correlation of observed impacts on HRU and costs with changes in obesity-related clinical outcomes.

Methods

Study Design

This retrospective longitudinal cohort study used a data set based on new Noom Weight registrants between July 31, 2018, and July 31, 2020, including self-reported demographic data recorded at the time of registration (eg, age, sex, height, and weight) and activity data (eg, body weight measurements, food intake, and physical activity) recorded longitudinally thereafter. This data set was linked with a cohort of patients across Eversana EHR and open insurance claims data, which included anonymized patient identifiers, vital signs, and health care provider visits with associated diagnoses and procedures for American patients, including those with commercial insurance, Medicare, and Medicaid coverage. Eversana's EHR data set is an aggregation and standardization of EHR data into a common data model for >120 million American patients. The data are derived from >2000 outpatient or ambulatory health centers, >500 hospitals, >30 health systems (including academic medical centers), and >50 unique electronic medical record platform providers across all 50 states in the United States. All database records are statistically deidentified and certified to be fully compliant with the patient confidentiality requirements set forth in the Health Insurance Portability and Accountability Act of 1996. A Health Insurance Portability and Accountability Act–compliant health data privacy service (Datavant) was used to create anonymized encrypted codes, which allowed for data sets to be linked together without the use or exchange of identifiable information [30]. All linked data were deidentified, and an expert determination was completed before data were received for analysis. No identifiable Noom Weight user information was used or exchanged in this study. HRU, health care costs, and obesity-related clinical outcomes were compared between Noom Weight users and demographically similar non–Noom Weight users with overweight or obesity.

Ethics Approval

Noom Weight user data were collected from the Noom Weight database with prior approval from the Advarra Institutional Review Board (Pro00017565).

Cohorts

Noom Weight Users

Noom Weight users were required to have an initial treatment window of continuous Noom Weight use lasting at least 3 months. Continuous use was defined as opening the Noom Weight program at least once per week on average. Each user's unique index date was defined as the first day of Noom Weight use in the treatment window. If users recorded multiple eligible 3-month treatment windows, the earliest eligible window was used. Users were required to be US residents aged between 18 and 80 years and have a baseline BMI of ≥25 kg/m^2. A minimum of 12 months of medical records before and after the index date, as well as a minimum documented insurance claims activity of at least 1 claim in the 12-month pre–index date period and at least 1 claim in the 12-month post–index date period, were required. In addition, users included in the 24-month post–index date analysis were required to have a second claim in the 12- to 24-month window.

Users were excluded if they had a history of medical conditions that would significantly affect body weight or the ability to fully engage in a comprehensive lifestyle intervention during the study period, including AIDS, cancer (all types), end-stage organ failure, hemiplegia, paraplegia, uncontrolled HIV infection, pregnancy, or wasting syndrome. Patients were also excluded if they had surgeries or acute-onset conditions affecting body weight, including bariatric surgery and cerebrovascular disease, at any time before the study up until before the end of the initial 3-month treatment window. Comorbidities were identified in the EHR using International Classification of Diseases, Ninth Revision, Clinical Modification and International Classification of Diseases, Tenth Revision, Clinical Modification codes.

Non–Noom Weight Users

A control cohort of non–Noom Weight users otherwise meeting the aforementioned inclusion and exclusion criteria defined for Noom Weight users were also identified using EHR and insurance claims data. The index date for non–Noom Weight users was defined as the date of the first qualifying BMI (≥25 kg/m^2) entry recorded in the EHR between July 31, 2018, and July 31, 2020.

Inverse Probability of Treatment Weighting Analysis

Baseline Covariates

Baseline covariates, including BMI, sex, age, and US census region were derived from Noom data for Noom Weight users and from EHR data for non–Noom Weight users. The type of insurance coverage was derived from insurance claims data for both cohorts. Covariates were balanced between the cohorts using inverse probability of treatment weighting (IPTW) before analyses.

HRU Determination

HRU was determined from all submitted insurance claims for any service. Claims were categorized based on the recorded place of service and type of claim, including inpatient visits, length of inpatient stay (in days), outpatient visits (including the number of clinic, office, and outpatient hospital visits), telehealth visits, other or unknown visits, surgeries, total prescriptions, and obesity-specific prescriptions. Unique visits were counted as single events regardless of the extent of services rendered during the visit, and total prescriptions included the total count of all prescribed medications. For each service type, the number of uses per patient, as well as the number of uses per patient among only those patients with ≥1 use, were determined at 12 and 24 months after the index date.

Health Care Costs

Health care costs were determined based on remitted insurance claims and included all unique entries with valid Current Procedural Terminology codes, Healthcare Common Procedure Coding System codes, or the National Drug Code. In cases where remitted amounts were not available, costs were imputed using the median remitted amount for similarly coded claims, aggregated on the claimant's insurance type, age group, sex, and state of residence. Prescription costs included only paid claims; submitted claims that were not reimbursed were excluded. Obesity-specific prescription costs included all medications approved for short-term or chronic weight management or those commonly prescribed off-label. Costs per patient were calculated at 12 and 24 months for each service type among all patients, as well as among only those patients with ≥1 use of each service type. All costs were reported in US dollars and adjusted for inflation to 2021 US dollars using the medical consumer price index inflation factors from the Federal Reserve Economic Data repository [31].

Statistical Analysis

Propensity score matching was conducted with IPTW to balance the Noom Weight and non–Noom Weight cohorts with respect to age, sex, geographic region, insurance plan, and BMI. Stabilized weights for reweighting were generated with the average treatment effect as the estimand. Summary statistics were expressed as mean and SD for continuous variables and frequency and percentage for categorical variables. Standardized mean differences (SMDs) were used to confirm covariate balance, with absolute SMDs <0.10 indicating potential balance. Mean differences (MDs) between the cohorts at 12 and 24 months were reported for HRU and costs. Generalized linear models were used to report incidence rate ratios (IRRs) for each HRU service (using a Poisson distribution with a log link) and cost ratios (CRs) for the overall costs (using a gamma distribution with a log link). All analyses were conducted using R statistical software (version 3.6.1; R Foundation for Statistical Computing).

Subgroup Analysis

Subgroup analyses were conducted by stratifying cohorts according to the diagnosis of type 2 diabetes (T2D; yes vs no), the diagnosis of hypertension (yes vs no), index date BMI (≥35 kg/m^2 vs <35 kg/m^2), Noom Weight use duration (≥6 mo vs <6 mo), and Noom Weight engagement level (high vs low). Engagement was classified as *high* if the Noom Weight program was opened ≥6 days per week on average and classified as *low* if opened <6 days per week during the initial 3-month treatment period.

Results

Patient Demographics

A total of 114,691 Noom Weight users were represented in all 3 linked data sources, of whom 78,375 (68.34%) had valid index dates. After exclusions for comorbidities and inclusion criteria for index date BMI, index date age, Noom Weight use, and insurance claims activity were applied, of the 78,375 Noom Weight users, 43,047 (54.92%) were included for the 12-month analyses and 14,141 (18.04%) for the 24-month analyses. A total of 107,519 non–Noom Weight users were identified in both EHR and insurance claims data, of whom 95,005 (88.36%) had valid index dates. All inclusion and exclusion criteria were met by non–Noom Weight users for the 12-month (14,587/95,005, 15.35%) and 24-month (6487/95,005, 6.83%) analyses.

The baseline demographics of the study population are shown in Table 1 before and after IPTW. Before IPTW, the unweighted mean ages at baseline were 51.6 (SD 12.0) years for Noom Weight users and 52.7 (SD 14.3) years for non–Noom Weight users (SMD −0.077), and 82.75% (35,622/43,047) of the Noom Weight users and 54.67% (7975/14,587) of the non–Noom Weight users were female (SMD −0.635). After IPTW (ie, the sample analyzed for the study), the mean ages were equivalent between the cohorts (Noom Weight users: 51.9, SD 12.1 years; non–Noom Weight users: 51.9, SD 13.8 years; SMD 0.001), and the proportions of female users were identical at 75.6% (proportion after weighting) for both Noom Weight users and non–Noom Weight users (SMD 0.000). All other covariates were also well balanced after IPTW, with the proportion of balanced covariates (absolute SMDs <0.10) increasing from 23% to 100%. Relevant comorbid conditions before weighting are presented in Table 1.

Table 1. Baseline demographics before and after inverse probability of treatment weighting.

	Before weighting			After weighting		
	Noom Weight users (n=43,047)	Non–Noom Weight users (n=14,587)	SMD[a]	Noom Weight users[b] (n=40,334[c])	Non–Noom Weight users[b] (n=10,549[c])	SMD
Age (years), mean (SD)	51.6 (12.0)	52.7 (14.3)	−0.077	51.9 (12.1)	51.9 (13.8)	0.001
Sex, n (%)						
Female	35,622 (82.8)	7975 (54.7)	N/A[d]	N/A (75.6)	N/A (75.6)	N/A
Male	7425 (17.2)	6612 (45.3)	−0.635	N/A (24.4)	N/A (24.4)	0.000
Region, n (%)						
South	17,902 (41.6)	6860 (47)	−0.128	N/A (43)	N/A (43.2)	−0.002
North Central	10,371 (24.1)	3801 (26.1)	−0.070	N/A (24.6)	N/A (24.8)	−0.001
West	8738 (20.3)	2905 (19.9)	−0.013	N/A (20.1)	N/A (19.9)	0.003
Northeast	6033 (14)	1020 (7)	0.213	N/A (12.2)	N/A (12.2)	0.001
Unknown	3 (0)	1 (0)	0.160	N/A (0)	N/A (0)	0.000
Insurance type, n (%)						
Commercial	32,726 (76)	8736 (59.9)	0.103	N/A (71.9)	N/A (71.9)	0.000
Medicare	8267 (19.2)	4150 (28.4)	−0.218	N/A (21.6)	N/A (21.6)	−0.001
Medicaid	1798 (4.2)	1620 (11.1)	−0.260	N/A (6)	N/A (5.9)	0.000
Others	256 (0.6)	81 (0.6)	0.125	N/A (0.6)	N/A (0.5)	0.001
BMI (kg/m^2), mean (SD)	33.1 (6.2)	32.5 (6.5)	0.100	33 (6.1)	33.1 (7)	−0.024

[a]SMD: standardized mean difference.

[b]Except for BMI and age, only percentages are reported for categorical variables after weighting to show the balance in distributions across the 2 cohorts.

[c]Effective sample sizes after weighting.

[d]N/A: not applicable.

HRU Assessment

Noom Weight users had statistically significantly lower HRU than non–Noom Weight users in the majority of places of service in both the 12-month (Table 2) and 24-month (Table 3) follow-up periods. Most notably, at 12 months after the index date, the average number of outpatient visits per person was 3.83 (SD 6.76) among Noom Weight users compared with 4.61 (SD 7.26) among non–Noom Weight users (MD −0.78, 95% CI −0.93 to −0.62; IRR 0.83, 95% CI 0.80-0.86; *P*<.001) and at 24 months after the index date was 8.16 (SD 11.77) visits among Noom Weight users compared with 8.74 (SD 12.04) visits among non–Noom Weight users (MD −0.58, 95% CI −1.00 to −0.17; IRR 0.93, 95% CI 0.89-0.98; *P*<.001). Fewer inpatient visits were recorded for Noom Weight users at 12 months (MD −0.03, 95% CI −0.04 to −0.03; IRR 0.53, 95% CI 0.47-0.60; *P*<.001) and 24 months (MD −0.04, 95% CI −0.06 to −0.02; IRR 0.68, 95% CI 0.58-0.79; *P*<.001) after the index date, and fewer surgeries were recorded at 12 months (MD −0.01, 95% CI −0.00 to −0.01; IRR 0.44, 95% CI 0.34-0.56; *P*<001) and 24 months (MD −0.01, 95% CI −0.01 to 0.00; IRR 0.67, 95% CI 0.51-0.86; *P*=.004) after the index date. Noom Weight users also had fewer prescriptions than non–Noom Weight users at 12 months (MD −1.39, 95% CI −1.76 to −1.03; IRR 0.92, 95% CI 0.90-0.94; *P*<.001) and 24 months (MD −3.13, 95% CI −4.25 to −2.00; IRR 0.92, 95% CI 0.89-0.95; *P*<.001) after the index date. The number of obesity-specific prescriptions was slightly higher among Noom Weight users than among non–Noom Weight users at 12 months after the index date (MD 0.08, 95% CI 0.01-0.16; IRR 1.07, 95% CI 1.01-1.14; *P*=.03), as was the number of telehealth visits (MD 0.02, 95% CI 0.01-0.04; IRR 1.50, 95% CI 1.15-1.97; *P*=.003), although significant differences did not persist at 24 months after the index date (*P*=.53 and *P*=.51, respectively). Additional analyses limited to patients with at least 1 encounter of each service type showed lower outpatient service use at 12 months after the index date as well as fewer prescriptions at 12 and 24 months after the index date for Noom Weight users compared with non–Noom Weight users (Table S1 in Multimedia Appendix 1).

Table 2. Health care resource utilization rates by service type at 12 months after the index date.

Service type	Noom Weight users (n=40,334[a]), mean (SD)	Non–Noom Weight users (n=10,549[a]), mean (SD)	Comparison between cohorts			
			Mean difference (95% CI)	*P* value	Incidence rate ratio (95% CI)	*P* value
Inpatient visits	0.04 (0.27)	0.07 (0.40)	−0.03 (−0.04 to −0.03)	<.001	0.53 (0.47 to 0.60)	<.001
Inpatient days	0.08 (0.76)	0.15 (1.09)	−0.07 (−0.09 to −0.05)	<.001	0.52 (0.45 to 0.61)	<.001
Telehealth visits	0.07 (0.81)	0.05 (0.56)	0.02 (0.01 to 0.04)	<.001	1.50 (1.15 to 1.97)	.003
Outpatient visits						
All	3.83 (6.76)	4.61 (7.26)	−0.78 (−0.93 to −0.62)	<.001	0.83 (0.80 to 0.86)	<.001
Clinic	0.17 (1.86)	0.18 (1.99)	−0.01 (−0.05 to 0.03)	.57	0.94 (0.77 to 1.15)	.57
Office	2.76 (5.55)	3.25 (5.75)	−0.49 (−0.61 to −0.37)	<.001	0.85 (0.82 to 0.88)	<.001
Hospital	0.90 (2.58)	1.17 (3.02)	−0.28 (−0.34 to −0.21)	<.001	0.77 (0.72 to 0.81)	<.001
Other visits[b]	0.86 (2.72)	0.95 (3.67)	−0.09 (−0.16 −0.01)	.02	0.91 (0.84 to 0.98)	.02
Surgeries	0.01 (0.09)	0.01 (0.14)	−0.01 (−0.01 to 0.00)	<.001	0.44 (0.34 to 0.56)	<.001
Prescriptions						
All	16.63 (15.73)	18.02 (17.23)	−1.39 (−1.76 to −1.03)	<.001	0.92 (0.90 to 0.94)	<.001
Obesity specific	1.21 (3.38)	1.12 (3.34)	0.08 (0.01 to 0.16)	.03	1.07 (1.01 to 1.14)	.03

[a]Effective sample size.

[b]Includes visits of unlisted or unknown types.

Table 3. Health care resource utilization rates by service type at 24 months after the index date.

Service type	Noom Weight users (n=11,438[a]), mean (SD)	Non–Noom Weight users (n=4485[a]), mean (SD)	Comparison between cohorts			
			Mean difference (95% CI)	*P* value	Incidence rate ratio (95% CI)	*P* value
Inpatient visits	0.09 (0.51)	0.13 (0.55)	−0.04 (−0.06 to −0.02)	<.001	0.68 (0.58 to 0.79)	<.001
Inpatient days	0.20 (1.52)	0.29 (1.43)	−0.08 (−0.13 to −0.04)	<.001	0.71 (0.59 to 0.85)	<.001
Telehealth visits	0.14 (1.53)	0.13 (1.42)	0.02 (−0.04 to 0.07)	.51	1.15 (0.76 to 1.74)	.52
Outpatient visits						
All	8.16 (11.77)	8.74 (12.04)	−0.58 (−1.00 to −0.17)	.006	0.93 (0.89 to 0.98)	.005
Clinic	0.34 (2.62)	0.38 (2.58)	−0.03 (−0.12 to 0.06)	.48	0.92 (0.72 to 1.17)	.48
Office	5.87 (9.57)	6.12 (9.61)	−0.25 (−0.59 to 0.09)	.14	0.96 (0.91 to 1.01)	.14
Hospital	1.94 (4.71)	2.24 (4.69)	−0.30 (−0.46 to −0.14)	<.001	0.87 (0.80 to 0.94)	<.001
Other visits[b]	1.80 (4.70)	1.71 (5.42)	0.09 (−0.08 to 0.27)	.31	1.05 (0.95 to 1.17)	.32
Surgeries	0.02 (0.14)	0.02 (0.18)	−0.01 (−0.01 to 0.00)	.004	0.67 (0.51 to 0.86)	.002
Prescriptions						
All	35.37 (30.47)	38.50 (32.62)	−3.13 (−4.25 to −2.00)	<.001	0.92 (0.89 to 0.95)	<.001
Obesity specific	2.62 (6.50)	2.55 (6.59)	0.08 (−0.16 to 0.31)	.53	1.03 (0.94 to 1.13)	.53

[a]Effective sample size.

[b]Includes visits of unlisted or unknown types.

HRU: Subgroup Analysis

When comparing Noom Weight users and non–Noom Weight users, subgroups without T2D or hypertension and with BMI <35 kg/m^2 had lower use of more service types than subgroups with T2D or hypertension and with BMI ≥35 kg/m^2, respectively (Tables S2-S4 in Multimedia Appendix 1); for example, although fewer outpatient visits were recorded among Noom Weight users than among non–Noom Weight users in both subgroups with and without T2D at 12 months after the index date, significant differences (reductions) were also observed for Noom Weight users compared with non–Noom Weight users in inpatient visits, inpatient days, surgeries, prescriptions, and obesity-specific prescriptions only in the subgroup without T2D (all *P*<.05; Table S2 in Multimedia Appendix 1). Similarly, relatively fewer significant differences between Noom Weight and non–Noom Weight users were observed for the subgroup with hypertension (all *P*<.05; Table S3 in Multimedia Appendix 1) and with BMI ≥35 kg/m^2 (all *P*<.05; Table S4 in Multimedia Appendix 1) in the respective subgroup analyses. The differences between the subgroups were more pronounced at 24 months after the index date.

More than three-quarters (33,810/44,416, 76.12%) of the Noom Weight users were categorized as *high engaged*, with the remaining users (10,606/44,416, 23.88%) categorized as *low engaged*. High-engaged Noom Weight users had significantly fewer prescriptions (overall and obesity specific) than low-engaged users at 12 months (overall MD −0.95, 95% CI −1.40 to −0.50; obesity-specific MD −0.16, 95% CI −0.26 to −0.07) and at 24 months (overall MD −2.79, 95% CI −4.41 to −1.17; obesity-specific MD −0.52, 95% CI −0.86 to −0.18) after the index date, and both engagement levels had significantly fewer inpatient visits, inpatient days, outpatient visits, and prescriptions than non–Noom Weight users at 12 months after the index date (Table S5 in Multimedia Appendix 1). These differences remained significant at 24 months after the index date for high-engaged Noom Weight users; for low-engaged Noom Weight users, only the differences in inpatient visits and outpatient visits remained significant at 24 months after the index date, and increases in obesity-specific prescriptions among low-engaged Noom Weight users compared with non–Noom Weight users were also noted at 12 and 24 months after the index date.

The mean duration of Noom Weight use was 8.67 (SD 5.70) months among all Noom Weight users, with 46.22% (20,530/44,416) using Noom Weight for <6 months and 53.78% (23,888/44,416) using Noom Weight for ≥6 months. Noom Weight users with ≥6 months of use had fewer prescriptions than users with <6 months of use at 12 months after the index date (MD −0.64, 95% CI −1.04 to −0.24), but a significant difference did not persist at 24 months (Table S6 in Multimedia Appendix 1). The pattern of significant differences for Noom Weight users of both durations was similar to that for Noom Weight engagement level at 12 months after the index date, with fewer inpatient visits, inpatient days, outpatient visits, surgeries, and prescriptions than for non–Noom Weight users. This pattern of significant differences was unchanged at 24 months after the index date for Noom Weight users with ≥6 months of use; for Noom Weight users with <6 months of use, only inpatient visits, inpatient days, a subset of outpatient visits (outpatient hospital visits only), and prescriptions were significantly lower than those for non–Noom Weight users.

Health Care Costs

Noom Weight users had lower overall health care costs at 12 months after the index date, with average expenditures of US $3433.89 (SD $10,397.96) per person compared with US $3884.28 (SD $13,661.66) per person for non–Noom Weight users (MD −450.39, 95% CI −706.28 to −194.50; CR 0.91, 95% CI 0.85-0.97; *P*<.001; Table 4). At 24 months, average overall costs for Noom Weight users were US $7367.97 (SD $19,748.80) per person compared with US $8587.03 (SD $29,190.01) per person for non–Noom Weight users (MD −1219.06, 95% CI −2061.56 to −376.55; CR 0.86, 95% CI 0.78-0.95; *P*=.005; Table 5). Expenditures for inpatient services, outpatient services, and overall prescriptions were lower for Noom Weight users than for non–Noom Weight users at 12 months, whereas telehealth expenditures were slightly higher. Of these, the reductions in outpatient expenditures, overall prescriptions, and overall costs remained statistically significant through 24 months. The additional analysis limited to patients with at least 1 encounter of each service type (Table S7 in Multimedia Appendix 1) showed significantly lower overall and obesity-specific prescription costs at both time points as well as significantly lower outpatient costs at 12 months for Noom Weight users compared with non–Noom Weight users (all *P*<.05).

Table 4. Health care costs by service type at 12 months after the index date.

Service type	Noom Weight users (US $; n=40,334[a]), mean (SD)	Non–Noom Weight users (US $; n=10,549[a]), mean (SD)	Mean difference (95% CI)	*P* value
Inpatient services	24.19 (368.01)	44.29 (497.70)	−20.10 (−30.08 to −10.12)	<.001
Telehealth services	6.08 (85.27)	3.52 (46.48)	2.56 (1.37 to 3.76)	<.001
Outpatient services				
All	492.50 (1360.32)	616.83 (1779.75)	−124.33 (−159.76 to −88.89)	<.001
Clinic	16.03 (134.04)	16.25 (151.81)	−0.22 (−2.74 to 2.29)	.86
Office	268.33 (835.97)	348.77 (1167.06)	−80.43 (−103.24 to −57.63)	<.001
Hospital	208.14 (978.29)	251.81 (1193.71)	−43.67 (−68.23 to −19.11)	<.001
Other services[b]	47.61 (1117.32)	42.32 (243.35)	5.29 (−6.95 to 17.52)	.40
Prescriptions				
All	2863.51 (10,160.60)	3177.33 (13,482.12)	−313.82 (−565.42 to −62.21)	.02
Obesity-specific	430.81 (2997.12)	466.96 (3047.48)	−36.15 (−101.33 to 29.02)	.28
Overall (all service types)[c]	3433.89 (10,397.96)	3884.28 (13,661.66)	−450.39 (−706.28 to −194.50)	<.001

[a]Effective sample size.

[b]Includes costs of unlisted or unknown types.

[c]The overall cost ratio was 0.91 (95% CI 0.85-0.97; *P*=.004), based on a gamma regression model and after cases with US $0 costs were removed.

Table 5. Health care costs by service type at 24 months after the index date.

Service type	Noom Weight users (US $; n=11,438[a]), mean (SD)	Non–Noom Weight users (US $; n=4485[a]), mean (SD)	Mean difference (95% CI)	*P* value
Inpatient services	54.65 (581.75)	62.62 (428.17)	−7.96 (−23.70 to 7.78)	.32
Telehealth services	14.04 (208.52)	9.06 (94.49)	4.97 (0.17 to 9.78)	.04
Outpatient services				
All	999.78 (2148.88)	1080.37 (2061.25)	−80.60 (−151.10 to −10.09)	.03
Clinic	35.10 (239.53)	38.80 (307.65)	−3.70 (−11.97 to 4.57)	.38
Office	572.34 (1448.81)	598.52 (1399.89)	−26.17 (−71.92 to 19.57)	.26
Hospital	392.33 (1335.99)	443.05 (1250.56)	−50.72 (−95.66 to −5.79)	.03
Other services[b]	84.43 (477.39)	80.38 (515.89)	4.05 (−8.55 to 16.65)	.53
Prescriptions				
All	6215.07 (19,439.55)	7354.59 (28,966.20)	−1139.52 (−1972.21 to −306.83)	.007
Obesity specific	918.06 (5578.53)	1149.87 (6679.35)	−231.81 (−459.45 to −4.16)	.05
Overall (all service types)[c]	7367.97 (19,748.80)	8587.03 (29,190.01)	−1219.06 (−2061.56 to −376.55)	.005

[a]Effective sample size.

[b]Includes costs of unlisted or unknown types.

[c]The overall cost ratio was 0.86 (95% CI 0.78-0.95; *P*=.004), based on a gamma regression model and after cases with US $0 costs were removed.

Health Care Costs: Subgroup Analysis

The results of the subgroup analyses for costs showed patterns similar to those for HRU. Overall, significantly lower costs were seen for Noom Weight users compared with non–Noom Weight users in more service types among cases without T2D (vs cases with T2D), without hypertension (vs cases without hypertension), and with BMI <35 kg/m^2 (vs cases with BMI ≥35 kg/m^2; all *P*<.05; Tables S8-S10 in Multimedia Appendix 1). Despite lower HRU among high-engaged versus low-engaged Noom Weight users and Noom Weight users with longer versus shorter duration of use, no significant corresponding differences in costs were observed between these groups (Tables S11 and S12 in Multimedia Appendix 1). Compared with non–Noom Weight users, high-engaged Noom Weight users had significantly lower costs at 12 months after the index date for inpatient and outpatient visits as well as lower overall costs, and significantly lower prescription costs and overall costs at 24 months after the index date. Low-engaged Noom Weight users had fewer differences in costs compared with non–Noom Weight users, with significantly lower costs for inpatient and outpatient visits as well as lower overall costs at 12 months after the index date, and no significant differences at 24 months after the index date. The pattern of significant cost differences for Noom Weight use ≥6 months and Noom Weight use <6 months compared with non–Noom Weight users was similar to that for high-engaged Noom Weight users and low-engaged Noom Weight users, respectively, at 12 and 24 months after the index date.

Discussion

Principal Findings

We showed that HRU is lower for Noom Weight users than for non–Noom Weight users at 12 and 24 months after the index date. Per user, 0.03 fewer inpatient visits, 0.83 fewer outpatient visits, 0.01 fewer surgeries, and 1.39 fewer prescriptions were recorded among Noom Weight users compared with non–Noom Weight users at 12 months after the index date. At 24 months after the index date, 0.04 fewer inpatient visits, 0.58 fewer outpatient visits, 0.01 fewer surgeries, and 3.13 fewer prescriptions were recorded among Noom Weight users compared with non–Noom Weight users. Noom Weight users had higher use of telehealth services at 12 months after the index date (MD 0.02/user), perhaps because of increased connectivity to digital health services owing to their use of Noom Weight or because of increased health responsibility as a result of the program [32]. There were also a greater number of obesity-specific prescriptions for Noom Weight users compared with non–Noom Weight users at 12 months (MD 0.08/user), which may be related to more health-conscious behavior [32] among newly registered Noom Weight users, potentially leading to higher rates of prescriptions. A statistically significant difference did not persist at 24 months.

The results also showed significantly lower health care costs for Noom Weight users compared with non–Noom Weight users at both 12 months and 24 months after the index date. Overall costs for Noom Weight users were US $450 lower per person at 12 months and US $1219 lower per person at 24 months compared with overall costs for individuals who did not use

Noom Weight. Furthermore, extending similar findings at 12 months, outpatient services costs (MD US $80/person) and prescription costs (MD US $1139/person) were lower for Noom Weight users than for non–Noom Weight users at 24 months after the index date.

Overall, our findings demonstrate significantly lower HRU and costs at 12 and 24 months for Noom Weight users compared with demographically similar non–Noom Weight users, with greater impact on HRU and costs observed for Noom Weight users without T2D, without hypertension, with BMI <35 kg/m^2, with higher Noom Weight engagement, and with longer duration of Noom Weight use.

Limitations

This was an observational study, which therefore does not permit causal associations to be drawn between Noom Weight use and HRU and cost outcomes. Another important limitation was the restricted sample size owing to the linking of 3 separate databases. Users were required to be present in all 3 data sources for inclusion, which sharply reduced the size of the available population. This also adds a risk of bias because the underlying systematic exclusions owing to missing data that may have affected patients in any 1 database would have been projected across all 3 databases, including those not previously affected by them. The requirement for Noom Weight users to use the program for 3 months may have biased this cohort toward including more health-conscious and motivated users, although it should be noted that this engagement criterion is similar to that used for previously studied Noom Weight populations and that this study's inclusion requirement to use the program at least 10 times in total during this time period is relatively low [25-27,29,33,34]. In addition, this study included only US residents, which may limit its generalizability. However, previous work has shown the comparable effectiveness of Noom Weight use for weight loss across different regions and income levels [16,28]. This may suggest similar cross-national effects on HRU and costs, which would be more affected by access to health care and existing HRU patterns in each country than by the differential impact of Noom Weight use. Furthermore, the study cohorts described here included mostly women and comprised individuals aged <80 years, further limiting the generalizability of the results.

Some potential imbalances between the cohorts may not have been accounted for in our IPTW analyses. Potential racial imbalances could not be accounted for because the Noom Weight user and non–Noom Weight user cohorts had either nonspecific or missing information for race, which prevented reweighting on this variable. Preexisting comorbidities were also not included in reweighting to permit subgroup analyses based on comorbid conditions. However, these were nevertheless reasonably well balanced in the reweighted cohorts. There may also be other confounding variables affecting HRU and costs that were not identified or accounted for in our analyses (eg, education level and income bracket). The potential impact of other common weight loss interventions used concurrently, such as weight loss programs and antiobesity medications, on the study results also requires further investigation.

In particular, an important limitation is some concurrent use of antiobesity medications, which raises the question of whether the effects were driven by the use of medications or the use of Noom Weight; this question cannot be definitively answered by this study. The data suggest that it is unlikely that this was a confound that primarily drove the results observed because there was no significant difference in antiobesity medication use between the Noom Weight user and non–Noom Weight user groups at 24 months after the index date despite significant differences in other types of HRU and costs, and a subgroup analysis of individuals with at least 1 prescription or health care visit showed no difference in obesity-specific prescriptions between the groups at all time points. However, any impact of antiobesity medications is not measured or ruled out by this study. Future research, especially with causal designs, should elucidate when and to what magnitude the concurrent use of antiobesity medications influences other types of HRU and costs. Furthermore, the study excluded individuals who had bariatric or cerebrovascular disease at any time before the study up until before the end of the initial 3-month treatment window. Future studies should examine how receiving bariatric surgery concurrently with Noom Weight use affects HRU and costs, especially on longer time scales, and whether the use of Noom Weight could be associated with motivation to undergo bariatric surgery.

Open insurance claims data were used, which allowed the assessment of direct medical costs in all care settings (eg, inpatient and outpatient) and provided large sample sizes covering patients with diverse backgrounds and medical needs. However, there are limitations associated with the use of open claims data. Open claims databases effectively capture patient activity longitudinally, but they do not necessarily capture all patient claims activity within a given time period. HRU involving service providers not included in the database will not be captured, giving a potentially incomplete picture of HRU and costs, biasing the results if certain types of HRU are less well represented and potentially excluding otherwise eligible patients for claims inactivity if they have unobserved claims. We applied a minimum claim activity criterion of 1 claim per 12-month period during the study period to mimic a continuous enrollment criterion that would be applied to a closed claims data set. Although this was a low threshold that preserved sample size, it may have introduced some bias toward patients more likely to file claims and therefore patients who were potentially sicker. As not all submitted claims in open claims databases are remitted, missing values were imputed to estimate costs. Imputation may potentially over- or underestimate true costs and systematically bias any subcategory of HRU that is particularly affected by missing data. Finally, because open claims databases are based on a large convenience sample that is not random, there may be potential biases or issues with generalizability.

Discrete surgical visits could be readily determined from insurance claims data for HRU analyses. However, individual Current Procedural Terminology codes for activities within each surgery were frequently not available and were aggregated under a master code for the entire procedure. This prevented meaningful cost assessments for surgeries, which would require the enumeration of the specific line items, and therefore surgical costs were only captured as a subset of overall costs.

Comparison With Prior Work

Noom Weight has previously been shown to be an effective treatment for obesity, frequently producing weight loss exceeding 5% of initial body weight [23,27,34,35] in as little as 8 weeks [36] and persisting for up to 52 weeks [25]. However, Noom Weight's impact and the impact of mHealth technologies generally on HRU and costs among users with overweight or obesity compared with a demographically similar control group have not been previously reported. Therefore, this study contributes to filling a substantial gap in the literature regarding the limited data on health care costs and HRU associated with digital programs. In the following paragraphs, we compare these findings to those of the few publications reporting on the impact of nonsurgical (ie, behavioral, not including bariatric surgery) weight loss on HRU and costs.

One study compared health care costs over 3 years for 4790 users of an employer-sponsored digital weight loss program with those for a propensity-matched control group (n=4790) who did not use the program [37]. Overall costs for those who used the program were US $771 lower per person over 3 years compared with those for nonusers. Specifically, the program was associated with lower outpatient (US $609/person) and inpatient (US $162/person) costs over 3 years. Our results compare favorably, showing even greater cost savings because, compared with non–Noom Weight users, Noom Weight users had lower overall costs of US $1219 per person over 2 years. In the study by Horstman et al [37], cost savings were mostly concentrated in outpatient costs, whereas we found comparatively lower outpatient cost savings over 2 years (US $78/person). This could be because of the limited availability of EHR data in our study in contrast to the health plan data used in the study by Horstman et al [37]. Future studies should collect full-scale health and insurance plan data in addition to the open insurance claims data used in this study.

An investigation by Ding et al [38] on the impact of nonsurgical weight loss on health care costs used insurance claims and EHR data for 20,488 adults with obesity in the IBM MarketScan Explorys Claims-Electronic Medical Record Data Set and found statistically significantly reduced costs for patients with >5% weight loss compared with those maintaining steady weight [38]. This aligns with our finding that Noom Weight users who were in this dedicated weight management program exhibited lower costs than the non–Noom Weight group. Furthermore, Ding et al [38] reported smaller absolute cost reductions after 2 years than in the first year alone, although the study did not directly compare the magnitude and significance of costs from 1 year to 2 years. In this study, cost reductions increased from year 1 through year 2 in absolute value. Although the 2 studies are not directly comparable, this raises the possibility that the cost impact of Noom Weight use may be longer lasting than that observed with nonsurgical weight loss interventions in general. This is consistent with the typical trend of long-term weight regain (potentially correlating with increased costs) among those with nonsurgical weight loss in the absence of intensive lifestyle interventions such as Noom Weight [8]. Future research should test this explanation.

The degree of weight loss among Noom Weight users is closely tied to the level of user engagement, with greater weight loss among patients who more frequently read articles, log data, and interact with coaches [25,27]. Similar results have also been reported with other weight loss programs [39]. In our study, higher Noom Weight engagement was also associated with lower HRU in terms of the number of prescriptions claimed (Table 3). High-engaged Noom Weight users claimed 0.95 (almost 1 unit) fewer prescriptions than low-engaged Noom Weight users through 12 months, increasing to 2.79 fewer prescriptions through 24 months. This was also true for obesity-specific prescriptions, which were fewer for high-engaged Noom Weight users than for low-engaged Noom Weight users at 12 months (MD −0.16) and 24 months (MD −0.52). Although this did not translate into statistically significantly lower costs for high-engaged Noom Weight users compared with low-engaged Noom Weight users, high Noom Weight engagement was associated with statistically significantly lower overall costs of −US $462 per user (95% CI −US $775.62 to −US $148.39) compared with non–Noom Weight use at 12 months, as well as lower overall costs of −US $1446 (95% CI −US $2469 to −US $422) and lower prescription costs of −US $1366 (95% CI −US $2377 to −US $355) compared with non–Noom Weight use at 24 months (Table S4 in Multimedia Appendix 1). In comparison, costs for these service types were not statistically significantly different for low-engaged Noom Weight users compared with non–Noom Weight users at 12 or 24 months.

In addition to subgroup analyses based on Noom Weight engagement, we also performed subgroup analyses according to BMI (<35 kg/m^2 vs ≥35 kg/m^2), T2D diagnosis, and hypertension diagnosis. Costs were significantly lower for the Noom Weight group versus the non–Noom Weight group for more HRU types in samples without T2D (vs samples with T2D), without hypertension (vs samples with hypertension), and with BMI <35 kg/m^2 (vs samples with BMI ≥35 kg/m^2). This could be because these conditions incur substantial health care costs; for example, poor glycemic control, as seen in T2D, is related to higher total health care, hospitalization, and medication costs [40]; in another study, the presence of hypertension substantially increased health care costs [41]. Therefore, cost differences between the Noom Weight user and non–Noom Weight user groups are likely starker without these conditions than with these conditions.

Previous work suggests some potential mechanisms for the results found in this study. We speculate that Noom Weight users may have shown significant cost savings compared with non–Noom Weight users because weight management efforts reduced the incidence of chronic conditions and their associated medical costs [42-44]. We also speculate that Noom Weight's educational content on healthy behaviors could result in improved medication adherence, especially because a previous study found that Noom Weight users' health responsibility (eg, taking interest in, and responsibility for, their overall physical health) improved over the course of the program [32,45].

However, because this study did not test causal pathways, future research should test and identify potential mechanisms.

Conclusions

To our knowledge, this is the first study using real-world data to show the economic impact of the use of an mHealth intervention by a cohort of users with overweight and obesity compared with a control cohort not using the mHealth intervention. We show lower HRU and costs for users of the Noom Weight mHealth program compared with non–Noom Weight users over a 2-year follow-up period. Comprehensively examining all service types, we found that inpatient visits, outpatient visits, surgical visits, and prescriptions were lower for Noom Weight users than for non–Noom Weight users for up to 24 months after initiating Noom Weight. Costs per Noom Weight user were statistically significantly lower by US $80 for outpatient services, US $1139 for prescriptions, and US $1219 overall at 24 months, which could correspond to savings of approximately US $609 per person per year during this period. These cost estimates compare favorably with those of previously studied programs. By linking Noom Weight data, EHR data, and insurance claims data, we were able to conduct several subgroup analyses for HRU and costs, including analyses based on T2D diagnosis or hypertension diagnosis, duration of Noom Weight use, user engagement level, and index date BMI. Further research is required to establish the relationship between changes in weight and BMI, as well as in comorbidities, with changes in HRU and costs, including the impact of the differential levels of weight loss. In addition, because this study focused on direct health care costs only, future research should investigate the impact of mHealth interventions on indirect costs (eg, productivity costs) as well.

Acknowledgments

This study was funded by Noom Inc; however, Noom Inc (apart from the authors employed by Noom Inc) had no role to play in the study’s design, execution, analysis, or the decision to publish the results.

Authors' Contributions

ESM, AF, and AM were responsible for the conceptualization of the study. ESM and AF were responsible for the methodology. ASH, AF, CNM, MB, PB, KS, and AM reviewed and edited the manuscript. OEM was responsible for data curation. KG was responsible for data cleaning and analysis. MS and MW were responsible for data cleaning and analysis supervision as well as writing the protocol. The Introduction and Discussion sections were written by MS and MW. MW reviewed the literature. AZ was the team director and supervisor.

Conflicts of Interest

ESM, AF, ASH, CNM, MB, PB, KS, and AM are employees of Noom Inc and have received a salary and stock options as employees. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Tables displaying results from subgroup analyses.
[DOCX File , 89 KB - mhealth_v11i1e47473_app1.docx]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
2. Fryar CD, Carroll MD, Afful J. Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960–1962 through 2017–2018. NCHS Health E-Stats. 2020. URL: https://www.cdc.gov/nchs/data/hestat/obesity-adult-17-18/obesity-adult.htm#Citation [accessed 2023-08-02]
3. Finkelstein EA, Fiebelkorn IC, Wang G. National medical spending attributable to overweight and obesity: how much, and who's paying? Health Aff (Millwood) 2003;Suppl Web Exclusives:W3-219-26. [doi: 10.1377/hlthaff.w3.219] [Medline: 14527256]
4. Cawley J, Biener A, Meyerhoefer C, Ding Y, Zvenyach T, Smolarz BG, et al. Direct medical costs of obesity in the United States and the most populous states. J Manag Care Spec Pharm 2021 Mar;27(3):354-366. [doi: 10.18553/jmcp.2021.20410] [Medline: 33470881]
5. Ward ZJ, Bleich SN, Long MW, Gortmaker SL. Association of body mass index with health care expenditures in the United States by age and sex. PLoS One 2021 Mar 24;16(3):e0247307 [FREE Full text] [doi: 10.1371/journal.pone.0247307] [Medline: 33760880]
6. Dieleman JL, Cao J, Chapin A, Chen C, Li Z, Liu A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020 Mar 03;323(9):863. [doi: 10.1001/jama.2020.0734]

7. Wang Y, Beydoun MA, Min J, Xue H, Kaminsky LA, Cheskin LJ. Has the prevalence of overweight, obesity and central obesity levelled off in the United States? Trends, patterns, disparities, and future projections for the obesity epidemic. Int J Epidemiol 2020 Jun 01;49(3):810-823 [FREE Full text] [doi: 10.1093/ije/dyz273] [Medline: 32016289]
8. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
9. Kirk SF, Penney TL, McHugh TL, Sharma AM. Effective weight management practice: a review of the lifestyle intervention evidence. Int J Obes (Lond) 2012 Feb;36(2):178-185. [doi: 10.1038/ijo.2011.80] [Medline: 21487396]
10. Wadden TA, Butryn ML, Wilson C. Lifestyle modification for the management of obesity. Gastroenterology 2007 May;132(6):2226-2238. [doi: 10.1053/j.gastro.2007.03.051] [Medline: 17498514]
11. Look AHEAD Research Group, Wing RR. Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial. Arch Intern Med 2010 Sep 27;170(17):1566-1575 [FREE Full text] [doi: 10.1001/archinternmed.2010.334] [Medline: 20876408]
12. Burgess E, Hassmén P, Pumpa KL. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. Clin Obes 2017 Jun;7(3):123-135. [doi: 10.1111/cob.12183] [Medline: 28296261]
13. Withall J, Jago R, Fox KR. Why some do but most don't. Barriers and enablers to engaging low-income groups in physical activity programmes: a mixed methods study. BMC Public Health 2011 Jun 28;11(1):507 [FREE Full text] [doi: 10.1186/1471-2458-11-507] [Medline: 21711514]
14. Lee IM, Djoussé L, Sesso HD, Wang L, Buring JE. Physical activity and weight gain prevention. JAMA 2010 Mar 24;303(12):1173-1179 [FREE Full text] [doi: 10.1001/jama.2010.312] [Medline: 20332403]
15. Tsai AG, Asch DA, Wadden TA. Insurance coverage for obesity treatment. J Am Diet Assoc 2006 Oct;106(10):1651-1655 [FREE Full text] [doi: 10.1016/j.jada.2006.07.012] [Medline: 17000198]
16. Han M, Rhee SY. Effect of adherence to smartphone app use on the long-term effectiveness of weight loss in developing and OECD countries: retrospective cohort study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e13496 [FREE Full text] [doi: 10.2196/13496] [Medline: 34255708]
17. Lewis E, Huang HC, Hassmén P, Welvaert M, Pumpa KL. Adding telephone and text support to an obesity management program improves behavioral adherence and clinical outcomes. A randomized controlled crossover trial. Int J Behav Med 2019 Dec 11;26(6):580-590. [doi: 10.1007/s12529-019-09815-1] [Medline: 31512155]
18. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky LA, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
19. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May 15;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
20. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020 Feb;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
21. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
22. Khokhar B, Jones J, Ronksley PE, Armstrong MJ, Caird J, Rabi D. Effectiveness of mobile electronic devices in weight loss among overweight and obese populations: a systematic review and meta-analysis. BMC Obes 2014 Oct 24;1:22 [FREE Full text] [doi: 10.1186/s40608-014-0022-4] [Medline: 26217509]
23. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016 Sep 08;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
24. Managing overweight and obesity in adults: systematic evidence review from the obesity expert panel, 2013. National Institutes of Health National Heart Lung and Blood Institute. 2013. URL: https://www.nhlbi.nih.gov/sites/default/files/media/docs/obesity-evidence-review.pdf [accessed 2021-06-17]
25. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR Mhealth Uhealth 2021 Nov 08;9(11):e30622 [FREE Full text] [doi: 10.2196/30622] [Medline: 34747706]
26. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo JT, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
27. DeLuca L, Toro-Ramos T, Michaelides A, Seng E, Swencionis C. Relationship between age and weight loss in Noom: quasi-experimental study. JMIR Diabetes 2020 Jun 04;5(2):e18363 [FREE Full text] [doi: 10.2196/18363] [Medline: 32497017]

28. Yang Q, Mitchell ES, Ho AS, DeLuca L, Behr H, Michaelides A. Cross-national outcomes of a digital weight loss intervention in the United States, Canada, United Kingdom and Ireland, and Australia and New Zealand: a retrospective analysis. Front Public Health 2021 Jun 10;9:604937 [FREE Full text] [doi: 10.3389/fpubh.2021.604937] [Medline: 34178911]
29. Lee J, Bae S, Park D, Kim Y, Park J. The effectiveness of a monetary reimbursement model for weight reduction via a smartphone application: a preliminary retrospective study. Sci Rep 2020 Sep 24;10(1):15714. [doi: 10.1038/s41598-020-72908-5]
30. Bernstam EV, Applegate RJ, Yu A, Chaudhari D, Liu T, Coda A, et al. Real-world matching performance of deidentified record-linking tokens. Appl Clin Inform 2022 Aug 14;13(4):865-873 [FREE Full text] [doi: 10.1055/a-1910-4154] [Medline: 35896508]
31. Consumer price index for all urban consumers: medical care in U.S. city average (CPIMEDSL). FRED. URL: https://fred.stlouisfed.org/series/CPIMEDSL [accessed 2022-08-08]
32. Behr H, Earl S, Ho AS, Lee J, Mitchell ES, McCallum M, et al. Changes in health-promoting behaviors and their association with weight loss, retention, and engagement on a digital program: prospective study. Nutrients 2022 Nov 14;14(22):4812 [FREE Full text] [doi: 10.3390/nu14224812] [Medline: 36432498]
33. Han M, Rhee SY. Effectiveness of mobile health applications for 5% body weight reduction in obese and overweight adults. J Obes Metab Syndr 2021 Dec 30;30(4):354-364 [FREE Full text] [doi: 10.7570/jomes21014] [Medline: 34853189]
34. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the diabetes prevention program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
35. May CN, Ho AS, Yang Q, McCallum M, Iyengar NM, Comander A, et al. Comparing outcomes of a digital commercial weight loss program in adult cancer survivors and matched controls with overweight or obesity: retrospective analysis. Nutrients 2021 Aug 24;13(9):2908 [FREE Full text] [doi: 10.3390/nu13092908] [Medline: 34578787]
36. Kim M, Kim Y, Go Y, Lee S, Na M, Lee Y, et al. Multidimensional cognitive behavioral therapy for obesity applied by psychologists using a digital platform: open-label randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e14817 [FREE Full text] [doi: 10.2196/14817] [Medline: 32352391]
37. Horstman CM, Ryan DH, Aronne LJ, Apovian CM, Foreyt JP, Tuttle HM, et al. Return on investment: medical savings of an employer-sponsored digital intensive lifestyle intervention, weight loss. Obesity (Silver Spring) 2021 Apr 23;29(4):654-661 [FREE Full text] [doi: 10.1002/oby.23117] [Medline: 33759385]
38. Ding Y, Fan X, Blanchette CM, Smolarz BG, Weng W, Ramasamy A. Economic value of nonsurgical weight loss in adults with obesity. J Manag Care Spec Pharm 2021 Jan;27(1):37-50. [doi: 10.18553/jmcp.2020.20036] [Medline: 33164723]
39. Painter S, Ditsch G, Ahmed R, Hanson NB, Kachin K, Berger J. Retrofit weight-loss outcomes at 6, 12, and 24 months and characteristics of 12-month high performers: a retrospective analysis. JMIR Mhealth Uhealth 2016 Aug 22;4(3):e101 [FREE Full text] [doi: 10.2196/mhealth.5873] [Medline: 27549134]
40. Mata-Cases M, Rodríguez-Sánchez B, Mauricio D, Real J, Vlacho B, Franch-Nadal J, et al. The association between poor glycemic control and health care costs in people with diabetes: a population-based study. Diabetes Care 2020 Apr;43(4):751-758. [doi: 10.2337/dc19-0573] [Medline: 32029636]
41. Francis BH, Song X, Andrews LM, Purkayastha D, Princic N, Sedgley R, et al. Progression to type 2 diabetes, healthcare utilization, and cost among pre-diabetic patients with or without comorbid hypertension. Curr Med Res Opin 2011 Apr 05;27(4):809-819. [doi: 10.1185/03007995.2011.554806] [Medline: 21306287]
42. Pasanisi F, Contaldo F, de Simone G, Mancini M. Benefits of sustained moderate weight loss in obesity. Nutr Metab Cardiovasc Dis 2001 Dec;11(6):401-406. [Medline: 12055705]
43. Dietrich M, Jialal I. The effect of weight loss on a stable biomarker of inflammation, C-reactive protein. Nutr Rev 2005 Jan;63(1):22-28. [doi: 10.1111/j.1753-4887.2005.tb00107.x] [Medline: 15730232]
44. Burini RC, Anderson E, Durstine JL, Carson JA. Inflammation, physical activity, and chronic disease: an evolutionary perspective. Sports Med Health Sci 2020 Mar;2(1):1-6 [FREE Full text] [doi: 10.1016/j.smhs.2020.03.004] [Medline: 35783338]
45. Conn VS, Ruppar TM, Enriquez M, Cooper P. Medication adherence interventions that target subjects with adherence problems: systematic review and meta-analysis. Res Social Adm Pharm 2016 Mar;12(2):218-246 [FREE Full text] [doi: 10.1016/j.sapharm.2015.06.001] [Medline: 26164400]

Abbreviations

CR: cost ratio
EHR: electronic health record
HRU: health care resource utilization
IPTW: inverse probability of treatment weighting
IRR: incidence rate ratio
MD: mean difference
mHealth: mobile health

SMD: standardized mean difference
T2D: type 2 diabetes

Edited by L Buis; submitted 22.03.23; peer-reviewed by N Mungoli, Y Ding; comments to author 27.04.23; revised version received 15.05.23; accepted 12.07.23; published 24.08.23.

Please cite as:
Mitchell ES, Fabry A, Ho AS, May CN, Baldwin M, Blanco P, Smith K, Michaelides A, Shokoohi M, West M, Gotera K, El Massad O, Zhou A
The Impact of a Digital Weight Loss Intervention on Health Care Resource Utilization and Costs Compared Between Users and Nonusers With Overweight and Obesity: Retrospective Analysis Study
JMIR Mhealth Uhealth 2023;11:e47473
URL: https://mhealth.jmir.org/2023/1/e47473
doi:10.2196/47473
PMID:37616049

Corrigenda and Addenda

Correction: WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial

Hai-Tong Sun[1,2*], BSc; Xiao-Ru Fan[1,2*], BSc; Yu-Zhou Gu[3], MSc; Yong-Heng Lu[4], BSc; Jia-Ling Qiu[1,2], MSc; Qing-Ling Yang[1,2], MSc; Jing-Hua Li[1,2], PhD; Jing Gu[1,2], PhD; Chun Hao[1,2], PhD

[1]Department of Medical Statistics, School of Public Health, Sun Yat-Sen University, Guangzhou, China

[2]Sun Yat-Sen Global Health Institute, Institute of State Governance, Sun Yat-Sen University, Guangzhou, China

[3]Guangzhou Center for Disease Control and Prevention, Guangzhou, China

[4]Lingnan Community Support Center, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Chun Hao, PhD
Department of Medical Statistics, School of Public Health, Sun Yat-Sen University
74 Zhongshan 2nd Rd Yuexiu District
Guangzhou, 510080
China
Phone: 86 87332517
Email: haochun@mail.sysu.edu.cn

Related Article:

Correction of: https://mhealth.jmir.org/2023/1/e44513

(JMIR Mhealth Uhealth 2023;11:e48961) doi:10.2196/48961

In “WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial” (JMIR mHealth and uHealth 2023;11:e48961), the following errors were corrected:

1. In the originally published article, the title appeared as follows:

> *WeChat-Based HIV e-Report, a New Manner for HIV Serostatus Request and Disclosure and Their Associated Factors Among Men Who Have Sex With Men: Prospective Subgroup Analysis of Randomized Controlled Trails*

This has been corrected to:

> *WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial*

2. In the Abstract: Background section of the originally published article, the second sentence appeared as follows:

> *However, the reliability of common methods for HIV serostatus requests and disclosure is unsatisfactory.*

This has been corrected to:

> *However, the reliability of common methods for HIV serostatus request and disclosure is inadequate.*

3. In the Abstract: Objectives section of the originally published article, the second sentence appeared as follows:

> *Additionally, the study aimed to explore its correlates with HIV serostatus requesting and disclosure receiving behavior.*

This has been corrected to:

> *Additionally, the study aimed to explore its correlation with HIV serostatus requesting and disclosure receiving behavior.*

4. In the Abstract: Methods section of the originally published article, the third sentence appeared as follows:

> *Participants completed web-based questionnaires at baseline and at the month 3 follow-up, which covered sociodemographic characteristics, HIV-related information, HIV serostatus requests, HIV serostatus disclosure receiving, and HIV e-report usage.*

This has been corrected to:

> *Participants completed web-based questionnaires at baseline and at the month 3 follow-up, which covered sociodemographic characteristics, HIV-related*

information, HIV serostatus requests, receiving HIV serostatus disclosures, and HIV e-report usage.

5. In the Abstract: Results section of the originally published article, the second sentence appeared as follows:

> *For HIV serostatus requests, 13.1% (27/205) and 10.5% (16/153) of participants started to use HIV e-reports to ask the HIV serostatus of regular and casual male sex partners, respectively. Of the regular and casual male sex partners, 27.3% (42/154) and 16.5% (18/109), respectively, chose HIV e-reports to disclose HIV serostatus. Compared to MSM who did not have HIV e-reports, those who said* […]

This has been corrected to:

> *In all, 13.1% (27/205) and 10.5% (16/153) of participants started to use HIV e-reports to request the HIV serostatus from regular and casual male sex partners, respectively. Moreover, 27.3% (42/154) and 16.5% (18/109) of the regular and casual male sex partners, respectively, chose HIV e-reports to disclose their HIV serostatus. Compared to MSM who did not have HIV e-reports, those who had HIV e-reports and stated [...]*

6. In the Abstract: Results section of the originally published article, the last sentence appeared as follows:

> *Whereas no factor was associated with HIV serostatus disclosure received from partners.*

This has been corrected to:

> *However, no factor was associated with receiving an HIV serostatus disclosure from partners.*

7. In the Abstract: Conclusions section of the originally published article, the first sentence appeared as follows:

> *The HIV e-report has been accepted by the MSM community in Guangzhou and could be applied as a new optional way for HIV serostatus request and disclosure.*

This has been corrected to:

> *The HIV e-report has been accepted by the MSM community in Guangzhou and could be applied as a new optional approach for HIV serostatus requests and disclosures.*

8. In the Introduction section of the originally published article, the second paragraph appeared as follows:

> *Studies indicated that over half of MSM used verbal communication and guessing for HIV serostatus requesting and disclosing among MSM [5, 6]. While taking HIV tests together is a reliable manner to confirm partners' HIV status, some individuals doubt the reliability of HIV self-testing, and self-test kits may not always be readily available. Deception of HIV serostatus in verbal information is prevalent [7] and difficult to confirm.*

This has been corrected to:

> *Studies indicated that over half of MSM used verbal communication and guessing for HIV serostatus request and disclosure among MSM [5,6]. While taking HIV tests together is a reliable approach to confirm partners' HIV status, some individuals doubt the reliability of HIV self-testing, and self-test kits may not always be readily available. Deception of HIV serostatus in verbal information is prevalent [7] and difficult to confirm.*

9. In the Introduction section of the originally published article, the third paragraph appeared as follows:

> *WeChat, a popular social media app with over 1.2 billion active users [8], similar to Twitter or the mix of WhatsApp and Facebook, is an ubiquitous daily use app in China [9]. WeChat miniprograms are subapps within the WeChat ecosystem. It has great potential for health intervention research [10]. In Guangzhou city, a unique and well-established WeChat miniprogram of the HIV testing service system in China is developed by Guangzhou Centers for Disease Control and Prevention (CDC) and MSM community-based organization Lingnan Partners Community Support Center (hereinafter called "Lingnan Center") [11].*

This has been corrected to:

> *WeChat, a popular social media app with over 1.2 billion active users [8], similar to Twitter or the mix of WhatsApp and Facebook, isa ubiquitous daily-use appin China [9]. WeChat miniprograms are subapps within the WeChat ecosystem. It has great potential for health intervention research [10]. In Guangzhou city, a unique and well-established WeChat miniprogram of the HIV testing service system in China has been developed by Guangzhou Centers for Disease Control and Prevention (CDC) and the MSM community-based organization Lingnan Partners Community Support Center (hereinafter called "Lingnan Center") [11].*

10. In the Introduction section of the originally published article, the last paragraph appears as follows:

> *The objective of this study is to describe the usage of the HIV e-report after it was available in Guangzhou and investigate whether it is associated with promoting HIV serostatus requests, and disclosure-related behaviors among this high-risk population.*

This has been corrected to:

> *The objective of this study is to describe the usage of the HIV e-report after it was available in Guangzhou and investigate whether it is associated with promoting HIV serostatus requests and disclosure-related behaviors among this high-risk population.*

11. In the Methods: Recruitment of Participants section of the originally published article, the second sentence of the third paragraph appears as follows:

After 3 months, alters participants would receive WeChat messages which contain the link to the follow-up questionnaire.

This has been corrected as follows.

After 3 months, alters would receive WeChat messages which contain the link to the follow-up questionnaire.

12. In the Methods: Measures: Sociodemographic Characteristics section of the originally published article, the first sentence appeared as follows:

All background characteristics of alter participants were collected in the baseline questionnaire.

This has been corrected as follows:

All background characteristics of alters were collected in the baseline questionnaire.

13. In the Methods: Measures: Sociodemographic Characteristics section of the originally published article, the last sentence appeared as follows:

We dichotomize age by 25 years, income by 5000 RMB (US $700) according to median.

This has been corrected as follows:

We dichotomized age by 25 years and income by 5000 RMB (US $700) according to the median.

14. In the Methods: Statistical Analysis section of the originally published article, the last sentence of the second paragraph appeared as follows:

Another bar graph was used to depict the manner of HIV serostatus disclosure receiving at the month 3 follow-up.

This has been corrected as follows:

Another bar graph was used to depict the manner of receiving HIV serostatus disclosures at the month 3 follow-up.

15. In the Results: Characteristics of Participants section of the originally published article, the third paragraph appeared as follows:

A total of 79% (282/357) of participants had intervened by HIV-related programs in the past 3 months. HIV stigma scores ranged from 8 to 23, and at a high level (median 19, IQR 17-22) overall. On average, participants' social norm (median 3, IQR 2.67-3) inclined to a positive direction. Further details on participants characteristics were presented in Table 1.

This has been corrected as follows:

A total of 79% (282/357) of participants took part in HIV-related programs in the past 3 months. HIV stigma scores ranged from8 to 23 and were at a high level (median 19, IQR 17-22) overall. On average, participants' social norm (median 3, IQR 2.67-3) inclined to a positive direction. Further details on participants' characteristics were presented in Table 1.

16. In the Results: HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving section of the originally published article, the section subtitle appeared as follows:

HIV e-Reports Emerging as the New Manner of HIV Serostatus Request and Disclosure Receiving

This has been corrected as follows:

HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving

17. In the Results: HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving section of the originally published article, the second paragraph appeared as follows:

At month 3 follow-up, for all 357 participants, 57.4% (205/357) of them had regular male sex partners, 42.9% (153/357) of them had casual male sex partners, and 73.4% (262/357) of them had any kind of male sex partners in the past 3 months.

This has been corrected as follows:

At month 3 follow-up, for all 357 participants, 57.4% (205/357) of them had regular male sex partners, 42.9% (153/357) of them had casual male sex partners, and 73.4% (262/357) of them had either kind of male sex partner in the past 3 months.

18. In the Results: HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving section of the originally published article, the third sentence of the third paragraph appeared as follows:

Therefore, 2 new request ways for HIV serostatus using the HIV e-report emerged; namely "I requested by sending my own HIV e-report" and "I requested by asking for partner's HIV e-report."

This has been corrected as follows:

Therefore, 2 new request approaches for HIV serostatus using the HIV e-report emerged; namely "I requested by sending my own HIV e-report" and "I requested by asking for partner's HIV e-report."

19. In the Results: HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving section of the originally published article, the fourth sentence of the third paragraph appeared as follows:

The proportions of these 2 ways were 10.7% (22/205) and 2.4% (5/205) toward regular male sex partners, and 7.2% (11/153) and 3.3% (5/153) toward casual male sex partners, respectively.

This has been corrected as follows:

The proportions of these 2 approaches were 10.7% (22/205) and 2.4% (5/205) toward regular male sex partners, and 7.2% (11/153) and 3.3% (5/153) toward casual male sex partners, respectively.

20. In the Methods: HIV e-Reports Emerging as the New Approach for HIV Serostatus Requests and Disclosure Receiving section of the originally published article, the title of Table 2 appeared as follows:

> *Table 2. HIV serostatus request and disclosure receiving behaviors from different male sex partners among alters at month 3 follow-up.*

This has been corrected as follows:

> *Table 2. HIV serostatus request and disclosure receiving behaviors toward different male sex partners among alters at month 3 follow-up.*

21. In the Results: Factors Associated With HIV Serostatus Requests and Receiving Disclosure section of the originally published article, the subsection title appeared as follows:

> *Associated Factors With HIV Serostatus Request and Disclosure Receiving*

This has been corrected as follows:

> *Factors Associated With HIV Serostatus Requests and Receiving Disclosures*

22. In the Factors Associated With HIV Serostatus Requests and Receiving Disclosure section of the originally published article, the last paragraph appeared as follows:

> *All variables listed in Table 2 were not associated with HIV serostatus disclosure receiving (not tabulated).*

This has been corrected as follows:

> *All variables listed in Table 2 were not associated with receiving HIV serostatus disclosures (not tabulated).*

23. In the Discussion: Principal Results section of the originally published article, the first two sentences appeared as follows:

> *e-Report is emerging as a new manner for HIV serostatus request and disclosure for the HIV risk population. MSM chose HIV e-report as the web-based way to disclose their own HIV serostatus or to request partner's HIV serostatus with authenticity when it was available in Guangzhou.*

This has been corrected as follows:

> *e-Reports are a new approach for HIV serostatus request and disclosure for the HIV risk population. MSM chose HIV e-report as theweb-based approach to disclose their own HIV serostatus or to request partner's HIV serostatus with authenticity when it was available in Guangzhou.*

24. In the Discussion: Principal Results section of the originally published article, the fourth sentence appeared as follows:

> *To the best of our knowledge, this is the first study that mentioned the HIV e-report and explored its association with HIV disclosure–related behaviors.*

This has been corrected as follows:

> *To the best of our knowledge, this is the first study to discuss the use of HIV e-reports and explore its association.*

25. In the Discussion: Principal Results section of the originally published article, the fifth sentence appeared as follows:

> *HIV e-report, codeveloped by MSM community itself, could be considered a novel approach to promote mutual HIV status disclosure before engaging in sexual behaviors among HIV high-risk population, and being capable to be replicated in other countries and regions based on the ability of building information platforms.*

This has been corrected as follows:

> *HIV e-report, codeveloped by the MSM community itself, could be considered a novel approach to promote mutual HIV status disclosure before engaging in sexual behaviors among HIV high-risk population, and being capable to be replicated in other countries and regions based on the ability of building information platforms.*

26. In the Discussion: Principal Results section of the originally published article, the second sentence of the second paragraph appeared as follows:

> *As e-report is a new modality in the HIV research area, studies to investigate the association between HIV e-report, HIV serostatus request, and disclosure behaviors have rarely been reported.*

This has been corrected as follows:

> *As e-report is a new modality in the HIV research area, studies to investigatethe association between HIV e-reports and HIV serostatus request and disclosure behaviors have rarely been reported.*

27. In the Discussion: Principal Results section of the originally published article, the sixth sentence of the second paragraph appeared as follows:

> *However, it is important to emphasize that HIV e-report is not a substitute for condom use.*

This has been corrected as follows:

> *However, it is important to emphasize that the HIV e-report is not a substitute for condom use.*

28. In the Discussion: Principal Results section of the originally published article, the first sentence of the third paragraph appeared as follows:

> *After HIV e-report was available, a new portion of MSM had applied e-report to request sex partner's HIV serostatus (13.4%, 35/262) and had received sex partner's e-report as HIV serostatus disclosure (26.2%, 51/195).*

This has been corrected as follows:

> *After the HIV e-report was available, a new proportion of MSM had used the e-report to request their sex partner's HIV serostatus (13.4%, 35/262)*

and had received their sex partner's e-report as an HIV serostatus disclosure (26.2%, 51/195).

29. In the Discussion: Principal Results section of the originally published article, the third sentence of the third paragraph appeared as follows:

> *MSM designed it because they feel sending out their own HIV e-report is the most natural and credible way to request male sex partners' HIV serostatus as well as disclosure of their own HIV serostatus.*

This has been corrected as follows:

> *MSM designed it because they feel sending out their own HIV e-report is the most natural and credible way to request male sex partners' HIV serostatusas well as disclose their own HIV serostatus.*

30. In the Discussion: Principal Results section of the originally published article, the fifth sentence of the fourth paragraph appeared as follows:

> *Only a few studies have investigated HIV serostatus request behavior, and our finding data contribute to the literature [5,6].*

This has been corrected as follows:

> *Only a few studies have investigated HIV serostatus request behavior, and our data contribute to the literature [5,6].*

31. In the Discussion: Principal Results section of the originally published article, the fourth sentence of the fifth paragraph appeared as follows:

> *We did not identify any factors associated with HIV serostatus disclosure receiving.*

This has been corrected as follows:

> *We did not identify any factors associated with receiving an HIV serostatus disclosure.*

32. In the Discussion: Principal Results section of the originally published article, the fourth sentence of the sixth paragraph appeared as follows:

> *The possible reason may be that disclosure receiving is a passive behavior that is primarily influenced by the characteristics of the person who disclosed their status rather than the recipient.*

This has been corrected as follows:

> *The possible reason may be that receiving a disclosure is a passive behavior that is primarily influenced by the characteristics of the person who disclosed their status rather than the recipient.*

33. In the Discussion: Principal Results section of the originally published article, the last sentence of the sixth paragraph appeared as follows:

> *Active coping strategies toward MSM their own, such as promoting the active behaviors of request through expanding the use of HIV e-report, should be promoted.*

This has been corrected as follows:

> *Active coping strategies used by MSM, such as promoting the active behavior of requests by expanding the use of HIV e-reports, should be promoted.*

34. In the Discussion: Limitations section of the originally published article, the third sentence appeared as follows:

> *Second, there might be selection bias since participants were recruited through HIV testers from a local MSM-friendly clinic in Guangzhou and questionnaires were conducted on the mobile app, which led participants trend to be young and well-educated.*

This has been corrected as follows:

> *Second, there might be selection bias since participants were recruited through HIV testers from a local MSM-friendly clinic in Guangzhou and questionnaires were conducted on the mobile app, which led to participants being younger and well educated.*

35. In the Discussion: Limitations section of the originally published article, the last sentence appeared as follows:

> *Though we find several factors associated with HIV serostatus request, the causation between them needs further study.*

This has been corrected as follows:

> *Though we found several factors associated with HIV serostatus request, the causation between them needs further study.*

36. In the Discussion: Conclusions section of the originally published article, the first sentence appeared as follows:

> *This study indicated that the HIV e-report, the health service tool coproduced by community members, has become acceptable and could be used as a new optional manner for HIV serostatus request and disclosure among sexually transmitted infections high-risk population.*

This has been corrected as follows:

> *This study indicated that the HIV e-report, the health service tool coproduced by community members, has become acceptable and could be used as a new optional approach for HIV serostatus request and disclosure among populations at high risk of sexually transmitted infections.*

37. In the Discussion: Conclusions section of the originally published article, the last sentence appeared as follows:

> *It is anticipated that e-report manner are able to have an extended spectrum of coverage to reach more target populations and ultimately accelerating the decline of infectious disease transmission.*

This has been corrected as follows:

> *It is anticipated that the e-report approach will have an extended spectrum of coverage to reach more*

target populations and ultimately accelerate the decline of infectious disease transmission.

The correction will appear in the online version of the paper on the JMIR Publications website on May 17, 2023, together with the publication

Submitted 12.05.23; this is a non–peer-reviewed article;accepted 12.05.23; published 17.05.23.

Please cite as:
Sun HT, Fan XR, Gu YZ, Lu YH, Qiu JL, Yang QL, Li JH, Gu J, Hao C
Correction: WeChat-Based HIV e-Report, a New Approach for HIV Serostatus Requests and Disclosures Among Men Who Have Sex With Men: Prospective Subgroup Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2023;11:e48961
URL: https://mhealth.jmir.org/2023/1/e48961
doi:10.2196/48961
PMID:37196344

Letter to the Editor

Methodological Considerations for a Diabetes Family-Based eHealth Intervention

Jan Brož[1*], PhD; Matthew D Campbell[2,3,4*], PhD; Pavlína Krollová[1*], MD; Juraj Michalec[1*]

[1]Department of Internal Medicine, Second Faculty of Medicine, Charles University, Prague, Czech Republic
[2]School of Nursing and Health Sciences, University of Sunderland, Sunderland, United Kingdom
[3]Wellcome-MRC Institute of Metabolic Science, University of Cambridge, Cambridge, United Kingdom
[4]Leeds Institute of Cardiovascular and Medicine, University of Leeds, Leeds, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Jan Brož, PhD
Department of Internal Medicine
Second Faculty of Medicine
Charles University
V Úvalu 84
Prague, 150 00
Czech Republic
Phone: 420 734738858
Fax: 420 22 443 4019
Email: zorb@seznam.cz

Related Articles:

Comment on: http://www.jmir.org/2023/1/e40420/

Comment in: http://mhealth.jmir.org/2023/1/e52204/

(JMIR Mhealth Uhealth 2023;11:e48652) doi:10.2196/48652

KEYWORDS
public health; type 2 diabetes mellitus; intervention; randomized controlled trial; community health center

We read with interest the article by Feng et al [1], “The Effectiveness of an eHealth Family-Based Intervention Program in Patients With Uncontrolled Type 2 Diabetes Mellitus (T2DM) in the Community Via WeChat: Randomized Controlled Trial,” published in this journal.

This trial aimed to assess the effectiveness of an eHealth family-based health education intervention for patients with T2DM to improve their glucose control, risk perception, and self-care behaviors.

After 1 year of intervention, patients in the intervention arm showed significantly lower hemoglobin A_{1c} (HbA_{1c}) values and improved several diabetes control–related skills (eg, general diet, special diet, blood sugar testing, foot care, risk knowledge, and personal control).

The authors concluded that the eHealth family-based intervention improved glucose control and self-care activities among patients with T2DM by aiding the implementation of interventions to enhance T2DM risk perceptions among family members. The intervention is generalizable for patients with T2DM using health management systems in community health centers. We applaud the authors in the preparation and execution of the study but have several questions that we feel would benefit the article’s readership.

In our observational study focused on metabolic control in patients with type 1 and type 2 diabetes treated with insulin in the Czech Republic and the Slovak Republic, we found a high clinical inertia resulting in a minimal and clinically insignificant difference in the mean HbA_{1c} within 3 years. Thus, we believe any new intervention targeting a long-term stabilized balance between the health carers’ therapeutic approach and the corresponding patient response is essential and can lead to substantial positive results [2,3].

Based on the 7-point scale, the differences in the individual skills observed are not very large. From the authors’ perspective, which intervention had the most significant effect on changes in HbA_{1c}?

Patients in the intervention group met with their physicians every 3 months. Is information available on the frequency of doctor visits in the control group?

The improvements in HbA_{1c} values in the intervention arm are substantial. They even correspond to possible major changes in treatment (eg, initiation of insulin therapy) [4]. Were treatment changes monitored during the study, as these could be responsible for the HbA_{1c} improvement?

It is also known that patients with higher HbA_{1c} levels benefit more from changes in therapy [4]. A subanalysis of HbA_{1c} changes in correlation with their baseline levels would further contribute to the discussion of the generalizability of the intervention design.

We respectfully suggest considering these remarks, especially if a study continuation is planned.

Acknowledgments

The article was supported by the Ministry of Health, Czech Republic – conceptual development of research organization, Motol University Hospital, Prague, Czech Republic, 00064203.

Conflicts of Interest

None declared.

References

1. Feng Y, Zhao Y, Mao L, Gu M, Yuan H, Lu J, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11:e40420 [FREE Full text] [doi: 10.2196/40420] [Medline: 36939825]
2. Brož J, Janíčková Žďárská D, Urbanová J, Brabec M, Doničová V, Štěpánová R, et al. Current level of glycemic control and clinical inertia in subjects using insulin for the treatment of type 1 and type 2 diabetes in the Czech Republic and the Slovak Republic: results of a multinational, multicenter, observational survey (DIAINFORM). Diabetes Ther 2018 Oct;9(5):1897-1906 [FREE Full text] [doi: 10.1007/s13300-018-0485-2] [Medline: 30094784]
3. Reach G, Pechtner V, Gentilella R, Corcos A, Ceriello A. Clinical inertia and its impact on treatment intensification in people with type 2 diabetes mellitus. Diabetes Metab 2017 Dec;43(6):501-511 [FREE Full text] [doi: 10.1016/j.diabet.2017.06.003] [Medline: 28754263]
4. Brož J, Janíčková Žďárská D, Štěpánová R, Kvapil M. Addition of basal insulin to oral antidiabetic agents in patients with inadequately controlled type 2 diabetes leads to improved HbA1c levels: metabolic control, frequency of hypoglycemia, and insulin titration analysis as results of a prospective observational study (BALI Study). Diabetes Ther 2019 Apr;10(2):663-672 [FREE Full text] [doi: 10.1007/s13300-019-0584-8] [Medline: 30788806]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
T2DM: type 2 diabetes mellitus

Edited by T Leung; submitted 02.05.23; this is a non–peer-reviewed article;accepted 28.08.23; published 18.09.23.

Please cite as:
Brož J, Campbell MD, Krollová P, Michalec J
Methodological Considerations for a Diabetes Family-Based eHealth Intervention
JMIR Mhealth Uhealth 2023;11:e48652
URL: https://mhealth.jmir.org/2023/1/e48652
doi:10.2196/48652
PMID:37721787

Letter to the Editor

Authors' Reply: Methodological Considerations for a Diabetes Family-Based eHealth Intervention

Yuheng Feng[1], BA; Xiaohong Li[1], PhD

Department of Health Policy and Management, School of Public Health, Fudan University, Shanghai, China

Corresponding Author:
Xiaohong Li, PhD
Department of Health Policy and Management
School of Public Health
Fudan University
PO Box 177, 130 Dong' an Road
Shanghai
China
Phone: 86 13918213586
Email: lixh@fudan.edu.cn

Related Articles:

Comment on: http://www.jmir.org/2023/1/e48652/

Comment on: http://www.jmir.org/2023/1/e40420/

(JMIR Mhealth Uhealth 2023;11:e52204) doi:10.2196/52204

KEYWORDS

public health; type 2 diabetes mellitus; intervention; randomized controlled trial; community health center

We greatly appreciate the authors' interest and comments [1] on our eHealth family-based intervention program [2]. We hope that our responses are beneficial to the article's readership.

Intervention Program Development

It is a common phenomenon that patients with type 2 diabetes mellitus (T2DM) have poor glucose control [3]. Clinical inertia, including physician related, health care system related, and patient related, is the main reason [4]. Our study developed an eHealth family-based intervention program, which decreased patient-related inertia. We interviewed endocrinologists and community physicians and found that physician and health care inertia existed. Some patients with T2DM with very poor glucose control failed to visit tertiary or secondary hospitals to adjust their medication.

Changes in Self-Care Activities

Changes in self-care activities were not substantial. There are two possible reasons. First, most patients with a long disease course had baseline self-care activities that were better than patients with a short disease course. So it was difficult to substantially enhance self-care activities. Second, the intervention intensity was not very high because we aimed to develop a generalizable labor-saving intervention model. Although differences were not substantial, it indicates that the eHealth family-based intervention program is effective.

The general diet and blood sugar test had the most significant effect on changes in hemoglobin A_{1c} (HbA_{1c}). During the intervention implementation, family members mainly concentrated on how to help patients with T2DM maintain a healthy diet. Regarding glucose self-monitoring, family members can help them test regularly for the glucose status and prompt them to follow the physician's suggestion.

Frequency of Doctor Visits in Control Group

According to the *National Standard for Basic Public Health Services*, community physicians need to follow up with patients registered in the community system once every 3 months. Our patients were recruited from this system. The frequency of doctor visits in the control group was the same as in the intervention group.

Monitoring of Treatment Changes

Due to workforce deficiencies in community health centers caused by the COVID-19 outbreak, we did not adjust oral medication or insulin. However, treatment changes should be monitored carefully. More detailed information could better explain how the intervention influenced the patients' glucose control.

Subanalysis of HbA1c Changes in Correlation With Their Baseline Levels

We considered baseline HbA_{1c} levels as a covariable. The covariance analysis indicated that patients with T2DM with worse baseline HbA_{1c} levels get better intervention effectiveness, similar to Brož et al's [5] study, which indicates that priorities should be given to patients with T2DM and poorer glucose control.

Future Studies

To reduce the physician-based and health care–based inertia, measures targeted toward the health care system and physicians should be implemented. Future studies could be focused on physician-based and health care–based factors. More attention could be paid to the mechanism of improving the referral system and community physicians' skills based on the medical alliance modes [6]. The feedback mechanism of two-way referral also should be improved, which will help the community physicians know whether patients with T2DM follow suggestions and provide targeted health education to those who do not follow suggestions.

Conflicts of Interest

None declared.

References

1. Brož Y, Campbell MD, Krollová P, Michalec J. Comment: methodological considerations for a diabetes family-based eHealth intervention. JMIR Mhealth Uhealth 2023;11:e48652 [FREE Full text] [doi: 10.2196/48652]
2. Feng Y, Zhao Y, Mao L, Gu M, Yuan H, Lu J, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11:e40420 [FREE Full text] [doi: 10.2196/40420] [Medline: 36939825]
3. Brož J, Janíčková Žďárská D, Urbanová J, Brabec M, Doničová V, Štěpánová R, et al. Current level of glycemic control and clinical inertia in subjects using insulin for the treatment of type 1 and type 2 diabetes in the Czech Republic and the Slovak Republic: results of a multinational, multicenter, observational survey (DIAINFORM). Diabetes Ther 2018 Oct;9(5):1897-1906 [FREE Full text] [doi: 10.1007/s13300-018-0485-2] [Medline: 30094784]
4. Reach G, Pechtner V, Gentilella R, Corcos A, Ceriello A. Clinical inertia and its impact on treatment intensification in people with type 2 diabetes mellitus. Diabetes Metab 2017 Dec;43(6):501-511 [FREE Full text] [doi: 10.1016/j.diabet.2017.06.003] [Medline: 28754263]
5. Brož J, Janíčková Žďárská D, Štěpánová R, Kvapil M. Addition of basal insulin to oral antidiabetic agents in patients with inadequately controlled type 2 diabetes leads to improved HbA1c levels: metabolic control, frequency of hypoglycemia, and insulin titration analysis as results of a prospective observational study (BALI Study). Diabetes Ther 2019 Apr;10(2):663-672 [FREE Full text] [doi: 10.1007/s13300-019-0584-8] [Medline: 30788806]
6. Cai Y, Wen C, Tang L, Liu P, Xu Y, Hu S, et al. Exploration and consideration of the medical alliance modes. Iran J Public Health 2018 Aug;47(8):1160-1165 [FREE Full text] [Medline: 30186788]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
T2DM: type 2 diabetes mellitus

Edited by T Leung; submitted 25.08.23; this is a non–peer-reviewed article;accepted 30.08.23; published 18.09.23.

Please cite as:
Feng Y, Li X
Authors' Reply: Methodological Considerations for a Diabetes Family-Based eHealth Intervention
JMIR Mhealth Uhealth 2023;11:e52204
URL: https://mhealth.jmir.org/2023/1/e52204
doi:10.2196/52204
PMID:37721796

Letter to the Editor

Concerns on Generalizability

Yongjian Lin[1], MD

[1]Department of Gastrointestinal and Gland Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning, China

Corresponding Author:
Yongjian Lin, MD
Department of Gastrointestinal and Gland Surgery
First Affiliated Hospital of Guangxi Medical University
No 6 Shuangyong Road
Qingxiu District
Nanning, 530021
China
Phone: 86 1 387 887 0525
Fax: 86 1 387 887 0525
Email: linyongjian163@163.com

Related Articles:

Comment on: http://mhealth.jmir.org/2023/1/e41099/

Comment in: http://mhealth.jmir.org/2023/1/e51852/

(JMIR Mhealth Uhealth 2023;11:e50280) doi:10.2196/50280

KEYWORDS

mHealth app; mobile health; mHealth; app; prediabetes; traditional Chinese medicine; TCM; health-related quality of life; body constitution; meridian energy

We read the study by Chung et al [1], which evaluated the effectiveness of a mobile health (mHealth) app based on traditional Chinese medicine (TCM) in patients with prediabetes. This randomized controlled trial showed that the TCM mHealth app improved physical energy, fitness, and quality of life in patients with prediabetes. However, the small sample sizes, short follow-up time, and multiple comparisons might limit the generalizability of the findings. After carefully reading this article, we present the suggestions below.

First, as the authors described in the *Methods* section, participants were randomized by a computer-generated randomization list into 3 groups—the TCM mHealth app, the ordinary mHealth app, or the control group—rather than propensity score matching [2]. In addition, most patients have individual polymorphisms in the real world, and although baseline patient characteristics are almost impossible to match in clinical studies (ie, clinical characteristics among patients in this study who received the TCM mobile), there were no significant differences between the general app and control groups. For baseline data, variable transformation, a nonparametric test using rank, or an approximate *t* test could be considered when performing a comparison between 2 small sample means if their overall variances are not equal. Therefore, the results of this study may not truly reflect patients in the real world.

Second, the statistical analysis presents descriptive characteristics as percentages or as mean (SD), as appropriate. A paired *t* test was used to examine the changes in outcome variables within groups and a 1-way ANOVA was used for comparisons among groups. However, there was no description of these data. According to Bridge and Sawilowsky [3], the Wilcoxon rank-sum test is recommended if the population characteristics are unknown, such as yang-deficiency, yin-deficiency, phlegm-stasis, body energy, and physical and mental component scores, and if the hypothesis being tested is a shift in means (or another location parameter).

Third, an interaction is an action that occurs when 2 or more objects interact with each other [4]. Since many variables are described in the first table of Chung et al's [1] paper, prespecified subgroup analyses based on these variables are necessary. Subgroup analyses based on age, gender, BMI, body composition, and blood pressure were not performed in this study. We suggest that prespecified subgroup analyses be conducted in the TCM mHealth app group, which might allow for more accurate conclusions.

In conclusion, we thank the authors for this excellent work, which provides important evidence for the integration of TCM concepts into an mHealth app for patients with prediabetes. However, we believe that the conclusions of the study would have been stronger if the abovementioned issues had been addressed.

Conflicts of Interest

None declared.

References

1. Chung H, Tai C, Chang P, Su W, Chien L. The effectiveness of a traditional Chinese medicine-based mobile health app for individuals with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 20;11:e41099 [FREE Full text] [doi: 10.2196/41099] [Medline: 37338977]
2. Gillespie IA, Floege J, Gioni I, Drüeke TB, de Francisco AL, Anker SD, on behalf the ARO Steering Committee collaborators. Propensity score matching and persistence correction to reduce bias in comparative effectiveness: the effect of cinacalcet use on all-cause mortality. Pharmacoepidemiol Drug Saf 2015 Jul;24(7):738-747 [FREE Full text] [doi: 10.1002/pds.3789] [Medline: 26011775]
3. Bridge PD, Sawilowsky SS. Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and and Wilcoxon rank-sum test in small samples applied research. J Clin Epidemiol 1999 Mar;52(3):229-235. [doi: 10.1016/s0895-4356(98)00168-1] [Medline: 10210240]
4. Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, et al. Derivation of genetic interaction networks from quantitative phenotype data. Genome Biol 2005;6(4):R38 [FREE Full text] [doi: 10.1186/gb-2005-6-4-r38] [Medline: 15833125]

Abbreviations

mHealth: mobile health
TCM: traditional Chinese medicine

Edited by T Leung; submitted 25.06.23; this is a non–peer-reviewed article;accepted 28.08.23; published 21.09.23.

Please cite as:
Lin Y
Concerns on Generalizability
JMIR Mhealth Uhealth 2023;11:e50280
URL: https://mhealth.jmir.org/2023/1/e50280
doi:10.2196/50280
PMID:37733387

Letter to the Editor

Authors’ Reply: Concerns on Generalizability

Hsueh-Wen Chung[1], MS; Chen-Jei Tai[2,3], MD, PhD; Polun Chang[4], PhD; Wen-Lin Su[5,6], MD, PhD; Li-Yin Chien[7], MPH, ScD

[1]Department of Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[2]Department of Obstetrics and Gynecology, School of Medicine, College of Medicine, Taipei Medical University, Taipei City, Taiwan
[3]Tai's Traditional Chinese Medicine Clinic, Taipei City, Taiwan
[4]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei City, Taiwan
[5]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan
[6]School of Medicine, Tzu Chi University, Hualien, Taiwan
[7]Institute of Community Health Care, College of Nursing, National Yang Ming Chiao Tung University, Taipei City, Taiwan

Corresponding Author:
Li-Yin Chien, MPH, ScD
Institute of Community Health Care
College of Nursing
National Yang Ming Chiao Tung University
No. 155, Section 2, Li-Nong Street
Beitou District
Taipei City, 112304
Taiwan
Phone: 1 886 2 2826 71
Fax: 1 886 2 2823 8614
Email: lychien@nycu.edu.tw

Related Articles:

Comment on: http://mhealth.jmir.org/2023/1/e50280/

Comment on: http://mhealth.jmir.org/2023/1/e41099/

(JMIR Mhealth Uhealth 2023;11:e51852) doi:10.2196/51852

KEYWORDS

mHealth app; mobile health; mHealth; app; prediabetes; traditional Chinese medicine; TCM; health-related quality of life; body constitution; meridian energy

We appreciate the thoughtful comments by Lin [1] on our study [2]. We have mentioned that the limitations of our study include a small sample size and a short follow-up period. We also suggested that future studies should be conducted with a larger sample size and a longer follow-up period.

The first comment raised was regarding the use of a randomized controlled trial (RCT) design without incorporating propensity score (PS) matching. Kuss et al [3] indicated that PS cannot take into account factors that are unknown or were not measured and, therefore, is more suitable for observational studies. An RCT is the only design that can ensure equal distributions of unknown confounding factors, and it enables the making of causal statements on treatment effects. Although in recent years a few studies have suggested the use of PS in RCTs [4], we are still uncertain about its appropriateness for RCTs. Future studies may attempt to investigate its use in RCTs. Nonetheless, following your suggestion, we tried regression adjustment with PS. The results appeared to be similar to those we presented in the paper.

The second comment suggests that we provide descriptive data for both between- and within-group comparisons. In addition, in cases where population characteristics are unknown, the Wilcoxon rank-sum test could have been considered for analysis. In our paper, we presented data for between- and within-group comparisons in Multimedia Appendix 5. We checked the distribution of the outcome variables including yang-deficiency, yin-deficiency, phlegm-stasis body constitution, body energy, and physical and mental component scores; the distribution approximated a normal distribution in our study. According to the central limit theorem, when the sample size of each group is greater than 30, it is reasonable to assume that the distribution of the sample means approaches normality [5].

The third point raised pertains to the absence of a prespecified subgroup analysis in our study. Owing to the predetermined objectives, hypothesis, and statistical analysis methodology at the initial stages of the study, coupled with the limitation of a small sample size, we did not incorporate a prespecified subgroup analysis. It is suggested that future studies, with an increase in sample size, may consider carrying out such an analysis.

Once again, we appreciate this opportunity to clarify our study. Such dialogues enable the identification and discussion of more aspects of this important issue.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of the primary and secondary outcomes among the TCM mHealth app, ordinary mHealth app, and control groups (N=121).

[PDF File (Adobe PDF File), 159 KB - mhealth_v11i1e51852_app5.pdf]

References

1. Lin Y. Comment: concerns on generalizability. JMIR Mhealth Uhealth 2023. [doi: 10.2196/50280]
2. Chung H, Tai C, Chang P, Su W, Chien L. The effectiveness of a traditional Chinese medicine–based mobile health app for individuals with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 20;11:e41099 [FREE Full text] [doi: 10.2196/41099] [Medline: 37338977]
3. Kuss O, Blettner M, Börgermann J. Propensity score: an alternative method of analyzing treatment effects. Dtsch Arztebl Int 2016 Sep 05;113(35-36):597-603 [FREE Full text] [doi: 10.3238/arztebl.2016.0597] [Medline: 27658473]
4. Loux T, Huang Y. The uses of propensity scores in randomized controlled trials. Observational Studies 2023;9(1):77-85. [doi: 10.1353/obs.2023.0007]
5. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. Korean J Anesthesiol 2017 Apr;70(2):144-156 [FREE Full text] [doi: 10.4097/kjae.2017.70.2.144] [Medline: 28367284]

Abbreviations

PS: propensity score
RCT: randomized controlled trial

Edited by T Leung; submitted 28.08.23; this is a non–peer-reviewed article;accepted 28.08.23; published 21.09.23.

Please cite as:
Chung HW, Tai CJ, Chang P, Su WL, Chien LY
Authors' Reply: Concerns on Generalizability
JMIR Mhealth Uhealth 2023;11:e51852
URL: https://mhealth.jmir.org/2023/1/e51852
doi:10.2196/51852
PMID:37733412

Corrigenda and Addenda

Correction: Evaluation Criteria for Weight Management Apps: Validation Using a Modified Delphi Process

Noemí Robles[1,2,3*], PhD; Elisa Puigdomènech Puig[1,3,4*], MSc; Corpus Gómez-Calderón[5*], BA; Francesc Saigí-Rubió[6,7*], PhD; Guillem Cuatrecasas Cambra[8*], MD; Alberto Zamora[9,10*], MD, PhD; Montse Moharra[4,11*], BA; Guillermo Paluzié[9*], MD, PhD; Mariona Balfegó[8*], PhD; Carme Carrion[1,2,3,6*], PhD

[1]eHealth Lab Research Group, Universitat Oberta de Catalunya, Barcelona, Spain
[2]eHealth Center, Universitat Oberta de Catalunya, Barcelona, Spain
[3]Red de Investigación en Servicios de Salud en Enfermedades Crónicas, Barcelona, Spain
[4]Agència de Qualitat i Avaluació Sanitàries de Catalunya, Barcelona, Spain
[5]Marina, Salud, Alicante, Spain
[6]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[7]Interdisciplinary Research Group on ICTs, Universitat Oberta de Catalunya, Barcelona, Spain
[8]Clínica Sagrada Família, CPEN SL Servei d'Endocrinologia i Nutrició, Barcelona, Spain
[9]Corporació de Salut del Maresme i la Selva, Hospital de Blanes, Blanes, Spain
[10]Grup de Medicina Traslacional i Ciències de la Decisió, Universitat de Girona, Girona, Spain
[11]Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública, Barcelona, Spain
[*]all authors contributed equally

Corresponding Author:
Elisa Puigdomènech Puig, MSc
Agència de Qualitat i Avaluació Sanitàries de Catalunya
Carrer Roc Boronat 81-95
Barcelona, E08005
Spain
Phone: 34 93 5513476
Email: epuigdomenech@gencat.cat

Related Article:

Correction of: https://mhealth.jmir.org/2020/7/e16899/

(JMIR Mhealth Uhealth 2023;11:e47584) doi:10.2196/47584

In "Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review" (JMIR Mhealth Uhealth 2020;8(7):e16899) the authors noted one clarification that should be added:

The Acknowledgments section reads as:

> *All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764.*

And will be changed to:

> *All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764 co-funded by FEDER.*

The correction will appear in the online version of the paper on the JMIR Publications website on April 10, 2023 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.03.23; this is a non–peer-reviewed article;accepted 28.03.23; published 10.04.23.

Please cite as:
Robles N, Puigdomènech Puig E, Gómez-Calderón C, Saigí-Rubió F, Cuatrecasas Cambra G, Zamora A, Moharra M, Paluzié G, Balfegó M, Carrion C
Correction: Evaluation Criteria for Weight Management Apps: Validation Using a Modified Delphi Process
JMIR Mhealth Uhealth 2023;11:e47584
URL: https://mhealth.jmir.org/2023/1/e47584
doi:10.2196/47584
PMID:37037025

Corrigenda and Addenda

Correction: Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review

Elisa Puigdomenech Puig[1,2,3], MSc; Noemí Robles[2,3,4], PhD; Francesc Saigí-Rubió[5,6], PhD; Alberto Zamora[7,8], MD, PhD; Montse Moharra[1,9], BA; Guillermo Paluzie[7], MD, PhD; Mariona Balfegó[10], PhD; Guillem Cuatrecasas Cambra[10], MD; Pilar Garcia-Lorda[5], MD, PhD; Carme Carrion[2,3,4,5], PhD

[1]Agència de Qualitat i Avaluació Sanitàries de Catalunya, Barcelona, Spain
[2]Red de Investigación en Servicios de Salud en Enfermedades Crónicas, Barcelona, Spain
[3]eHealth Lab, Barcelona, Spain
[4]eHealth Center, Universitat Oberta de Catalunya, Barcelona, Spain
[5]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[6]Interdisciplinary Research Group on ICTs, Barcelona, Spain
[7]Corporació de Salut del Maresme i la Selva, Hospital de Blanes, Blanes, Spain
[8]Grup de Medicina Traslacional i Ciències de la Decisió, Departament de Ciències Mèdiques, Facultat de Medicina, Universitat de Girona, Girona, Spain
[9]CIBER Epidemiología y Salud Pública, Barcelona, Spain
[10]Clínica Sagrada Família, CPEN SL Servei d'Endocrinologia i Nutrició, Barcelona, Spain

Corresponding Author:
Carme Carrion, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona
Spain
Phone: 34 934 505 254
Email: mcarrionr@uoc.edu

Related Article:

Correction of: https://mhealth.jmir.org/2019/10/e12612/

(JMIR Mhealth Uhealth 2023;11:e47585) doi:10.2196/47585

In "Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review" (JMIR Mhealth Uhealth 2019;7(10):e12612) the authors noted one clarification that should be added.

In the Acknowledgments section it says:

> *All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764.*

It should say:

> *All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764 co-funded by FEDER.*

The correction will appear in the online version of the paper on the JMIR Publications website on April 10, 2023 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.03.23; this is a non–peer-reviewed article;accepted 28.03.23; published 10.04.23.

Please cite as:
Puigdomenech Puig E, Robles N, Saigí-Rubió F, Zamora A, Moharra M, Paluzie G, Balfegó M, Cuatrecasas Cambra G, Garcia-Lorda P, Carrion C
Correction: Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review
JMIR Mhealth Uhealth 2023;11:e47585
URL: https://mhealth.jmir.org/2023/1/e47585
doi:10.2196/47585
PMID:37037024

Corrigenda and Addenda

Correction: Predictors of Playing Augmented Reality Mobile Games While Walking Based on the Theory of Planned Behavior: Web-Based Survey

Hyeseung Elizabeth Koh[1,2], MA; Jeeyun Oh[1,2], PhD; Michael Mackert[1,2,3,4], PhD

[1]The Center for Health Communication, Moody College of Communication, The University of Texas at Austin, Austin, TX, United States
[2]Stan Richards School of Advertising and Public Relations, Moody College of Communication, The University of Texas at Austin, Austin, TX, United States
[3]Department of Population Health, Dell Medical School, The University of Texas at Austin, Austin, TX, United States
[4]School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Hyeseung Elizabeth Koh, MA
Stan Richards School of Advertising and Public Relations
Moody College of Communication
The University of Texas at Austin
300 West Dean Keeton, A1200, BMC 4.338
Austin, TX, 78712
United States
Phone: 1 512 471 1101
Fax: 1 512 471 7018
Email: kohhye@utexas.edu

Related Article:

Correction of: https://mhealth.jmir.org/2017/12/e191/

(JMIR Mhealth Uhealth 2023;11:e49937) doi:10.2196/49937

In "Predictors of Playing Augmented Reality Mobile Games While Walking Based on the Theory of Planned Behavior: Web-Based Survey" (JMIR Mhealth Uhealth 2017;5(12):e191) the authors noted one error.

In Table 6, the results for the factor “Enjoyment” in the third block were shifted one column to the right. The original table can be seen in Multimedia Appendix 1. This has been changed to read as follows:

Table 6. Regression results for intention to play a mobile game while walking in study 2 (N=197).

Predictors	*B*	Standard error (SE)	Beta	*t*	*P*	*sr*
Third Block						
Enjoyment	.23	.09	.15	2.55	.01	.12

The correction will appear in the online version of the paper on the JMIR Publications website on June 19, 2023, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Multimedia Appendix 1
Original published version of “Table 6. Regression results for intention to play a mobile game while walking in study 2 (N=197).”.
[DOCX File , 15 KB - mhealth_v11i1e49937_app1.docx]

Submitted 13.06.23; this is a non–peer-reviewed article;accepted 15.06.23; published 19.06.23.

Please cite as:
Koh HE, Oh J, Mackert M
Correction: Predictors of Playing Augmented Reality Mobile Games While Walking Based on the Theory of Planned Behavior: Web-Based Survey
JMIR Mhealth Uhealth 2023;11:e49937
URL: https://mhealth.jmir.org/2023/1/e49937
doi:10.2196/49937
PMID:37335989

Corrigenda and Addenda

Correction: Efficacy, Effectiveness, and Quality of Resilience-Building Mobile Health Apps for Military, Veteran, and Public Safety Personnel Populations: Scoping Literature Review and App Evaluation

Melissa Voth[1,2], BEd; Shannon Chisholm[2], BSc; Hannah Sollid[2], BSc; Chelsea Jones[1,3,4], PhD; Lorraine Smith-MacDonald[1,2], PhD; Suzette Brémault-Phillips[1,2], PhD

[1]Heroes in Mind, Advocacy and Research Consortium, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[2]Department of Occupational Therapy, Faculty of Rehabilitation, University of Alberta, Edmonton, AB, Canada

[3]Leiden University Medical Centre, Leiden University, Leiden, Netherlands

[4]Operational Stress Injury Clinic, Alberta Health Services, Edmonton, AB, Canada

Corresponding Author:
Chelsea Jones, PhD
Heroes in Mind, Advocacy and Research Consortium
Faculty of Rehabilitation Medicine
University of Alberta
1-94 Corbett Hall
8205 - 114 Street Edmonton
Edmonton, AB, T6G 2G4
Canada
Phone: 1 7804920404
Email: cweiman@ualberta.ca

Related Article:

Correction of: https://mhealth.jmir.org/2022/1/e26453

(JMIR Mhealth Uhealth 2023;11:e51609) doi:10.2196/51609

In "Efficacy, Effectiveness, and Quality of Resilience-Building Mobile Health Apps for Military, Veteran, and Public Safety Personnel Populations: Scoping Literature Review and App Evaluation ([JMIR Mhealth Uhealth 2022;10(1):e26453]) the authors made the following 4 corrections:

1. In the originally published article, in 8 instances the name of an app appeared as:

> *Resilience@Work/Mindarma*

This has been corrected as follows in all 8 instances:

> *Mindarma*

2. In the originally published article, in the Results: Study Findings: Evidence-Based Merit section, the last sentence appeared as follows:

> *Virtual Hope Box, eQuoo, and Resilience@Work/Mindarma were evaluated separately in their respective RCT studies.*

This has been corrected as follows:

> *Virtual Hope Box and eQuoo, were evaluated separately in their respective RCT studies. It was noted that Mindarma was utilized as a part of a mindfulness program for first responders [15].*

3. In the originally published article, in the Results: Study Findings: Mental Control, Emotional Regulation, Coping, and Self-efficacy section, the following sentence appeared:

> *Resilience@Work/Mindarma was the only app in this study that drew from acceptance and commitment therapy principles.*

This sentence has been deleted from the paper.

4. In the originally published article, in the Results: Study Findings: Effect of Apps on Resilience section, the following sentences appeared:

> *Similarly, Resilience@Work/Mindarma showed improved adaptive resilience and psychological flexibility [15]. Joyce et al [15] also found that this app significantly increased optimism and mindfulness practice among study participants. This study also found that the app increased use of emotional support from others and help-seeking behavior, which addresses the social support pillar of the pillars of mental resilience.*

These sentences have been deleted from the paper.

The correction will appear in the online version of the paper on the JMIR Publications website on August 28 2023, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

15. Joyce S, Shand F, Lal TJ, Mott B, Bryant RA, Harvey SB. Resilience@Work mindfulness program: results from a cluster randomized controlled trial with first responders. J Med Internet Res 2019 Feb 19;21(2):e12894 [FREE Full text] [doi: 10.2196/12894] [Medline: 30777846]

Submitted 04.08.23; this is a non–peer-reviewed article;accepted 08.08.23; published 28.08.23.

Please cite as:
Voth M, Chisholm S, Sollid H, Jones C, Smith-MacDonald L, Brémault-Phillips S
Correction: Efficacy, Effectiveness, and Quality of Resilience-Building Mobile Health Apps for Military, Veteran, and Public Safety Personnel Populations: Scoping Literature Review and App Evaluation
JMIR Mhealth Uhealth 2023;11:e51609
URL: https://mhealth.jmir.org/2023/1/e51609
doi:10.2196/51609
PMID:37639713

Corrigenda and Addenda

Correction: Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study

Kiran K G Ravindran[1,2], PhD; Ciro della Monica[1,2], PhD; Giuseppe Atzori[1,2], MSc, RPGST; Damion Lambert[1,2], MSc; Hana Hassanin[2,3,4], MBBS, Dr med; Victoria Revell[1,2], PhD; Derk-Jan Dijk[1,2], PhD

[1]Surrey Sleep Research Centre, School of Biosciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom
[2]UK Dementia Research Institute, Care Research and Technology Centre at Imperial College, London, and the University of Surrey, Guildford, United Kingdom
[3]Surrey Clinical Research Facility, School of Biosciences, Faculty of Health and Medical Sciences, Guildford, United Kingdom
[4]National Institute for Health Research - Royal Surrey Clinical Research Facility, Guildford, United Kingdom

Corresponding Author:
Kiran K G Ravindran, PhD
Surrey Sleep Research Centre, School of Biosciences, Faculty of Health and Medical Sciences
University of Surrey
Surrey Clinical Research Building
Egerton Road
Guildford, GU27XP
United Kingdom
Phone: 44 01483683709
Email: k.guruswamyravindran@surrey.ac.uk

Related Article:

Correction of: https://mhealth.jmir.org/2023/1/e46338

(JMIR Mhealth Uhealth 2023;11:e54856) doi:10.2196/54856

In “Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study” (JMIR Mhealth Uhealth 2023;11:e46338) the authors noted one error.

The *P* value was erroneously swapped between the reported R-squared values of two of the compared devices. This error occurs in two places, and the following corrections have been made:

In the Results section of the Abstract, the sentence

> *The deep sleep duration estimates of Somnofy correlated (r^2=0.60; P<.01) with electroencephalography slow wave activity (0.75-4.5 Hz) derived from PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.21; Emfit: r^2=0.11, P=.58).*

has been changed to

> *The deep sleep duration estimates of Somnofy correlated (r^2=0.60; P<.01) with electroencephalography slow wave activity (0.75-4.5 Hz) derived from PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.58; Emfit: r^2=0.11, P=.21).*

In the sub section “DS and EEG SWA in NREM Sleep” of Results, the sentence

> *Therefore, we investigated whether DS as detected by the CSTs was associated with SWA. Somnofy DS duration was significantly correlated (r^2=0.6; P<.01) with the average SWA detected via PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.21; Emfit: r^2=0.11, P=.58; Figure 3).*

has been revised to

> *Therefore, we investigated whether DS as detected by the CSTs was associated with SWA. Somnofy DS duration was significantly correlated (r^2=0.6; P<.01) with the average SWA detected via PSG, whereas for the undermattress devices, this correlation was not significant (WSA: r^2=0.0096, P=.58; Emfit: r^2=0.11, P=.21; Figure 3).*

The correction will appear in the online version of the paper on the JMIR Publications website on December 5, 2023 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.11.23; this is a non–peer-reviewed article;accepted 24.11.23; published 05.12.23.

Please cite as:
G Ravindran KK, della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, Dijk DJ
Correction: Three Contactless Sleep Technologies Compared With Actigraphy and Polysomnography in a Heterogeneous Group of Older Men and Women in a Model of Mild Sleep Disturbance: Sleep Laboratory Study
JMIR Mhealth Uhealth 2023;11:e54856
URL: https://mhealth.jmir.org/2023/1/e54856
doi:10.2196/54856
PMID:38052028

Original Paper

Exploring Digital Biomarkers of Illness Activity in Mood Episodes: Hypotheses Generating and Model Development Study

Gerard Anmella[1,2,3,4,5*], MD; Filippo Corponi[6*], MD; Bryan M Li[6*], MSc; Ariadna Mas[1,2,3,4,5], PsyD; Miriam Sanabra[1,2,3,4,5], PsyD, PhD; Isabella Pacchiarotti[1,2,3,4,5], MD, PhD; Marc Valentí[1,2,3,4,5], MD, PhD; Iria Grande[1,2,3,4,5], MD, PhD; Antoni Benabarre[1,2,3,4,5], MD, PhD; Anna Giménez-Palomo[1,2,3,4,5], MD; Marina Garriga[1,2,3,4,5], MD, PhD; Isabel Agasi[1], RNC; Anna Bastidas[1,4,5], RNC; Myriam Cavero[1,2,3,4,5], MD, PhD; Tabatha Fernández-Plaza[1], MD; Néstor Arbelo[1,3,4,5,7], MD; Miquel Bioque[1,3,4,5,7], MD, PhD; Clemente García-Rizo[1,3,4,5,7], MD, PhD; Norma Verdolini[1,2,3,4,5], MD, PhD; Santiago Madero[1,3,4,5,7], MD; Andrea Murru[1,2,3,4,5], MD, PhD; Silvia Amoretti[1,2,3,4,5], PsyD, PhD; Anabel Martínez-Aran[1,2,3,4,5], PsyD, PhD; Victoria Ruiz[1], RNC; Giovanna Fico[1,2,3,4,5], MD; Michele De Prisco[1,2,3,4,5], MD; Vincenzo Oliva[1,2,3,4,5], MD; Aleix Solanes[2,4,5,8], MSc, PhD; Joaquim Radua[2,4,5,8,9,10], MD, PhD; Ludovic Samalin[11,12], MD, PhD; Allan H Young[13], MD, PhD; Eduard Vieta[1,2,3,4,5], MD, PhD; Antonio Vergari[6], MSc, PhD; Diego Hidalgo-Mazzei[1,2,3,4,5], MD, PhD

[1]Department of Psychiatry and Psychology, Institute of Neuroscience, Hospital Clínic de Barcelona, Barcelona, Catalonia, Spain

[2]Bipolar and Depressive Disorders Unit, Digital Innovation Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

[3]Biomedical Research Networking Centre Consortium on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

[4]Department of Medicine, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain

[5]Institute of Neurosciences (UBNeuro), University of Barcelona, Barcelona, Catalonia, Spain

[6]School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

[7]Barcelona Clinic Schizophrenia Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

[8]Imaging of Mood- and Anxiety-Related Disorders (IMARD) Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

[9]Early Psychosis: Interventions & Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, United Kingdom

[10]Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[11]Department of Psychiatry, Centre Hospitalier Universitaire (CHU) Clermont-Ferrand, University of Clermont Auvergne, Centre National de la Recherche Scientifique (CNRS), Clermont Auvergne INP, Institut Pascal (UMR 6602), Clermont-Ferrand, France

[12]Association Française de Psychiatrie Biologique et Neuropsychopharmacologie (AFPBN), Paris, France

[13]Centre for Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Diego Hidalgo-Mazzei, MD, PhD
Department of Psychiatry and Psychology
Institute of Neuroscience
Hospital Clínic de Barcelona
Villarroel St, 170
Barcelona, Catalonia, 08036
Spain
Phone: 34 932275400 ext 4189
Email: dahidalg@clinic.cat

Abstract

Background: Depressive and manic episodes within bipolar disorder (BD) and major depressive disorder (MDD) involve altered mood, sleep, and activity, alongside physiological alterations wearables can capture.

Objective: Firstly, we explored whether physiological wearable data could predict (aim 1) the severity of an acute affective episode at the intra-individual level and (aim 2) the polarity of an acute affective episode and euthymia among different individuals.

Secondarily, we explored which physiological data were related to prior predictions, generalization across patients, and associations between affective symptoms and physiological data.

Methods: We conducted a prospective exploratory observational study including patients with BD and MDD on acute affective episodes (manic, depressed, and mixed) whose physiological data were recorded using a research-grade wearable (Empatica E4) across 3 consecutive time points (acute, response, and remission of episode). Euthymic patients and healthy controls were recorded during a single session (approximately 48 h). Manic and depressive symptoms were assessed using standardized psychometric scales. Physiological wearable data included the following channels: acceleration (ACC), skin temperature, blood volume pulse, heart rate (HR), and electrodermal activity (EDA). Invalid physiological data were removed using a rule-based filter, and channels were time aligned at 1-second time units and segmented at window lengths of 32 seconds, as best-performing parameters. We developed deep learning predictive models, assessed the channels' individual contribution using permutation feature importance analysis, and computed physiological data to psychometric scales' items normalized mutual information (NMI). We present a novel, fully automated method for the preprocessing and analysis of physiological data from a research-grade wearable device, including a viable supervised learning pipeline for time-series analyses.

Results: Overall, 35 sessions (1512 hours) from 12 patients (manic, depressed, mixed, and euthymic) and 7 healthy controls (mean age 39.7, SD 12.6 years; 6/19, 32% female) were analyzed. The severity of mood episodes was predicted with moderate (62%-85%) accuracies (aim 1), and their polarity with moderate (70%) accuracy (aim 2). The most relevant features for the former tasks were ACC, EDA, and HR. There was a fair agreement in feature importance across classification tasks (Kendall W=0.383). Generalization of the former models on unseen patients was of overall low accuracy, except for the intra-individual models. ACC was associated with "increased motor activity" (NMI>0.55), "insomnia" (NMI=0.6), and "motor inhibition" (NMI=0.75). EDA was associated with "aggressive behavior" (NMI=1.0) and "psychic anxiety" (NMI=0.52).

Conclusions: Physiological data from wearables show potential to identify mood episodes and specific symptoms of mania and depression quantitatively, both in BD and MDD. Motor activity and stress-related physiological data (EDA and HR) stand out as potential digital biomarkers for predicting mania and depression, respectively. These findings represent a promising pathway toward personalized psychiatry, in which physiological wearable data could allow the early identification and intervention of mood episodes.

(JMIR Mhealth Uhealth 2023;11:e45405) doi:10.2196/45405

KEYWORDS

depression; mania; bipolar disorder; major depressive disorder; machine learning; deep learning; physiological data; digital biomarker; wearable; Empatica E4

Introduction

Mood disorders, including bipolar disorder (BD) and major depressive disorder (MDD), are ranked among the top 25 leading causes of disease burden worldwide [1] and are associated with recurrent depressive and manic episodes. Manic episodes are characterized by increased activity and self-esteem, reduced need for sleep, and expansive mood and behavior, whereas during depressive episodes, patients experience decreased energy and activity, sadness, low self-esteem, and social withdrawal [2-4]. These changes in mood, sleep, and activity during mood episodes translate to changes in physiological data that novel research-grade wearables can capture with high precision in real time [5,6]. Linking these digital signals with illness activity could potentially identify digital biomarkers [7].

Biomarkers are characteristics that are measured as an indicator of pathogenic processes (disease-associated biomarkers) or responses to an exposure or intervention (drug-related biomarkers) [8]. These can include molecular, histological, radiographic, or physiological characteristics. Digital biomarkers are objective, quantifiable, and physiological, and behavioral measures are collected using digital devices that are portable, wearable, implantable, or digestible [9]. Traditional biomarkers can be invasive and expensive to measure and are difficult to collect over time, thus giving an incomplete view of the complexity and dynamism of the disease. Alternatively, digital biomarkers are usually noninvasive, modular, and cheaper to measure, and they provide access to continuous and longitudinal measurements, both qualitative and quantitative. Moreover, they offer novel ways of measuring health status by providing perspectives into diseases that were unavailable before, which can supplement and enhance conclusions from traditional biomarkers [10]. Digital biomarkers have the potential to redefine diagnosis, improve the accuracy of diagnostic methods, enhance monitoring, and personalize interventions [11], leading to precision medicine, especially in psychiatric diseases [12].

In the last decade, there has been an exponential growth in the number of digital biomarker studies in the health domain, especially in cardiovascular and respiratory diseases [9]. Wearables are the most common type of digital devices used in digital biomarker studies, especially those incorporating accelerometer sensors that measure physical activity [13]. Wearable devices include wristbands, smartwatches, smart shirts, smart rings, smart electrodes, smart headsets, smart glasses, and so on. Wrist-worn devices are the most common type of wearable device in mental health studies and have shown to be effective in diagnosing anxiety and depression. However, none of the studies used it for treatment. The most commonly used category of data for model development was physical activity data, followed by sleep and heart rate (HR) data [14]. There are several areas in health care in which wearable devices have shown potential, including monitoring, diagnosis, treatment, and rehabilitation of diseases. Even though wearables have shown accurate activity-tracking measurements and are acceptable for users [15], including feasibility studies in people with mental health problems [16], their implementation in usual clinical practice is still challenging [17].

Wearables collecting actigraphy, the noninvasive method of monitoring human rest and activity [18], can capture altered sleep rhythms in remitted BD [19] and also depressive symptoms [20]. In addition, actigraphy data from wearables have shown to accurately predict mood disorder diagnoses and symptom change [21]. Moreover, wearables collecting blood pulse have shown differences in HR variability (HRV) between BD and healthy controls (HCs) [22], as well as between affective states in BD [23]. In addition, people with bipolar and unipolar depression and suicidal behavior have long shown autonomic alterations that can be captured as hyporeactive electrodermal activity (EDA) [24,25], and in recent years, research-grade wearables have incorporated sensors allowing continuous EDA collection [26]. With these upgrades, in the latest years, it is now feasible to monitor mood changes in patients with MDD [27] and also predict the presence and severity of depressive states in BD and MDD with promising accuracy using wearable physiological data [28]. Despite these promising results, the specific roles of these digital signals and their longitudinal potential to measure illness activity and treatment response in mood disorders are still unknown.

The conjuncture of advances in machine learning [29] and the improved precision of wearable devices [30] may help identify physiological patterns of illness activity in mood disorders. Firstly, considering this promising background, we explored whether physiological wearable data could predict the severity of an acute affective episode at the intra-individual level (aim 1) and the polarity of an acute affective episode and euthymia among different individuals (aim 2). Secondarily, we explored which physiological data were related to prior predictions, generalization across patients, and associations between affective symptoms and physiological data.

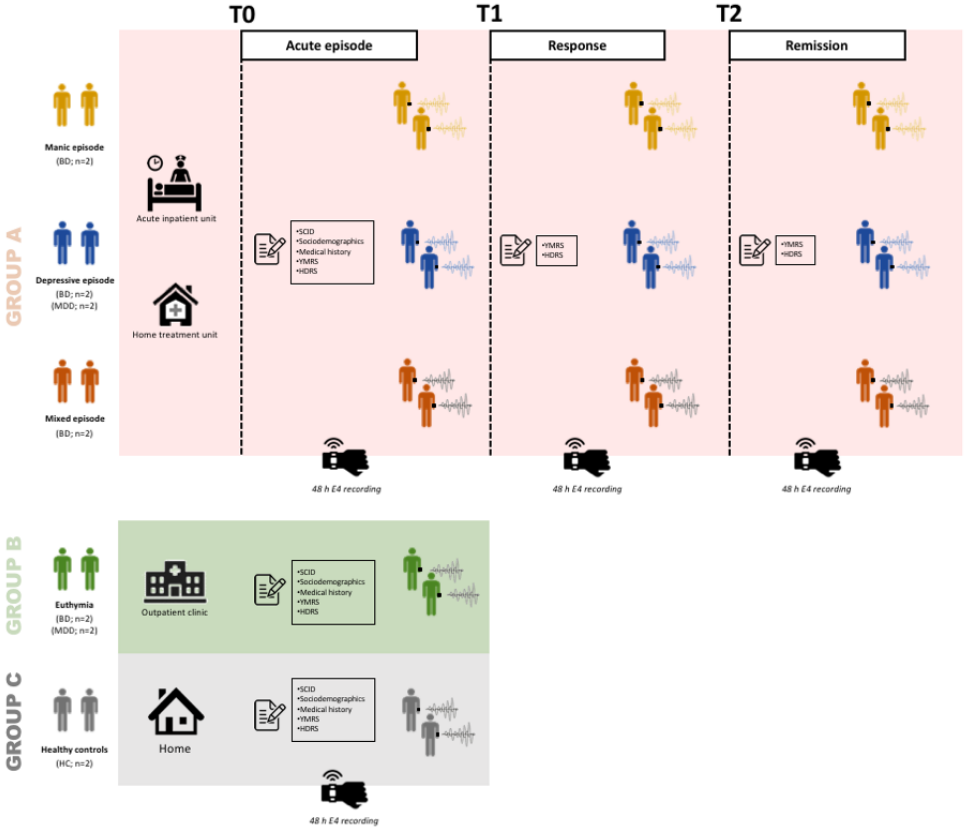
Methods

Study Design

A prospective exploratory observational study with 3 independent groups (Figure 1): group A, patients on acute affective episodes, manic episodes in BD (n=2), major depressive episodes in BD (n=2) and MDD (n=2), and mixed features manic episodes in BD (n=2); group B, euthymic patients with BD (n=2) and MDD (n=2); and group C, HC (n=7). Potential participants were identified at the outpatient and the acute inpatient or hospitalization at home units by their clinicians (ie, psychiatrists). Physiological data were recorded across 3 consecutive time points for group A: T0-acute (T0): current acute affective episodes according to the Diagnostic and Statistical Manual of Mental Disorders–5 (DSM-5); T1-response (T1): symptom response, as more than 30% improvement in the Young Mania Rating Scale (YMRS) score or the 17-item Hamilton Depression Rating Scale (HDRS) score; and T2-remission (T2): symptomatic remission, with YMRS and HDRS score ≤7 [31]). Euthymic patients (group B) and HCs (group C) were recorded during a single session.

The inclusion criteria were as follows: (1) aged above 18 years; (2) having a diagnosis according to the DSM-5 [32] criteria confirmed with the Structured Clinical Interview for DSM-5 Disorders [33]; and (3) willingness and ability to give consent (reconfirmed upon clinical remission). In addition, euthymic patients (group B) should also (4) score ≤7 on the YMRS and HDRS for at least 8 weeks [31]. HC (group C) should present no current or previous psychiatric disorder according to the DSM-5 criteria and confirmed using the Structured Clinical Interview for DSM-5 Disorders, excluding nicotine substance use disorder. Exclusion criteria for all groups were as follows: (1) concomitant severe cardiovascular or neurological medical conditions with a potential autonomic dysfunction, ongoing cardiovascular arrhythmia, or pacemaker; (2) comorbid current substance use disorder according to the DSM-5 criteria, excluding nicotine substance use disorder; (3) comorbid current psychiatric disorder with great interference of symptoms (eg, obsessive compulsive disorder with ritualized behaviors); (4) current pharmacological treatment with β-blockers or other pharmacological treatments affecting the autonomic nervous system; and (5) ongoing pregnancy.

Figure 1. Study design and recordings. BD: bipolar disorder; HC: healthy controls; HDRS: Hamilton Depression Rating Scale; MDD: major depressive disorder; SCID: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission; YMRS: Young Mania Rating Scale.



Assessments

The following sociodemographic variables were collected: age, sex, DSM-5 psychiatric diagnoses [32], medical and psychiatric comorbidities, years of illness duration, first-degree relative with mental illness, and drug misuse habits. Psychopathological assessments were conducted using the YMRS [34,35] for manic symptoms and the 17-item HDRS [36,37] for depressive symptoms. Clinical assessments were performed during a single session for euthymic patients (group B) and HCs (group C) and at 3 consecutive time points (T0-acute, T1-response, and T2-remission) for patients on acute affective episodes (group A), as described in Figure 1.

Research-Grade Wearable Device for Recording

When choosing a wearable device for a research project, there are several factors that should be considered, including (1) the signals of interest to be captured (eg, stress-related and actigraphy); (2) the users who will be studied (eg, inpatients, outpatients, and HCs); (3) the pragmatic needs of the study (eg, budget, battery life, placement of the devices, and confidentiality of participants); (4) establishing assessment procedures (eg, stress elicitation task, resting, and sleep); and (5) performing qualitative and quantitative analyses on resulting data (eg, visually inspecting the data registered, quantifying data loss, assessing the quality of data, and comparing the data of different wearable devices) [38]. Considering the previous points, the E4 wristband from Empatica [39] was the preferred wearable device for the purpose of our study for several reasons. First, the E4 has shown accuracy in measuring HR, HRV [40], and EDA compared with laboratory conditions [41], as well as for sleep staging [42]. As previously mentioned, these physiological parameters have been shown to be altered in mood disorders and mood episodes [19-23,25-28]. Second, the E4 has been validated in scientific research for detecting emotional arousal, stress [43,44], and mental effort [45] using the aforementioned physiological signals. Furthermore, the E4 has proven to be useful in predicting depressive symptoms in MDD with low relative errors [46,47], predicting self-reported depressive states [48], and identifying and quantifying the severity of anxiety states [49]. In patients with BD, the E4 has shown to be useful in distinguishing manic from euthymic mood states [50,51]. Third, the inpatients included in the study were in a highly restricted setting, which would not allow the use of user-dependent wearables or devices providing external communication (eg, an internet connection). This requirement was fulfilled by the E4 device. Finally, the data recorded by the E4 are of high precision and quality [40,41], with minimal data loss when performing the analyses (see the *Results* section).

Recording Procedure of Physiological Data

For each recording, patients and HCs were provided with an E4 wristband [39] (Multimedia Appendix 1) for approximately 48 hours (limited by battery life). The research team collected the wearables after each session. Individuals' behavior was not externally influenced in any manner, further to the requirement of wearing the wristband. Patients with acute affective episodes (group A), during their psychiatric admission in the inpatient unit, were not allowed to leave the hospital at any point until discharge, as it is the standard practice with inpatients. T0-acute, T1-response, and T2-remission recordings were usually carried out in this setting. This was not the case with patients at the hospitalization at home or outpatient units (a minority of all cases), in which patients were not subject to mobility restrictions. In all cases, both for patients and HCs, participants were asked to wear the wristband during their daily life, with little to no interference in their behavior. They were also asked to put the wristband themselves at the beginning of the recording while researchers checked for adequate contact between the sensors and the skin wrist. Participants received instructions to remove the device when taking a shower to preserve the integrity of the device.

The E4 wristband has sensors that collect physiological data at different sampling rates. The physiological data signals from each recording session were collected from the following channels and sampling rates as raw data: 3D acceleration (ACC) in space over time on an x-, y-, and z-axis (ACC, 32 Hz); EDA (4 Hz); skin temperature (TEMP, 4 Hz); and blood volume pulse (BVP, 64 Hz); or in a processed format: interbeat intervals (IBIs, the time between 2 consecutive heart ventricular contractions) and HR (1 Hz). The BVP signal is obtained using a photoplethysmography sensor that measures volume changes in the blood. Empatica uses 2 algorithms on the BVP signal to construct an IBI with which HR (and HRV) can be calculated. The 2 algorithms are optimized to detect heartbeats and discard beats that contain artifacts [39,40].

Preprocessing of Physiological Data

Owing to the naturalistic setting of the recording sessions, the data obtained from the E4 wristband are inherently noisy. For instance, some patients show low levels of compliance during an affective episode (eg, mania), which can lead to poor skin contact from the device, hence inaccurate readings for certain channels, or complete removal of the wearable device, resulting in unusable data. To that end, we removed invalid physiological data enforcing the rules-based filter by Kleckner et al [52] and an additional rule to remove HR values that exceed the physiologically plausible range (25-250 bpm) to quality control the raw data and remove physiologically impossible recordings (Table 1). Quality controlling physiological data from wearable devices is common practice, as this type of data is particularly noisy, and failing to quality control the data favors spurious correlations, and previous works have advised against imputing data in this scenario [53].

We did not use IBI data because of the disproportionately high number of missing values (approximately 70%) relative to data from different channels [54], especially because it is only a derivation of BVP. Therefore, we did not calculate HRV features. In sum, a total of 7 channels from the E4 device (ACC_X, ACC_Y, ACC_Z, BVP, EDA, HR, and TEMP) were used as physiological data to build the prediction models. Different time units (μ) and window lengths (w) were explored during tuning, and the best combination was selected. Because the sampling rate varied across different channels, the recordings were time aligned. If a channel's sampling rate was higher than 1 Hz, that channel was downsampled by taking the average value across samples within μ. We compared different time units (μ=1, 2, 4, 32, and 64 Hz), and we used 1 Hz because it showed the best performance; therefore, a time unit μ=1 second was set across all channels. Upon time alignment, each recording was then segmented into a predefined number of segments using a tunable window length (w), taking values in real-time seconds (s) (only powers of 2, specifically from 2^0 [1 s] to 2^{11} [2048 s], were explored for computational convenience). Of note, by tuning the hyperparameter w, an interesting pattern appeared across tasks, whereby a value of 2^5 (ie, 32 s) emerged as an optimal point, whereas smaller or higher values were associated with a deterioration in validation performance (U-shaped performance); therefore, μ=1 Hz and w=2^5 (32) seconds were used for analyses as the best-performing algorithm (Multimedia Appendix 2).

To obtain an equal number of segments from each class for model evaluation, we randomly selected 20 segments from each session and stored them as a held-out test set, which was never observed by the model during either training or validation. We then randomly assigned the remaining segments to the train and validation sets with ratios of 80% and 20%, respectively. Each segment was normalized (scaled to [0, 1]) using the per-channel global (across all segments) minimum and maximum values derived from the train set.

Table 1. Rules-based filter for invalid physiological data.

Rules	Filter for invalid data	Range
1	To prevent "floor" artifacts (eg, electrode loses contact with skin) and "ceiling" artifacts (circuit is overloaded)—EDA[a] not in a valid range	0.05 to 60 μS[b]
2	EDA changes too quickly—EDA slope not in a valid range	−10 to +10 μS/second
3	Skin temperature suggests the EDA sensor is not being worn—skin temperature not in a valid range	30 to 40 °C
4[c]	HR[d] not in a valid range	25 to 250 bpm[e]
5	Transitional data surrounding segments identified as invalid via the preceding rules—account for transition effects	Within 5 seconds

[a]EDA: electrodermal activity.

[b]μS: microsiemens.

[c]Addition to the algorithm used by Kleckner et al [52].

[d]HR: heart rate.

[e]bpm: beats per minute.

Data Analyses

Tasks

The recording segments produced with the preprocessing steps described earlier were used in supervised learning experiments as input to the supervised models. For aim 1, models were trained on 3-class classification tasks (T0-acute, T1-response, and T2-remission) for each individual on an acute affective episode (manic BD, depressed BD, depressed MDD, and mixed BD). For aim 2, one model was trained on a 7-class classification task (manic BD, depressed BD, mixed BD, depressed MDD, euthymic BD, euthymic MDD, and HCs).

Segments from each class under a given task were extracted in the same number to obtain perfectly balanced classes. As sets were designed to be perfectly balanced, we adopted accuracy as our primary metric but also reported the F_1-score, precision, and recall and computed the area under the receiver operating characteristic (AUROC) curves. It should be noted that ours is a multiclass setting, but as we had perfectly balanced sets, micro-, macro-, and weighted averages coincided. For the AUROC curves, the one-vs-rest multiclass strategy was adopted, also known as one-vs-all, which amounts to computing a receiver operating characteristic (ROC) curve for each class, so that at a given step, a given class is regarded as positive and the remaining classes are lumped together as a single negative class.

As part of our exploratory data analysis, to quantify the association between physiological data and affective symptoms measured by the YMRS and HDRS scale items, their normalized mutual information (NMI) was computed.

For each task, with the exception of the one about distinguishing members of a group of only HCs, as we were interested in testing the degree to which a model can generalize to different individuals, unseen during training, and sharing the same psychiatric label (diagnosis and psychopathological status), we prepared a test set of segments from recordings collected from an independent group of individuals. Therefore, the model was tested on this extra, independent holdout set to obtain an estimate of the out-of-sample generalization performance.

Model

We elected a Bidirectional Long Short-Term Memory (BiLSTM) model [55] as our model architecture. BiLSTM is a type of recurrent neural network (RNN), a class of deep learning model specifically designed to handle sequence data such as time series. RNNs process streams of data one time step at a time, and they store information regarding previous time steps in a hidden unit, such that the model output at each time step is informed by the current time step as well as by previous ones. Long short-term memory (LSTM) units represent an improvement over vanilla RNNs, as they address gradient instability by modeling the hidden state with cells that decide what to keep in memory and what to discard. This feature makes LSTM more efficient in capturing long-range dependencies. In contrast to a simple LSTM, BiLSTM reads the input sequence in 2 directions, from start to end and from end to start, thereby allowing for a richer representation. Although other deep learning architectures suitable for time series have been developed (more recently, the transformer [56]), as the aim of this work was exploratory rather than benchmarking different models, we contented ourselves with a single popular architectural choice for time series. By the same token, we used a simple shallow BiLSTM with 128 hidden units and tanh activation, followed by a single dense layer with softmax activation, to output the possible classes. The BiLSTM model was trained using the Adam optimizer [57] for 120 epochs with a learning rate of 0.001 and a batch size of 32 to minimize the cross-entropy between the ground-truth distribution over classes and the probability distribution of belonging to such classes outputted by the last network layer. To reduce overfitting, dropout [58] and early stopping were used. The choice of hyperparameters was based on a random search that yielded the best performance in the validation set.

Permutation Feature Importance

To assess the channels' individual impact on the test set performance in the aforementioned tasks, we adopted a perturbation-based approach. For each channel at a time, we randomly permuted its values in the test set segments and computed the difference in performance relative to the baseline model. We chose this approach because it has a straightforward interpretation and provides a highly compressed, global insight into the importance of the channels. Agreement on channels' relevance across different tasks was measured using the Kendall W.

Code and Data Availability

The codebase was written in Python (version 3.8; Python Software Foundation), where the deep learning models were implemented in TensorFlow and developed on a single NVIDIA RTX 2080Ti. The repository for this study can be found on the internet [59].

Ethics Approval and Confidentiality

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and Good Clinical Practice and the Hospital Clinic Ethics and Research Board (HCB/2021/104). All participants provided written informed consent before their inclusion in the study. All data were collected anonymously and stored encrypted in servers complying with all General Data Protection Regulation and Health Insurance Portability and Accountability Act regulations.

Results

Overview

A total of 35 sessions from 12 patients (manic, depressed, mixed, and euthymic) and 7 HCs (mean age 39.7, SD 12.6 years; 6/19, 32% female) were analyzed, totaling 1512 hours recorded. The median percentage of data per recording session dropped from further analysis of quality control was 11.05 (range 2.50-34.21). A clinical demographic overview of the study sample is presented in Table 2.

Table 2. Clinical demographic overview of the study sample.

Diagnosis	Age (years)	Sex	HDRS[a] score			YMRS[b] score		
			T0[c]	T1[d]	T2[e]	T0	T1	T2
Manic BD[f]	40	Male	5	4	4	24	8	2
Manic BD[g]	21	Male	3	5	4	23	15	1
Depressed BD[h]	33	Male	23	6	4	0	0	0
Depressed BD[g,h]	36	Male	17	12	3	2	4	2
Mixed BD	30	Female	8	4	4	30	20	5
Mixed BD[g]	40	Male	11	2	1	29	10	3
Depressed MDD[i]	57	Male	33	13	7	7	2	0
Depressed MDD[g]	45	Male	27	11	7	4	1	1
Euthymic BD	54	Male	3	—[j]	—	0	—	—
Euthymic BD[g]	61	Male	1	—	—	3	—	—
Euthymic MDD	60	Female	4	—	—	0	—	—
Euthymic MDD[g]	60	Male	3	—	—	0	—	—
HC[k]	32	Female	0	—	—	0	—	—
HC[g]	34	Male	0	—	—	0	—	—
HC	28	Female	0	—	—	1	—	—
HC	29	Male	0	—	—	2	—	—
HC	31	Male	2	—	—	1	—	—
HC	32	Female	1	—	—	3	—	—
HC	31	Female	0	—	—	1	—	—

[a]HDRS: Hamilton Depression Rating Scale.
[b]YMRS: Young Mania Rating Scale.
[c]T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes or only register for euthymic patients and healthy controls.
[d]T1: symptoms' response.
[e]T2: symptomatic remission.
[f]BD: bipolar disorder.
[g]The recording segments extracted from the marked subjects were used to check the models' ability to generalize to clinically similar subjects, unseen during training.
[h]All registers performed at the hospitalization at home or outpatient units.
[i]MDD: major depressive disorder.
[j]Euthymic patients and healthy controls were recorded during a single session (T0).
[k]HC: healthy control.

Aim 1: Prediction of the Severity of an Acute Affective Episode at the Intra-individual Level

The 3-class classification tasks (T0-acute, T1-response, T2-remission; accuracy expected by chance: 1/3=33%) to predict the severity of an acute affective episode showed accuracies ranging from 62% (depressed BD) to 85% (depressed MDD). The generalization models on unseen patients showed accuracies ranging from 28% (depressed MDD) to 57% (manic BD; Table 3). The confusion matrix is shown in Multimedia Appendix 3. This means that the model showed moderate to high accuracies for classifying the severity of each acute affective episode, with the best prediction models classifying individuals with depressed MDD and manic BD. However, generalization of the models was of very low accuracy for depressed MDD and mixed BD (by chance; approximately 30%), of low accuracy (slightly above chance; >40%) for mixed BD, and of moderate accuracy (>55%) for manic BD.

The permutation importance analysis for the classification tasks for aims 1 and 2 is shown in Figure 2. Kendall W was 0.383, indicating fair agreement in feature importance across both intra- and inter-individual classification tasks. ACC was the most relevant channel for predicting mania, whereas EDA and HR, followed by TEMP, were the most relevant channels for predicting both BD and unipolar depression (aim 1). The BVP channel did not change performance for either better or worse (Figure 2).

Table 3. Prediction of the severity of an acute affective episode: model and generalization on unseen patients.

Individuals with affective episodes and performance metric	Model	Generalization
Manic BD[a]		
Accuracy[b] (%)	70	56.67
F_1-score	0.6978	0.5279
Precision	0.6979	0.5381
Recall	0.7000	0.5667
AUROC[c]	0.6980	0.5432
Depressed BD		
Accuracy[b] (%)	61.67	41.67
F_1-score	0.6171	0.3968
Precision	0.6273	0.4085
Recall	0.6167	0.4167
AUROC	0.6115	0.4067
Mixed BD		
Accuracy[b] (%)	63.33	30
F_1-score	0.6333	0.2576
Precision	0.6333	0.3004
Recall	0.6333	0.3068
AUROC	0.6333	0.3012
Depressed MDD[d]		
Accuracy[b] (%)	85	28.33
F_1-score	0.8492	0.2451
Precision	0.8774	0.2581
Recall	0.8500	0.2833
AUROC	0.8672	0.2856

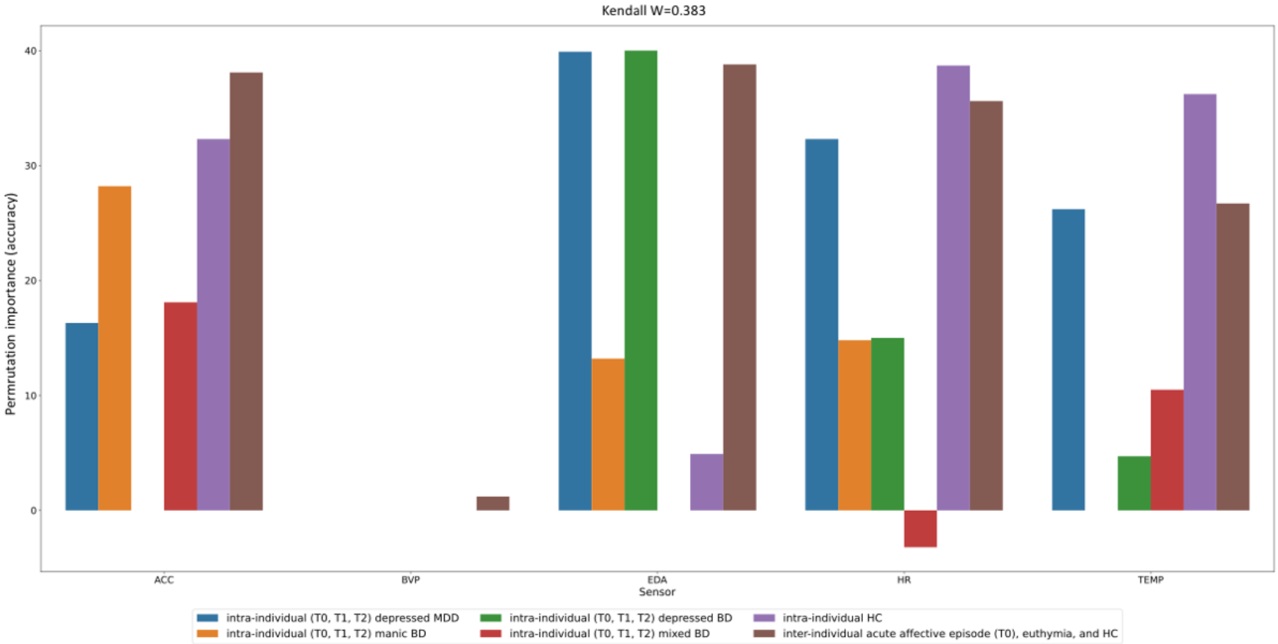
[a]BD: bipolar disorder.

[b]Accuracy expected by chance for a 3-class classification task is 1/3=33%. Thus, accuracies above 33% suggest that the model can predict outcomes better than random guessing, and higher values for accuracy indicate better predictive capacity of the model. Note that the test set was designed to have the same number of samples in each class. This is reflected in the values of F_1-score, precision, and recall being very close to each other and to that of accuracy.

[c]AUROC: area under the receiver operating characteristic.

[d]MDD: major depressive disorder.

Figure 2. Permutation importance analysis. The height of the bars shows the change in accuracy at test time upon scrambling a channel through a random permutation of its values. A positive (negative) permutation importance value means that scrambling that channel results in a drop (increase) in accuracy relatively to the baseline where original (nonpermuted) values were used across all channels, that is, the channel's permutation deteriorates (improves) the performance. A "0" permutation importance value indicates that a random permutation of the channel's values does not affect accuracy in either direction. For instance, electrodermal activity (EDA) shows a positive change in accuracy of 40% for the intra-individual depressed BD severity prediction model; this means that removing this channel from the model would result in a decrease of prediction accuracy of 40%—from 62% to 22%—thus EDA is highly relevant for that model. Different colors correspond to the different tasks being investigated. ACC: acceleration; BD: bipolar disorder; BVP: blood volume pulse; HC: healthy controls; HR: heart rate; MDD: major depressive disorder; TEMP: temperature; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission.



Aim 2: Prediction of the Polarity of an Acute Affective Episode and Euthymia Among Different Individuals

The 7-class classification task (accuracy expected by chance: 1/7=14%) to predict the polarity of affective episodes and euthymia showed an accuracy of 70%. The best classifications were depressed and euthymic MDD, followed by depressed BD, and the worst was manic BD, followed by HCs. The generalization model showed an accuracy of 15.7% (slightly above chance). The classification task for 7 HCs showed an accuracy of 50% (Table 4). The confusion matrix is shown in Multimedia Appendix 4. Thus, both models showed predictions above chance, but their generalization was poor. Moreover, the model including patients with acute affective episodes obtained higher accuracy (70%) than the model including 7 HCs (50%). This increased prediction capacity suggests that psychopathological symptoms during acute affective episodes may translate into physiological alterations that are not present in HCs.

The most relevant channels for predicting the polarity of affective episodes, euthymia, and HCs among different individuals (aim 2) were EDA, followed by ACC, HR, and TEMP (all channels showed >30% permutation importance). The BVP channel permutation importance was approximately 0%. These results were highly similar for the classification task of 7 HCs, but EDA showed only 4.9% permutation importance (Figure 2).

Table 4. Prediction of the polarity of an acute affective episode and euthymia among different individuals: model and generalization on unseen patients.

Individuals with affective episodes and performance metric	Model	Generalization
6 patients (acute affective episodes and euthymia) and 1 HC[a]		
Accuracy[b] (%)	70	15.7
F_1-score	0.6927	0.1516
Precision	0.6889	0.1513
Recall	0.6934	0.1517
AUROC[c]	0.6900	0.1510
7 HCs		
Accuracy[b] (%)	50	—[d]
F_1-score	0.4923	—
Precision	0.4911	—
Recall	0.4988	—
AUROC	0.4998	—

[a]HC: healthy control.

[b]Accuracy expected by chance for a 3-class classification task is 1/3=33%. Thus, accuracies above 33% suggest that the model can predict outcomes better than random guessing, and higher values for accuracy indicate better predictive capacity of the model. Note that the test set was designed to have the same number of samples in each class. This is reflected in the values of F_1-score, precision, and recall being very close to each other and to that of accuracy.

[c]AUROC: area under the receiver operating characteristic.

[d]As we were interested in predicting affective psychopathology, we tested the degree to which a model can generalize to different individuals for each task except for the one about distinguishing members of a group of only HCs.

Symptom Association With Physiological Data

The tile plots for the NMI between physiological data and the YMRS and HDRS scale items for the former intra-individual (aim 1) and between-individuals (aim 2) classification tasks are shown in Figures 3 and 4, respectively. TEMP had the highest association with psychometric scales (NMI approximately 1.0), and BVP had the lowest consistency (NMI scores oscillating from 0 to 1).

Figure 3. Tile plots for the normalized mutual information analysis between physiological data and psychometric scales' items: intra-individual level. For each scales' item the mutual information (MI) with respect to each of the channels was measured and scaled to 0 to 1 dividing by the maximum MI value for that item. Values of zero indicate no associations, values of 1 indicate the maximum recorded MI across all channels for an individual item. ACC_X: x-axis acceleration; ACC_Y: y-axis acceleration; ACC_Z: z-axis acceleration; BD: bipolar disorder; BVP: blood volume pulse; EDA: electrodermal activity; HDRS: Hamilton Depression Rating Scale; HR: heart rate; MDD: major depressive disorder; TEMP: temperature; YMRS: Young Mania Rating Scale.

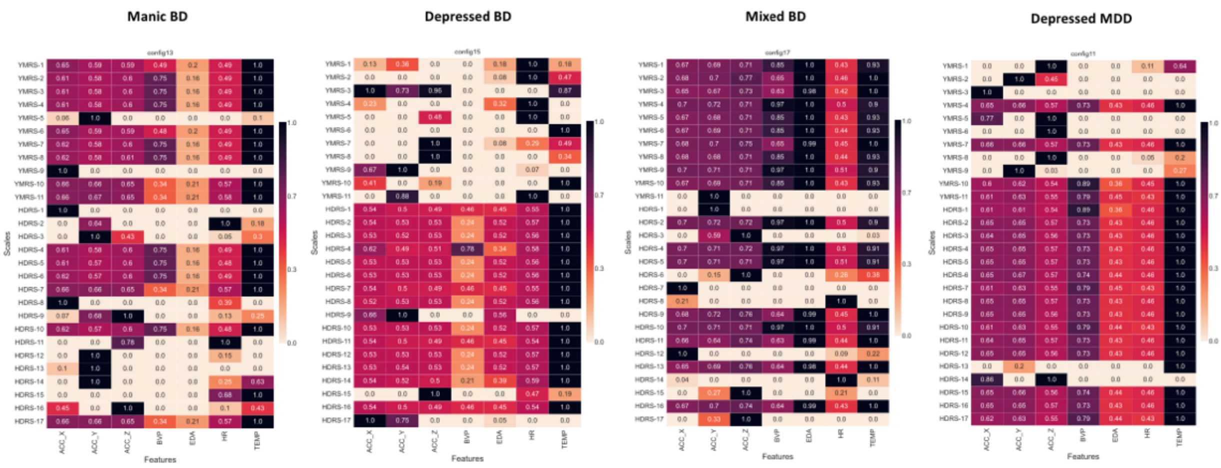
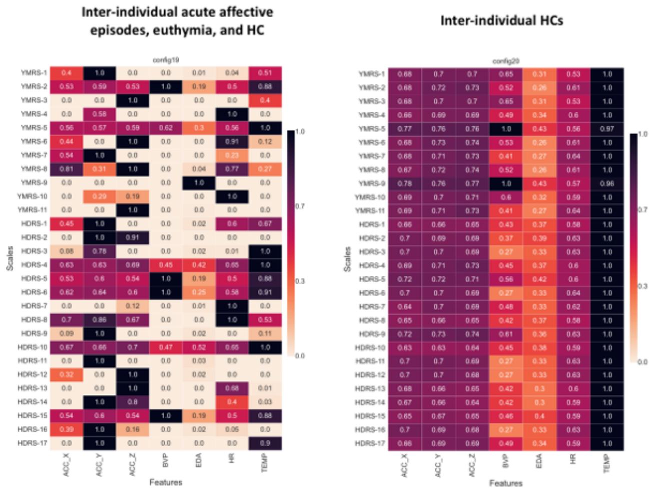


Figure 4. Tile plot for the normalized mutual information analysis between physiological data and psychometric scales' items: between-individual level. For each scales' item, the mutual information (MI) with respect to each of the channels was measured and scaled to 0 to 1 dividing by the maximum MI value for that item. Values of "0" indicate no associations; values of 1 indicate the maximum recorded MI across all channels for an individual item. ACC_X: x-axis acceleration; ACC_Y: y-axis acceleration; ACC_Z: z-axis acceleration; BVP: blood volume pulse; EDA: electrodermal activity; HC: healthy controls; HDRS: Hamilton Depression Rating Scale; HR: heart rate; TEMP: temperature; YMRS: Young Mania Rating Scale.



Intra-individual NMI Analysis

Motor activity (ACC) channels were highly associated with manic symptoms (NMI>0.6), and stress-related channels (EDA and HR) with depressive symptoms (NMI from 0.4 to 1.0), as shown in Figure 3.

Between-Individuals NMI Analysis

"Increased motor activity" (YMRS item 2 [YMRS2]) was associated with ACC (NMI>0.55), "aggressive behavior" (YMRS9) with EDA (NMI=1.0), "insomnia" (HDRS4-6) with ACC (NMI~0.6), "motor inhibition" (HDRS8) with ACC (NMI~0.75), and "psychic anxiety" (HDRS10) with EDA (NMI=0.52), as shown in Figure 4.

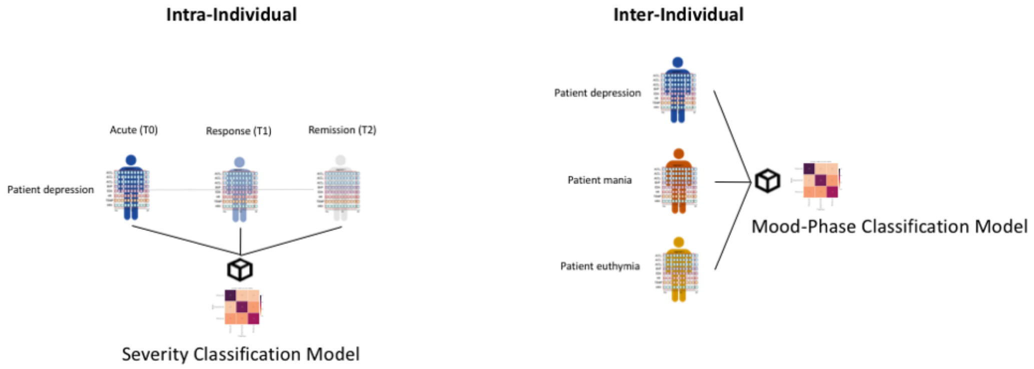
Discussion

Principal Findings

Although other studies have used raw physiological data to predict mental health status, this is the first study to present a novel fully automated method for the analysis of raw physiological data from a research-grade wearable device, including a rules-based filter for invalid physiological data, whereas all other studies presented methods that required manual interventions at some point in the pipeline [46,47,51,60], thus hindering the replicability and scalability of results. Moreover, our preprocessing pipeline is strictly based on the best-performing algorithm for analysis (ie, not arbitrarily decided), whereas other studies decided arbitrary cutoff points for analyzing raw physiological data (eg, ACC data recorded at 32 Hz sampling rates analyzed arbitrarily in 1-min epochs [50]). Our method may allow other research teams to use a viable supervised learning pipeline for time-series analyses for a popular research-grade wristband [39]. In addition, our work integrates physiological digital data from all sensors captured by a research-grade wearable, and we assessed the relevance of each channel (ACC, TEMP, BVP, HR, and EDA) in the prediction models. In contrast, other studies have focused on specific digital signals, such as actigraphy [50], or used combinations of digital signals (such as actigraphy and EDA) and predesigned features (eg, amplitude of skin conductance response peaks) [51] but arbitrarily disregarded other digital signals, such as TEMP, or derived features, such as HRV. Furthermore, we aimed to distinguish the severity of mania and depression in a progressive and longitudinal manner according to the usual clinical resolution of mood episodes. We believe that the potential quantification of affective episodes is harder but a clinically more relevant task that may allow a more accurate and precise understanding of the disease rather than a mere dichotomous (acute vs remission) classification, as done in previous studies [50,51]. In addition, we included in the same work analyses at the intra-individual level and between different individuals, analyses targeting specific mood symptoms and generalization of the models on unseen patients. We believe that the use of different analysis methods allows us to examine the data from complementary perspectives to answer specific research questions. In addition, these different approaches may reveal random associations or artifacts that would stay hidden without replication. On the basis of these exploratory results, we propose hypotheses for future testing [61] in current and other similar projects.

Note that both (1) intra- and (2) inter-individual analyses approach different research questions: the (1) intra-individual analytical approach looks at the course of an index episode within a single patient and examines whether different states (from the acute phase to response and remission) can be distinguished from each other; on the other hand, the (2) inter-individual analytical approach takes a cross-sectional view and studies the degree to which different mood disorder states (comprising the full spectrum from depression to mixed state, mania, and euthymia) can be separated. Both analyses try to identify digital biomarkers of illness activity using physiological data collected with a wristband. However, intra-individual analyses look for a fine-grained quantification of illness activity that may allow the identification of low-severity mood states (or prodromal phases) in comparison with moderate to severe ones. Conversely, inter-individual analyses could potentially distinguish between mood phases (mania vs depression) or cases from HCs but may not be suitable for assessing the severity of mood episodes, as represented in Figure 5. Studies in similar areas, such as brain computer interfaces for the rehabilitation of motor impairments [62] or seizure forecasting [63], emphasized the importance of the subject-wise approach (modeling each subject separately). In many instances, despite work on domain adaptation [64] to learn subject-invariant representations, a model has to be fine-tuned to the level of the single patient.

Figure 5. Severity versus Mood-Phase Classification Models: visual grounds for both intra- and inter-individual analyses. On the left, a severity classification model for a patient with depression (acute-response-remission phases). On the right, a mood-phase classification model (depression, mania, and euthymia). Note that on the left model, the same individual is compared at 3 different states (corresponding to a reduction in depressive psychopathology). Thus, individual-level characteristics (age, sex, and gait) should go through little to no variation across; should remain the same on the 3 longitudinal registers; and therefore, the shift in the covariate distribution should be relatively contained and not influence the classification of the model (capturing mood-relevant signals). In contrast, on the right, 3 different individuals at 3 different mood states are compared. In this case, the model would potentially distinguish between mood phases (mania vs depression), or cases from healthy controls, but may not be able to distinguish longitudinal changes in disease severity over the course of an index episode. In addition, in the latter model, subject-specific characteristics may be overlapped with mood-relevant signals, thus acting as confounders for the model. T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes; T1: symptoms' response; T2: symptomatic remission.



Studies comparing intra- and inter-individual models show that although intra-individual (cross-subject or patient-specific) models are trained on the data of a single subject, they perform better than intersubject (within-subject or generalized) models [65]. However, some studies have shown that hybrid models trained on multiple subjects and then fine-tuned on subject-specific data led to the best performance, without requiring as much data from a specific subject [66]. In intersubject studies, models generally see more data, as multiple subjects are included, but must contend with greater data variability, which introduces different challenges. In fact, there is both intra- and intersubject variability owing to time-variant factors related to the experimental setting and underlying psychological parameters. This impedes direct transferability or generalization among sessions and subjects [62]. To illustrate this, in a study aimed at evaluating a seizure detection model using physiological data and determining its application in a real-world setting, 2 procedures were applied: intra- and intersubject evaluation. Intrasubject evaluation focuses on the performance of the methodology when applied to data from a single patient, whereas intersubject evaluation assesses the performance of multiple patients with potentially different types of epilepsy and seizure manifestations [63].

Notably, the out-of-sample generalizations of both models differ vastly. Whereas the intra-individual model requires multiple seizures recorded per subject and will produce individualized models tailored to a single patient, the inter-individual model requires seizures recorded from multiple participants and will provide intersubject models to be used over wider populations. For this purpose, intersubject variability plays a key role: focal seizures have a multitude of possible clinical manifestations that can occur in sequence or in parallel and can be repeated or not occur at all, in a single seizure. For instance, preictal tachycardia appears to be a phenomenon that is not generalizable to patient cohorts. Furthermore, although there may oftentimes be little change in the semiology of seizures for a single patient, they can be very heterogeneous across populations. Intra-individual models optimized for each patient can robustly detect seizures in some patients with epilepsy, but they may fail, especially when the seizures have differing semiologies that are not represented in the training data for the model. Intersubject models perform worse than if trained in an individualized manner, at least in terms of either sensitivity or false-alarm rates [63]. This is equivalent to a study aimed at evaluating a model for mood episode detection and determining its application in a real-world setting. During acute affective episodes, a huge combination of symptoms can be present in 2 different patients [67,68], and recurrent longitudinal affective episodes in a single patient can present with a similar combination of symptoms, but this is not always the case [69-72]. At the intrasubject level, out-of-sample generalization would require multiple episodes of disease occurrence longitudinally in a single patient. In fact, similar studies with intra-individual models have achieved high detection accuracies with low sample sizes and better performance than intersubject classification [63,73]. In contrast, at the intersubject level, out-of-sample generalization does not require longitudinal episodes but only cross-sectional episodes in different patients. Therefore, both models serve different but complementary purposes to build a real-world model for the detection of prodromal affective symptoms. Future studies combining intra- and inter-individual analyses should determine which of these approaches may work best to identify affective episodes, giving guidance for the design of future studies in the field.

Clinically, the end goal is to have a model inferring mood states at the individual level, regardless of whether such a model is shared across subjects or if each subject has a tailored model. Although most digital biomarker research has focused on diagnosis classification, few studies have aimed to detect longitudinal symptom change. Developing methods to detect changes in mood symptoms has the potential to prompt just-in-time interventions to prevent full-blown affective relapses and clinical deterioration and evaluate the response to pharmacological treatments with objective measures [21].

In our sample, both intra- and inter-individual models for respectively assessing differences in severity of acute affective episodes over time (Table 3) and differences in the polarity of acute affective episodes, euthymia, and HCs (Table 4) showed accuracies considerably above chance. Although preliminary, these results indicate that there may be objective differences in digital signals (ie, digital biomarkers) according to the psychopathological severity of patients (intra-individual models) and that patients with BD or MDD may present particular patterns of digital signals for mood episodes of mania and depression (inter-individual models). However, with few patients and measurements per model, these digital biomarkers may be challenging to identify and even harder to generalize.

Motor activity (from ACC) was the most relevant digital signal for predicting the severity of mania and mixed mania (but not for unipolar or bipolar depression) and also for predicting the polarity of acute affective episodes between individuals (Figure 2). In line with our results, other research groups have found that wearable motor activity data can distinguish mania from remission in patients with BD at the intra-individual level [50]. Moreover, other studies have shown that motor activity data could identify mood episodes and euthymia among different individuals, including mania versus euthymia [51], depression versus HCs [60], and mania versus depression versus HCs [74]. In fact, “activation,” which comprises having objective (motor activity) and related subjective (energy) levels emerging from underlying physiological changes, has been widely recognized as a key feature from mania [75]. Previous literature proposes that mood and activation represent distinct dimensions of BD [76] with distinct intervention approaches [77]. In addition, dysregulation of patterns of activity has been observed in BD both in acute phases and euthymia and has been proposed as a potential biomarker for BD [78]. However, it should be noted that mania may be better characterized by differences in robustness, variability, predictability, or complexity of activation rather than mean levels of activity [75], so future analyses should explore which characteristics of motor activity are key for the former predictions.

In contrast, “stress-related” digital signals (EDA and HR) were the most relevant for predicting the severity of both unipolar and bipolar depression (but not mania or mixed mania) and were also prominent for predicting the polarity of acute affective episodes between individuals (Figure 2). In fact, when looking at psychic anxiety as a symptom (item 10 from HDRS), EDA and HR showed strong associations (Figure 4). Moreover, EDA showed relevance for predicting the polarity of affective episodes between individuals but did not differentiate between HCs (38% vs 4.9%), as shown in Figure 2. This suggests that EDA may be a specific marker for psychopathological alterations that are not present in HCs. Furthermore, skin TEMP (a proposed marker of stress) was also a relevant physiological signal for predicting the severity of unipolar and bipolar depression (Figure 2). These findings are in line with previous literature [26,79-82] and reinforce the hypothesis that stress plays a key role in people with depression. Whereas patients with manic episodes usually lack insight into their symptoms, patients with depression are usually aware of their altered state and bear much distress and anxiety [83], which may be translated into physiological alterations, as suggested in our findings.

Generalizations of the former models on unseen patients were of overall low accuracy, which may be due to high psychopathological and individual heterogeneity, as well as external factors. Although mood episodes share many psychopathological aspects, they can present with multiple combinations of symptoms [68,76,84]. Each digital signal may provide information on a specific symptom dimension (altered motor activity, sleep disturbances, and stress-related symptoms) rather than the entire affective episode (manic, depressive, or mixed). We hypothesized that training the models with a larger sample, including patients with different symptom combinations for each affective episode, will result in more precise generalizations. Thus, exploring how patients cluster according to physiological data might help toward a dimensional (rather than categorical) disease classification. Deep learning is a promising approach for clustering high-dimensional, unstructured data [85], and new methods have been proposed specifically for data from wearable devices (multivariate time series) [86,87]. Apart from polymorphic psychopathological presentations in mood episodes, there is high between-subject heterogeneity in physiological data. For instance, skin TEMP, HR, and EDA vary within a physiological range in the same individual according to external (ie, atmospheric humidity or ambient TEMP) or internal factors (ie, hydration, diet, caffeine intake, and drugs) [52], and there are also individual-level patterns (eg, specific gaits, circadian rhythms, basal skin TEMP, or HR). This calls for ad-hoc techniques to disentangle between-patient heterogeneity from mood-related signals [88] and consider the role of potential confounders in the models (eg, drugs, medical comorbidities, physical activity, atmospheric conditions, and diet). Notwithstanding, generalizations of the intra-individual models for manic BD and depressed BD were above chance, in contrast to the generalization of the inter-individual model (almost by chance). This may suggest that individual heterogeneity is partially controlled for when comparing the same individual at different time points. This way, physiological changes may be more related to psychopathology rather than simply to individual characteristics (eg, gait, sex, and age) However, intra-individual comparisons do not control for external factors (eg, humidity, atmospheric TEMP, exercise, or hydration), which should be considered and controlled for.

When exploring the association between affective symptoms and physiological data, skin TEMP showed the highest association with psychometric scales (NMI approximately 1.0; Figures 3 and 4). Skin TEMP has been proposed as an objective physiological marker of stress [89,90], and it has been shown that people with mood disorders present objective reductions in peripheral skin TEMP (due to vasoconstriction) after stress-oriented interventions [91]. Moreover, skin TEMP from wearable data has been used to study circadian rhythms in patients with mood disorders, showing alterations in their chronobiology [92]. Even so, thermoregulatory dysfunction has been proposed in a subgroup of patients with BD [93]. However, the skin TEMP continuously recorded with wearables has been relatively understudied in mood disorders, and further efforts should be made in this direction.

Regarding the most relevant inputs for the previous models, physiological data related to specific symptom dimensions (eg, ACC with motor activity and EDA and HR variation with stress response or anxiety) seemed to be more relevant signals for predicting mood episode severity and polarity rather than more raw data, such as BVP with nearly 0% permutation importance in all models (Figures 2-4), which do not seem to have a direct clinical translation to physiological alterations related to mental health symptoms. We hypothesized that complex features with potential clinical translation (ie, indicating stress response or autonomic dysfunction), such as HRV [22,23,94], which is calculated from BVP, and EDA reactivity, calculated from EDA [26], may be of greater value than second-to-second changes in motor activity (ACC), EDA, pulse (BVP), and TEMP. We hypothesized that adding derived features as input to the models will probably result in better predictions, as shown by other research groups when identifying mood states in BD using the same wristband device [51]. Therefore, we are currently exploring derived features from raw data (ie, statistical, time-domain, and frequency-domain features) [53], assessing EDA reactivity by extracting information on the tonic and phasic components of skin conductance using novel automated methods [18,53,95], and performing stress elicitation to assess potential alterations (hyporeactivity) in the phasic component of EDA during mood episodes [26]. Finally, considering the sleep and circadian rhythm disturbances in mood disorders in both euthymia [19,96] and acute phases [97-99], we are exploring automated methods to separate sleep from wake times [87,100,101]. Our goal is to evaluate sleep disturbances and differences in physiological signals during sleep and wake periods during mood episodes [77].

Limitations

We acknowledge several limitations in this study. First, the limited sample size for model development does not allow us to make strong claims about generalization performance [102]. However, most recordings were longer than 40 hours and each patient on an acute mood episode was recorded longitudinally at 3 time points (acute, response, and remission). In fact, our data set in terms of recording hours is well above other data sets modeled with deep learning in health care settings: the deep convolutional approach proposed by Musallam et al [103] was applied to 60 hours of electroencephalogram recordings [104]. In addition, the wearable device used (E4), allows fine-grained collection of digital physiological data (from 1 Hz to 64 Hz) for precision longitudinal time-series analyses. Regarding sample size in terms of the number of subjects, previous endeavors used as few as 12 subjects [46]. Unfortunately, this type of data, that is, recorded with a research-grade wearable device on a population with a psychiatric condition (arguably interfering with compliance to instructions), is expensive and time-consuming to collect. Second, potential confounding variables such as sex, age, pharmacological treatments, exercise, or BMI were not controlled for, and some of the study sample was not matched by age and sex. This may have biased the results, as those variables have been found to affect motor activity data, especially in between-group comparisons [60]. The within-subject design allows partial mitigation of both the weakness of a small sample size and the influence of confounders, so the models can capture mood-related signals. Therefore, we performed intra-individual comparisons across consecutive time points. In fact, the generalization of intra-individual models obtained substantially better accuracies, showing glimpses of capturing the severity of manic and depressive psychopathology.

Future works will further explore the capabilities of advanced automated machine learning models for identifying affective illness activity and the role of confounders in this association. Of particular interest are the application of clustering algorithms [87], exploring derived features (HRV [94] and EDA reactivity [26]), the role of wake and sleep periods [77,105], and the potential of physiological data to predict treatment responses and detect prodromal signs of mood episodes [106]. Future projects will include (1) studying the role of psychotic symptoms in patients with affective disorders, as well as in patients with schizophrenia; (2) assessing the role of smartphone-based derived data, including ecologic momentary assessments and passive data [107-109], in patients with BD using the SIMPLe smartphone app [110,111]; and (3) investigating the potential of combining physiological wearable data with peripheral biomarkers [112,113] and speech features [114-118].

Conclusions

Physiological wearable data may have the potential to identify and predict the severity of mania and depression in mood disorders as well as specific symptoms quantitatively. Motor activity appears to be the most relevant digital biomarker for predicting mania, whereas stress-related digital biomarkers (EDA and HR) appear to be the most relevant for predicting both bipolar and unipolar depression. In the context of biomarkers in mood disorders, these findings represent a promising pathway toward personalized psychiatry, in which clinical decisions and treatments could be supported by passive continuous and objective digital data.

Acknowledgments

The authors acknowledge the contribution of all the participants of the study.

GA is supported by a Rio Hortega 2021 grant (CM21/00017) from the Spanish Ministry of Health financed by the Instituto de Salud Carlos III (ISCIII) and cofinanced by Fondo Social Europeo Plus (FSE+). FC and BML are supported by the United Kingdom Research and Innovation (grant EP/S02431X/1), UK Research and Innovation (UKRI) Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. A Mas is supported by an Agència de Gestió d'Ajudes Universitàries i de Investigació (AGAUR)—PANDÈMIES 2020 grant (PI047003) from the Generalitat de Catalunya. MS is supported by a grant from the Baszucki Brain Research Fund. IG thanks the support of the Spanish Ministry of Science and Innovation (PI19/00954) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación y el Fondos Europeos de la Unión Europea (FEDER, FSE, Next Generation EU/Plan de Recuperación Transformación y Resiliencia_PRTR); the ISCIII; the CIBER of Mental Health (CIBERSAM); and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), Centres de Recerca de Catalunya (CERCA) Programme or Generalitat de Catalunya as well as the Fundació Clínic per la Recerca Biomèdica (Pons Bartran 2022-FRCB_PB1_2022). AG-P is supported by a Rio Hortega 2021 grant (CM21/00094) from the Spanish Ministry of Health financed by ISCIII and cofinanced by Fondo Social Europeo Plus (FSE+). MB thanks the Spanish Ministry of Health and ISCIII (PI20/01066). NV thanks the Biomedicine International Training Research Programme for Excellent Clinician-Scientists (BITRECS) project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754550 and from "La Caixa" Foundation (ID 100010434), under the agreement LCF/PR/GN18/50310006. SM is supported by the grant "Contracte de Recerca Emili Letang-Josep Font" provided by Hospital Clínic de Barcelona. A Murru thanks the support of the Spanish Ministry of Science and Innovation (PI19/00672) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER). SA has been supported by a Sara Borrell contract (CD20/00177), funded by ISCIII and cofunded by the European Social Fund "Investing in your future." AM-A thanks the support of the Spanish Ministry of Science and Innovation (PI18/00789, PI21/00787) integrated into the Plan Nacional de I+D+I and cofinanced by ISCIII-Subdirección General de Evaluación and the FEDER; the ISCIII; the CIBERSAM; the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), the CERCA Programme, and the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00177. GF is supported by a fellowship from "La Caixa" Foundation (ID 100010434)—fellowship code—LCF/BQ/DR21/11880019. JR is supported by a Miguel Servet II contract (CPII19/00009), funded by ISCIII and cofunded by the European Social Fund "Investing in your future." EV thanks the support of the Spanish Ministry of Science, Innovation and Universities (PI15/00283, PI18/00805, PI19/00394, CPII19/00009) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación and the FEDER; the ISCIII; the CIBERSAM; the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), and the CERCA Programme or Generalitat de Catalunya. The authors would like to thank the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357. DH-M is supported by a Juan Rodés JR18/00021 granted by the ISCIII.

This project was funded by the ISCIII (FIS PI21/00340, TIMEBASE Study), cofunded by the Euopean Union, as well as a Baszucki Brain Research Fund grant (PI046998) from the Milken Foundation. The ISCIII or the Milken Foundation had no further role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Data Availability

The data supporting the findings of this study are available upon request from the corresponding author.

Authors' Contributions

GA and DH-M were responsible for study planning, project conception, and coordination. A Mas, MS, IP, MV, IG, A Benabarre, AG-P, MG, IA, A Bastidas, MC, TF-P, NA, MB, CG-R, NV, SM, SA, AM-A, and VR were responsible for recruitment. FC, BML, AV, MDP, VO, AS, and JR were responsible for data analysis. GA, FC, BML, and DH-M were responsible for manuscript preparation. All authors revised the final manuscript.

Conflicts of Interest

GA has received continuing medical education (CME)–related honoraria or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck and Otsuka, and Angelini. IP has received CME-related honoraria, or consulting fees from ADAMED, Janssen-Cilag, and Lundbeck. IG has received grants and served as consultant, advisor or CME speaker for the following identities: Angelini, Casen Recordati, Ferrer, Janssen Cilag, and Lundbeck, Lundbeck-Otsuka, Luye, SEI Healthcare. AG-P has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Casen Recordati and Angelini. MC has received grants and served as consultant, advisor or CME speaker for the following entities: Lundbeck, Esteve, Pfizer. NA has received CME-related financing from Janssen-Cilag, Lundbeck, Adamed, Pfizer, Angelini and Boston Scientific. MB has been a consultant for, received grant/research support and honoraria from, and been on the speakers/advisory board of has received honoraria from talks and/or consultancy of Adamed, Angelini, Casen-Recordati, Exeltis, Ferrer, Janssen, Lundbeck, Neuraxpharm, Otsuka, Pfizer and Sanofi. NV has received financial support for CME activities and travel funds from the following entities: Angelini, Janssen-Cilag, Lundbeck, Otsuka. SM has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck/Otsuka, and Angelini. A Murru has received grants and served as consultant, advisor or CME speaker for the following entities: Angelini, Idorsia, Lundbeck, Pfizer, Takeda. LS has received CME-related honoraria, or consulting fees from Boehringer -Ingelheim, Janssen, Lundbeck/Otsuka, Sanofi-Aventis. AHY has received honoraria for lectures and advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders. EV has received research support from or served as consultant, adviser or speaker for AB-Biotics, Abbott, Abbvie, Adamed, Angelini, Biogen, Celon, Dainippon Sumitomo Pharma, Ferrer, Gedeon Richter, GH Research, Glaxo SmithKline, Janssen, Lundbeck, Organon, Otsuka, Rovi, Sage pharmaceuticals, Sanofi-Aventis, Shire, Sunovion, Takeda, and Viatris. DH-M has received CME-related honoraria and served as consultant for Abbott, Angelini, Ethypharm Digital Therapy and Janssen-Cilag. All authors report no financial or other relationship relevant to the subject of this article.

Multimedia Appendix 1
Empatica E4.
[PNG File , 667 KB - mhealth_v11i1e45405_app1.png]

Multimedia Appendix 2
Validation set performance (accuracy) as a function of time alignment (Hz) and window length (w).
[PNG File , 150 KB - mhealth_v11i1e45405_app2.png]

Multimedia Appendix 3
Confusion matrix for the prediction of the severity of an acute affective episode: models and generalization. BD: bipolar disorder; MDD: major depressive disorder.
[PNG File , 296 KB - mhealth_v11i1e45405_app3.png]

Multimedia Appendix 4
Confusion matrix for the prediction of the polarity of affective episodes, euthymia, and healthy controls: models and generalization. BD: bipolar disorder; HC: healthy controls; MDD: major depressive disorder; T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes.
[PNG File , 341 KB - mhealth_v11i1e45405_app4.png]

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
2. Vieta E, Berk M, Schulze TG, Carvalho AF, Suppes T, Calabrese JR, et al. Bipolar disorders. Nat Rev Dis Primers 2018 Mar 08;4:18008. [doi: 10.1038/nrdp.2018.8] [Medline: 29516993]
3. Carvalho AF, Firth J, Vieta E. Bipolar disorder. N Engl J Med 2020 Jul 02;383(1):58-66. [doi: 10.1056/NEJMra1906193] [Medline: 32609982]
4. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. Nat Rev Dis Primers 2016 Sep 15;2:16065. [doi: 10.1038/nrdp.2016.65] [Medline: 27629598]
5. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
6. Hidalgo-Mazzei D, Young AH, Vieta E, Colom F. Behavioural biomarkers and mobile mental health: a new paradigm. Int J Bipolar Disord 2018 May 06;6(1):9 [FREE Full text] [doi: 10.1186/s40345-018-0119-7] [Medline: 29730832]
7. Sheikh M, Qassem M, Kyriacou PA. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. Front Digit Health 2021 Apr 07;3:662811 [FREE Full text] [doi: 10.3389/fdgth.2021.662811] [Medline: 34713137]
8. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) resource. Food and Drug Administration, National Institutes of Health. Silver Spring, MD, USA: Food and Drug Administration; 2016. URL: https://www.fdanews.com/ext/resources/files/2020/11-24-20-BEST.pdf?1606261388 [accessed 2023-02-15]
9. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, Zrubka Z. Digital biomarker-based studies: scoping review of systematic reviews. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35722 [FREE Full text] [doi: 10.2196/35722] [Medline: 36279171]
10. Babrak LM, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and digital biomarkers: two worlds apart? Digit Biomark 2019 Aug 16;3(2):92-102 [FREE Full text] [doi: 10.1159/000502000] [Medline: 32095769]
11. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]

12. Salagre E, Vieta E. Precision psychiatry: complex problems require complex solutions. Eur Neuropsychopharmacol 2021 Nov;52:94-95. [doi: 10.1016/j.euroneuro.2021.07.003] [Medline: 34325190]
13. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Abid MM, Péntek M, Gulácsi L, et al. Digital biomarker-based interventions: systematic review of systematic reviews. J Med Internet Res 2022 Dec 21;24(12):e41042 [FREE Full text] [doi: 10.2196/41042] [Medline: 36542427]
14. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. J Med Internet Res 2023 Jan 19;25:e42672 [FREE Full text] [doi: 10.2196/42672] [Medline: 36656625]
15. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
16. de Angel V, Adeleye F, Zhang Y, Cummins N, Munir S, Lewis S, et al. The feasibility of implementing remote measurement technologies in psychological treatment for depression: mixed methods study on engagement. JMIR Ment Health 2023 Jan 24;10:e42866 [FREE Full text] [doi: 10.2196/42866] [Medline: 36692937]
17. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
18. de Looff P, Duursma R, Noordzij M, Taylor S, Jaques N, Scheepers F, et al. Wearables: an R package with accompanying shiny application for signal analysis of a wearable device targeted at clinicians and researchers. Front Behav Neurosci 2022 Jun 23;16:856544 [FREE Full text] [doi: 10.3389/fnbeh.2022.856544] [Medline: 35813597]
19. Geoffroy PA, Scott J, Boudebesse C, Lajnef M, Henry C, Leboyer M, et al. Sleep in patients with remitted bipolar disorders: a meta-analysis of actigraphy studies. Acta Psychiatr Scand 2015 Feb;131(2):89-99. [doi: 10.1111/acps.12367] [Medline: 25430914]
20. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e24872 [FREE Full text] [doi: 10.2196/24872] [Medline: 34694233]
21. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019 Feb 01;2:3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
22. Faurholt-Jepsen M, Kessing LV, Munkholm K. Heart rate variability in bipolar disorder: a systematic review and meta-analysis. Neurosci Biobehav Rev 2017 Feb;73:68-80. [doi: 10.1016/j.neubiorev.2016.12.007] [Medline: 27986468]
23. Faurholt-Jepsen M, Brage S, Kessing LV, Munkholm K. State-related differences in heart rate variability in bipolar disorder. J Psychiatr Res 2017 Jan;84:169-173 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.10.005] [Medline: 27743529]
24. Iacono WG, Lykken DT, Peloquin LJ, Lumry AE, Valentine RH, Tuason VB. Electrodermal activity in euthymic unipolar and bipolar affective disorders. A possible marker for depression. Arch Gen Psychiatry 1983 May;40(5):557-565. [doi: 10.1001/archpsyc.1983.01790050083010] [Medline: 6838333]
25. Sarchiapone M, Gramaglia C, Iosue M, Carli V, Mandelli L, Serretti A, et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: a systematic review and narrative synthesis. BMC Psychiatry 2018 Jan 25;18(1):22 [FREE Full text] [doi: 10.1186/s12888-017-1551-4] [Medline: 29370787]
26. Greco A, Valenza G, Lanata A, Rota G, Scilingo EP. Electrodermal activity in bipolar patients during affective elicitation. IEEE J Biomed Health Inform 2014 Nov;18(6):1865-1873. [doi: 10.1109/JBHI.2014.2300940] [Medline: 25375684]
27. Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. JMIR Mhealth Uhealth 2021 Mar 08;9(3):e24365 [FREE Full text] [doi: 10.2196/24365] [Medline: 33683207]
28. Tazawa Y, Liang KC, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Feb;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
29. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022 Jun 09;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
30. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in Health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
31. Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, et al. The International Society for Bipolar Disorders (ISBD) task force report on the nomenclature of course and outcome in bipolar disorders. Bipolar Disord 2009 Aug;11(5):453-473. [doi: 10.1111/j.1399-5618.2009.00726.x] [Medline: 19624385]
32. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; May 18, 2013.
33. First MB, Williams JB, Karg RS, Spitzer RL. Structured clinical interview for DSM-5, research version. American Psychiatric Association. 2015 Nov 05. URL: https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5 [accessed 2023-02-15]

34. Colom F, Vieta E, Martínez-Arán A, Garcia-Garcia M, Reinares M, Torrent C, et al. [Spanish version of a scale for the assessment of mania: validity and reliability of the Young Mania Rating Scale]. Med Clin (Barc) 2002 Sep 28;119(10):366-371. [doi: 10.1016/s0025-7753(02)73419-2] [Medline: 12372167]
35. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
36. Ramos-Brieva JA, Cordero-Villafafila A. A new validation of the Hamilton Rating Scale for Depression. J Psychiatr Res 1988;22(1):21-28. [doi: 10.1016/0022-3956(88)90024-6] [Medline: 3397906]
37. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23(1):56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
38. Kleckner IR, Feldman MJ, Goodwin MS, Quigley KS. Framework for selecting and benchmarking mobile devices in psychophysiological research. Behav Res Methods 2021 Apr;53(2):518-535 [FREE Full text] [doi: 10.3758/s13428-020-01438-9] [Medline: 32748241]
39. Empatica E4. Empatica. URL: https://www.empatica.com/en-gb/research/e4/ [accessed 2023-02-15]
40. Schuurmans AA, de Looff P, Nijhof KS, Rosada C, Scholte RH, Popma A, et al. Validity of the Empatica E4 wristband to measure Heart Rate Variability (HRV) parameters: a comparison to Electrocardiography (ECG). J Med Syst 2020 Sep 23;44(11):190 [FREE Full text] [doi: 10.1007/s10916-020-01648-w] [Medline: 32965570]
41. Menghini L, Gianfranchi E, Cellini N, Patron E, Tagliabue M, Sarlo M. Stressing the accuracy: wrist-worn wearable sensor validation over different conditions. Psychophysiology 2019 Nov;56(11):e13441. [doi: 10.1111/psyp.13441] [Medline: 31332802]
42. Li Q, Li Q, Cakmak AS, Da Poian G, Bliwise DL, Vaccarino V, et al. Transfer learning from ECG to PPG for improved sleep staging from wrist-worn wearables. Physiol Meas 2021 May 13;42(4):1088/1361-6579/abf1b0 [FREE Full text] [doi: 10.1088/1361-6579/abf1b0] [Medline: 33761477]
43. Alinia P, Sah RK, McDonell M, Pendry P, Parent S, Ghasemzadeh H, et al. Associations between physiological signals captured using wearable sensors and self-reported outcomes among adults in alcohol use disorder recovery: development and usability study. JMIR Form Res 2021 Jul 21;5(7):e27891 [FREE Full text] [doi: 10.2196/27891] [Medline: 34287205]
44. Ollander S, Godin C, Campagne A, Charbonnier S. A comparison of wearable and stationary sensors for stress detection. In: Proceedings of the 2016 International Conference on Systems, Man, and Cybernetics. 2016 Presented at: SMC '16; October 9-12, 2017; Budapest, Hungary p. 4362-4366 URL: https://ieeexplore.ieee.org/abstract/document/7844917 [doi: 10.1109/smc.2016.7844917]
45. Romine W, Schroeder N, Banerjee T, Graft J. Toward mental effort measurement using electrodermal activity features. Sensors (Basel) 2022 Sep 28;22(19):7363 [FREE Full text] [doi: 10.3390/s22197363] [Medline: 36236461]
46. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In: proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction. 2017 Presented at: ACII '17; October 23-26, 2017; San Antonio, TX, USA p. 325-332 URL: https://www.computer.org/csdl/proceedings/acii/2017/12OmNCcbEdf [doi: 10.1109/acii.2017.8273620]
47. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, et al. Monitoring changes in depression severity using wearable and mobile sensors. Front Psychiatry 2020 Dec 18;11:584711 [FREE Full text] [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
48. Choi J, Lee S, Kim S, Kim D, Kim H. Depressed mood prediction of elderly people with a wearable band. Sensors (Basel) 2022 May 31;22(11):4174 [FREE Full text] [doi: 10.3390/s22114174] [Medline: 35684797]
49. Shaukat-Jali R, van Zalk N, Boyle DE. Detecting subclinical social anxiety using physiological data from a wrist-worn wearable: small-scale feasibility study. JMIR Form Res 2021 Oct 07;5(10):e32656 [FREE Full text] [doi: 10.2196/32656] [Medline: 34617905]
50. Jakobsen P, Stautland A, Riegler MA, Côté-Allard U, Sepasdar Z, Nordgreen T, et al. Complexity and variability analyses of motor activity distinguish mood states in bipolar disorder. PLoS One 2022 Jan 21;17(1):e0262232 [FREE Full text] [doi: 10.1371/journal.pone.0262232] [Medline: 35061801]
51. Côté-Allard U, Jakobsen P, Stautland A, Nordgreen T, Fasmer OB, Oedegaard KJ, et al. Long–short ensemble network for bipolar manic-euthymic state recognition based on wrist-worn sensors. IEEE Pervasive Comput 2022 Apr 1;21(2):20-31 [FREE Full text] [doi: 10.1109/mprv.2022.3155728]
52. Kleckner IR, Jones RM, Wilder-Smith O, Wormwood JB, Akcakaya M, Quigley KS, et al. Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. IEEE Trans Biomed Eng 2018 Jul;65(7):1460-1467 [FREE Full text] [doi: 10.1109/TBME.2017.2758643] [Medline: 28976309]
53. Föll S, Maritsch M, Spinola F, Mishra V, Barata F, Kowatsch T, et al. FLIRT: a feature generation toolkit for wearable data. Comput Methods Programs Biomed 2021 Nov;212:106461 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106461] [Medline: 34736174]
54. How is IBI.csv obtained? Empatica E4. 2021 Jun 17. URL: https://support.empatica.com/hc/en-us/articles/201912319-How-is-IBI-csv-obtained [accessed 2022-07-05]
55. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997 Nov;45(11):2673-2681 [FREE Full text] [doi: 10.1109/78.650093]

56. Vaswani A, Shazeer N, Parmar N, Erban R, Uszkoreit J, Jones L, et al. Attention is all you need. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems. 2017 Presented at: NeurIPS '17; December 4-9, 2017; Long Beach, CA, USA.
57. Kingma DP, Ba LJ. Adam: a method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations. 2015 May Presented at: ICLR' 15; May 7-9, 2015; San Diego, CA, USA URL: https://dare.uva.nl/search?identifier=a20791d3-1aff-464a-8544-268383c33a75 [doi: 10.5860/choice.189890]
58. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov RR. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014 Jan 01;15(1):1929-1958 [FREE Full text]
59. Anmella G, Corponi F, Li BM, Mas A, Sanabra M, Pacchiarotti I, et al. INTREPIBD/JMIR2023: Code for JMIR mHealth and uHealth paper "Exploring digital biomarkers of illness activity in mood episodes: hypotheses generating and model development study". GitHub. 2023. URL: https://github.com/INTREPIBD/JMIR2023 [accessed 2023-04-19]
60. Jakobsen P, Garcia-Ceja E, Riegler M, Stabell LA, Nordgreen T, Torresen J, et al. Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. PLoS One 2020 Aug 24;15(8):e0231995 [FREE Full text] [doi: 10.1371/journal.pone.0231995] [Medline: 32833958]
61. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A 2018 Mar 13;115(11):2600-2606 [FREE Full text] [doi: 10.1073/pnas.1708274114] [Medline: 29531091]
62. Saha S, Baumert M. Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. Front Comput Neurosci 2019 Jan 21;13:87 [FREE Full text] [doi: 10.3389/fncom.2019.00087] [Medline: 32038208]
63. Böttcher S, Bruno E, Epitashvili N, Dümpelmann M, Zabler N, Glasstetter M, et al. Intra- and inter-subject perspectives on the detection of focal onset motor seizures in epilepsy patients. Sensors (Basel) 2022 Apr 26;22(9):3318 [FREE Full text] [doi: 10.3390/s22093318] [Medline: 35591007]
64. Özdenizci O, Wang YE, Koike-Akino T, ErdoĞmuŞ D. Learning invariant representations from EEG via adversarial inference. IEEE Access 2020;8:27074-27085 [FREE Full text] [doi: 10.1109/access.2020.2971600] [Medline: 33747669]
65. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: a systematic review. J Neural Eng 2019 Aug 14;16(5):051001. [doi: 10.1088/1741-2552/ab260c] [Medline: 31151119]
66. Page A, Shea C, Mohsenin T. Wearable seizure detection using convolutional neural networks with transfer learning. In: Proceedings of the 2016 International Symposium on Circuits and Systems. 2016 Presented at: ISCAS '16; May 22-25, 2016; Montreal, Canada p. 1086-1089 URL: https://ieeexplore.ieee.org/abstract/document/7527433 [doi: 10.1109/iscas.2016.7527433]
67. Corponi F, Anmella G, Verdolini N, Pacchiarotti I, Samalin L, Popovic D, et al. Symptom networks in acute depression across bipolar and major depressive disorders: a network analysis on a large, international, observational study. Eur Neuropsychopharmacol 2020 Jun;35:49-60. [doi: 10.1016/j.euroneuro.2020.03.017] [Medline: 32409261]
68. Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiatr Scand 2011 Dec;124(6):495-496. [doi: 10.1111/j.1600-0447.2011.01744.x] [Medline: 21838736]
69. Andrade-González N, Álvarez-Cadenas L, Saiz-Ruiz J, Lahera G. Initial and relapse prodromes in adult patients with episodes of bipolar disorder: a systematic review. Eur Psychiatry 2020 Feb 12;63(1):e12 [FREE Full text] [doi: 10.1192/j.eurpsy.2019.18] [Medline: 32093795]
70. Solomon DA, Leon AC, Coryell WH, Endicott J, Li C, Fiedorowicz JG, et al. Longitudinal course of bipolar I disorder: duration of mood episodes. Arch Gen Psychiatry 2010 Apr;67(4):339-347 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.15] [Medline: 20368510]
71. Mignogna KM, Goes FS. Characterizing the longitudinal course of symptoms and functioning in bipolar disorder. Psychol Med 2022 Jun 14:1-11. [doi: 10.1017/S0033291722001489] [Medline: 35699120]
72. Nandi A, Beard JR, Galea S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. BMC Psychiatry 2009 Jun 01;9:31 [FREE Full text] [doi: 10.1186/1471-244X-9-31] [Medline: 19486530]
73. Bin Heyat MB, Akhtar F, Abbas SJ, Al-Sarem M, Alqarafi A, Stalin A, et al. Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. Biosensors (Basel) 2022 Jun 17;12(6):427 [FREE Full text] [doi: 10.3390/bios12060427] [Medline: 35735574]
74. Krane-Gartiser K, Henriksen TE, Morken G, Vaaler A, Fasmer OB. Actigraphic assessment of motor activity in acutely admitted inpatients with bipolar disorder. PLoS One 2014 Feb 20;9(2):e89574 [FREE Full text] [doi: 10.1371/journal.pone.0089574] [Medline: 24586883]
75. Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, et al. Activation in bipolar disorders: a systematic review. JAMA Psychiatry 2017 Feb 01;74(2):189-196. [doi: 10.1001/jamapsychiatry.2016.3459] [Medline: 28002572]
76. Martino DJ, Valerio MP, Parker G. The structure of mania: an overview of factorial analysis studies. Eur Psychiatry 2020 Feb 10;63(1):e10 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.18] [Medline: 32093802]
77. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with Bipolar Disorder. JAMA Psychiatry 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]

78. Shou H, Cui L, Hickie I, Lameira D, Lamers F, Zhang J, et al. Dysregulation of objectively assessed 24-hour motor activity patterns as a potential marker for bipolar I disorder: results of a community-based family study. Transl Psychiatry 2017 Aug 22;7(8):e1211 [FREE Full text] [doi: 10.1038/tp.2017.136] [Medline: 28892068]
79. Kircanski K, Williams LM, Gotlib IH. Heart rate variability as a biomarker of anxious depression response to antidepressant medication. Depress Anxiety 2019 Jan;36(1):63-71 [FREE Full text] [doi: 10.1002/da.22843] [Medline: 30311742]
80. Schiweck C, Piette D, Berckmans D, Claes S, Vrieze E. Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. Psychol Med 2019 Jan;49(2):200-211. [doi: 10.1017/S0033291718001988] [Medline: 30134999]
81. Kim AY, Jang EH, Kim S, Choi KW, Jeon HJ, Yu HY, et al. Automatic detection of major depressive disorder using electrodermal activity. Sci Rep 2018 Nov 19;8(1):17030 [FREE Full text] [doi: 10.1038/s41598-018-35147-3] [Medline: 30451895]
82. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review. Int J Med Inform 2023 May;173:105026 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105026] [Medline: 36893657]
83. Stone LB, McCormack CC, Bylsma LM. Cross system autonomic balance and regulation: associations with depression and anxiety symptoms. Psychophysiology 2020 Oct;57(10):e13636 [FREE Full text] [doi: 10.1111/psyp.13636] [Medline: 33460174]
84. Corponi F, Anmella G, Pacchiarotti I, Samalin L, Verdolini N, Popovic D, et al. Deconstructing major depressive episodes across unipolar and bipolar depression by severity and duration: a cross-diagnostic cluster analysis on a large, international, observational study. Transl Psychiatry 2020 Jul 19;10(1):241 [FREE Full text] [doi: 10.1038/s41398-020-00922-2] [Medline: 32684621]
85. Aljalbout E, Golkov V, Siddiqui Y, Strobel M, Cremers D. Clustering with deep learning: taxonomy and new methods. arXiv 2018 Jan 23 [FREE Full text] [doi: 10.48550/arxiv.1801.07648]
86. Ienco D, Interdonato R. Deep multivariate time series embedding clustering via attentive-gated autoencoder. In: Proceedings of the 24th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. 2020 May Presented at: PAKDD' 2020; May 11–14, 2020; Singapore p. 318-329 URL: https://link.springer.com/chapter/10.1007/978-3-030-47426-3_25#citeas [doi: 10.1007/978-3-030-47426-3_25]
87. Park S, Lee SW, Han S, Cha M. Clustering insomnia patterns by data from wearable devices: algorithm development and validation study. JMIR Mhealth Uhealth 2019 Dec 05;7(12):e14473 [FREE Full text] [doi: 10.2196/14473] [Medline: 31804187]
88. Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 2022 Jun 03;376(6597):1070-1074. [doi: 10.1126/science.abm2461] [Medline: 35653486]
89. Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JH, Olivier B, et al. The effect of stress on core and peripheral body temperature in humans. Stress 2013 Sep;16(5):520-530. [doi: 10.3109/10253890.2013.807243] [Medline: 23790072]
90. Herborn KA, Graves JL, Jerem P, Evans NP, Nager R, McCafferty DJ, et al. Skin temperature reveals the intensity of acute stress. Physiol Behav 2015 Dec 01;152(Pt A):225-230 [FREE Full text] [doi: 10.1016/j.physbeh.2015.09.032] [Medline: 26434785]
91. Klainin-Yobas P, Ignacio J, He HG, Lau Y, Ngooi BX, Koh SQ. Effects of a stress-management program for inpatients with mental disorders: a feasibility study. Biol Res Nurs 2016 Mar;18(2):213-220. [doi: 10.1177/1099800415595877] [Medline: 26183182]
92. Serrano-Serrano AB, Marquez-Arrico JE, Navarro JF, Martinez-Nicolas A, Adan A. Circadian characteristics in patients under treatment for substance use disorders and severe mental illness (schizophrenia, major depression and bipolar disorder). J Clin Med 2021 Sep 25;10(19):4388 [FREE Full text] [doi: 10.3390/jcm10194388] [Medline: 34640406]
93. Murphy PJ, Frei MG, Papolos D. Alterations in skin temperature and sleep in the fear of harm phenotype of pediatric bipolar disorder. J Clin Med 2014;3(3):959-971 [FREE Full text] [doi: 10.3390/jcm3030959] [Medline: 25530872]
94. Stautland A, Jakobsen P, Fasmer OB, Osnes B, Torresen J, Nordgreen T, et al. Heart rate variability as biomarker for bipolar disorder. medRxiv 2022 Feb 15 (forthcoming). [doi: 10.1101/2022.02.14.22269413]
95. Hernando-Gallego F, Luengo D, Artes-Rodriguez A. Feature extraction of galvanic skin responses by nonnegative sparse deconvolution. IEEE J Biomed Health Inform 2018 Sep;22(5):1385-1394. [doi: 10.1109/JBHI.2017.2780252] [Medline: 29990244]
96. Meyer N, Faulkner SM, McCutcheon RA, Pillinger T, Dijk DJ, MacCabe JH. Sleep and circadian rhythm disturbance in remitted schizophrenia and bipolar disorder: a systematic review and meta-analysis. Schizophr Bull 2020 Mar 10;46(5):1126-1143 [FREE Full text] [doi: 10.1093/schbul/sbaa024] [Medline: 32154882]
97. Lewis KS, Gordon-Smith K, Forty L, Di Florio A, Craddock N, Jones L, et al. Sleep loss as a trigger of mood episodes in bipolar disorder: individual differences based on diagnostic subtype and gender. Br J Psychiatry 2017 Sep;211(3):169-174 [FREE Full text] [doi: 10.1192/bjp.bp.117.202259] [Medline: 28684405]
98. Murru A, Guiso G, Barbuti M, Anmella G, Verdolini N, Samalin L, BRIDGE-II-Mix Study Group. The implications of hypersomnia in the context of major depression: results from a large, international, observational study. Eur

Neuropsychopharmacol 2019 Apr;29(4):471-481 [FREE Full text] [doi: 10.1016/j.euroneuro.2019.02.011] [Medline: 30846287]
99. Steinan MK, Scott J, Lagerberg TV, Melle I, Andreassen OA, Vaaler AE, et al. Sleep problems in bipolar disorders: more than just insomnia. Acta Psychiatr Scand 2016 May;133(5):368-377 [FREE Full text] [doi: 10.1111/acps.12523] [Medline: 26590799]
100. Liu J, Zhao Y, Lai B, Wang H, Tsui KL. Wearable device heart rate and activity data in an unsupervised approach to personalized sleep monitoring: algorithm validation. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e18370 [FREE Full text] [doi: 10.2196/18370] [Medline: 32755887]
101. Wei J, Boger J. Sleep detection for younger adults, healthy older adults, and older adults living with dementia using wrist temperature and actigraphy: prototype testing and case study analysis. JMIR Mhealth Uhealth 2021 Jun 01;9(6):e26462 [FREE Full text] [doi: 10.2196/26462] [Medline: 34061038]
102. Vos G, Trinh K, Sarnyai Z, Azghadi M. Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices. arXiv 2022 Sep 30. [doi: 10.48550/arxiv.2209.15146]
103. Musallam YK, AlFassam NI, Muhammad G, Amin SU, Alsulaiman M, Abdul W, et al. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. Biomed Signal Process Control 2021 Aug;69:102826 [FREE Full text] [doi: 10.1016/j.bspc.2021.102826]
104. Kaya M, Binli MK, Ozbay E, Yanar H, Mishchenko Y. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. Sci Data 2018 Oct 16;5:180211 [FREE Full text] [doi: 10.1038/sdata.2018.211] [Medline: 30325349]
105. Zhang Y, Folarin AA, Sun S, Cummins N, Bendayan R, Ranjan Y, RADAR-CNS Consortium. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 2021 Apr 12;9(4):e24604 [FREE Full text] [doi: 10.2196/24604] [Medline: 33843591]
106. Van Assche E, Antoni Ramos-Quiroga J, Pariante CM, Sforzini L, Young AH, Flossbach Y, et al. Digital tools for the assessment of pharmacological treatment for depressive disorder: state of the art. Eur Neuropsychopharmacol 2022 Jul;60:100-116. [doi: 10.1016/j.euroneuro.2022.05.007] [Medline: 35671641]
107. Dunster GP, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. Neuropsychopharmacology 2021 Jan;46(1):197-208 [FREE Full text] [doi: 10.1038/s41386-020-00830-5] [Medline: 32919408]
108. Gillett G, Saunders KE. Remote monitoring for understanding mechanisms and prediction in psychiatry. Curr Behav Neurosci Rep 2019 May 2;6(2):51-56 [FREE Full text] [doi: 10.1007/s40473-019-00176-3]
109. Kessing LV, Faurholt-Jepsen M. Mood instability - a new outcome measure in randomised trials of bipolar disorder? Eur Neuropsychopharmacol 2022 May;58:39-41. [doi: 10.1016/j.euroneuro.2022.02.005] [Medline: 35219178]
110. García-Estela A, Cantillo J, Angarita-Osorio N, Mur-Milà E, Anmella G, Pérez V, et al. Real-world implementation of a smartphone-based psychoeducation program for bipolar disorder: observational ecological study. J Med Internet Res 2022 Feb 02;24(2):e31565 [FREE Full text] [doi: 10.2196/31565] [Medline: 35107440]
111. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]
112. Stanislaus S, Faurholt-Jepsen M, Vinberg M, Poulsen HE, Kessing LV, Coello K. Associations between oxidative stress markers and patient-reported smartphone-based symptoms in patients newly diagnosed with bipolar disorder: an exploratory study. Eur Neuropsychopharmacol 2022 Sep;62:36-45. [doi: 10.1016/j.euroneuro.2022.07.002] [Medline: 35896055]
113. Anmella G, Sanabra M, Mas-Musons A, Hidalgo-Mazzei D. Combining digital with peripheral biomarkers in bipolar disorder. Eur Neuropsychopharmacol 2022 Oct;63:71-72. [doi: 10.1016/j.euroneuro.2022.07.187] [Medline: 36081269]
114. Zhang J, Pan Z, Gui C, Xue T, Lin Y, Zhu J, et al. Analysis on speech signal features of manic patients. J Psychiatr Res 2018 Mar;98:59-63. [doi: 10.1016/j.jpsychires.2017.12.012] [Medline: 29291581]
115. Guidi A, Schoentgen J, Bertschy G, Gentili C, Landini L, Scilingo EP, et al. Voice quality in patients suffering from bipolar disease. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:6106-6109. [doi: 10.1109/EMBC.2015.7319785] [Medline: 26737685]
116. Weiner L, Doignon-Camus N, Bertschy G, Giersch A. Thought and language disturbance in bipolar disorder quantified via process-oriented verbal fluency measures. Sci Rep 2019 Oct 03;9(1):14282 [FREE Full text] [doi: 10.1038/s41598-019-50818-5] [Medline: 31582814]
117. Carrillo F, Mota N, Copelli M, Ribeiro S, Sigman M, Cecchi G, et al. Emotional intensity analysis in bipolar subjects. arXiv 2016 Jun 07 [FREE Full text]
118. Carrillo F, Sigman M, Fernández Slezak DF, Ashton P, Fitzgerald L, Stroud J, et al. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. J Affect Disord 2018 Apr 01;230:84-86. [doi: 10.1016/j.jad.2018.01.006] [Medline: 29407543]

Abbreviations

ACC: acceleration
AUROC: area under the receiver operating characteristic
BD: bipolar disorder
BiLSTM: Bidirectional Long Short-Term Memory
BVP: blood volume pulse
DSM-5: Diagnostic and Statistical Manual of Mental Disorders–5
EDA: electrodermal activity
HC: healthy control
HDRS: Hamilton Depression Rating Scale
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
LSTM: long short-term memory
MDD: major depressive disorder
NMI: normalized mutual information
RNN: recurrent neural network
ROC: receiver operating characteristic
T0: current acute Diagnostic and Statistical Manual of Mental Disorders–5 affective episodes
T1: symptoms' response
T2: symptomatic remission
TEMP: temperature
YMRS: Young Mania Rating Scale

Edited by L Buis; submitted 29.12.22; peer-reviewed by OR Patil, Y Zhang, V Gupta, S Tedesco; comments to author 25.01.23; revised version received 20.02.23; accepted 07.03.23; published 04.05.23.

Please cite as:
Anmella G, Corponi F, Li BM, Mas A, Sanabra M, Pacchiarotti I, Valentí M, Grande I, Benabarre A, Giménez-Palomo A, Garriga M, Agasi I, Bastidas A, Cavero M, Fernández-Plaza T, Arbelo N, Bioque M, García-Rizo C, Verdolini N, Madero S, Murru A, Amoretti S, Martínez-Aran A, Ruiz V, Fico G, De Prisco M, Oliva V, Solanes A, Radua J, Samalin L, Young AH, Vieta E, Vergari A, Hidalgo-Mazzei D
Exploring Digital Biomarkers of Illness Activity in Mood Episodes: Hypotheses Generating and Model Development Study
JMIR Mhealth Uhealth 2023;11:e45405
URL: https://mhealth.jmir.org/2023/1/e45405
doi:10.2196/45405
PMID:36939345

Review

Associations Between Social Cognitive Determinants and Movement-Related Behaviors in Studies Using Ecological Momentary Assessment Methods: Systematic Review

Kelsey M Bittel[1], MS; Kate Y O'Briant[1], BS; Rena M Ragaglia[1], BS; Lake Buseth[1], BS; Courtney Murtha[1], BS; Jessica Yu[1], BS; Jennifer M Stanely[1], BS; Brynn L Hudgins[1], MS; Derek J Hevel[1], PhD; Jaclyn P Maher[1], PhD

Department of Kinesiology, University Of North Carolina Greensboro, Greensboro, NC, United States

Corresponding Author:
Kelsey M Bittel, MS
Department of Kinesiology
University Of North Carolina Greensboro
1408 Walker Avenue
Greensboro, NC, 27412
United States
Phone: 1 8147690663
Email: bittekelsey@gmail.com

Abstract

Background: The social cognitive framework is a long-standing framework within physical activity promotion literature to explain and predict movement-related behaviors. However, applications of the social cognitive framework to explain and predict movement-related behaviors have typically examined the relationships between determinants and behavior across macrotimescales (eg, weeks and months). There is more recent evidence suggesting that movement-related behaviors and their social cognitive determinants (eg, self-efficacy and intentions) change across microtimescales (eg, hours and days). Therefore, efforts have been devoted to examining the relationship between social cognitive determinants and movement-related behaviors across microtimescales. Ecological momentary assessment (EMA) is a growing methodology that can capture movement-related behaviors and social cognitive determinants as they change across microtimescales.

Objective: The objective of this systematic review was to summarize evidence from EMA studies examining associations between social cognitive determinants and movement-related behaviors (ie, physical activity and sedentary behavior).

Methods: Studies were included if they quantitatively tested such an association at the momentary or day level and excluded if they were an active intervention. Using keyword searches, articles were identified across the PubMed, SPORTDiscus, and PsycINFO databases. Articles were first assessed through abstract and title screening followed by full-text review. Each article was screened independently by 2 reviewers. For eligible articles, data regarding study design, associations between social cognitive determinants and movement-related behaviors, and study quality (ie, Methodological Quality Questionnaire and Checklist for Reporting Ecological Momentary Assessment Studies) were extracted. At least 4 articles were required to draw a conclusion regarding the overall associations between a social cognitive determinant and movement-related behavior. For the social cognitive determinants in which a conclusion regarding an overall association could be drawn, 60% of the articles needed to document a similar association (ie, positive, negative, or null) to conclude that the association existed in a particular direction.

Results: A total of 24 articles including 1891 participants were eligible for the review. At the day level, intentions and self-efficacy were positively associated with physical activity. No other associations could be determined because of conflicting findings or the small number of studies investigating associations.

Conclusions: Future research would benefit from validating EMA assessments of social cognitive determinants and systematically investigating associations across different operationalizations of key constructs. Despite the only recent emergence of EMA to understand social cognitive determinants of movement-related behaviors, the findings indicate that daily intentions and self-efficacy play an important role in regulating physical activity in everyday life.

Trial Registration: PROSPERO CRD42022328500; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=328500

(JMIR Mhealth Uhealth 2023;11:e44104) doi:10.2196/44104

KEYWORDS
motivation; psychosocial; physical activity; sedentary behavior; ambulatory assessment; mobile phone

Introduction

Background

The World Health Organization recommends that adults engage in at least 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity physical activity (PA; or an equivalent combination of both) per week and limit the amount of time spent engaging in sedentary behavior (SB) [1]. Despite these recommendations, approximately 28% of adults do not meet PA guidelines [2]. Furthermore, on average, adults engage in 8.2 hours per day of SB [3]. This represents a considerable health burden as individuals who are physically inactive and do not meet PA guidelines have a 20% to 30% increased chance of premature death versus individuals who are active and meet PA guidelines [2,4]. Physical inactivity and SB can also lead to chronic diseases and are a contributing factor to 35 pathological and clinical conditions (eg, obesity, type 2 diabetes, and cardiovascular diseases) [5]. Considering the risks associated with physical inactivity and SB, understanding the factors that influence movement-related behaviors such as PA and SB is paramount.

There are a number of different theoretical approaches to understand and explain PA and SB engagement (eg, the humanistic framework, the dual process framework, and ecological models) [6]. Social cognitive framework represents one of the most used and long-standing theoretical frameworks within the movement-related behavior literature to explain and predict behavioral engagement [6]. Social cognitive framework emphasizes key determinants of behavior as individuals' cognitions about the anticipated outcomes of a behavior and their ability to engage in a behavior [7,8]. In line with these theories, individuals will focus their efforts toward (eg, intentions) and subsequently engage in a behavior if their beliefs about the behavior are positive (eg, outcome expectations) and they are confident in their ability to engage in the behavior (eg, self-efficacy) [9-13]. There are 2 prominent theories that are considered part of the social cognitive framework are Social Cognitive Theory and the Theory of Planned Behavior [9,10].

Specifically, Social Cognitive Theory identifies self-efficacy—an individual's belief in their ability to engage in a behavior—and outcome expectations—the perception of consequences (positive or negative) of an individual's action—as key determinants of behavior [8,10]. The Theory of Planned Behavior posits that intention, or one's willingness to try to engage in a behavior, is the main antecedent of behavior [7,9]. The Theory of Planned Behavior also proposes that intention formation is predicted by three factors: (1) attitudes or the extent to which one has positive or negative evaluations of a behavior, (2) subjective norms or the perceived social pressure to engage in a behavior, and (3) perceived behavioral control or the extent to which a person has the capacity and free will to engage in a behavior [7,9]. Furthermore, both theories acknowledge that facilitators and barriers can help or hinder engagement in a behavior, respectively [7-10]. Therefore, these theories outline social cognitive determinants hypothesized to influence movement-related behaviors.

Several reviews have indicated that social cognitive determinants predict PA behavior in observational studies. However, there is experimental evidence indicating that the extent to which these determinants predict changes in behavior is generally much weaker [14-16]. Within the last decade, evidence has accumulated suggesting that PA and SB are independent health behaviors [17,18], and studies have applied social cognitive frameworks to explain and predict SB. Findings across SB studies appear to mirror those of PA studies, with evidence for relationships between social cognitive determinants and behavior indicating stronger relationships in observational studies than in experimental studies [19-21].

A potential reason for the limited effectiveness of social cognitive determinants in explaining and predicting PA and SB is that the timescale in which these relationships are assessed is not the timescale in which social cognitive determinants influence decisions to engage in a behavior [10,22]. Most of the research investigating associations between social cognitive determinants and PA or SB tends to assess these constructs infrequently and ask participants to report their usual level of social cognitive determinants or behavior over macrotimescales (eg, weeks and months) [21,23]. However, PA and SB are repeat-occurrence behaviors, meaning that these behaviors are typically engaged in multiple times per week or even multiple times per day [24]. Furthermore, the day is an elemental structure in human life. Days are easily defined and universally experienced because of the light-dark cycles of the sun and associated sleep-wake cycles. Moreover, people self-regulate and restore self-regulatory resources throughout these cycles [25]. The changing contexts of people's daily lives could also influence daily and within-day motivation and movement-related behaviors. Therefore, methods that capture typical levels of movement-related behaviors and social cognitive determinants on a macrotimescale may overlook important information regarding decisions to engage in a bout of PA or SB across microtimescales (eg, hours and days). Previous research has documented that social cognitive determinants, PA, and SB are dynamic, varying within individuals across time and space [26-28]. Investigating associations between social cognitive determinants and movement-related behaviors in the content of daily life across microtimescales can elucidate the motivational determinants of movement-related behaviors and potentially enhance intervention efforts.

Recent evidence of the dynamic and time-varying nature of PA and SB has increased in part from advances in methodology to assess behavior and its determinants. Ecological momentary assessment (EMA) has gained popularity over the last decade in the movement-related behavior literature as a methodology to capture fluctuations in behavior and social cognitive determinants across microtimescales in naturalistic settings [29,30]. EMA is a real-time data capture methodology that repeatedly assesses individuals on a phenomenon of interest in their natural environment (eg, motivation and behavior) [31]. EMA is useful for assessing phenomena that change across time and space, such as movement-related behaviors and their determinants. For instance, an individual's engagement in PA behaviors may change over the course of the day, as may their feelings of confidence (ie, self-efficacy) in engaging in PA. To study these fluctuations in behavior as well as how these 2 constructs might covary, a smartphone-based EMA protocol could assess self-efficacy at predetermined (eg, every 2 hours) or randomly occurring (eg, anytime between 8 AM and 8 PM) times throughout the day. In addition, monitoring via an accelerometer could be conducted. In an EMA protocol, when participants receive a notification to complete a questionnaire through an app or website on their smartphone that measures self-efficacy to engage in PA over the following 2 hours, they are asked to briefly stop what they are doing to complete the questionnaire. Responses on the smartphone are date- and time-stamped to facilitate easy pairing of the smartphone questionnaire and accelerometer data in the 2-hour window (referenced in the self-efficacy assessment) after the EMA prompt. As these notifications happen repeatedly, researchers are able to capture these constructs and their associations as individuals go about their day-to-day activities, allowing them to examine these associations across the changing contexts of everyday life.

Therefore, EMA methodology can reduce recall biases and enhance ecological validity by evaluating a phenomenon of interest close in time to when it occurs in real-world settings as individuals go about their normal day-to-day lives [29]. Today, EMA protocols can be delivered through various media (ie, apps on mobile devices, internet-based questionnaires, and SMS text messages) that can provide time stamps of participant responses. As noted, this can facilitate the pairing of EMA responses with other time-stamped data sources such as accelerometers and allows for the investigation of the temporal sequence of relationships between key constructs. Previous research has established the feasibility and validity of using smartphone-based EMA to assess PA and SB as well as their determinants in diverse populations across their life span [28,32-35].

Objectives

The use of EMA to capture and understand PA and SB has increased over the past decade. As a result, recent reviews have summarized EMA findings regarding various determinants of PA and SB, including affective states [36] and environmental contexts [37]. However, there is yet to be a review summarizing the associations between social cognitive determinants and movement-related behaviors from studies using EMA methodologies. This systematic review aimed to summarize the literature regarding within-day and day-level associations between social cognitive determinants and movement-related behaviors using within-day or daily EMA methodologies. The decision to focus exclusively on within-day and day-level associations was based on the repeat-occurrence nature of movement-related behaviors and the fact that assessment schedules occurring less frequently than the day level may not be sensitive to the changing contexts in everyday life and the factors driving decisions to engage in occasions of PA or SB. This rationale is bolstered by the fact that the day represents a natural and fundamental reoccurring event in human life.

Methods

This systematic review was conducted and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [38] and was registered in PROSPERO (CRD42022328500).

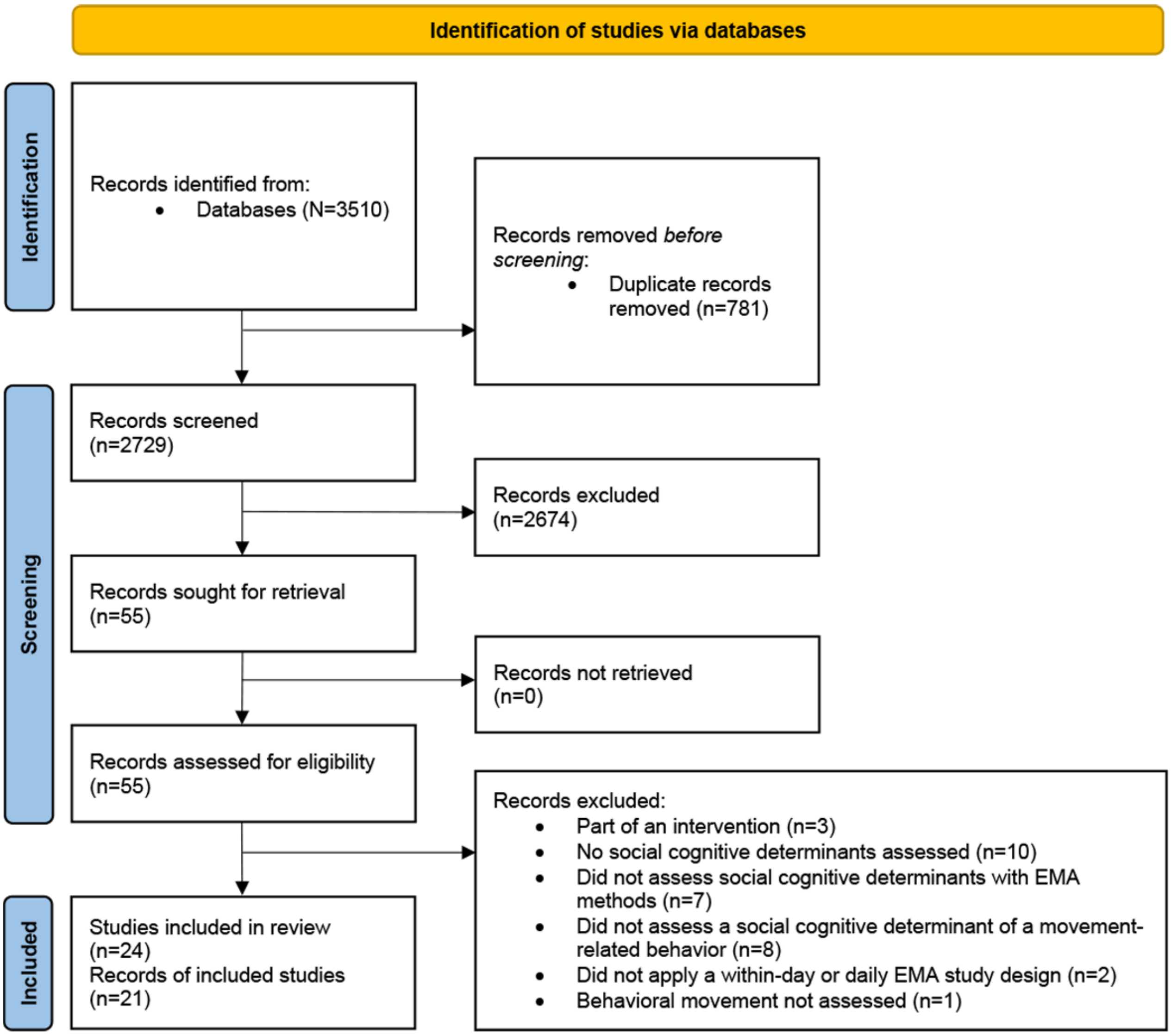
Inclusion and Exclusion Criteria

Articles that met the following criteria were included in the review: (1) human-participant research, (2) available in English, (3) quantitative data available for at least one association between a social cognitive determinant (ie, the independent variable) and movement-related behavior (ie, the dependent variable), and (4) within-day or daily EMA study design. Social cognitive determinants were defined as constructs specified within Social Cognitive Theory and the Theory of Planned Behavior, 2 popular social cognitive frameworks in the movement-related behavior literature [7-9,22,39]. Therefore, social cognitive determinants of interest for this systematic review included intentions, attitudes, subjective norms, perceived behavioral control, self-efficacy, outcome expectations, risk perceptions, barriers, facilitators, goals, and plans regarding movement-related behaviors. To focus on naturally occurring associations between social cognitive determinants and movement-related behaviors, articles were excluded if they used an active experimental design. In addition, articles were excluded if they were not published in a peer-reviewed scholarly journal.

Search Process

Literature searches were conducted on May 25, 2022, in the PubMed, SPORTDiscus, and PsycINFO databases to identify relevant articles that used EMA methods to examine associations between social cognitive determinants and movement-related behaviors. No date restrictions were applied in the searches. This process is shown in the PRISMA flow diagram in Figure 1. In total, 3 sets of search terms were used to identify potentially relevant articles. The first set of search terms used general terms, including the following: ("physical activity" OR "exercise" OR "sedentary behavior" OR "movement behavior" OR "physical exercise" OR "sitting") AND ("ecological momentary assessment" OR "EMA" OR "daily diary" OR "experience sampling"). The second set of search terms included the general terms (from the first search) along with specific terms related to psychological determinants: ("social cognitive" OR "motivation" OR "psychosocial" OR "behavioral cognitions"). The third set of search terms included the general terms from the first search along with specific terms related to social cognitive determinants: ("self-efficacy" OR "outcome expectation" OR "intention" OR "attitude" OR "subjective norm" OR "control" OR "risk perception" OR "barriers" OR "facilitators" OR "goal" OR "plan"). See Multimedia Appendix 1 for the complete search term queries in each database. These specific social cognitive terms were selected by identifying constructs outlined within the Theory of Planned Behavior and Social Cognitive Theory, 2 of the most prominent social cognitive frameworks [6]. The first search using general terms was completed so as to not miss any EMA movement-related behavior articles that may have assessed social cognitive determinants but did not have them as their focus. All articles were collected in Zotero (Corporation for Digital Scholarship) and uploaded to Rayyan (Rayyan Systems Inc), a web tool used to provide support for systematic reviews [40].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the literature search on social cognitive determinants, movement-related behaviors, and ecological momentary assessment (EMA) methods.



Article Screening and Coding

A total of 9 reviewers were involved in the process of screening and extracting information from the eligible articles. All reviewers received training on EMA studies of movement-related behaviors and social cognitive determinants from an expert in this content area (JPM). Following training, all reviewers screened the same 50 abstracts to determine whether the full text should be examined for eligibility. All reviewers then met with the content area expert to discuss and resolve discrepancies. Following this, reviewers independently screened assigned abstracts. One reviewer (KMB) screened all abstracts, and one additional reviewer (KYO, RMR, LB, CM, JY, JMS, BLH, or DJH) was assigned to each abstract. Following screening of all abstracts, discrepancies were discussed and resolved between each pair of reviewers with the content area expert.

The articles retained after title and abstract screening were gathered as full texts. In total, 2 reviewers (KMB and JPM) reviewed the full articles independently to determine eligibility and came together to discuss and resolve any discrepancies. A flowchart using the PRISMA 2020 guidelines shows the screening process for eligible articles (Figure 1) [41].

After identifying the articles to be included in this systematic review, 4 reviewers (KMB, RMR, KYO, and BLH) independently extracted relevant information from 8 articles, met with the content area expert (JPM) to discuss and resolve any discrepancies, and then independently coded the remaining articles. One reviewer (KMB) extracted relevant information from all the articles, and one additional reviewer (RMR, KYO, or BLH) was assigned to each article. Finally, the reviewers and content area expert met to discuss and resolve any discrepancies regarding article information extraction.

For each eligible article, the extracted information focused on study participant characteristics, study design characteristics, study main findings, and methodological quality assessments according to 2 established instruments. First, the Methodological Quality Questionnaire (MQQ) [42] assessed overall study quality based on 9 dimensions (ie, theoretical or conceptual definition, research design, sampling design, sample, evidence of reliability and validity, data analysis, implications for practice, and implications for policy). MQQ scores can range from 0 to 27. Second, the Checklist for Reporting Ecological Momentary Assessment Studies (CREMAS) [43] assessed quality with regard to reporting EMA methodology on 5 dimensions that outline specific criteria that have to be included in the title, introduction, methods, results, and discussion. Across these 5 dimensions, there were 16 items (ie, title and keywords, rationale, training, technology, wave duration, monitoring period, prompting design, prompt frequency, design features, attrition, prompt delivery, latency, compliance rate, missing data, limitations, and conclusions).

CREMAS scores can range from 0 to 16, with 1 point per item addressed. For both the MQQ and CREMAS, higher scores indicate better methodological quality.

Analysis

The associations between social cognitive determinants and movement-related behaviors were assessed using the guidelines developed by Sallis et al [44]. An association was supported if 60% to 100% of the articles reported such an association. No association was supported if 0% to 33% of the articles reported an association. An indeterminate or inconclusive association was supported if 34% to 59% of the articles reported an association. Statistical significance (P<.05) and parameter estimates (and 95% CIs, if reported) were used to determine whether any association between a social cognitive determinant and movement-related behaviors existed and the direction of the association (ie, positive or negative), respectively. All findings from individual articles were presented; however, at least 4 articles were needed to make an assessment regarding an overall association between a given social cognitive determinant and movement-related behavior in this systematic review. The choice to conduct a systematic review of results was based on the substantial diversity of study designs, operationalizations of social cognitive determinants and movement-related behaviors, and analyses to test associations (eg, multilevel linear regression, multilevel logistic regression, multilevel negative binomial model, and time-varying effect modeling).

Results

Overview

A total of 3510 articles were identified across all the database searches. After duplicates were removed (781/3510, 22.25%), the remaining articles were screened by title and abstract (2729/3510, 77.75%). After screening, 2.02% (55/2729) of the articles were identified for full-text retrieval. Of those 55 articles, 31 (56%) were excluded, leaving 24 (44%) articles from 21 unique studies to be included in this systematic review (Figure 1). The publication year, sample and study characteristics, and methodological quality scores of each article are presented in Table 1.

Table 1. Summary of article characteristics (N=24).

Study	Sample size, N	Age (years), mean (SD)	Sex or gender (% of female and women participants)	Race and ethnicity (% of White or non-Hispanic White participants)	Population	EMA[a] delivery medium	EMA protocol design	Social cognitive determinants of interest	Behavioral constructs of interest	Behavioral assessment method	CREMAS[b] rating	MQQ[c] rating
Arigo et al [45]	75	51.61 (5.43)	100	73	Adult women	Website	5 prompts per day for 10 days; single wave; signal contingent	Intentions	PA[d]	Accelerometer	10	21
Cook et al [46]	55	20-69[e]	15	58	Adults with HIV	Website	1 prompt per day for 30 days; single wave; signal contingent	Self-efficacy	PA	Accelerometer	9	16
Maher et al[f] [47]	116	40. 3 (9.6)	74.2	46.6	Adults	Smartphone	8 prompts per day for 4 days; multiple waves; signal contingent	Intentions, self-efficacy, and outcome expectations	PA	Accelerometer	15	22
Maher et al[f] [48]	116	40.3 (9.6)	74.2	46.6	Adults	Smartphone	8 prompts per day for 4 days; multiple waves; signal contingent	Intentions	PA	Accelerometer	15	27
Pickering et al[f] [49]	116	40.3 (9.6)	74.2	33	Adults	Smartphone	8 prompts per day for 4 days; multiple waves; signal contingent	Intentions, self-efficacy, and outcome expectations	PA	Accelerometer	13	25
Maher and Dunton[g] [50]	103	72 (7)	62	74	Older adults	Smartphone	6 prompts per day for 10 days; single wave; signal contingent	Intentions and self-efficacy	SB[h]	Accelerometer	15	27
Maher and Dunton[g] [51]	103	72.4 (7.44)	62.5	67.3	Older adults	Smartphone	6 prompts per day for 10 days; single wave; signal contingent	Intentions and self-efficacy	PA and SB	Accelerometer	15	27
Reifsteck et al [52]	17	21.82 (0.64)	47.06	82.35	College student athletes	Smartphone	4 prompts per day for 7 days; single wave; signal contingent	Intentions, self-efficacy, and outcome expectations	PA	Accelerometer	14	18

Study	Sample size, N	Age (years), mean (SD)	Sex or gender (% of female and women participants)	Race and ethnicity (% of White or non-Hispanic White participants)	Population	EMA[a] delivery medium	EMA protocol design	Social cognitive determinants of interest	Behavioral constructs of interest	Behavioral assessment method	CREMAS[b] rating	MQQ[c] rating
Zenk et al [53]	97	25-65[e]	100	0	Adult women	Website	5 prompts per day[i] for 7 days; single wave; signal contingent	Barriers	PA and SB	Accelerometer	10	19
Anderson [54]	76	40.29 (13.69)	57.9	85.5	Adults	Website	1 prompt per day for 14 days; single wave; interval contingent	Planning	PA	Self-report	11	27
Carraro and Gaudreau [55]	97	20.45 (4.61)	68	71	College students	Website	1 prompt per day[i] for 6 days; single wave; interval contingent	Action planning and coping planning	PA	Self-report	13	25
Dunton et al [56]	23	60.65 (8.22)	70	91	Older adults	PDA	4 prompts per day for 14 days; single wave; interval contingent	Self-efficacy	PA	Self-report	14	14
McDonald et al [57]	7	62.71	71.4	Not collected[j]	Older adults	PDA	2 prompts per day for a minimum of 2 months (maximum of 7 months); single wave; interval contingent	Intentions and perceived behavioral control	PA	Accelerometer	11	14
Bermudez et al [58]	111	62.32 (10.85)	14.4	Not collected	Adults with cardiac disease	PDA	1 prompt per day for 21 days; single wave; event contingent	Intentions, perceived behavioral control, subjective norms, and explicit attitudes	PA	Accelerometer	8	16
Berli et al [59]	120	Women: 43.39 (12.67); men: 45.07 (13.92)	50	Not collected	Adult dual-smoker couples	Smartphone	1 prompt per day for 28 days; single wave; event contingent	Intentions, self-efficacy, and action planning	PA	Accelerometer	12	23

Study	Sample size, N	Age (years), mean (SD)	Sex or gender (% of female and women participants)	Race and ethnicity (% of White or non-Hispanic White participants)	Population	EMA[a] delivery medium	EMA protocol design	Social cognitive determinants of interest	Behavioral constructs of interest	Behavioral assessment method	CREMAS[b] rating	MQQ[c] rating
Bond et al [60]	21	48.5 (2.8)	81	71.4	Adults who had undergone bariatric surgery	PDA	1 prompt per day for 6 days; single wave; event contingent	Intentions	PA	Self-report	11	24
Borowski et al [61]	61	41.4 (9.9)	88.5	90.2	Adults	Website	1 prompt per day for 7 days; single wave; event contingent	Barriers	PA	Self-report	10	21
Conroy et al [62]	63	Not collected	58.7	87	College students	Website	1 prompt per day for 14 days; single wave; event contingent	Intentions	PA	Self-report	14	23
Conroy et al [63]	128	21.3 (1.1)	58.6	89	College students	Website	1 prompt per day for 14 days; single wave; event contingent	Intentions	SB	Accelerometer	12	23
Curtis et al [64]	185	61.7 (5.9)	55.1	Not collected	Older adults	Website	1 prompt per day for 7 days; single wave; event contingent	Self-efficacy	PA	Self-report	10	18
Maher and Conroy [65]	100	74.2 (8.2)	67	99	Older adults	PDA	2 prompts per day for 14 days; single wave; event contingent	Intentions, self-efficacy, action planning, and coping planning	SB	Accelerometer and self-report	12	23
Rebar et al [66]	103	21.57 (2.97)	52	Not collected	College students	Website	1 prompt per day for 7 days; single wave; event contingent	Intentions	PA	Self-report	11	21
Schwaninger et al [67]	198	Women: 45.31 (13.51); men: 47.29 (13.94)	50	Not collected	Adult couples	Smartphone	1 prompt per day for 14 days; single wave; event contingent	Self-efficacy	PA	Accelerometer	8	15

Study	Sample size, N	Age (years), mean (SD)	Sex or gender (% of female and women participants)	Race and ethnicity (% of White or non-Hispanic White participants)	Population	EMA[a] delivery medium	EMA protocol design	Social cognitive determinants of interest	Behavioral constructs of interest	Behavioral assessment method	CREMAS[b] rating	MQQ[c] rating
Zhaoyang et al [68]	135	65.71 (9.83)	56.3	86.67	Older adult couples with knee OA[k]	PDA	2 prompts per day for 22 days; single wave; event contingent	Self-efficacy	PA	Accelerometer	13	25

[a]EMA: ecological momentary assessment.

[b]CREMAS: Checklist for Reporting Ecological Momentary Assessment Studies. Lowest score=0; highest score=16.

[c]MQQ: Methodological Quality Questionnaire. Lowest score=0; highest score=27.

[d]PA: physical activity.

[e]Only age range was reported.

[f]Articles used data from the same study.

[g]Articles used data from the same study.

[h]SB: sedentary behavior.

[i]Although the study assessed the constructs of interest within days, the data were aggregated to the day level for analysis.

[j]Participants were not asked about this variable.

[k]OA: osteoarthritis.

Study Sample Characteristics

The analytic sample size of each study ranged from 7 to 198 participants, with a mean sample size of 90.05 (SD 50.70). All articles except for the one by Reifsteck et al [52] (23/24, 96%) reported on studies with samples in which at least 50% of the participants identified as women. The samples reported in 21% (5/24) of the articles were college or university students [52,55,62,63,66]. The samples reported in 50% (12/24) of the articles focused on adults [45-49,53,54,58-61,67]. The samples reported in 29% (7/24) of the articles were older adults (aged ≥50 years) [50,51,56,57,64,65,68]. The mean age of the participants from the 21 unique studies was 47.65 (SD 17.84) years. None of the studies included samples of children or adolescents. Of the 18 articles that reported on participants' racial or ethnic identities, 4 (22%; n=3, 75% using data from the same study) [47-49] indicated that participants identifying as White or non-Hispanic White did not make up most of the sample [47-49,53].

EMA Protocol

Of the 24 articles, 21 (88%) reported on studies that collected data over 1 wave. A total of 12% (3/24) of the articles (using data from the same study) reported that data were collected across 3 measurement waves spaced over 1 year [47-49]. Regardless of the number of waves, the monitoring period in each study ranged from 4 to 30 days, except for 4% (1/24) of the articles, which sampled participants daily for 2 to 7 months [57]. Regarding EMA prompting design, of the 24 articles, 12 (50%) reported on studies that had 1 prompt per day [46,54,55,58-64,66,67], 3 (12%) reported on studies that had 2 prompts per day (ie, one in the morning and one in the evening) [57,65,68], and the remaining articles (9/24, 38%) reported on studies that delivered between 4 and 8 prompts per day [45,47-53,56].

In total, 38% (9/24) of the articles indicated that EMA prompting was a signal-contingent design (ie, occurring at randomly prompted times) [45-53]. A total of 17% (4/24) of the articles reported on studies that used interval-contingent design (ie, occurring at fixed times) [52,54,56,57], whereas 46% (11/24) of the articles reported on studies that used event-contingent design (eg, self-initiated EMA questionnaire after a specific event) [55,58-65,67,68].

Regarding design considerations to reduce participant burden, the studies reported in 17% (4/24) of the articles did not ask about social cognitive determinants at every EMA prompt to limit the number of items in each prompt [47-49,56]. The studies reported in 8% (2/24) of the articles customized the time of the prompts based on each participant's wake and sleep schedules [45,57], whereas another study had participants select their own start day to fit their work schedule [61].

EMA protocols were delivered via smartphone (8/24, 33%), PDA (6/24, 25%), or website (10/24, 42%). Among the articles that reported on studies that used a smartphone to deliver EMA protocols, commercially available apps including Personal Analytics Companion (1/24, 4%) [66], movisensXS (2/24, 8%) [50,51], and MyExperience (3/24, 12%) [47-49] were used. A total of 8% (2/24) of the articles reported on studies that had participants complete the study protocol on a smartphone but did not specify the platform used to deliver the EMA prompts [59,67]. PDA devices comprised handheld computers (3/24, 12%) [56,60,68], tablets (2/24, 8%) [58,65], and a wrist-worn Patient-Reported Outcomes Diary device (1/24, 4%) [57]. Articles on studies that used websites either did not specify the website [45,52,53,55,62-64] or did specify it using Prolific [54], Qualtrics [61], or REDCap (Research Electronic Data Capture; Vanderbilt University) [46].

All articles (24/24, 100%) reported participants' compliance or response rates within the EMA protocol. These rates ranged from 56.9% to 95%, indicating general compliance with the EMA methodology.

Social Cognitive Determinants

Intentions (15/24, 62%) [45,47-51,55,57-60,62,63,65,66] and self-efficacy (14/24, 58%) [46,47,49-51,56-59,64-68] were the most frequently assessed social cognitive determinants. Of the 14 articles categorized as reporting on studies that assessed self-efficacy [57,58], 2 (14%) operationalized their construct of interest as perceived behavioral control; however, perceived behavioral control is generally considered synonymous with self-efficacy [9,69], so the findings were combined into 1 category for this review. Other social cognitive determinants assessed included planning (4/24, 17%) [52,54,59,65], outcome expectations (3/24, 12%) [47,49,66], barriers (2/24, 8%) [53,61], and explicit attitudes (1/24, 4%) [58].

Intentions and self-efficacy were primarily assessed using 1 to 2 items, but 4% (1/24) of the articles reported on studies that used 4 items to assess self-efficacy [46]. The studies reported in 29% (7/24) of the articles adapted social cognitive determinant items from validated scales [46,54,55,58,59,61,65], but almost all (5/7, 71%) reduced the number of items or adapted the time frame of the items for delivery as part of their EMA protocol. Common behavioral targets assessed with social cognitive determinant items included engaging in PA [46,53-56,58-62,64,66-68], engaging in a specified duration of PA [45,47-52,57], or limiting SB to a specified total amount of time [53,63,65,66].

Movement-Related Behaviors

Methods for assessing movement-related behaviors included device-based (15/24, 62%) and self-reported (8/24, 33%) assessments or both (1/24, 4%) [65]. Common devices used included the ActiGraph GT3x or GT3x+ (7/24, 29%) [45,55,58,59,63,66,67], ActiGraph GT2M (3/24, 12%) [47-49], ActiGraph GT1M (1/24, 4%) [53], Fitbit Alta HR (1/24, 4%) [46], and activPAL 3 (3/24, 12%) [50,51,65]. A total of 4% (1/24) of the articles reported on studies that provided participants with either the ActiGraph GT1M or GT3x+ device [68], and 4% (1/24) used the Patient-Reported Outcomes Diary device [57]. In total, 75% (18/24) of the articles reported on studies that required a minimum threshold of valid wear time for data to be included in the analysis [45-53,56,58,59,61,63,65-68]. Minutes of moderate- to vigorous-intensity PA (MVPA) were the most common operationalization of PA (12/24, 50%) [45,47-49,53,55,56,58,59,63,66,67], and minutes of sedentary time were the most common operationalization of SB (5/24, 21%) [50,51,53,55,65] using device-based measures.

Among the self-reported measures of movement-related behaviors, the studies reported in 25% (6/24) of the articles adapted items from existing validated measures, including the International Physical Activity Questionnaire–Short Form by Sjöström et al [63,65], Godin Leisure-Time Exercise Questionnaire by Godin and Shephard [52,54,61], and a measure of sedentary time in older adults developed by Gardiner et al [65]. Other assessments of daily movement-related behaviors included checklists in which participants indicated the activities they had participated in that day and for how long [56,64]. Bond et al [60] and Rebar et al [66] assessed the daily duration of MVPA using items created for the study.

Methodological Quality

The average MQQ score was 21.38 (SD 4.31; range 14-27), with 21% (5/24) of the articles scoring 27 (the highest possible score). The interrater reliability for the MQQ scores was 98.3%, which is an acceptable level of agreement [55]. On the basis of the 9 criteria of the MQQ, the most frequently omitted information pertained to evidence of reliability and validity provided for the data collected (12/24, 50%) [45,46,53,56-58,60-63,66,67] followed by implications for policy (11/24, 46%) [45,46,53,55-59,61,64,67].

The average CREMAS score was 11.92 (SD 2.21; range 8-15). The interrater reliability for the CREMAS scores was 93.75%. Commonly omitted elements of the CREMAS among applicable articles included participant training procedures, latency, and missing data analysis. Of the 20 articles in which latency was applicable (eg, interval- and signal-contingent designs), none reported latency. A total of 54% (13/24) of the articles did not report any missing data analyses [45,46,52-54,58-61,64-67]. In total, 50% (12/24) of the articles did not report any training to familiarize participants with the EMA protocol [46,49,53-56,58,61,64,66-68]. Except for the studies reported in 12% (3/24) of the articles [47-49], wave duration (ie, the number of data collection waves in a study) was not applicable as all other studies collected data over a single wave.

PA and Social Cognitive Determinants

This systematic review identified studies examining associations between specific social cognitive determinants (ie, intentions, self-efficacy, outcome expectations, planning, perceived barriers, and attitudes) and PA; however, the availability of data regarding these associations differed at the momentary and daily levels. The findings are summarized in the following sections.

Intentions

Momentary Associations

The studies reported in 25% (6/24) of the articles assessed associations between intentions and PA at the momentary level, which is shown in Table 2. Of those 6 articles, 3 (50%) focused on direct relationships between momentary intentions and subsequent PA. Among college student athletes, Reifsteck et al [52] documented a positive association between intentions and behavior such that, on occasions when participants reported stronger-than-usual intentions, they engaged in more device-based MVPA over the following 3 hours. Similarly, Arigo et al [45] found a weak positive association between the number of intended minutes and minutes of device-based MVPA over the following 3 hours among adults, although the results were not significant. However, this positive association between intentions and behavior became stronger and more significant on occasions when participants experienced less contentment or body satisfaction than usual. Conversely, Pickering et al [49] found a null association between momentary intentions and subsequent device-based MVPA among adults. However, on occasions when adults had higher self-efficacy to engage in PA than was typical for them, momentary intentions were positively associated with subsequent MVPA. As only 12% (3/24) of the articles reported on studies that investigated momentary intention–PA relationships and we determined a priori that there must be at least 4 articles present to draw a conclusion regarding an overall association, this cannot be made at this time.

In total, 12% (3/24) of the articles focused specifically on time-varying moderators of intention-PA relationships, with all articles (3/3, 100%) reporting on studies that used device-based measures of PA. Time of day and day of the week were moderators investigated by Maher et al [47,51]. Time-varying effect models applied by Maher et al [47] revealed that intentions to be physically active positively predicted subsequent MVPA in the mornings and evenings but not in the afternoons. On weekdays, intentions were unrelated to subsequent PA on weekends. Using a similar approach, Maher and Dunton [51] found that, among older adults, intentions to be active were positively associated with subsequent time spent upright (ie, standing or stepping) during the morning, afternoon, and evening on both weekdays and weekends, although the magnitude of the associations changed throughout the day. Finally, Maher et al [48] investigated affect states and physical context as moderators of the intention-PA coupling and found that individuals were more likely to follow through with their intentions to be physically active on occasions when they reported greater positive affect than was typical for them (at the same time that they reported their intentions). Owing to the range of time-varying moderators investigated, conclusions regarding consistent moderators of intention-PA relationships at the momentary level cannot be drawn.

Table 2. Associations between social cognitive determinants and physical activity and sedentary behavior.[a]

Social cognitive determinant and timescale of associations	Positive association	Negative association	Null association	Moderators	Overall association
Physical activity					
Intentions predicting physical activity					
Momentary associations	• Reifsteck et al [52] • Arigo et al [45]	• No articles reported a negative association	• Pickering et al [49]	• Self-efficacy: Pickering et al [49] • Time of day: Maher et al [47] and Maher and Dunton [51] • Day of the week: Maher et al [47] and Maher and Dunton [51] • Positive affect: Maher et al [48]	N/A[b]
Daily associations	• Conroy et al [62] • Berli et al [59] • Bermudez et al [58] • Bond et al [60] • McDonald et al [57]	• McDonald et al [57]	• Rebar et al [66] • McDonald et al [57]	• Ego depletion: Rebar et al [66]	+[c]
Self-efficacy predicting physical activity					
Momentary associations	• Cook et al [46] • Dunton et al [56]	• No articles reported a negative association	• Reifsteck et al [52] • Cook et al [46] • Pickering et al [49]	• Intentions: Pickering et al [49] • Time of day: Maher et al [47] and Maher and Dunton [51] • Day of the week: Maher et al [47] and Maher and Dunton [51]	?[d]
Daily associations	• Berli et al [59] • Schwaninger et al [67] • Curtis et al [64] • McDonald et al [57] • Zhaoyang et al [68]	• Bermudez et al [58] • McDonald et al [57]	• Bermudez et al [58] • McDonald et al [57]	• Age: Curtis et al [64]	+
Outcome expectations predicting physical activity					
Momentary associations	• Reifsteck et al [52]	• No articles reported a negative association	• Maher et al [47] • Pickering et al [49]	• No articles assessed moderators	N/A
Planning predicting physical activity					
Daily associations	• Carraro and Gaudreau [55] • Anderson [54] • Berli et al [59]	• No articles reported a negative association	• Carraro and Gaudreau [55]	• Typical plans: Anderson [54] • Goal conflict: Carraro and Gaudreau [55]	N/A
Barriers predicting physical activity					
Daily associations	• No articles reported a positive association	• Borowski et al [61] • Zenk et al [53]	• Zenk et al [53]	• No articles assessed moderators	N/A

Social cognitive determinant and timescale of associations	Positive association	Negative association	Null association	Moderators	Overall association
Attitudes predicting physical activity					
Daily associations	• No articles reported a positive association	• No articles reported a negative association	• Bermudez et al [58]	• No articles assessed moderators	N/A
Sedentary behavior					
Intentions predicting sedentary behavior					
Momentary associations	• No articles reported a positive association	• Maher and Dunton [50]	• No articles reported a null association	• Time of day: Maher and Dunton [51] • Day of the week: Maher and Dunton [51]	N/A
Daily associations	• No articles reported a positive association	• Conroy et al [63]	• No articles reported a null association	• No articles assessed moderators	N/A
Self-efficacy predicting sedentary behavior					
Momentary associations	• No articles reported a positive association	• Maher and Dunton [50]	• No articles reported a null association	• Time of day: Maher and Dunton [51] • Day of the week: Maher and Dunton [51]	N/A
Planning predicting sedentary behavior					
Daily associations	• No articles reported a positive association	• Maher and Conroy [65]	• No articles reported a null association	• No articles assessed moderators	N/A
Barriers predicting sedentary behavior					
Daily associations	• No articles reported a positive association	• No articles reported a negative association	• Zenk et al [53]	• No articles assessed moderators	N/A

[a]A total of 4 articles were needed to make an overall association.

[b]N/A: not applicable; <4 articles on the social cognitive determinant, so an overall association cannot be determined.

[c]+: Positive association (≥60% of the studies showing an association).

[d]?: inconclusive (34%-59% of the studies showing an association).

Daily Associations

The studies reported in 25% (6/24) of the articles assessed the association between intentions and PA at the day level. Of these 6 articles, 4 (67%) documented positive associations between daily intentions and behavior regardless of whether intentions were assessed upon waking [60] or the previous evening [59]. Furthermore, these associations were consistent across different operationalizations of PA, including self-reported MVPA [60,62] and device-based MVPA [57-59]. Conversely, Rebar et al [66] found a null association between intentions to exercise the following day and self-reported exercise; however, on nights when university students experienced more ego depletion (limited cognitive and physical capabilities), they were more likely to successfully enact those exercise intentions the following day. McDonald et al [57] observed older adults in the months leading up to and following retirement and found that intentions did not consistently predict PA in all participants. The findings indicate that intentions to be active were positively associated with (1/24, 4%), negatively associated with (1/24, 4%), or not related to (5/24, 21%) likelihood of engaging in a PA bout depending on the participant. On the basis of the criteria by Sallis et al [44], the findings indicate an overall positive association between intentions and subsequent PA at the day level (ie, ≥60 of the articles; 4/6, 67% reported a consistent positive association); however, further investigation of moderators of these daily associations is warranted.

Self-efficacy

Momentary Associations

The studies reported in 25% (6/24) of the articles assessed associations between self-efficacy and PA at the momentary level, which is shown in Table 2. Of these 6 articles, 4 (67%) focused on direct relationships between momentary self-efficacy and subsequent PA. Among older adults, Dunton et al [56] found a positive association between momentary self-efficacy and subsequent self-reported MVPA. Similarly, Cook et al [46] found a positive association between momentary self-efficacy and device-based MVPA, although a null association was found between momentary self-efficacy and total steps.

Conversely, both Pickering et al [49] and Reifsteck et al [52] found null associations between momentary self-efficacy and subsequent device-based MVPA among adults and college student athletes, respectively. Pickering et al [49] did find that momentary intentions moderated momentary self-efficacy–behavior relationships such that, on occasions when adults had stronger intentions to engage in PA than usual, momentary self-efficacy was positively associated with subsequent MVPA. On the basis of the criteria by Sallis et al [44] and because of mixed findings across these articles, an inconclusive overall association was found between momentary self-efficacy and PA (ie, 12/24, 50% of the articles reported a consistent positive association).

The studies reported in 8% (2/24) of the articles specifically explored moderators of self-efficacy–PA relationships at the momentary level. Maher et al [47,51] investigated differences in self-efficacy and PA associations by time of day and day of the week. Among adults, self-efficacy was positively associated with device-based MVPA on weekday evenings [47]. At no other time on weekdays or weekends was self-efficacy associated with subsequent PA. Among older adults, self-efficacy was positively associated with device-based time spent upright (standing or stepping) all day on weekdays and during the mornings and afternoons but not in the evenings on weekends [51]. These findings point to potential temporal moderators of self-efficacy–PA relationships at the momentary level; however, as only 8% (2/24) of the articles reported on studies that investigated these temporal processes, firm conclusions cannot be drawn.

Daily Associations

The studies reported in 25% (6/24) of the articles examined associations between self-efficacy and PA at the day level. Of these 6 articles, 4 (67%) documented positive associations between self-efficacy and subsequent PA on a given day. Findings were consistent across self-reported [64] and device-based [59,67,68] MVPA as well as ratings of self-efficacy that occurred retrospectively (ie, self-efficacy over the previous 24 hours [64], the previous evening [59,67], and upon waking [68]). In addition, Curtis et al [64] did document moderation by age such that, for older adults aged >70 years, daily self-efficacy was positively associated with higher levels of self-reported PA, whereas daily self-efficacy was not associated with PA for participants aged 51 to 69 years. Throughout the retirement transition, McDonald et al [57] found that, for 2 participants (out of 7), stronger feelings of perceived behavioral control (which is argued to be synonymous with self-efficacy [9]) were associated with a greater likelihood of engaging in a device-captured PA bout on the same day. However, 1 participant had a negative association, and 4 participants had no association between these constructs. Contrary to the relationships hypothesized within social cognitive frameworks, Bermudez et al [58] documented a null association among perceived behavioral control, device-based MVPA, and light-intensity PA at the day level. Furthermore, the same study documented a negative association between usual levels of perceived behavioral control and light-intensity PA, where higher usual levels (as opposed to on a given day) of perceived behavioral control were associated with lower levels of light-intensity PA. On the basis of the criteria by Sallis et al [44], the findings indicate a positive association between self-efficacy and PA at the day level (ie, ≥60% of the articles—4/6, 67%—reported consistent positive associations); however, further investigation of moderators of these daily associations is necessary.

Outcome Expectations

Momentary Associations

The studies reported in 12% (3/24) of the articles assessed outcome expectations and PA at the momentary level, shown in Table 2. Among college student athletes, Reifsteck et al [52] found that, on occasions when outcome expectations regarding PA were higher than usual, individuals engaged in more device-based MVPA over the following 3 hours. However, using an identical measure of outcome expectations and PA as in the study by Reifsteck et al [52], Pickering et al [49] found null associations between momentary outcome expectations and PA among adults. Furthermore, Maher et al [47] found that momentary outcome expectations were not associated with subsequent PA regardless of time of day or day of the week. As only 12% (3/24) of the articles reported on studies that investigated momentary outcome expectation–PA relationships and we determined a priori that there must be at least 4 articles present to draw a conclusion regarding an overall association, this cannot be made at this time.

Daily Associations

None of the articles reported on studies that examined associations between daily outcome expectations and PA.

Planning

Momentary Associations

None of the articles reported on studies that examined associations between planning and PA at the momentary level.

Daily Associations

The studies reported in 12% (3/24) of the articles examined associations between daily planning and PA. Of these 3 articles, 1 (33%) reported on a study that investigated general planning [54], 1 (33%) reported on a study that investigated action planning [59], and 1 (33%) focused on action and coping planning [55]. General planning and action planning were both found to be positively associated with self-reported [54,55] and device-based [59] daily MVPA; however, daily coping planning and self-reported PA were not associated [55]. As only 12% (3/24) of the articles reported on studies that investigated daily planning–PA relationships and we determined a priori that there must be at least 4 articles present to draw a conclusion regarding an overall association, this cannot be made at this time.

Regarding moderators, Anderson [54] found that usual levels of planning moderated associations between daily planning and PA such that, for individuals who tended to have weaker PA planning, on days when they reported stronger-than-usual plans to be active, they engaged in more self-reported MVPA. Carraro and Gaudreau [55] found that a time-varying factor—daily academic goal conflict—moderated associations between daily action planning and PA such that daily action planning was positively associated with daily self-reported PA on days during which individuals experienced lower academic goal conflict. Although these studies suggest possible time-invariant and time-varying moderators of daily planning–PA relationships, because of the limited number of studies investigating such moderators, conclusions cannot be drawn at this time.

Perceived Barriers

Momentary Associations

None of the articles reported on studies that examined associations between momentary barriers and PA.

Daily Associations

Borowski et al [61] documented a negative association between daily barriers to exercise (eg, no time, feeling tired, and air or noise pollution) and self-reported PA, and for each additional barrier reported, participants engaged in 27% fewer minutes of PA that day. Zenk et al [53] examined associations between different types of barriers and device-based PA and found that the extent to which African American women endorsed poor weather as a barrier to PA was associated with less PA but that environmental (eg, no sidewalk or no indoor facilities) and social (eg, no one to exercise with and safety or crime concerns) barriers were not associated with PA. Owing to the limited number of articles, an overall association between daily barriers and PA cannot be determined at this time.

Attitudes

Momentary Associations

None of the articles reported on studies that examined associations between momentary attitudes and PA.

Daily Associations

Bermudez et al [58] found that neither affective nor instrumental attitudes on a given day were associated with device-based MVPA or light-intensity PA. As only 4% (1/24) of the studies investigated this topic, an overall association between attitudes and PA cannot be determined at this time.

SB and Social Cognitive Determinants

Overview

This systematic review identified studies examining associations between specific social cognitive determinants (ie, intentions, self-efficacy, planning, and perceived barriers) and SB, but the availability of data regarding these associations differed at the momentary and daily levels. However, because of the limited number of studies investigating such associations, overall associations between each social cognitive determinant and SB cannot be determined at this time. Nevertheless, we report our findings in the following sections.

Intentions

Momentary Associations

Maher and Dunton [50] found that, on occasions when older adults had stronger intentions than usual to limit their SB, they subsequently engaged in less device-based SB in the following 2-hour period. Using the same data set, Maher and Dunton [51] found that, on weekdays, momentary intentions to limit SB negatively predicted subsequent SB across the entire day, but on weekends, intentions only negatively predicted SB in the morning, afternoon, and early evening.

Daily Associations

Conroy et al [63] examined day-level associations between end-of-day intentions to limit SB and next-day SB among university students and found that, on days when university students had stronger-than-usual intentions to limit their SB, they subsequently engaged in less self-reported SB the following day.

Self-efficacy

Momentary Associations

Maher et al [50,51] have published 2 articles examining associations between momentary self-efficacy and SB. Maher and Dunton [50] found that, on occasions when older adults had stronger self-efficacy to limit SB, they subsequently engaged in less device-based SB over the following 2 hours. Investigation of day of the week and time of day as moderators of this association revealed that self-efficacy to limit SB was negatively associated with SB throughout the day on weekdays but that, on weekends, self-efficacy to limit SB was only associated with subsequent SB in the afternoon [51].

Daily Associations

None of the articles reported on studies that examined associations between daily self-efficacy and SB.

Planning

Momentary Associations

None of the articles reported on studies that examined associations between momentary plans and SB.

Daily Associations

Maher and Conroy [65] investigated daily associations between planning (ie, a composite score of action and coping planning) and SB and found that, on mornings when plans were stronger than usual to limit SB, older adults engaged in less device-based SB that day.

Perceived Barriers

Momentary Associations

None of the articles reported on studies that examined associations between barriers and SB.

Daily Associations

Zenk et al [53] found null associations between weather, environment, and social barriers to engaging in PA and device-based SB on a given day among African American women.

Discussion

Principal Findings

This review is the first to summarize the available EMA evidence of associations between social cognitive determinants and movement-related behaviors at the momentary and daily levels. Although this review included 24 articles comprising 21 unique studies, there were limited studies investigating each individual social cognitive determinant's relationship with subsequent behavior, especially after accounting for the timescale of assessment (ie, momentary level vs day level). The largest evidence base and, therefore, the strongest conclusions in the systematic review pertain to relationships between intentions and PA and between self-efficacy and PA. Overall, synthesizing the available evidence contributes to our preliminary understanding of the impact of social cognitive determinants on subsequent movement-related behaviors in real-world environments, identifies gaps in the literature to direct future movement-related behavior EMA research, and can begin to inform intervention efforts that are designed to deliver contextually relevant motivation content during periods of opportunity and vulnerability.

This systematic review suggests that positive associations exist among intentions, self-efficacy, and PA at the day level. It is not surprising that intentions and self-efficacy, which are prominent constructs posited to directly influence behavior within social cognitive frameworks [70,71], emerged as consistent and positive predictors of PA at the day level in this systematic review. However, articles that reported on studies investigating associations among intentions, self-efficacy, and PA at the momentary level (6/24, 25% and 6/24, 25%, respectively) revealed mixed findings. For instance, although a sufficient number of studies investigated associations between momentary self-efficacy and subsequent PA to determine an overall association, conflicting findings across the studies resulted in an inconclusive overall association. Several time-varying moderators were documented across the articles at the momentary level. It is possible that associations between momentary social cognitive determinants and subsequent PA may be affected by contextual factors that change across time and space as individuals navigate their daily lives, whereas day-level associations may be less affected by immediate contextual features of one's current environment.

This systematic review suggests that more research is needed to better understand associations between social cognitive determinants and subsequent movement-related behaviors in the context of everyday life. For instance, results from this systematic review suggest that no studies used daily or within-day EMA methodology to examine relationships between PA facilitators (eg, optimal weather) and subsequent PA behavior or between outcome expectations and attitudes and subsequent SB. In addition, there were not enough eligible articles that reported on studies investigating relationships between any social cognitive determinant and SB to draw conclusions. In addition, although some social cognitive determinants such as daily planning appeared to indicate consistent associations with PA across the studies, <4 studies investigated this association, which prevented a conclusion from being drawn. Further complicating the state of the literature in this area is that a variety of measures were used to assess both social cognitive determinants and behavior. Such diversity may have contributed to the discrepant findings across the studies. For instance, among the studies investigating associations between momentary self-efficacy and PA (where an inconclusive overall association was determined; only 2/4, 50% indicated a positive association), Dunton et al [56] assessed participants' confidence in engaging in PA and found a positive association with PA, whereas Pickering et al [49] assessed participants' confidence in engaging in PA despite possible barriers and found a null association. However, even across the diverse assessments used, we were able to find consistent trends in relationships between some social cognitive determinants and behaviors (eg, daily associations between intentions and PA), increasing the confidence in our conclusions on those relationships as the diversity of measures reduces the likelihood that the effect is the result of measurement bias. Greater consistency across the studies in the assessment of social cognitive determinants and behavior would allow for more precise conclusions on associations between social cognitive determinants and movement-related behavior and would be more appropriate to quantitatively estimate associations in a meta-analysis.

Limitations of This Review

Although studies using EMA methods can uncover more nuanced associations between social cognitive determinants and movement-related behaviors by collecting ecologically valid and intensive longitudinal data, the limitations of the studies included in this review should be noted. First, the findings synthesized in this review may not apply to all developmental periods across the life span or individuals of racially or ethnically minoritized backgrounds. No studies included in this review focused on child or adolescent samples, and almost all studies (12/14, 86% that reported race and ethnicity) featured samples in which most participants identified as White or non-Hispanic White. Given that PA levels decline rapidly throughout late childhood and adolescence and racial and ethnic minorities experience physical inactivity–related health disparities [72], EMA may be a critical methodological tool to understand relationships between motivation and behavior in these vulnerable populations. In addition, the studies featured in this review comprised insufficiently active or sedentary individuals as well as those who were sufficiently active; therefore, it is likely that the results are aggregated across individuals at various stages of the behavior change continuum. There is evidence suggesting that social cognitive determinants may differentially regulate behavior across the behavior change continuum, which is an important direction to explore in EMA work to understand movement-related behaviors [24,73]. Further influencing the generalizability of the findings is the lack of missing data analysis in several studies included in this review (12/24, 50%). Such an analysis is essential in EMA studies to determine if data are missing at random or in systematic patterns, which may influence the extent to which documented associations translate to all occasions or all people [43,74].

Most EMA studies included in this review (16/24, 67%) created measures to assess social cognitive determinants or behavior. Although such an approach is common in EMA research as fewer items are used to limit participant burden and fatigue that may result from repeated, intensive assessments [29], the items used in the studies in this review rarely presented psychometric data to establish the validity or reliability of the measures, and the studies often appeared to create items based on face validity. More rigorous and systematic approaches are necessary to develop EMA items to assess social cognitive determinants and movement-related behaviors (for a more in-depth discussion, see the study by Reichert et al [75]).

Furthermore, the limitations of the systematic review itself should be addressed. Although this review assessed the quality of the studies and study reporting through the MQQ and CREMAS, respectively, the information was not used in the interpretation of the results. Neither quality assessment tool specifies thresholds for low-, medium-, or high-quality articles. Therefore, we chose to present the raw scores on each assessment tool. On the basis of those scores, of the 24 articles included in the review, only 5 (21%) scored <18 on the MQQ, and 7 (29%) scored <11 on the CREMAS, which would indicate that those articles satisfied less than two-thirds of the criteria specified within the respective quality assessment tools.

In addition, the inclusion and exclusion criteria of this review did not place any restrictions on the participant sample size necessary for inclusion. A small participant sample size can affect the likelihood of obtaining significant results as well as the magnitude of the association documented. Only 17% (4/24) of the articles included in this review reported on studies that had <50 participants—a threshold determined through simulation to provide adequate power for variance, SE, and fixed-effects estimates at both the between- and within-person levels [76,77]. However, the findings presented in this review primarily focus on within-person findings (or the extent to which social cognitive determinants predict subsequent behavior on a given day or moment). Therefore, power for the analysis is derived from the number of occasions in the analytic sample, and because of the intensive assessments that are a hallmark of EMA, studies typically generate many observations per person. For instance, the study by McDonald et al [57] had the smallest participant sample size of the studies included in the review (N=7) but collected daily assessments from 2 to 7 months, resulting in between 87 and 196 observations per participant. Furthermore, the size of the associations was not considered to fully interpret the relationships documented across the studies as few (7/24, 29%) reported effect sizes. As the volume of studies in this area increases, a meta-analysis should be conducted that accounts for the sample size and effect sizes of relevant studies to further characterize momentary and daily associations between social cognitive determinants and movement-related behaviors.

Future Directions

PA and SB are considered independent health behaviors with different processes regulating each behavior and different health consequences associated with each [78]; however, most of the research investigating relationships between social cognitive determinants and movement-related behaviors focuses on PA. Given the prevalence of excessive SB as well as recent calls to design movement-related behavior interventions to initially focus on reducing SB and over time build to engaging in MVPA [79,80], EMA studies specifically addressing social cognitive determinant–SB relationships are necessary to develop and refine theoretical frameworks to explain and predict SB.

Across both behaviors, intentions and self-efficacy were the most investigated social cognitive determinants. This is not surprising as the Theory of Planned Behavior and Social Cognitive Theory posit that intentions and self-efficacy are proximal determinants of behavior, respectively [7,9,10,39]. However, based on preliminary evidence from this review, momentary self-efficacy and daily planning may also be important motivational determinants of behavior, but this is based on a limited number of studies. Interestingly, the relationships between planning and PA seemed to depend on the type of planning used (eg, action planning vs coping planning), but it is unclear why these differences might exist across microtimescales given the consistent finding across macrotimescales [81]. Perhaps associations between coping planning and movement-related behavior on microtimescales depend on whether an anticipated barrier is encountered in one's daily experiences. Similarly, although many studies (12/24, 50%) indicated that self-efficacy was assessed, further inspection of items revealed subtle differences in the type of self-efficacy assessed across the studies (eg, task self-efficacy and barrier self-efficacy), which may be differentially related to behavior and are likely important to consider in future EMA research [82,83].

Findings from this review provide evidence that both time-varying and time-invariant factors can moderate associations between social cognitive determinants and subsequent movement-related behaviors at the momentary and daily levels. A recent systematic review of studies investigating the PA intention-behavior coupling across macrotimescales (eg, weeks and months) found that consistent moderators of the intention-behavior coupling included motivational factors such as intention stability, intention commitment, low goal conflict, affective attitude, anticipated regret, perceived behavioral control or self-efficacy, and exercise identity [69]. Although some EMA studies have investigated some of these motivational factors as moderators [49,55], more work is needed to establish whether between-person moderators across macrotimescales also serve as within-person moderators across microtimescales. In addition, to date, only a handful of studies have investigated the moderating role of affective states and physical and social contexts on social cognitive determinant–behavior relationships [48]. There is recent evidence suggesting that these time-varying contextual factors can influence behavior, and it may be that this influence on behavior is through motivational processes [36,37]. Future research should continue to investigate context-sensitive moderators as this work is an essential first step in theory refinement that is sensitive to the contexts that people encounter in their daily lived experiences [84].

This systematic review focused on observational EMA studies to better understand the naturally occurring relationships between social cognitive determinants and subsequent behavior.

However, using intensive assessment methods such as EMA and accelerometry to understand how social cognitive determinant–movement-related behavior relationships change over the course of an intervention may be important for developing more effective behavioral interventions. For instance, Basen-Engquist et al [85] had endometrial cancer survivors complete 3- , 10- , and 12-day EMA protocols spaced over the course of a 6-month PA intervention. Although the authors aggregated data across these 3 time points to determine momentary associations among self-efficacy, outcome expectations, and PA, such data could reveal the extent to which associations between social cognitive determinants and subsequent behavior change over the course of an intervention. Such data could also reveal the extent to which behavior change content delivered at a specific point in the intervention is able to affect social cognitive determinant–behavior relationships. Collecting EMA data regarding social cognitive determinant–behavior relationships may help identify potent intervention content in the context of everyday life.

Conclusions

This systematic review synthesized EMA-derived associations between social cognitive determinants and subsequent movement-related behavior over microtimescales. Overall, based on the available evidence, social cognitive determinants do regulate movement-related behaviors in the context of everyday life. Specifically, daily intentions and self-efficacy appeared to have a consistent and positive link with PA behavior across the studies using self-reported and device-based measures of behavior. Future research is necessary to determine whether these associations extend to the momentary level, investigate associations among a broader range of social cognitive determinants and movement-related behaviors in diverse populations across the life span, and explore time-varying and time-invariant moderators of these associations. In addition, efforts should be devoted to developing more rigorous study designs, including validating EMA social cognitive determinant assessments and conducting missing data analyses. Ultimately, this systematic review provides foundational knowledge for understanding the motivational determinants of movement-related behaviors within people and is essential for directing future research regarding movement-related behaviors in the context of daily life. In addition, such inquiries are necessary to inform the development of interventions to promote active lifestyles in the context of everyday life.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search term queries.
[DOCX File , 13 KB - mhealth_v11i1e44104_app1.docx]

References

1. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
2. Physical activity fact sheet. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2021-08-30]
3. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]
4. Kohl 3rd HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305 [FREE Full text] [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
5. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211 [FREE Full text] [doi: 10.1002/cphy.c110025] [Medline: 23798298]
6. Rhodes RE, McEwan D, Rebar AL. Theories of physical activity behaviour change: a history and synthesis of approaches. Psychol Sport Exerc 2019 May;42:100-109. [doi: 10.1016/j.psychsport.2018.11.010]
7. Ajzen I, Schmidt P. Changing behavior using the theory of planned behavior. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020:17-31.
8. Luszczynska A, Schwarzer R. Changing behavior using social cognitive theory. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020:32-45.
9. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
10. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]

11. Ghahremani L, Niknami S, Nazari M. The prediction of physical activity intention and behavior in elderly male residents of a nursing home: a comparison of two behavioral theories. Iran J Med Sci 2012 Mar;37(1):23-31 [FREE Full text] [Medline: 23115427]
12. Phipps DJ, Hannan TE, Rhodes RE, Hamilton K. A dual-process model of affective and instrumental attitudes in predicting physical activity. Psychol Sport Exerc 2021 May;54:101899. [doi: 10.1016/j.psychsport.2021.101899]
13. Young MD, Plotnikoff RC, Collins CE, Callister R, Morgan PJ. Social cognitive theory and physical activity: a systematic review and meta-analysis. Obes Rev 2014 Dec;15(12):983-995. [doi: 10.1111/obr.12225] [Medline: 25428600]
14. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [FREE Full text] [doi: 10.2105/AJPH.2010.194381] [Medline: 21330590]
15. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474 [FREE Full text] [doi: 10.1037/a0032853] [Medline: 23730717]
16. Rhodes RE, Dickau L. Experimental evidence for the intention-behavior relationship in the physical activity domain: a meta-analysis. Health Psychol 2012 Nov;31(6):724-727. [doi: 10.1037/a0027290] [Medline: 22390739]
17. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
18. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018 Oct 5;6:288 [FREE Full text] [doi: 10.3389/fpubh.2018.00288] [Medline: 30345266]
19. Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]
20. Prapavessis H, Gaston A, DeJesus S. The theory of planned behavior as a model for understanding sedentary behavior. Psychol Sport Exerc 2015 Jul;19:23-32. [doi: 10.1016/j.psychsport.2015.02.001]
21. Rollo S, Gaston A, Prapavessis H. Cognitive and motivational factors associated with sedentary behavior: a systematic review. AIMS Public Health 2016 Nov 28;3(4):956-984 [FREE Full text] [doi: 10.3934/publichealth.2016.4.956] [Medline: 29546206]
22. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
23. Hagger MS, Chatzisarantis NL, Biddle SJ. A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: predictive validity and the contribution of additional variables. J Sport Exerc Psychol 2002;24(1):3-32. [doi: 10.1123/jsep.24.1.3]
24. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. JAMA 2018 Aug 21;320(7):639-640 [FREE Full text] [doi: 10.1001/jama.2018.6621] [Medline: 29852046]
25. Kühnel J, Bledow R, Feuerhahn N. When do you procrastinate? Sleep quality and social sleep lag jointly predict self-regulatory failure at work. J Organ Behav 2016 Oct;37(7):983-1002. [doi: 10.1002/job.2084]
26. Chevance G, Baretta D, Heino M, Perski O, Olthof M, Klasnja P, et al. Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. PLoS One 2021 May 14;16(5):e0251659 [FREE Full text] [doi: 10.1371/journal.pone.0251659] [Medline: 33989338]
27. Conroy DE, Elavsky S, Hyde AL, Doerksen SE. The dynamic nature of physical activity intentions: a within-person perspective on intention-behavior coupling. J Sport Exerc Psychol 2011 Dec;33(6):807-827 [FREE Full text] [doi: 10.1123/jsep.33.6.807] [Medline: 22262706]
28. Dunton GF, Liao Y, Kawabata K, Intille S. Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. Front Psychol 2012 Jul 30;3:260 [FREE Full text] [doi: 10.3389/fpsyg.2012.00260] [Medline: 22866046]
29. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
30. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4(1):35-52. [doi: 10.1023/A:1023657221954]
31. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994 Sep 1;16(3):199-202. [doi: 10.1093/abm/16.3.199]
32. Knell G, Gabriel KP, Businelle MS, Shuval K, Wetter DW, Kendzor DE. Ecological momentary assessment of physical activity: validation study. J Med Internet Res 2017 Jul 18;19(7):e253 [FREE Full text] [doi: 10.2196/jmir.7602] [Medline: 28720556]
33. Maher JP, Rebar AL, Dunton GF. Ecological momentary assessment is a feasible and valid methodological tool to measure older adults' physical activity and sedentary behavior. Front Psychol 2018 Aug 15;9:1485 [FREE Full text] [doi: 10.3389/fpsyg.2018.01485] [Medline: 30158891]

34. Maher JP, Sappenfield K, Scheer H, Zecca C, Hevel DJ, Kennedy-Malone L. Feasibility and validity of assessing low-income, African American older adults' physical activity and sedentary behavior through ecological momentary assessment. J Meas Phys Behav 2021;4(4):343-352. [doi: 10.1123/jmpb.2021-0024]
35. Romanzini CL, Romanzini M, Batista MB, Barbosa CC, Shigaki GB, Dunton G, et al. Methodology used in ecological momentary assessment studies about sedentary behavior in children, adolescents, and adults: systematic review using the checklist for reporting ecological momentary assessment studies. J Med Internet Res 2019 May 15;21(5):e11967 [FREE Full text] [doi: 10.2196/11967] [Medline: 31094349]
36. Liao Y, Shonkoff ET, Dunton GF. The acute relationships between affect, physical feeling states, and physical activity in daily life: a review of current evidence. Front Psychol 2015 Dec 23;6:1975 [FREE Full text] [doi: 10.3389/fpsyg.2015.01975] [Medline: 26779049]
37. Papini NM, Yang CH, Do B, Mason TB, Lopez NV. External contexts and movement behaviors in ecological momentary assessment studies: a systematic review and future directions. Int Rev Sport Exerc Psychol (forthcoming) 2020 Dec 17:1-31. [doi: 10.1080/1750984x.2020.1858439]
38. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
39. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Hoboken, NJ, USA: Prentice Hall; 1986.
40. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
41. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
42. Acosta S, Garza T, Hsu HY, Goodson P. Assessing quality in systematic literature reviews: a study of novice rater training. SAGE Open 2020 Jul 07;10(3):215824402093953. [doi: 10.1177/2158244020939530]
43. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
44. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. Med Sci Sports Exerc 2000 May;32(5):963-975. [doi: 10.1097/00005768-200005000-00014] [Medline: 10795788]
45. Arigo D, Hevel D, Bittel K, Maher JP. Within-person examination of the exercise intention-behavior gap among women in midlife with elevated cardiovascular disease risk. Psychol Sport Exerc 2022 May;60:102138. [doi: 10.1016/j.psychsport.2022.102138] [Medline: 35531355]
46. Cook P, Jankowski C, Erlandson KM, Reeder B, Starr W, Flynn Makic MB. Low- and high-intensity physical activity among people with HIV: multilevel modeling analysis using sensor- and survey-based predictors. JMIR Mhealth Uhealth 2022 Apr 14;10(4):e33938 [FREE Full text] [doi: 10.2196/33938] [Medline: 35436236]
47. Maher JP, Dzubur E, Huh J, Intille S, Dunton GF. Within-day time-varying associations between behavioral cognitions and physical activity in adults. J Sport Exerc Psychol 2016 Aug;38(4):423-434 [FREE Full text] [doi: 10.1123/jsep.2016-0058] [Medline: 27634288]
48. Maher JP, Rhodes RE, Dzubur E, Huh J, Intille S, Dunton GF. Momentary assessment of physical activity intention-behavior coupling in adults. Transl Behav Med 2017 Dec;7(4):709-718 [FREE Full text] [doi: 10.1007/s13142-017-0472-6] [Medline: 28155108]
49. Pickering TA, Huh J, Intille S, Liao Y, Pentz MA, Dunton GF. Physical activity and variation in momentary behavioral cognitions: an ecological momentary assessment study. J Phys Act Health 2016 Mar;13(3):344-351. [doi: 10.1123/jpah.2014-0547] [Medline: 26284314]
50. Maher JP, Dunton GF. Editor's choice: dual-process model of older adults' sedentary behavior: an ecological momentary assessment study. Psychol Health 2020 May;35(5):519-537 [FREE Full text] [doi: 10.1080/08870446.2019.1666984] [Medline: 31550923]
51. Maher JP, Dunton GF. Within-day time-varying associations between motivation and movement-related behaviors in older adults. Psychol Sport Exerc 2020 Mar;47:101522. [doi: 10.1016/j.psychsport.2019.04.012]
52. Reifsteck EJ, Anderson SN, Newton JD, Maher JP. A practical guide and empirical example for implementing ecological momentary assessment in sport psychology research with athletes. Sport Exerc Perform Psychol 2021 Aug;10(3):408-422. [doi: 10.1037/spy0000252]
53. Zenk SN, Horoi I, Jones KK, Finnegan L, Corte C, Riley B, et al. Environmental and personal correlates of physical activity and sedentary behavior in African American women: an ecological momentary assessment study. Women Health 2017 Apr;57(4):446-462. [doi: 10.1080/03630242.2016.1170093] [Medline: 27014957]
54. Anderson AR. Is a better planned day a healthier day? A daily diary study. Psychol Health 2021 Oct;36(10):1217-1234. [doi: 10.1080/08870446.2020.1841763] [Medline: 33155509]

55. Carraro N, Gaudreau P. Predicting physical activity outcomes during episodes of academic goal conflict: the differential role of action planning and coping planning. Pers Soc Psychol Bull 2015 Sep;41(9):1291-1305. [doi: 10.1177/0146167215594121] [Medline: 26174696]
56. Dunton GF, Atienza AA, Castro CM, King AC. Using ecological momentary assessment to examine antecedents and correlates of physical activity bouts in adults age 50+ years: a pilot study. Ann Behav Med 2009 Dec;38(3):249-255 [FREE Full text] [doi: 10.1007/s12160-009-9141-4] [Medline: 20052568]
57. McDonald S, Vieira R, Godfrey A, O'Brien N, White M, Sniehotta FF. Changes in physical activity during the retirement transition: a series of novel n-of-1 natural experiments. Int J Behav Nutr Phys Act 2017 Dec 08;14(1):167 [FREE Full text] [doi: 10.1186/s12966-017-0623-7] [Medline: 29221449]
58. Bermudez T, Bolger N, Bierbauer W, Bernardo A, Fleisch-Silvestri R, Hermann M, et al. Physical activity after cardiac rehabilitation: explicit and implicit attitudinal components and ambivalence. Health Psychol 2021 Aug;40(8):491-501. [doi: 10.1037/hea0001109] [Medline: 34618497]
59. Berli C, Lüscher J, Luszczynska A, Schwarzer R, Scholz U. Couples' daily self-regulation: the health action process approach at the dyadic level. PLoS One 2018 Oct 29;13(10):e0205887 [FREE Full text] [doi: 10.1371/journal.pone.0205887] [Medline: 30372470]
60. Bond DS, Thomas JG, Ryder BA, Vithiananthan S, Pohl D, Wing RR. Ecological momentary assessment of the relationship between intention and physical activity behavior in bariatric surgery patients. Int J Behav Med 2013 Mar;20(1):82-87 [FREE Full text] [doi: 10.1007/s12529-011-9214-1] [Medline: 22203518]
61. Borowski S, Savla J, Zvonkovic AM. Impact of flexible work arrangements, self-efficacy, and barriers on daily physical activity among university staff. J Phys Act Health 2021 May 01;18(5):594-602. [doi: 10.1123/jpah.2020-0099] [Medline: 33811186]
62. Conroy DE, Elavsky S, Doerksen SE, Maher JP. A daily process analysis of intentions and physical activity in college students. J Sport Exerc Psychol 2013 Oct;35(5):493-502 [FREE Full text] [doi: 10.1123/jsep.35.5.493] [Medline: 24197717]
63. Conroy DE, Maher JP, Elavsky S, Hyde AL, Doerksen SE. Sedentary behavior as a daily process regulated by habits and intentions. Health Psychol 2013 Nov;32(11):1149-1157 [FREE Full text] [doi: 10.1037/a0031629] [Medline: 23477579]
64. Curtis RG, Windsor TD, Mogle JA, Bielak AA. There's more than meets the eye: complex associations of daily pain, physical symptoms, and self-efficacy with activity in middle and older adulthood. Gerontology 2017;63(2):157-168. [doi: 10.1159/000450786] [Medline: 27756059]
65. Maher JP, Conroy DE. A dual-process model of older adults' sedentary behavior. Health Psychol 2016 Mar;35(3):262-272. [doi: 10.1037/hea0000300] [Medline: 26690644]
66. Rebar AL, Dimmock JA, Rhodes RE, Jackson B. A daily diary approach to investigate the effect of ego depletion on intentions and next day behavior. Psychol Sport Exerc 2018 Nov;39:38-44. [doi: 10.1016/j.psychsport.2018.07.010]
67. Schwaninger P, Berli C, Lüscher J, Scholz U. Cultivation or enabling? Day-to-day associations between self-efficacy and received support in couples. Soc Sci Med 2021 Oct;287:114330. [doi: 10.1016/j.socscimed.2021.114330] [Medline: 34455336]
68. Zhaoyang R, Martire LM, Sliwinski MJ. Morning self-efficacy predicts physical activity throughout the day in knee osteoarthritis. Health Psychol 2017 Jun;36(6):568-576 [FREE Full text] [doi: 10.1037/hea0000479] [Medline: 28277696]
69. Rhodes RE, Cox A, Sayar R. What predicts the physical activity intention-behavior gap? A systematic review. Ann Behav Med 2022 Jan 01;56(1):1-20. [doi: 10.1093/abm/kaab044] [Medline: 34231844]
70. Maddock JE, Suess C, Bratman GN, Smock C, Kellstedt D, Gustat J, et al. Development and validation of self-efficacy and intention measures for spending time in nature. BMC Psychol 2022 Mar 03;10(1):51 [FREE Full text] [doi: 10.1186/s40359-022-00764-1] [Medline: 35241177]
71. Strecher VJ, DeVellis BM, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health Educ Q 1986;13(1):73-92. [doi: 10.1177/109019818601300108] [Medline: 3957687]
72. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
73. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
74. Little RJ, Rubin DB. Missing data in experiments. In: Little RJ, Rubin DB, editors. Statistical Analysis with Missing Data. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2002:24-40.
75. Reichert M, Giurgiu M, Koch E, Wieland LM, Lautenbach S, Neubauer AB, et al. Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychol Sport Exerc 2020 Sep;50:101742 [FREE Full text] [doi: 10.1016/j.psychsport.2020.101742] [Medline: 32831643]
76. Maas CJ, Hox JJ. Sufficient sample sizes for multilevel modeling. Methodology 2005 Jan;1(3):86-92. [doi: 10.1027/1614-2241.1.3.86]
77. McNeish DM, Stapleton LM. The effect of small sample size on two-level model estimates: a review and illustration. Educ Psychol Rev 2016;28(2):295-314. [doi: 10.1007/s10648-014-9287-x]

78. Katzmarzyk PT. Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? Diabetes 2010 Nov;59(11):2717-2725 [FREE Full text] [doi: 10.2337/db10-0822] [Medline: 20980470]
79. Dogra S, Copeland JL, Altenburg TM, Heyland DK, Owen N, Dunstan DW. Start with reducing sedentary behavior: a stepwise approach to physical activity counseling in clinical practice. Patient Educ Couns 2022 Jun;105(6):1353-1361. [doi: 10.1016/j.pec.2021.09.019] [Medline: 34556383]
80. Ussery EN, Fulton JE, Galuska DA, Katzmarzyk PT, Carlson SA. Joint prevalence of sitting time and leisure-time physical activity among US adults, 2015-2016. JAMA 2018 Nov 20;320(19):2036-2038 [FREE Full text] [doi: 10.1001/jama.2018.17797] [Medline: 30458482]
81. Carraro N, Gaudreau P. Spontaneous and experimentally induced action planning and coping planning for physical activity: a meta-analysis. Psychol Sport Exerc 2013 Mar;14(2):228-248. [doi: 10.1016/j.psychsport.2012.10.004]
82. Beauchamp MR, Crawford KL, Jackson B. Social cognitive theory and physical activity: mechanisms of behavior change, critique, and legacy. Psychol Sport Exerc 2019 May;42:110-117. [doi: 10.1016/j.psychsport.2018.11.009]
83. Zhang CQ, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. Health Psychol 2019 Jul;38(7):623-637. [doi: 10.1037/hea0000728] [Medline: 30973747]
84. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Jan;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
85. Basen-Engquist K, Carmack CL, Li Y, Brown J, Jhingran A, Hughes DC, et al. Social-cognitive theory predictors of exercise behavior in endometrial cancer survivors. Health Psychol 2013 Nov;32(11):1137-1148 [FREE Full text] [doi: 10.1037/a0031712] [Medline: 23437853]

Abbreviations

CREMAS: Checklist for Reporting Ecological Momentary Assessment Studies
EMA: ecological momentary assessment
MQQ: Methodological Quality Questionnaire
MVPA: moderate- to vigorous-intensity physical activity
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
REDCap: Research Electronic Data Capture
SB: sedentary behavior

Edited by L Buis; submitted 08.11.22; peer-reviewed by J de Brito, H Zihao, T Martino; comments to author 03.01.23; revised version received 26.02.23; accepted 27.02.23; published 07.04.23.

Please cite as:
Bittel KM, O'Briant KY, Ragaglia RM, Buseth L, Murtha C, Yu J, Stanely JM, Hudgins BL, Hevel DJ, Maher JP
Associations Between Social Cognitive Determinants and Movement-Related Behaviors in Studies Using Ecological Momentary Assessment Methods: Systematic Review
JMIR Mhealth Uhealth 2023;11:e44104
URL: https://mhealth.jmir.org/2023/1/e44104
doi:10.2196/44104
PMID:37027185

Original Paper

Reliability and Validity of Noncognitive Ecological Momentary Assessment Survey Response Times as an Indicator of Cognitive Processing Speed in People's Natural Environment: Intensive Longitudinal Study

Raymond Hernandez[1], PhD; Claire Hoogendoorn[2,3], PhD; Jeffrey S Gonzalez[2,3], PhD; Haomiao Jin[1,4], PhD; Elizabeth A Pyatak[5], PhD; Donna Spruijt-Metz[1,6,7], PhD; Doerte U Junghaenel[1,7], PhD; Pey-Jiuan Lee[1], MS; Stefan Schneider[1,7], PhD

[1]Center of Economic and Social Research, University of Southern California, Los Angeles, CA, United States

[2]Ferkauf Graduate School of Psychology, Yeshiva University, Bronx, NY, United States

[3]Fleischer Institute for Diabetes and Metabolism, Division of Endocrinology, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, United States

[4]School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[5]Chan Division of Occupational Science and Occupational Therapy, University of Southern California, Los Angeles, CA, United States

[6]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[7]Department of Psychology, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Raymond Hernandez, PhD
Center of Economic and Social Research
University of Southern California
635 Downey Way, VPD
Los Angeles, CA, 90089
United States
Phone: 1 213 821 1899
Email: hern939@usc.edu

Abstract

Background: Various populations with chronic conditions are at risk for decreased cognitive performance, making assessment of their cognition important. Formal mobile cognitive assessments measure cognitive performance with greater ecological validity than traditional laboratory-based testing but add to participant task demands. Given that responding to a survey is considered a cognitively demanding task itself, information that is passively collected as a by-product of ecological momentary assessment (EMA) may be a means through which people's cognitive performance in their natural environment can be estimated when formal ambulatory cognitive assessment is not feasible. We specifically examined whether the item response times (RTs) to EMA questions (eg, mood) can serve as approximations of cognitive processing speed.

Objective: This study aims to investigate whether the RTs from noncognitive EMA surveys can serve as approximate indicators of between-person (BP) differences and momentary within-person (WP) variability in cognitive processing speed.

Methods: Data from a 2-week EMA study investigating the relationships among glucose, emotion, and functioning in adults with type 1 diabetes were analyzed. Validated mobile cognitive tests assessing processing speed (Symbol Search task) and sustained attention (Go-No Go task) were administered together with noncognitive EMA surveys 5 to 6 times per day via smartphones. Multilevel modeling was used to examine the reliability of EMA RTs, their convergent validity with the Symbol Search task, and their divergent validity with the Go-No Go task. Other tests of the validity of EMA RTs included the examination of their associations with age, depression, fatigue, and the time of day.

Results: Overall, in BP analyses, evidence was found supporting the reliability and convergent validity of EMA question RTs from even a single repeatedly administered EMA item as a measure of average processing speed. BP correlations between the Symbol Search task and EMA RTs ranged from 0.43 to 0.58 ($P<.001$). EMA RTs had significant BP associations with age ($P<.001$), as expected, but not with depression ($P=.20$) or average fatigue ($P=.18$). In WP analyses, the RTs to 16 slider items and all 22 EMA items (including the 16 slider items) had acceptable (>0.70) WP reliability. After correcting for unreliability in multilevel models, EMA RTs from most combinations of items showed moderate WP correlations with the Symbol Search task (ranged from 0.29 to 0.58; *P*<.001) and demonstrated theoretically expected relationships with momentary fatigue and the time of day. The associations between EMA RTs and the Symbol Search task were greater than those between EMA RTs and the Go-No Go task at both the BP and WP levels, providing evidence of divergent validity.

Conclusions: Assessing the RTs to EMA items (eg, mood) may be a method of approximating people's average levels of and momentary fluctuations in processing speed without adding tasks beyond the survey questions.

(JMIR Mhealth Uhealth 2023;11:e45203) doi:10.2196/45203

KEYWORDS

cognitive performance; processing speed; ecological momentary assessment; ambulatory assessment; type 1 diabetes; survey response times; paradata; chronic illness; smartphone; mobile health; mHealth; mobile phone

Introduction

Background

Various illnesses are risk factors for decreased cognitive performance, including type 1 diabetes (T1D), type 2 diabetes, depression, and cardiovascular disease [1-4], making the assessment of cognition in these populations important. Although the measurement of the cognitive performance of individuals with various chronic illnesses in their real-world environments is potentially useful (eg, as it provides ecologically valid measures) [5], formal ambulatory cognitive assessment is at times infeasible because of limited resources and time demands already placed on participants from other ambulatory assessment tasks. We attempt to capitalize on the growing use of the ecological momentary assessment (EMA) methodology in behavioral health research [6-8] and propose a novel approach for assessing a central cognitive measure, processing speed, using EMA survey paradata. Paradata are data about the response process, such as response times (RTs) when answering survey questions [9,10], that can be passively captured alongside survey responses. Measuring cognitive performance via EMA paradata could create novel opportunities for researchers to examine respondents' real-time cognitive performance in studies in which formal ambulatory cognitive testing cannot be readily implemented. This, in turn, could allow for a more frequent investigation of the antecedents, correlates, and consequences of changes in processing speed across a wider range of individuals and populations with chronic conditions.

Several behavioral health-focused EMA studies have used formal ambulatory cognitive tests in various populations, including individuals with T1D [11-13], breast cancer survivors [14], and people with fibromyalgia [15]. Formal ambulatory cognitive assessments have been viewed as a gold standard for capturing people's cognitive performance in their natural environment and overcome several limitations of traditional cognitive testing, including the ability to represent cognitive performance in real-world settings, increased frequency with which tests can be administered, and the ability to capture changes over short time frames [5]. However, formal ambulatory cognitive assessments often require more time to complete than other EMA measures. For instance, assessing a single aspect of cognition often requires 45 to 60 seconds [5,16], whereas the measurement of constructs such as stress requires the completion of a single item. Therefore, to administer formal ambulatory cognitive tests, researchers must at times limit the number of survey items included to keep the overall time to complete EMA surveys manageable. They also often require a costly setup and the use of specific programs or apps, which can be obstacles to implementation for many researchers. The difficulties in implementing formal ambulatory assessments limit our ability to more frequently investigate cognitive performance in everyday real-world settings in populations with chronic conditions. When considering T1D specifically, ambulatory cognitive performance has rarely been assessed [12,13,16], limiting our understanding of time-varying correlates and ultimately our understanding of the multifactorial pathways connecting diabetes to cognitive performance and decline.

Cognitive performance may potentially be inferred from EMA survey paradata (eg, RTs to mood items) and thus could be a means to approximate people's momentary cognitive functioning in their natural environments without the time demand of additional formal cognitive testing. Paradata have been investigated as a means of approximating the cognitive aspects underlying survey responding in traditional (non-EMA) survey studies. For instance, a study examining survey response behaviors at older ages found that the time of survey initiation and time of survey completion were related to mild cognitive impairment [8]. In another study, survey RTs and answer changes were operationalized as indicators of cognitive effort [9]. However, to date, very few studies have investigated the potential use of paradata in EMA surveys as indicators of cognitive performance [10]. One study found a moderate association between the time to complete EMA surveys and processing speed [10] but did not examine the effect of the types of EMA items or the degree of within-person (WP) reliability of EMA RTs.

This Study

The purpose of this study was to investigate whether the RTs to noncognitive EMA questions (eg, mood, stress, and activity done) can serve as approximate indicators of person-level differences and momentary WP fluctuations in processing speed. Aside from individuals' average processing speed (person-level differences), momentary fluctuations in processing speed may also be useful to assess via EMA survey paradata. For instance, intraindividual cognitive variability increases with age, even among those who remain cognitively healthy [17], and variability is a risk factor for mild cognitive impairment and Alzheimer disease and related dementias [18-22], even after adjusting for average performance [18]. In addition, WP variability in processing speed, as measured by formal cognitive tests, has been shown to be affected by momentary factors, including caffeine consumption [23], social context [24], and fatigue [14]. We acknowledge that RTs to survey items have been proposed to contain different types of information, including the level of cognitive effort invested [25]; processing speed [26]; and, for self-reports of current mood, the level of emotional clarity [27]. Therefore, survey item RTs likely reflect a combination of several factors. Thus, we did not expect a complete overlap between RTs and the results of mobile cognitive testing but did expect a substantial association.

We capitalized on preexisting data from an EMA study in which adults with T1D completed 2 weeks of EMA surveys together with smartphone-based mobile processing speed and sustained attention tests [16]. Cognitive tests provided validated processing speed and attention measures against which we compared the EMA survey RTs. The makeup of the sample, adults with T1D, allowed for analyses in a sample for which processing speed may be especially relevant. In individuals with T1D, previous studies have found relationships between blood glucose metrics and cognitive performance, including processing speed [11,28,29].

As the primary test of convergent validity, we examined the associations between the RTs to different subsets of EMA items and the scores of mobile processing speed tests. We hypothesized that if EMA survey RTs captured processing speed, slower RTs would be associated with worse performance on the formal processing speed test, both at the between-person (BP) and WP levels.

As secondary tests of convergent validity, we examined the associations between the RTs to different subsets of EMA items and depression symptoms, age, fatigue, and a diurnal cycle. In a review, individuals with major depression were found to have lower processing speed than controls [30], so we expected greater depression symptoms to be associated with slower mean EMA survey RTs. Aging has often been linked to decreased processing speed through various neurobiological pathways [31,32], so greater age was hypothesized to be associated with slower mean EMA survey RTs. We hypothesized that fatigue would have associations with EMA survey RTs at both the BP and WP levels. At the BP level, chronic fatigue syndrome has been associated with slower overall processing speed [33]. At the WP level, processing speed was slower among breast cancer survivors reporting higher than usual fatigue [14]. Finally, cognitive abilities have been found to be at the lowest level during early morning and nighttime, increasing throughout the day until evening [34]. EMA survey RTs were expected to follow a similar diurnal pattern.

To assess divergent validity, we tested the association between EMA item RTs and sustained attention ability. We anticipated that if EMA RTs are indicators of processing speed, they would have stronger associations with processing speed than with sustained attention ability. Although both processing speed and sustained attention ability are fundamental cognitive skills, they are distinct aspects of cognitive performance that are measured using different tests [5,35]. For instance, sustained attention tests are often scored for accuracy, whereas processing speed tests are scored for speed [36].

Methods

Study Design

The goal of the EMA study from which data were analyzed was to investigate the relationships among momentary emotion, function, and glucose, the full methodology of which has been outlined previously [16]. Participants were recruited from 3 clinical sites, and the inclusion criteria were as follows: age of >18 years, familiarity with using a smartphone, and sufficient visual acuity, cognitive ability, and manual dexterity to complete study tasks, such as processing speed tests [16]. Consent to participate was provided on the web through the REDCap (Research Electronic Data Capture; Vanderbilt University) e-consent framework [37]. Study procedures included the completion of baseline surveys; 2 weeks of phone-based EMA and cognitive testing with 5 to 6 assessments per day, wearing a continuous glucose monitor and accelerometer during the EMA period; and follow-up surveys.

Ethics Approval

The data collection procedures were approved by the University of Southern California institutional review board (proposal #HS-18-01014).

Measures

EMA Item RTs

RTs to the 22 noncognitive EMA survey items listed in Table 1 were examined as potential processing speed indicators. All items were derived from validated measures or used in prior EMA research [16]. The items were presented one at a time on study phone screens via the mobile EMA app [38]. The participants were not informed that their EMA survey RTs were being measured. Whether participants should be informed of the collection of paradata continues to be debated [39]. RTs were recorded in seconds (to three decimal places) for each item. Values of <0.2 seconds or >30 seconds were considered missing in analyses (5979/461,896, 1.29% of observations) because prior literature suggested that ultrafast EMA RTs are likely indicative of careless responding [40] and because RTs of >30 seconds were deemed outliers that may have been caused by disruptions in completing the survey. Log transformation was applied to the RTs to create more normal distributions.

Table 1. Ecological momentary assessment items from which response times were analyzed.

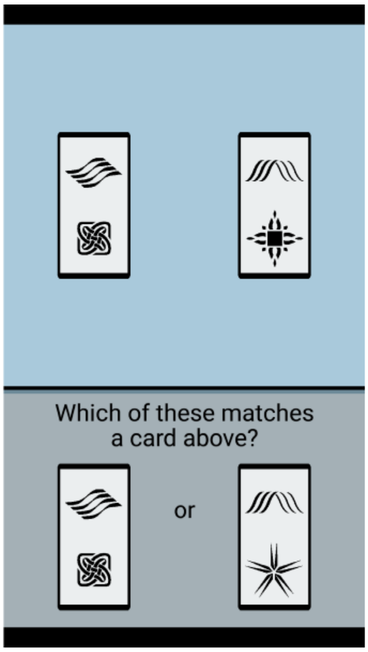
Item type and question or questions	Response option or options
16 slider scale items	
Positive affect	0 (not at all) to 100 (extremely)
Right now, how content do you feel?	
Right now, how happy do you feel?	
Right now, how excited do you feel?	
Right now, how enthusiastic do you feel?	
Negative affect	0 (not at all) to 100 (extremely)
Right now, how disappointed do you feel?	
Right now, how sad do you feel?	
Right now, how upset do you feel?	
Right now, how anxious do you feel?	
Activity engagement (with reference to the activity the participant reporting doing right before the survey)	
How well were you able to do this activity?	0 (unable) to 100 (extremely well)
How satisfied are you with the way you did this activity?	0 (not satisfied) to 100 (extremely satisfied)
How important is this activity to you?	0 (not important) to 100 (extremely important)
Stress	
How stressed are you right now?	0 (not at all stressed) to 100 (extremely stressed)
How stressed do you feel about your diabetes or diabetes management right now?	0 (not at all stressed) to 100 (extremely stressed)
Right now, how tense do you feel?	0 (not at all) to 100 (extremely)
Fatigue	
At this moment, how tired do you feel?	0 (not at all) to 100 (extremely)
Pain	
At this moment, how much bodily pain do you have?	0 (none) to 100 (extreme pain)
3 multiple-choice items	
Activity done	
What were you doing right before starting this survey?	10 choices (eg, work and relaxing)
Where activity was done	
Where were you when doing this activity?	5 choices (eg, home)
Subjective blood sugar level	
How does your blood sugar feel right now?	Likert 0-4: very low, low, just right, high, and very high
3 checkbox items	
With whom	
Who were you with when doing this activity?	8 choices (eg, alone and friend)
Diabetes intrusiveness	
Did your diabetes get in the way of doing this activity?	4 choices (eg, no and yes because of my devices)
Eat or drink	
Did you eat or drink in the last 3 hours?	Ate, drank, and neither

Our primary set of noncognitive EMA survey RTs was the RTs to the 16 slider items, but other noncognitive EMA survey RT combinations were also examined. To compare the reliability and validity of EMA RTs across types of EMA items, we classified the items by response option type and further by item content domains, where possible, as presented in Table 1. We note that the EMA items were chosen to serve the goal from the overarching study of investigating the relationships between various measures and blood glucose metrics and were not deliberately crafted to ensure a sufficient range of item characteristics. The 16 slider items were designated as primary because they comprised the largest group of items with the same type of response options and thus seemed most likely to have the highest reliability among the EMA item groups. As there were relatively few multiple-choice and checkbox items, we did not create subgroups by item content for the items with these response options. Grouping was applied to the extent feasible to allow for the investigation of the impact of item characteristics on the validity and reliability of EMA RTs as indicators of processing speed. Given the prior findings that the content of survey items and the type of response options may affect RTs [41,42], we carefully examined the extent to which the association between EMA survey RTs and processing speed differed according to these item characteristics.

Processing Speed Test

At the end of each EMA survey, participants were prompted to complete the Symbol Search task, a phone-based mobile processing speed test [5]. This task has previously demonstrated construct validity and was found to correlate at r=0.74 with its standard laboratory version [5]. In this task, participants were presented with 2 cards at the top and bottom of the screen, each with 2 symbols (Figure 1). They were asked to choose the card at the bottom of the screen that matched the card on top as quickly as they could for 20 trials. One Symbol Search session was approximately 45 seconds long. When Symbol Search RTs were <0.2 or >5 seconds (9557/287,543, 3.32% of observations) or they were part of sessions with <70% accuracy (a cutoff for inattentive responding; 2956/287,543, 1.03% of observations) [43], they were considered missing (12,513/287,543, 4.35% of observations).

Figure 1. Processing speed test (Symbol Search) on study phone screen, where participants are asked to choose a card on the bottom that matches the card on top as quickly as they can for 20 trials.



Other Measures

The measures of fatigue, depression, age, and the time of day were also used for validity testing. The momentary fatigue item (ie, "At this moment, how tired do you feel?") was similar to that used in a previous EMA study [14]. Depression was measured using the Patient Health Questionnaire-8 (PHQ-8) [44] at baseline, along with age. The time of day when each ambulatory assessment was completed was automatically recorded using the EMA app. Sustained attention ability was assessed using an ambulatory cognitive assessment, the Go-No Go task [35], which was administered immediately before the Symbol Search task. In the Go-No Go task, the participants were shown a series of 75 images of a mountain or city, presented one at a time for 800 ms. They were asked to tap an indicated button when seeing an image of a city but to abstain from tapping when presented with an image of a mountain. The measure d prime (d') was computed as the sustained attention ability score, a metric that considers both the number of correct city taps and incorrect mountain taps using a signal detection approach [35].

Statistical Analyses

Analysis Strategy Overview

Multilevel modeling was used to evaluate (1) the WP and BP reliability of EMA RTs, (2) WP and BP correlations of EMA RTs with the Symbol Search task (primary convergent validity test), and (3) WP and BP associations of EMA RTs with other constructs (secondary convergent validity and divergent validity tests). Analyses of reliability and convergent validity were conducted in a parallel fashion for EMA RTs and for the Symbol Search task. Thus, the results of the reliability and validity testing of EMA RTs could be directly compared with the findings from the Symbol Search analyses. For instance, the magnitude of the correlation between EMA RTs and fatigue can be compared with the size of the association between the Symbol Search task and fatigue.

To examine the extent to which the characteristics of EMA items may impact the reliability and validity of EMA RTs, analyses were conducted using the RTs to all 22 EMA items, as well as using the RTs to subgroups of EMA items and the RTs to individual items. The groupings of item RTs were as presented in Table 1: 16 slider scale items, 4 positive affect slider items, 4 negative affect slider items, 3 activity engagement slider items, 3 stress slider items, 3 multiple-choice items, and 3 checkbox items. The 16 slider items consisted of the slider item subgroups and slider items addressing fatigue and pain in combination. Single-item EMA RTs from each of the 22 EMA items were also included in the analyses as the lower benchmarks of the reliability and validity of EMA RTs provided by this minimal information source and to examine the extent to which the various individual item RTs provided similar or markedly different information when used as an indicator of processing speed.

Reliability

Both the BP and WP reliabilities of EMA RTs and the Symbol Search task were estimated. BP reliability describes the consistency of a person's mean value across all measurement occasions of a given measure (ie, RTs) [45]. It can be calculated as BP reliability = Var(BP)/(Var(BP) + Var(WP)/n) [46], where Var(BP) is the BP variance in the average of scores across measurement occasions, Var(WP) is the variance of scores across measurement occasions within a person, and n is the number of measurement occasions. The equation implies that greater WP variation in EMA RTs decreases the consistency of the average of RTs across all measurement points and that a greater number of measurements (eg, more EMA surveys) increases reliability. An average of 70 surveys were completed over the 2 weeks of the study, and we estimated the BP reliabilities for increasing numbers of measurement occasions (from 2 to 70). This allowed the examination of how BP reliability increased as a function of the number of ambulatory assessments. The variance components Var(BP) and Var(WP) were estimated using a 2-level multilevel model, in which measurement occasions were nested in participants. BP intraclass correlation coefficients (ICCs) were also computed for each measure, which represent BP reliabilities associated with a single measurement occasion (ie, n=1).

To estimate the WP reliabilities of the measures, we capitalized on the fact that the RTs from multiple EMA items (and from multiple Symbol Search trials) were available at each measurement occasion (ie, at each EMA prompt). WP reliability is the consistency of the mean RT across EMA items within a single measurement occasion. The formula is WP reliability = $Var(WP_{occasion})/(Var(WP_{occasion}) + Var(WP_{trial})/i)$ [47], where $Var(WP_{occasion})$ is the variance within a person across different measurement occasions, $Var(WP_{trial})$ is the variance of RTs across the EMA items administered within a given measurement occasion, and *i* is the number of EMA items. The variance components $Var(WP_{occasion})$ and $Var(WP_{trial})$ [48] were estimated using 3-level multilevel models (EMA items or trials nested in measurement occasions nested in people).

Validity

As the primary convergent validity test, the WP and BP correlations of EMA RTs with the Symbol Search task were examined using bivariate multilevel models, in which both measures were entered as bivariate (ie, correlated) dependent variables. Specifically, 3-level models (items nested in measurement occasions nested in people) were specified, whereby the WP correlation was estimated at level 2 and the BP correlation was estimated at level 3. This had the advantage that the correlations at both levels were adjusted for the unreliability due to variance in RTs within measurement occasions (estimated at level 1). For exploratory purposes, 2-level models were also examined to allow for comparison with the results from the 3-level models. In these 2-level models, rather than estimating the variance in RTs within measurement occasions at level 1, we used the observed (manifest) average of RTs.

Secondary convergent validity and divergent validity tests were conducted similarly with 3-level multilevel models. As fatigue ratings and Go-No Go (sustained attention ability) varied both within and between individuals, we estimated both between-individual and within-individual correlations of fatigue and Go-No Go with EMA RTs (and, for comparison, with the Symbol Search task, examined in separate models). For the BP variables age and depression, we estimated only BP correlations with EMA RTs (and with the Symbol Search task, in separate models). The diurnal cycle of EMA RTs was also examined to test whether the pattern was consistent with previous research on the diurnal cycle of cognitive performance. A multilevel cosinor model [49] was used, in which EMA RTs for a measurement occasion were regressed on the sine and cosine of the hour (0-24 hours) during which the survey was conducted. A 3-level multilevel model (EMA items nested in measurement occasions nested in individuals) was used again, in which EMA RTs were regressed on the sine and cosine of the time of day at level 2 to estimate WP changes in EMA RTs by the time of day. For comparison, a cosinor model was also tested for the Symbol Search task. All reliability and validity analyses were conducted in M*plus* (version 8.8; Muthén & Muthén) [50] with the R package *MplusAutomation* [48] in the statistical software R (R Foundation for Statistical Computing) [51].

Results

Sample Characteristics

The analyses were conducted on data from 198 participants (Table 2). A total of ≥4 EMA prompts (together with Symbol Search assessments) were completed on 81.9% (2321/2834) of the data collection days pooled across all participants. The median EMA completion rate over the 2-week study period was 92%. The mean score on the PHQ-8 was 5.44 (SD 4.30), with scores of >9 indicating moderate or more severe depressive symptoms. Overall, 15.7% (31/198) of the participants had PHQ-8 scores of >9. In terms of fatigue, the mean level reported in EMA was 42.70 (SD 18.60), with ratings given on a scale ranging from 0 to 100.

Table 2. Demographic and health characteristics (n=198).

Characteristic	Values
Age (years), mean (SD; range)	39.8 (14.4; 18-75)
Gender, n (%)	
Men	89 (44.9)
Women	109 (55.1)
Ethnicity, n (%)	
White	57 (28.8)
Latino	81 (40.9)
African American	29 (14.6)
Multiethnic	14 (7.1)
Asian	7 (3.5)
Other	6 (3)
Not reported	4 (2)
Preferred language, n (%)	
English	177 (89.4)
Spanish	21 (10.6)
Employment status, n (%)	
Full time	70 (35.4)
Part time	23 (11.6)
Full-time homemaker	10 (5.1)
Student	18 (9.1)
Unemployed	27 (13.6)
Retired	15 (7.6)
Disabled	23 (11.6)
Other	8 (4)
Not reported	4 (2)
Education, n (%)	
High school graduate or less	50 (25.3)
Some college	68 (34.3)
Bachelor's degree	55 (27.8)
Graduate degree	22 (11.1)
Not provided	3 (1.5)
Annual household income (US $), n (%)	
<25,000	48 (24.2)
25,000-49,999	44 (22.2)
50,000-74,999	15 (7.6)
≥75,000	40 (20.2)
Not provided	51 (25.8)
Average blood glucose over at least 10 days of CGM[a] data (mg/dL; n=154), mean (SD; range)	183.6 (55.0; 98.5-419.8)
Time since T1D[b] diagnosis (years; n=195), mean (SD; range)	20.9 (12.6; 1-57)
Insulin delivery system, n (%)	
AID[c]	45 (22.7)

Characteristic	Values
Non-AID CGM	70 (35.4)
No CGM	83 (41.9)

[a]CGM: continuous glucose monitor.

[b]T1D: type 1 diabetes.

[c]AID: automated insulin delivery.

Reliability

The reliabilities of the Symbol Search task and EMA RTs are listed in Table 3. The Symbol Search task had a BP ICC of 0.71, suggesting acceptable BP reliability from 1 assessment. With 3 measurement occasions, the BP reliability for the Symbol Search task increased to 0.88. The WP reliability of the Symbol Search task was 0.76, indicating acceptable consistency (>0.7) [52] of the RTs within a single Symbol Search measurement occasion.

Table 3. Reliability of the response times to the Symbol Search task and different sets of ecological momentary assessment (EMA) items.

	20 SS[a] trials	16 slider items	4 positive affect items	4 negative affect items	3 activity engagement items	3 stress items	3 MC[b] items, 5-10 choices	3 checkbox items, 3-8 boxes	Slider, MC, and check box items
ICC[c]	0.71	0.56	0.44	0.43	0.42	0.44	0.31	0.39	0.56
Between-person reliability of the mean of 3 EMA surveys	0.88	0.79	0.70	0.70	0.69	0.70	0.57	0.66	0.79
Between-person reliability of the mean of 70 EMA surveys	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.99
Within-person reliability	0.76	0.82	0.58	0.59	0.60	0.49	0.08	0.22	0.80

[a]SS: Symbol Search (higher values indicate worse processing speed).

[b]MC: multiple-choice.

[c]ICC: intraclass correlation coefficient.

EMA RTs from the 16 slider items had a BP ICC of 0.56, which was lower than the BP Symbol Search ICC. With 3 measurement occasions, the BP reliability of EMA RTs from the 16 slider items increased to 0.79. The WP reliability of EMA RTs from the 16 slider items was 0.82, which was slightly higher than that of the Symbol Search task. The BP and WP reliability values of EMA RTs from all 22 EMA items were nearly identical to those from the 16 slider items.

In terms of RTs from other sets of EMA items comprising 3 to 4 questions, BP ICCs ranged from 0.31 for the 3 multiple-choice items to 0.44 for the 4 positive affect slider items. With 3 measurement occasions, only the BP reliability of the RTs for the 3 multiple-choice items was notably <0.70, with a value of 0.57. WP reliability for the EMA RTs of the slider item subgroups ranged from 0.49 to 0.60. RTs for the 3 multiple-choice and 3 checkbox items had much lower WP reliabilities, with values of 0.08 and 0.22, respectively.

Figure 2 depicts how the BP reliability for both the Symbol Search task and EMA RTs to 16 slider items varies as a function of the number of measurement occasions. For the Symbol Search task, just 1 measurement occasion is sufficient for a BP reliability of at least 0.70. RTs for 16 slider items crossed the threshold of 0.70 BP reliability upon the completion of 2 EMA surveys.

Figure 2. Between-person reliability of the Symbol Search task and ecological momentary assessment (EMA) response times to 16 slider items by the number of measurement occasions.

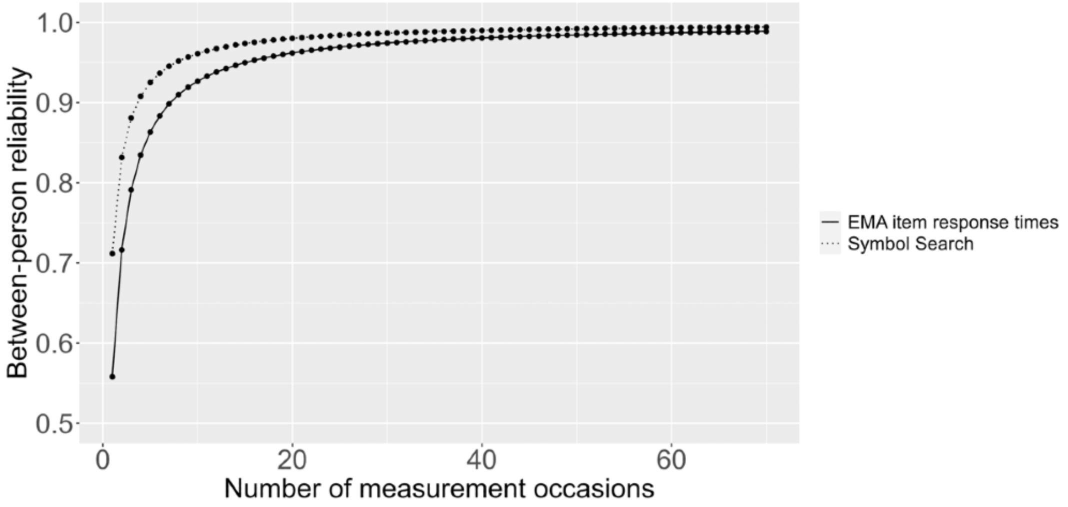
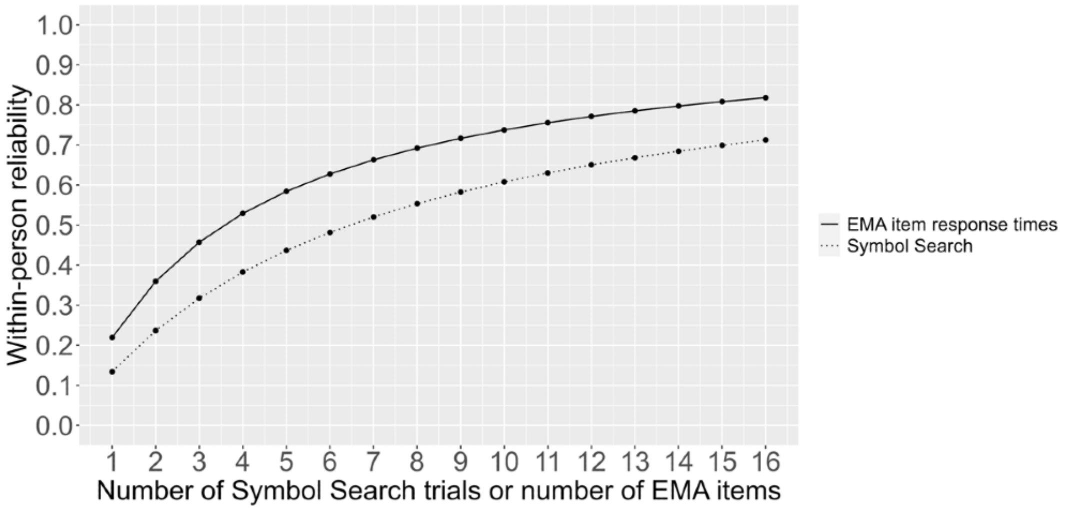


Figure 3 shows the WP variability of the Symbol Search task and RTs to 16 EMA slider items as a function of the number of Symbol Search trials or EMA items completed. Increasing the number of EMA RTs to items of the same type (here slider) steadily increased WP reliability, and the rate of increase roughly mirrored that of the Symbol Search task. For the Symbol Search task, 16 trials were required for a WP reliability of at least 0.70. EMA RTs for 16 slider items crossed the threshold of 0.70 WP reliability upon the completion of 9 items.

Figure 3. Within-person reliability of the Symbol Search task and ecological momentary assessment (EMA) response times to 16 slider items by the number of trials or the number of EMA items completed. Note that each Symbol Search session had 20 trials, but only the reliability of up to 16 trials was plotted in the figure to correspond with the 16 slider items.



BP reliabilities for the RTs to single EMA items are presented in Tables S1 and S2 in Multimedia Appendix 1. Note that using RTs from a single item does not allow for the calculation of WP reliability. Overall, the BP reliabilities of single items from a single EMA measurement occasion (ICC) ranged from 0.17 for the multiple-choice activity engagement item to 0.36 for the diabetes stress item. For the average of 3 measurement occasions, the BP reliabilities of the RTs to single EMA items ranged from 0.50 to 0.63. For the average of 7 measurement occasions, BP reliabilities were at least 0.7 for all items except the multiple-choice items.

Validity

Associations Between EMA RTs and the Symbol Search Task

At the BP level, the correlation between the Symbol Search task and EMA RTs was 0.58 when all EMA items were used, and correlations ranged from 0.49 to 0.57 when subsets of EMA items were used to estimate person-level average EMA RTs (Table 4). At the WP level, medium associations were found between the Symbol Search task and EMA RTs for all item sets except the multiple-choice items, with correlations ranging from r=0.29 (P<.001) to r=0.40 (P<.001); multiple-choice items had a larger correlation with the Symbol Search task (r=0.58, P<.001; refer to the "within-person correlations" category in Table 4).

Table 4. Between-person (person-level) correlations between response times from different items (columns) and other variables (rows) as calculated from 3 level models.

	20 SS[a] trials	16 slider items	4 PA[b] slider items	4 NA[c] slider items	3 activity slider items	3 Stress slider items	3 MC[d] items	3 check items	22 slider, MC, check
Between-person correlations									
SS									
r	1.00	0.55	0.54	0.55	0.49	0.52	0.57	0.55	0.58
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Fatigue									
r	−0.07	−0.03	−0.10	0.00	−0.03	−0.01	0.00	0.00	−0.03
P value	.18	.28	.08	.47	.37	.46	.49	.50	.37
Depression									
r	0.06	0.05	0.02	0.05	0.06	0.03	0.10	0.07	0.05
P value	.20	.26	.38	.23	.23	.32	.08	.17	.29
Age									
r	0.42	0.52	0.49	0.53	0.47	0.54	0.52	0.54	0.54
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
GNG[e]									
r	−0.15	0.17	0.16	0.1	0.11	0.2	−0.08	−0.03	0.1
P value	<.001	.02	<.001	.08	.07	<.001	.16	.33	.04
Within-person correlations									
SS									
r	1.00	0.35	0.30	0.29	0.35	0.35	0.58	0.40	0.37
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Fatigue									
r	0.14	0.08	0.06	0.08	0.05	0.06	0.05	0.05	0.07
P value	<.001	<.001	<.001	<.001	<.001	<.001	.11	<.001	<.001
GNG									
r	−0.04	0	−0.01	−0.03	0.01	0.02	0.02	−0.03	0
P value	<.001	.45	.14	.02	.18	.08	.29	.06	.41

[a]SS: Symbol Search (higher values indicate worse processing speed).

[b]PA: positive affect.

[c]NA: negative affect.

[d]MC: multiple-choice.

[e]GNG: Go-No Go (higher values indicate better sustained attention ability).

The WP and BP correlations between EMA RTs from single items and the Symbol Search task are presented in Table S3 in Multimedia Appendix 1. To summarize, BP correlations with the Symbol Search task ranged from 0.43 to 0.58, close in magnitude to the BP correlations with RTs from the larger EMA item sets. The WP correlations ranged from 0.09 to 0.21.

Secondary Convergent Validity and Divergent Validity Tests

At the BP level, neither the EMA RTs from different sets of items nor the Symbol Search task were significantly associated with average fatigue (P=.18) or depression ratings (P=.20), contrary to our hypothesis (Table 4). Older age was significantly positively correlated with worse Symbol Search RTs (r=0.42; P<.001); and age was similarly correlated with EMA RTs, with magnitudes ranging from 0.47 to 0.54 (P<.001). Consistent with our hypothesis, EMA RTs were more highly correlated with the Symbol Search scores (r values ranging from 0.49 to 0.58; P<.001) than with the Go-No Go scores (r values ranging from −0.08 to 0.20; 4 of 8 nonsignificant P values of .08, .07, .16, and .33).

At the WP level, the correlations were overall consistent with our hypotheses (Table 4). Worse Symbol Search RTs were significantly associated with greater momentary fatigue levels (r=0.14; P<.001). Slower RTs for EMA items were similarly associated with greater momentary fatigue, and this relationship was significant for RTs from all sets of EMA items (*r* values ranging from 0.05 to 0.08; *P*<.001), except for those from multiple-choice items (*P*=.11). EMA RTs were again more highly correlated with the Symbol Search scores (*r* values ranging from 0.29 to 0.58; *P*<.001) than with the Go-No Go scores (*r* values ranging from −0.03 to 0.02; 7 of 8 nonsignificant *P* values of .45, .14, .18, .08, .29, .06, and .41).

Tables S4 and S5 in Multimedia Appendix 1 show the BP and WP correlations among the study measures, as calculated from 2-level models instead of the 3-level models presented earlier. The greatest difference is that WP correlations between the Symbol Search task and EMA RTs were somewhat lower in the 2-level models (eg, *r*=0.27 in a 2-level model vs *r*=0.35 in a 3-level model for the 16 EMA slider items). RTs from multiple-choice EMA items showed the biggest difference in WP correlation with the Symbol Search task when comparing 2-level and 3-level models (*r*=0.18 in a 2-level model vs *r*=0.58 in a 3-level model).

The diurnal cycle of EMA RTs was examined and compared with that of the Symbol Search RTs. As shown in Figure 4, the average Symbol Search RTs were lowest around 3 PM to 4 PM and highest in the morning and evening. The standardized amplitude of the diurnal cycle was 0.34 *z* scores (SE 0.03; *P*<.001), which translated to RT fluctuations of 0.34 × 2 = 0.68 SDs within a day, corresponding to a medium to large effect size [53]. EMA RTs had a similar but less pronounced diurnal cycle (Figure 5). RTs to the 16 slider EMA items were, on average, slowest during the early morning and evening and fastest from 2 PM to 4 PM. The standardized amplitude of the diurnal cycle was 0.16 (SE 0.02; *P*<.001), meaning that RTs fluctuated by approximately 0.16 × 2 = 0.32 SDs (*z* scores) within a day, corresponding to a small effect size [53]. Diurnal plots for the other EMA item sets demonstrated similar trends (not shown here).

Figure 4. The mean Symbol Search response times (RTs) during typical waking hours (6 AM-12 AM), the period during which most surveys were completed. The black line is the predicted Symbol Search RT from the cosinor model, the band is the 95% CI of the predicted RTs, and the red dots are the observed averages of the Symbol Search RTs.

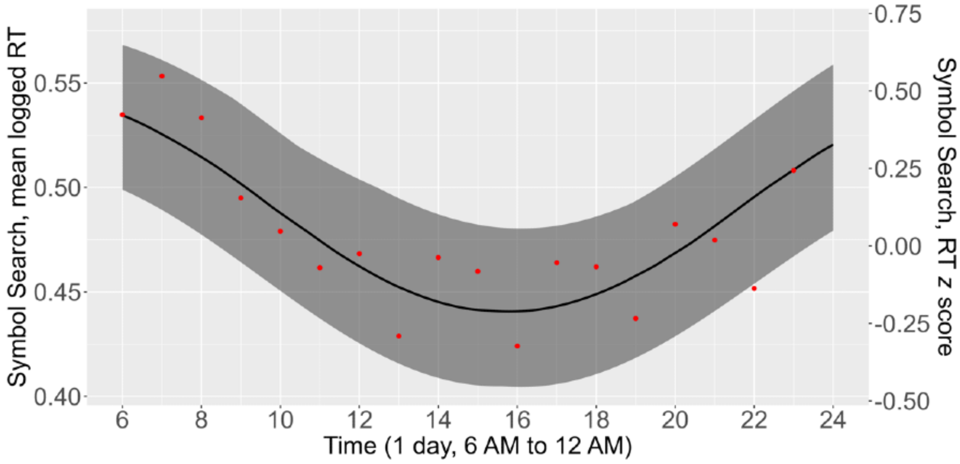
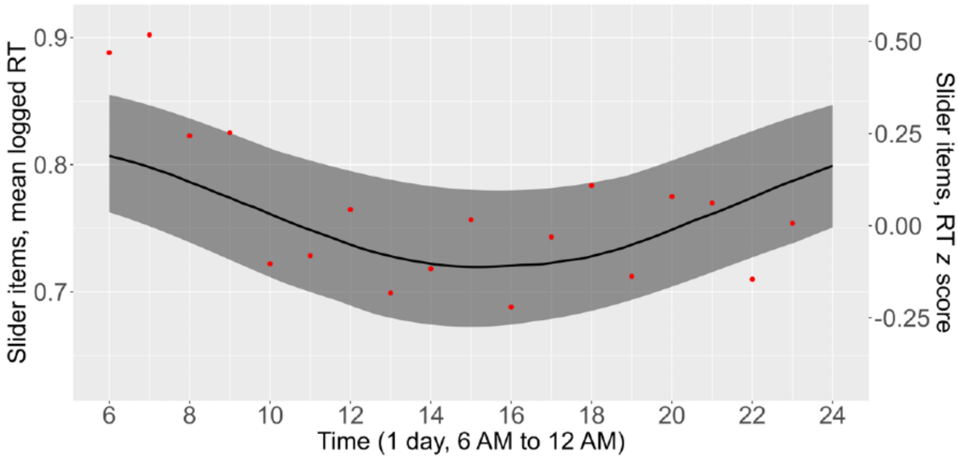


Figure 5. The mean response times (RTs) to 16 slider scale items during typical waking hours (6 AM-12 AM), the period during which most surveys were completed. The black line is the slider scale RT from the cosinor model, the band is the 95% CI of the predicted RTs, and the red dots are the observed averages of the slider scale RTs.



Discussion

Overview

Overall, our results suggest that EMA RTs can serve as approximate indicators of both momentary and average processing speeds. The EMA RTs for the items analyzed in this study were better indicators of average, as compared with momentary, processing speed. A formal processing speed test (Symbol Search) and EMA RTs had correlations of approximately 0.5 at the BP level and 0.3 at the WP level. These correlation sizes may be acceptable in research contexts in which investigators do not have the resources for administering formal cognitive testing. Correlations of these magnitudes may be sufficient to detect strong associations with processing speed, although weaker associations may be missed. The findings of the reliability and validity tests are described in greater detail in the subsequent sections.

Reliability

Overall, EMA RTs showed acceptable reliability under conditions (ie, number of items and measurement occasions) typical of many EMA studies. Furthermore, BP reliability for EMA RTs were similar to that for Symbol Search RTs, and WP reliability for RTs to the 16 EMA slider items slightly exceeded that for the Symbol Search task.

The BP reliability of EMA RTs from various item sets differed largely as a function of the number of EMA measurement occasions. With just 3 EMA measurement occasions, BP reliability was acceptable (approximately 0.70), except for RTs from the 3 multiple-choice and 3 checkbox items. The lower BP reliability in these item sets may have been due to their greater heterogeneity (eg, different item content and number of response options), leading to greater variability (ie, more error variance) in mean RTs. For single items, except for multiple-choice questions, the average RTs had a reliability of at least 0.70 with 7 EMA measurement occasions. This suggests that with a relatively small number of EMA measurement occasions, even RTs from single EMA items will likely have acceptable BP reliability. For 16 slider items, the completion of just 2 EMA surveys was sufficient to cross the threshold of 0.70 BP reliability, likely because it was a larger set of items that shared the same type of response options. With more than 16 items sharing similar response options, it would perhaps be possible to obtain a reliable assessment of EMA survey RTs from just 1 measurement occasion.

In terms of the WP reliability of EMA RTs (consistency of RTs within a single EMA measurement occasion), the number of EMA questions and the type of response options appeared to be major contributing factors. The RTs to the 16 slider items had a WP reliability of 0.82, considerably higher than the WP reliability of the RTs to the item sets with only 3 or 4 items. These smaller item sets had reliability values between 0.08 and 0.60, limiting their precision for capturing WP changes in processing speed with 2-level (and not 3-level) modeling. Of the item sets with 3 or 4 items, the more homogenous sets (slider items by topic) had much greater WP reliability than the heterogeneous sets (multiple-choice and checkbox questions with differing numbers of response options). WP reliability improved after combining RTs from various slider items, suggesting that item content was not as important to reliability as the response option type. When considering RTs from the heterogeneous response option item sets, in addition to the 16 slider items, the WP reliability was similar to that found for the 16 slider items alone. Thus, for the WP reliability of EMA RTs, considering RTs from more items may not always be beneficial, specifically when the additional items have different response options.

Validity

Although some findings were contrary to our hypotheses (ie, no relationship between EMA RTs and depression or fatigue), the results appeared overall supportive of the validity of EMA item RTs as an approximate measure of processing speed. In our primary convergent validity test at the BP level, EMA RTs had moderate to large correlations with the Symbol Search task, a magnitude expected if EMA RTs were indicators of processing speed. In terms of secondary convergent validity tests at the BP level, observed relationships with EMA RTs were sometimes contrary to our hypotheses but were typically very similar to the associations seen with the Symbol Search task. At the BP level, we hypothesized that slower EMA RTs would be associated with greater average fatigue and greater depressive symptoms. Neither of these relationships was confirmed. For fatigue, this may have been because previous research found associations between slower processing speed and chronic fatigue syndrome (more severe than typical fatigue) [33], but the mean level of fatigue reported in the EMA in our sample may have been less severe (mean 42.70, SD 18.60; scale of 0 to 100). In terms of depression, the proportion of people in our sample with any severity of depression (ie, PHQ-8 scores of >9) was 15.7% (31/198), which was greater than the 8.58% (17,040/198,678) found in a previous study on the general population [44]. Associations between depressive symptoms and EMA RTs were trending in the theoretically expected direction (ie, more depressive symptoms were associated with slower processing speed), but our sample may have been underpowered to detect small BP relationships. Symbol Search RTs did not show significant relationships with either fatigue (P=.18) or depression (P=.20). Significant BP correlations were found between age and both RTs to EMA items (P<.001) and the Symbol Search task (P<.001). Consistent with our hypothesis, EMA RTs had a greater BP association with the Symbol Search task than with the Go-No Go task. Overall, we interpret the BP correlations as being supportive of the validity of EMA item RTs as approximate measures of processing speed.

At the BP level, higher sustained attention ability was correlated with greater processing speed as measured by the Symbol Search task and was generally weakly associated with EMA RTs. Interestingly, at the BP level, better sustained attention ability was associated with faster Symbol Search RTs but with slower EMA RTs for 4 of the 8 EMA RT item sets. We can only speculate why this might be the case. Perhaps participants with greater sustained attention ability were better able to process information quickly when they were explicitly instructed to respond as fast as possible (in the Symbol Search task), whereas they may have more deliberately read the EMA items and more carefully considered their responses, leading to slower RTs in EMA.

At the WP level, the results were also overall supportive of the validity of EMA item RTs as measures of processing speed. Most importantly, moderate or larger correlations between different sets of EMA RTs and the Symbol Search task were observed. The diurnal cycle of EMA RTs roughly approximated the daily pattern for Symbol Search RTs. Slower EMA RTs were associated with greater fatigue for all item sets, except for multiple-choice questions. Notably, fatigue was associated with processing speed and EMA RTs at the WP level but not at the BP level. It is possible that participants in our sample did at times experience fatigue (within level) but not frequently enough that the average level of fatigue experienced was associated with decrements in processing speed on average (between level). Consistent with our hypothesis, EMA RTs had greater WP associations with the Symbol Search task than with the Go-No Go task.

Although associations between EMA RTs and the Symbol Search task were observed, the relationships were not strong enough to argue that they provided identical measures. From the outset, we did not advocate for RTs of EMA items to serve as a replacement for formal cognitive testing; rather, we sought to examine whether they can serve as rough processing speed indicators when formal tests are not available. The tasks of completing a formal processing speed test and completing EMA items differ in several aspects. For instance, one common formal processing speed test is searching for a figure that matches a given image. In this task, RTs conceptually capture perceptual speed [5], a component of processing speed [31]. RTs in EMA may be more likely to capture decisional speed when faced with a moderately complex task (eg, answering survey items), a conceptually related but different processing speed indicator [31]. As another example, formal processing speed tests often have explicit performance and speed expectations, whereas EMA surveys do not, particularly if participants are not aware that their time to answer questions is being measured. EMA RTs may, therefore, be more affected by distractions because participants may assume that they can attend to distractors and then return to answering EMA questions at their own pace. Given the differences between formal processing speed tests and EMA surveys, their RTs and, by extension, their measures of processing speed were unlikely to correspond exactly with one another. However, because perceptual speed and decision speed both fall under the umbrella of processing speed [31], some associations were expected.

Validities of EMA RTs With Low Reliabilities

The RTs to single EMA items appeared to lack WP validity, as evidenced by low WP correlations with the Symbol Search task, but they may have some degree of BP validity with sufficient EMA measurement occasions. With an average of 70 EMA surveys completed, the BP association between the RTs to single EMA items and RTs to the Symbol Search task ranged from 0.43 to 0.58. The results of reliability analyses suggested that slightly more than 7 EMAs may result in a BP reliability of at least 0.7, indicating that a relatively small number of EMA instances (eg, 2 days with 4 EMA surveys daily) is sufficient to recover high BP associations with the Symbol Search task.

The RTs to the 3 multiple-choice items notably had the highest WP correlation with the Symbol Search task (r=0.58) but also a WP reliability much lower than other item sets. Three-level modeling helped to compensate for this low reliability by removing the errors from item-level RT variance, and the result was a much higher correlation compared with when a 2-level model was used (r=0.18). The practical implication may be that EMA RTs with low WP reliability, such as those from a few multiple-choice items differing in content and the number of response options, would not be useful to model with the 2-level approach and requires 3-level modeling. However, with a greater number of parameters specified, 3-level versions of 2-level models require larger sample sizes.

Another implication of the relatively high WP correlation between the RTs to the multiple-choice items and the Symbol Search task may be that multiple-choice question RTs deserve further investigation as potential processing speed indicators, even with the low WP reliability found here. The 3 multiple-choice items asked about activity done before the EMA (from 10 choices), where the activity was done (from 5 choices), and the perceived level of blood glucose (from 4 choices). In a future study, it may be useful to more formally investigate the extent to which item content and the number of response options in multiple-choice questions affect relationships with a formal processing speed test.

Limitations

The RTs from only a small subset of possible question types were investigated here. For multiple-choice and checkbox items, there were not enough items to investigate the effect of the number of response choices or the topics covered by these response option types. Although the effect of question topic did not appear to exert a large impact on EMA RTs for the slider items in this study, we cannot say whether the content of EMA items influences the reliability or validity of RTs to multiple-choice and checkbox items.

Data from standard laboratory-administered cognitive assessments were not collected. Therefore, we could not examine the convergent validity of individual differences in EMA item RTs using a full-length laboratory assessment of processing speed. Although a previous study found a high correlation between standard laboratory and ambulatory assessments of processing speed [5], whether EMA item RTs are associated with laboratory-based processing speed tests needs to be examined.

The recommended outcome measure for the Symbol Search task, the median reaction time in accurate trials [5], was not used here. This median RT score provides only 1 processing speed measure per Symbol Search session, which does not allow for modeling the Symbol Search scores as a 3-level multilevel model (with items nested in survey sessions nested in people). To allow such modeling, the log-transformed RT for each Symbol Search trial was computed. In preliminary 3-level multilevel model analyses, the mean of the logged RTs of accurate trials (modeled at all levels) correlated with the median

RT of accurate trials (modeled at levels 2 and 3), *r*=0.99 (*P*<.001) at the BP level (level 3) and *r*=0.97 (*P*<.001) at the survey session level (level 2). The close correspondence between the 2 appeared to justify the use of mean RTs for accurate trials, instead of the median, to enable the modeling of the Symbol Search task at 3 levels.

This study was conducted with a sample of adults with T1D experiencing various stages of the COVID-19 pandemic, which may limit the generalizability of the results. For instance, study participants were more likely to complete EMA surveys at home during times of stricter social distancing requirements. The completion of EMA surveys at home may have less potential for exposure to environmental distractors, which may have reduced the variability in item RTs. As we only examined EMA RTs as processing speed indicators in adults with T1D, further research may be needed to investigate whether findings can be replicated in other populations. For instance, the causes of processing speed fluctuations are often chronic condition specific. In adults with T1D, acute hypoglycemia has been associated with decreased processing speed [28,29]. In adults with fibromyalgia, greater momentary experiences of pain has been associated with decreased processing speed [54]. The different causes of processing speed fluctuations may also impact the relationship between EMA RTs and scores on formal processing speed tests.

Conclusions

Overall, EMA RTs appeared to be reliable and valid indicators of average and momentary processing speeds. They were not correlated to the extent that EMA survey RTs can replace formal processing speed tests. Rather, EMA survey RTs may be serviceable as rough processing speed indicators when formal processing speed testing is not feasible and when the magnitude of associations with processing speed is large. The reliability and validity of EMA survey RTs as measures of processing speed differed according to the sets of items from which the RTs were extracted, implying that EMA items can potentially be intentionally crafted to have greater associations with processing speed. For instance, in a future study, factorial analyses or machine learning models could be used to identify the specific combinations of EMA items for which the pattern of RTs is the most predictive of scores from a formal processing speed test. Analysis of RTs from EMA items may be a method of assessing average and momentary processing speeds in people's natural environments, which does not require participants to complete additional tasks beyond answering EMA survey questions. RTs to noncognitive EMA items may be important to facilitate research on the impacts of processing speed on daily functioning, especially for populations with chronic conditions.

Acknowledgments

This research was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at the National Institutes of Health (NIH; NIH/NIDDK 1R01DK121298-01), the National Institute on Aging (NIA) at the NIH (NIH/NIA R01AG068190), and the New York Regional Center for Diabetes Translation Research (NIH/NIDDK P30 DK111022). The Southern California Clinical and Translational Science Institute (SC CTSI) supported the use of REDCap (Research Electronic Data Capture) in the study from which data were analyzed (NIH/NCATS UL1TR001855 and UL1TR000130). The Symbol Search cognitive test was developed in a project supported by the National Institute on Aging at the NIH (NIH/NIA 1U2CAG060408-01; principal investigator: Martin J Sliwinski).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of supplementary analyses.
[DOCX File , 30 KB - mhealth_v11i1e45203_app1.docx]

References

1. Hoogendoorn CJ, Qin J, Wang C, Roque N, Laurenceau J, Katz MJ, et al. Depressive symptoms mediate the relationship between diabetes and cognitive performance in a community-based sample of older adults. J Diabetes Complications 2022 Jul;36(7):108183. [doi: 10.1016/j.jdiacomp.2022.108183] [Medline: 35718600]
2. Xue M, Xu W, Ou Y, Cao X, Tan M, Tan L, et al. Diabetes mellitus and risks of cognitive impairment and dementia: a systematic review and meta-analysis of 144 prospective studies. Ageing Res Rev 2019 Nov;55:100944. [doi: 10.1016/j.arr.2019.100944] [Medline: 31430566]
3. Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol Bull 2013 Jan;139(1):81-132 [FREE Full text] [doi: 10.1037/a0028727] [Medline: 22642228]
4. Harrison S, Ding J, Tang E, Siervo M, Robinson L, Jagger C, et al. Cardiovascular disease risk models and longitudinal changes in cognition: a systematic review. PLoS One 2014;9(12):e114431 [FREE Full text] [doi: 10.1371/journal.pone.0114431] [Medline: 25478916]

5. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
6. Ehrmann D, Schmitt A, Priesterroth L, Kulzer B, Haak T, Hermanns N. Time with diabetes distress and glycemia-specific distress: new patient-reported outcome measures for the psychosocial burden of diabetes using ecological momentary assessment in an observational study. Diabetes Care 2022 Jul 07;45(7):1522-1531. [doi: 10.2337/dc21-2339] [Medline: 35613338]
7. Perski O, Keller J, Kale D, Asare BY, Schneider V, Powell D, et al. Understanding health behaviours in context: a systematic review and meta-analysis of ecological momentary assessment studies of five key health behaviours. Health Psychol Rev 2022 Dec 15;16(4):576-601 [FREE Full text] [doi: 10.1080/17437199.2022.2112258] [Medline: 35975950]
8. Smiley SL, Milburn NG, Nyhan K, Taggart T. A systematic review of recent methodological approaches for using ecological momentary assessment to examine outcomes in u.S. Based hiv research. Curr HIV/AIDS Rep 2020 Aug 27;17(4):333-342. [doi: 10.1007/s11904-020-00507-0] [Medline: 32594365]
9. Couper M. Review: web surveys: a review of issues and approaches. Public Opin Q 2000;64(4):464-494. [doi: 10.1086/318641]
10. McClain CA, Couper MP, Hupp AL, Keusch F, Peterson G, Piskorowski AD, et al. A typology of web survey paradata for assessing total survey error. Soc Sci Comput Rev 2018 Feb 27;37(2):196-213. [doi: 10.1177/0894439318759670]
11. Cox D, Kovatchev B, Gonder-Frederick L, Summers K, McCall A, Grimm K, et al. Relationships between hyperglycemia and cognitive performance among adults with type 1 and type 2 diabetes. Diabetes Care 2005 Jan;28(1):71-77. [doi: 10.2337/diacare.28.1.71] [Medline: 15616236]
12. Søholm U, Broadley M, Zaremba N, Divilly P, Nefs G, Mahmoudi Z, Hypo-RESOLVE Consortium. Investigating the day-to-day impact of hypoglycaemia in adults with type 1 or type 2 diabetes: design and validation protocol of the Hypo-METRICS application. BMJ Open 2022 Feb 01;12(2):e051651 [FREE Full text] [doi: 10.1136/bmjopen-2021-051651] [Medline: 35105572]
13. Mascarenhas Fonseca L, Strong RW, Singh S, Bulger JD, Cleveland M, Grinspoon E, et al. Glycemic variability and fluctuations in cognitive status in adults with type 1 diabetes (GluCog): observational study using ecological momentary assessment of cognition. JMIR Diabetes 2023 Jan 05;8:e39750 [FREE Full text] [doi: 10.2196/39750] [Medline: 36602848]
14. Small BJ, Jim HS, Eisel SL, Jacobsen PB, Scott SB. Cognitive performance of breast cancer survivors in daily life: role of fatigue and depressed mood. Psychooncology 2019 Nov 30;28(11):2174-2180 [FREE Full text] [doi: 10.1002/pon.5203] [Medline: 31418499]
15. Kratz AL, Whibley D, Kim S, Sliwinski M, Clauw D, Williams DA. Fibrofog in daily life: an examination of ambulatory subjective and objective cognitive function in fibromyalgia. Arthritis Care Res (Hoboken) 2020 Dec 09;72(12):1669-1677 [FREE Full text] [doi: 10.1002/acr.24089] [Medline: 31609548]
16. Pyatak EA, Hernandez R, Pham LT, Mehdiyeva K, Schneider S, Peters A, et al. Function and emotion in everyday life with type 1 diabetes (FEEL-T1D): protocol for a fully remote intensive longitudinal study. JMIR Res Protoc 2021 Oct 22;10(10):e30901 [FREE Full text] [doi: 10.2196/30901] [Medline: 34463626]
17. Bielak AA, Cherbuin N, Bunce D, Anstey KJ. Intraindividual variability is a fundamental phenomenon of aging: evidence from an 8-year longitudinal study across young, middle, and older adulthood. Dev Psychol 2014 Jan;50(1):143-151. [doi: 10.1037/a0032650] [Medline: 23586940]
18. Bielak AA, Hultsch DF, Strauss E, Macdonald SW, Hunter MA. Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. Neuropsychology 2010 Nov;24(6):731-741. [doi: 10.1037/a0019802] [Medline: 20853957]
19. Roalf DR, Quarmley M, Mechanic-Hamilton D, Wolk DA, Arnold SE, Moberg PJ. Within-individual variability: an index for subtle change in neurocognition in mild cognitive impairment. J Alzheimers Dis 2016 Aug 23;54(1):325-335. [doi: 10.3233/jad-160259]
20. Christensen H, Dear KB, Anstey KJ, Parslow RA, Sachdev P, Jorm AF. Within-occasion intraindividual variability and preclinical diagnostic status: is intraindividual variability an indicator of mild cognitive impairment? Neuropsychology 2005 May;19(3):309-317. [doi: 10.1037/0894-4105.19.3.309] [Medline: 15910117]
21. Dixon RA, Garrett DD, Lentz TL, MacDonald SW, Strauss E, Hultsch DF. Neurocognitive markers of cognitive impairment: exploring the roles of speed and inconsistency. Neuropsychology 2007 May;21(3):381-399. [doi: 10.1037/0894-4105.21.3.381] [Medline: 17484601]
22. Hultsch DF, MacDonald SW, Hunter MA, Levy-Bencheton J, Strauss E. Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults. Neuropsychology 2000 Oct;14(4):588-598. [doi: 10.1037/0894-4105.14.4.588]
23. Hunt M, Momjian A, Wong K. Effects of diurnal variation and caffeine consumption on Test of Variables of Attention (TOVA) performance in healthy young adults. Psychol Assess 2011 Mar;23(1):226-233. [doi: 10.1037/a0021401] [Medline: 21244169]
24. Daniëls NE, Bartels S, Verhagen S, Van Knippenberg R, De Vugt M, Delespaul P. Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned. Internet Interv 2020 Mar;19:100300 [FREE Full text] [doi: 10.1016/j.invent.2019.100300] [Medline: 31970080]

25. Höhne JK, Schlosser S, Krebs D. Investigating cognitive effort and response quality of question formats in web surveys using paradata. Field Methods 2017 Jul 20;29(4):365-382. [doi: 10.1177/1525822x17710640]
26. Roque N, Sliwinski M, Ram N. Questionnaire-based everyday reaction time: reliable, valid, and unobtrusive measure of cognition. Innov Aging 2020;4(Supplement_1):599. [doi: 10.1093/geroni/igaa057.2015]
27. Lischetzke T, Cuccodoro G, Gauger A, Todeschini L, Eid M. Measuring affective clarity indirectly: individual differences in response latencies of state. Emotion 2005 Dec;5(4):431-445. [doi: 10.1037/1528-3542.5.4.431] [Medline: 16366747]
28. Ewing FM, Deary IJ, McCrimmon RJ, Strachan MW, Frier BM. Effect of acute hypoglycemia on visual information processing in adults with type 1 diabetes mellitus. Physiol Behav 1998 Jul;64(5):653-660. [doi: 10.1016/s0031-9384(98)00120-6] [Medline: 9817577]
29. McAulay V, Deary I, Ferguson S, Frier B. Acute hypoglycemia in humans causes attentional dysfunction while nonverbal intelligence is preserved. Diabetes Care 2001 Oct;24(10):1745-1750. [doi: 10.2337/diacare.24.10.1745] [Medline: 11574436]
30. Nuño L, Gómez-Benito J, Carmona VR, Pino O. A systematic review of executive function and information processing speed in major depression disorder. Brain Sci 2021 Jan 22;11(2):147 [FREE Full text] [doi: 10.3390/brainsci11020147] [Medline: 33499360]
31. Salthouse TA. Aging and measures of processing speed. Biol Psychol 2000 Oct;54(1-3):35-54. [doi: 10.1016/s0301-0511(00)00052-1] [Medline: 11035219]
32. Eckert MA. Slowing down: age-related neurobiological predictors of processing speed. Front Neurosci 2011;5:25 [FREE Full text] [doi: 10.3389/fnins.2011.00025] [Medline: 21441995]
33. Robinson LJ, Gallagher P, Watson S, Pearce R, Finkelmeyer A, Maclachlan L, et al. Impairments in cognitive performance in chronic fatigue syndrome are common, not related to co-morbid depression but do associate with autonomic dysfunction. PLoS One 2019;14(2):e0210394 [FREE Full text] [doi: 10.1371/journal.pone.0210394] [Medline: 30721241]
34. Valdez P. Homeostatic and circadian regulation of cognitive performance. Biological Rhythm Res 2018 Jul 02;50(1):85-93. [doi: 10.1080/09291016.2018.1491271]
35. Fortenbaugh FC, DeGutis J, Germine L, Wilmer JB, Grosso M, Russo K, et al. Sustained attention across the life span in a sample of 10,000: dissociating ability and strategy. Psychol Sci 2015 Sep 07;26(9):1497-1510 [FREE Full text] [doi: 10.1177/0956797615594896] [Medline: 26253551]
36. Kyllonen P, Zu J. Use of response time for measuring cognitive ability. J Intell 2016 Nov 01;4(4):14. [doi: 10.3390/jintelligence4040014]
37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
38. ilumivu homepage. ilumivu. URL: https://ilumivu.com/ [accessed 2023-05-12]
39. Couper MP, Singer E. Informed consent for web paradata use. Surv Res Method 2013;7(1):57-67 [FREE Full text] [Medline: 24098312]
40. Jaso BA, Kraus NI, Heller AS. Identification of careless responding in ecological momentary assessment research: from posthoc analyses to real-time data monitoring. Psychol Methods 2022 Dec 16;27(6):958-981. [doi: 10.1037/met0000312] [Medline: 34582244]
41. Couper M, Kreuter F. Using paradata to explore item level response times in surveys. J Royal Statistical Society Series A Statistics Soc 2013;176(1):286. [doi: 10.1111/j.1467-985X.2012.01041.x]
42. Holdershaw J, Melnyk V, Gendall P, Wright M. Predicting behaviour: comparing the performance of factual versus attitudinal approaches. Intl J Social Res Methodol 2017 Nov 13;21(4):439-452. [doi: 10.1080/13645579.2017.1402156]
43. Liu Y, Schneider S, Orriens B, Meijer E, Darling JE, Gutsche T, et al. Self-administered web-based tests of executive functioning and perceptual speed: measurement development study with a large probability-based survey panel. J Med Internet Res 2022 May 09;24(5):e34347 [FREE Full text] [doi: 10.2196/34347] [Medline: 35532966]
44. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
45. Nezlek JB. A practical guide to understanding reliability in studies of within-person variability. J Res Personality 2017 Aug;69:149-155. [doi: 10.1016/j.jrp.2016.06.020]
46. Raykov T, Marcoulides GA. On multilevel model reliability estimation from the perspective of structural equation modeling. Structural Equation Model Multidisciplinary J 2006 Jan;13(1):130-141. [doi: 10.1207/s15328007sem1301_7]
47. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Pers Soc Psychol Bull 2006 Jul 02;32(7):917-929 [FREE Full text] [doi: 10.1177/0146167206287721] [Medline: 16738025]
48. Hallquist MN, Wiley JF. MplusAutomation: an R package for facilitating large-scale latent variable analyses in Mplus. Struct Equ Modeling 2018 Jan 19;25(4):621-638 [FREE Full text] [doi: 10.1080/10705511.2017.1402334] [Medline: 30083048]
49. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014 Apr 11;11:16 [FREE Full text] [doi: 10.1186/1742-4682-11-16] [Medline: 24725531]
50. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.

51. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
52. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ 2004 Sep;38(9):1006-1012. [doi: 10.1111/j.1365-2929.2004.01932.x] [Medline: 15327684]
53. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Cambridge, Massachusetts, United States: Academic Press; 1969.
54. Whibley D, Williams DA, Clauw DJ, Sliwinski MJ, Kratz AL. Within-day rhythms of pain and cognitive function in people with and without fibromyalgia: synchronous or syncopated? Pain 2022 Mar 01;163(3):474-482 [FREE Full text] [doi: 10.1097/j.pain.0000000000002370] [Medline: 34393201]

Abbreviations

BP: between-person
EMA: ecological momentary assessment
ICC: intraclass correlation coefficient
PHQ-8: Patient Health Questionnaire-8
REDCap: Research Electronic Data Capture
RT: response time
T1D: type 1 diabetes
WP: within-person

Edited by L Buis; submitted 20.12.22; peer-reviewed by D Ehrmann, L Mascarenhas Fonseca; comments to author 17.02.23; revised version received 14.03.23; accepted 27.04.23; published 30.05.23.

Please cite as:
Hernandez R, Hoogendoorn C, Gonzalez JS, Jin H, Pyatak EA, Spruijt-Metz D, Junghaenel DU, Lee PJ, Schneider S
Reliability and Validity of Noncognitive Ecological Momentary Assessment Survey Response Times as an Indicator of Cognitive Processing Speed in People's Natural Environment: Intensive Longitudinal Study
JMIR Mhealth Uhealth 2023;11:e45203
URL: https://mhealth.jmir.org/2023/1/e45203
doi:10.2196/45203
PMID:37252787

Original Paper

Using Mobile Ecological Momentary Assessment to Understand Consumption and Context Around Online Food Delivery Use: Pilot Feasibility and Acceptability Study

Si Si Jia[1], BSc, MND; Margaret Allman-Farinelli[2,3], PhD; Rajshri Roy[4], PhD; Philayrath Phongsavan[3,5], PhD; Karice Hyun[1,6], PhD; Alice Anne Gibson[3,7*], PhD; Stephanie Ruth Partridge[1,3*], PhD

[1]School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
[2]Nutrition and Dietetics, Sydney School of Nursing, The University of Sydney, Sydney, Australia
[3]Charles Perkins Centre, The University of Sydney, Sydney, Australia
[4]Discipline of Nutrition and Dietetics, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand
[5]Prevention Research Collaboration, Sydney School of Public Health, The University of Sydney, Sydney, Australia
[6]Concord Hospital, ANZAC Research Institute, The University of Sydney, Sydney, Australia
[7]Menzies Centre for Health Policy and Economics, The University of Sydney, Sydney, Australia
[*]these authors contributed equally

Corresponding Author:
Si Si Jia, BSc, MND
School of Health Sciences
Faculty of Medicine and Health
The University of Sydney
Camperdown NSW 2006
Sydney, 2006
Australia
Phone: 61 2 8627 1697
Email: sisi.jia@sydney.edu.au

Abstract

Background: Mobile ecological momentary assessment (EMA) is a powerful tool for collecting real-time and contextual data from individuals. As our reliance on online technologies to increase convenience accelerates, the way we access food is changing. Online food delivery (OFD) services may further encourage unhealthy food consumption habits, given the high availability of energy-dense, nutrient-poor foods. We used EMA to understand the real-time effects of OFD on individuals' food choices and consumption behaviors.

Objective: The primary aims of this pilot study were to assess the feasibility and acceptability of using EMA in young users of OFD and compare 2 different EMA sampling methods. The secondary aims were to gather data on OFD events and their context and examine any correlations between demographics, lifestyle chronic disease risk factors, and OFD use.

Methods: This study used EMA methods via a mobile app (mEMASense, ilumivu Inc). Existing users of OFD services aged 16 to 35 years in Australia who had access to a smartphone were recruited. Participants were randomly assigned to 1 of 2 groups: signal-contingent or event-contingent. The signal-contingent group was monitored over 3 days between 7 AM and 10 PM. They received 5 prompts each day to complete EMA surveys via the smartphone app. In contrast, the event-contingent group was monitored over 7 days and was asked to self-report any instance of OFD.

Results: A total of 102 participants were analyzed, with 53 participants in the signal-contingent group and 49 participants in the event-contingent group. Compliance rates, indicating the feasibility of signal-contingent and event-contingent protocols, were similar at 72.5% (574/792) and 73.2% (251/343), respectively. Feedback from the participants suggested that the EMA app was not easy to use, which affected their acceptability of the study. Participants in the event-contingent group were 3.53 (95% CI 1.52-8.17) times more likely to have had an OFD event captured during the study. Pizza (23/124, 18.5%) and fried chicken (18/124, 14.5%) comprised a bulk of the 124 OFD orders captured. Most orders were placed at home (98/124, 79%) for 1 person (68/124, 54.8%). Age (incidence rate ratio 0.95, 95% CI 0.91-0.99; *P*=.03) and dependents (incidence rate ratio 2.01, 95% CI

1.16-3.49; *P*=.01) were significantly associated with the number of OFD events in a week after adjusting for gender, socioeconomic status, diet quality score, and perceived stress levels.

Conclusions: This pilot study showed that EMA using an event-contingent sampling approach may be a better method to capture OFD events and context than signal-contingent sampling. The compliance rates showed that both sampling methods were feasible and acceptable. Although the findings from this study have gathered some insight on the consumption and context of OFD in young people, further studies are required to develop targeted interventions.

(JMIR Mhealth Uhealth 2023;11:e49135) doi:10.2196/49135

KEYWORDS

ecological momentary assessment; mobile applications; mobile apps; feasibility studies; online food delivery; smartphone; young adult; adolescent; food environment; consumer behavior; mobile phone

Introduction

Background

Unhealthy diets are a leading contributor to the global burden of disease [1]. Globally, there is an increasing trend toward the consumption of “out-of-home” foods, potentially exacerbating suboptimal diets. These foods include fast foods and takeaway or “take-out” foods from a variety of sources including restaurants, fast-food chains, convenience stores, coffee shops, and takeaway food outlets. An analysis of the UK National Diet and Nutrition Survey 2008-2012 found that 27% of adults had consumed meals outside the home once per week or more [2], and research from the United States estimated that out-of-home meals comprised 50% of household budgets in 2018 [3]. Results from Australia’s National Household Expenditure and Time Use Surveys have shown a steady incline in spending on out-of-home foods—increasing from 22.8% of total food budgets in 1989 to 26.5% in 2010 [4]. Rising consumption of out-of-home foods has been similarly observed in low- and middle-income countries including China [5,6] and Latin American countries [7].

However, these out-of-home meals are often high in fat, salt, and sugar and low in vitamins and minerals [8]. A systematic review of 15 prospective studies found that a high frequency of eating out and consuming out-of-home meals was associated with weight gain and a greater risk of becoming overweight or obese [9]. Moreover, studies have shown associations between frequent consumption of home-cooked meals and greater adherence to the Dietary Approaches to Stop Hypertension diet and Mediterranean diets as well as greater intake of fruits and vegetables [10]. Thus, the replacement of home-cooked meals with out-of-home meals may result in further detriments to dietary health.

In recent years, online food delivery (OFD) services such as Uber Eats, DoorDash, and JustEat have transformed the concept of out-of-home foods by allowing customers to order a variety of food and drink items straight from kitchen to doorstep [11]. With this added convenience, OFD potentially increases access to and consumption of out-of-home food. A study from the United Kingdom found that adults with access to the greatest number of food outlets available to them online had 71% greater odds of OFD use (odds ratio [OR] 1.71, 95% CI 1.09-2.68) compared with those with the lowest access [12].

A growing number of studies have shown that a high proportion of menu offerings on popular OFD platforms are poor in nutritional quality. In Thailand, a majority of the most popular menu items were considered unhealthy against the World Health Organization’s recommended daily intake values [13]. Of the 25 most popular menu items, 23 exceeded the recommended sodium intake for adults, and 80% of all sweet items offered were 1.5 times above the recommended daily intake for sugar [13]. Similarly, more than three-quarters of menu items were classified as discretionary “junk foods” in Australia [14] and New Zealand [15]. A cross-sectional study conducted in 3 international cities (Chicago, Melbourne, and Amsterdam) revealed “burgers,” “pizza,” and “Italian” were in the top 10 most advertised meals on the website or app of the OFD [16].

OFD services are now regularly used by millions of people with heightened use found among young people aged between 16 and 34 years [17,18], who are already experiencing escalating rates of weight gain [19]. Furthermore, in Australia, by the age of 16 years, approximately 80% of young people have a debit card [20] in their name. This is critical considering the ease of digital payment options and tools offered on OFD apps, which amplifies young people’s accessibility to takeaway foods. A study in China found that >47.8% of male university students and >30.7% of female university students have used OFD more than once per week [21]. Therefore, there is an urgent need to assess and understand the consumption and behaviors associated with the use of these OFD services.

Traditional dietary assessment methods such as 24-hour recalls, diet history, and food frequency questionnaires are reliant on participants retrospectively recalling their intake, whereas food records involve prospective recording, but both approaches are time-consuming and prone to inherent biases [22,23]. Recent studies have demonstrated that ecological momentary assessments (EMAs) can be a valid measure of dietary intake [24,25]. EMA is a data collection method that gathers real-time information from participants studying behaviors and contexts as they perform regular day-to-day activities in real-world settings [26]. Due to technological advances, researchers now use smartphones or mobile phones to conduct EMA studies, which have improved data collection and added flexibility to study designs [27]. A systematic review of 39 EMA studies conducted in young people aged 16 to 30 years indicated that EMA is an acceptable and feasible methodology to capture dietary intake and food consumption for this population [28].

Objectives

Although dietary intake generally can be spontaneous and difficult to capture, OFD events are likely to be even more sporadic. Thus, compared with traditional dietary assessment methods, EMA may be an alternate method for capturing OFD events. In addition, unlike traditional dietary assessment methods, the context including social and psychological factors surrounding the real-time OFD event can be obtained using EMA. To the best of our knowledge, no previous study has used EMA to study OFD use. Therefore, this study aimed to determine the feasibility and acceptability of EMA in a sample of young people who are users of OFD. A secondary aim was to investigate associations between the frequency of OFD use and demographic variables such as age, gender, and socioeconomic level as well as other lifestyle chronic disease risk factors including diet quality score, physical activity levels, stress levels, and sleep quality.

Methods

Recruitment

From June to October 2022, participants were recruited via flyers, social media advertising, and word of mouth. Eligible participants were (1) aged between 16 and 35 years; (2) users of OFD, defined as at least once in the past 3 months; (3) living in Australia; and (4) having access to a smartphone. This was to ensure that participants were not first-time users of OFD, as this study aimed to gain a specific understanding of the behaviors and demographics of OFD users. The target sample comprised at least 140 participants accounting for a 30% (42/140) dropout rate. Using quota sampling, the target was 28 participants in the 16 to 18 years age group, 28 participants in the 30 to 35 years age group, and 42 participants each in the 19 to 24 and 25 to 29 years age groups. Previous pilot EMA studies used nonprobability sampling methods to recruit a similar number of participants [25,29].

Ethical Considerations

This study was approved by The University of Sydney Human Research Ethics Committee (2022/006). Reporting followed an adapted Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist for Reporting EMA Studies (CREMAS) guidelines [30] (Multimedia Appendix 1).

Data Collection

After participants joined the study and provided consent, they were directed to a series of baseline surveys on REDCap (Research Electronic Data Capture; Vanderbilt University) to obtain data on participant demographics, usual OFD behaviors, physical activity levels, sleep quality, and perceived stress levels. The questions were adapted from the 7-item International Physical Activity Questionnaire—Short Form [31], Pittsburgh Sleep Quality Index [30], and Perceived Stress Scale [32].

Upon completing the baseline surveys on REDCap, participants were sent a unique mobile code to download and set up the EMA mobile app (mEMASense, ilumivu Inc) on their personal smartphone devices. This app is available for both the Android and Apple operating systems. Using a computer-generated code, participants were randomly allocated 1:1 to 1 of 2 groups, "Prompts" or "No Prompts," which determined the sampling approach used—respectively representative of the signal-contingent or event-contingent approach. A signal-contingent design involves sending prompts to participants randomly over the course of a given hour, day, or week [33]. In contrast, an event-contingent design allows participants to complete a prompt whenever they experience the event of interest [33].

Participants in each group received a separate set of instructions via REDCap to inform them of their group allocation and what was required from them in either the Prompts or No Prompts group. An external link was also sent to participants to complete the Australian Eating Survey [34], a validated food frequency questionnaire developed by researchers from the University of Newcastle, to capture usual dietary intake.

Data were collected over 1 monitoring period (1 wave). Figure 1 depicts a flow diagram of the overall data collection process, and Figure 2 provides an overview of the 2 varying EMA sampling protocols used in the study.

Figure 1. Flowchart of the data collection processes. mEMA: mobile-based ecological momentary assessment; REDCap: Research Electronic Data Capture.

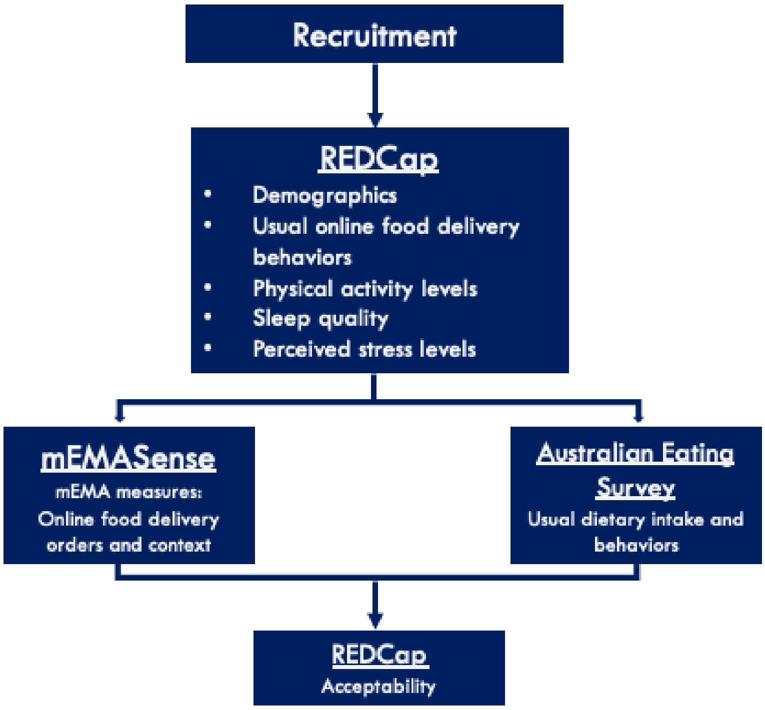
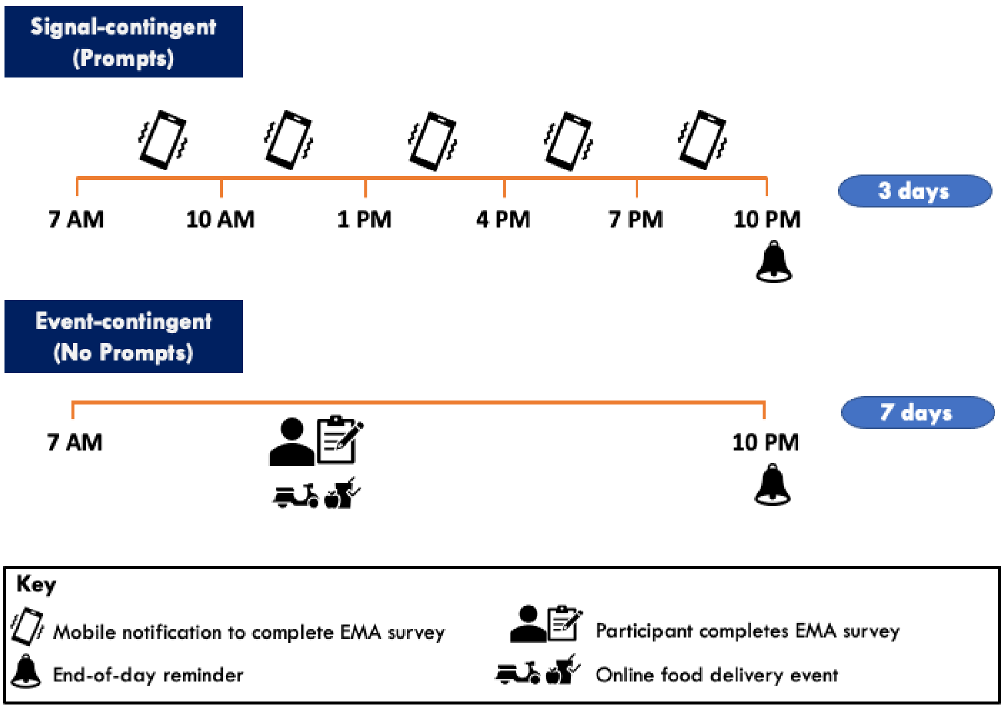


Figure 2. Overview of the 2 varying mobile-based ecological momentary assessment sampling protocols used in the study: signal-contingent (Prompts) group compared with the event-contingent (No Prompts) group. Participants in the No Prompts group were observed over 1 week (7 days), whereas those in the Prompts group were observed over 1 week with prompts sent on 3 separate days. EMA: ecological momentary assessment.



EMA Procedure

Signal-Contingent Sampling (Prompts Group)

Participants in the Prompts group were monitored over 1 week on 3 days, including 2 weekdays (Monday-Friday) and 1 day on the weekend (Saturday or Sunday) between 7 AM and 10 PM. To ensure that at least 1 weekday and 1 weekend day were captured, the EMA prompt days were assigned on the second, fourth, and sixth day from the day participants joined the study. Participants were unaware of this prompting schedule and were only informed that they would be prompted on 3 "random days" over the 7 days from their enrollment into the study. This was implemented to improve the external validity of the study by aiming to capture unanticipated OFD events.

Participants were prompted by the app via a notification at a randomly assigned time during the 5 predetermined intervals (7 AM-10 AM, 10 AM-1 PM, 1 PM-4 PM, 4 PM-7 PM, and 7 PM-10 PM). An end-of-day prompt was sent at 10 PM on each of the study days to record any OFD ordering event that was not previously captured. When prompted, participants were asked to complete the short survey on the EMA app that started with the question "In the past 10-15 minutes, were you eating or drinking?" If unavailable at the time of the prompt, participants had a 45-minute time window to respond with 2 reminders sent 15 minutes apart after the initial prompt. After this point, the survey became inaccessible until the next prompt and recording opportunity. In total, 6 EMA surveys were prompted on each of the 3 study days, totaling 18 EMA surveys per participant across the data collection period.

Event-Contingent Sampling (No Prompts Group)

Participants in the No Prompts group were instructed to record any online food order they placed over 1 week from their enrollment in the study. An end-of-day prompt was sent at 10 PM on each of the 7 days to remind and allow participants to record any OFD event not previously captured.

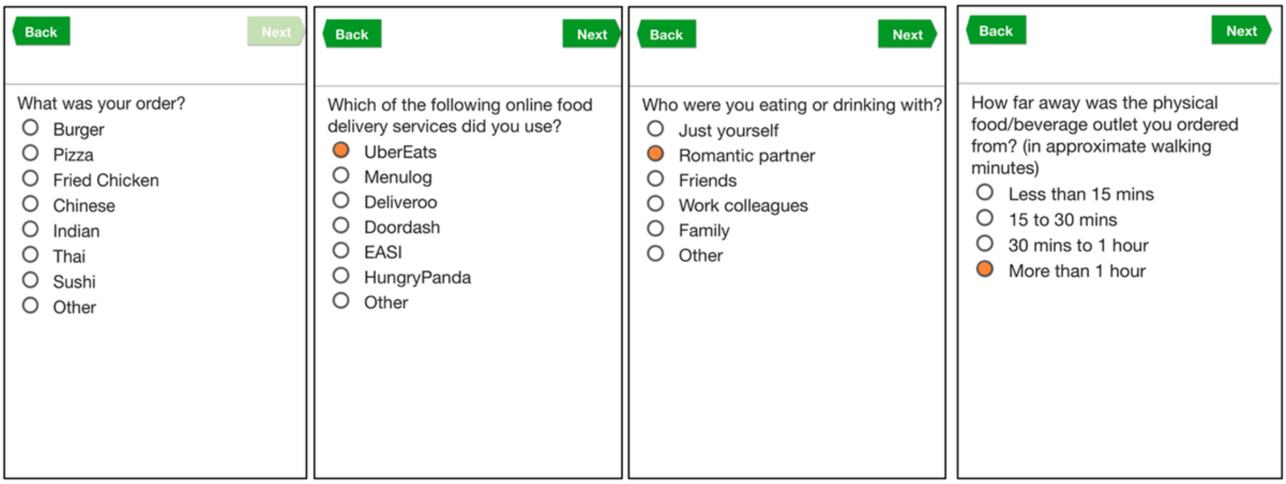
Participants Attending High School

Participants aged between 16 and 18 years who were attending high school in the Prompts group were monitored across 3 days, including 2 weekdays (Monday-Friday) during nonschool hours (4 PM and 10 PM) and 1 day on the weekend (Saturday or Sunday) between 7 AM and 10 PM. This was done to ensure minimal disruption to high school commitments. Individuals attending high school in the No Prompts (the "event-contingent") group were only required to answer the surveys during nonschool hours on weekdays (from 4 PM to 10 PM).

Capturing an OFD Order Event

Participants in both the Prompts and No Prompts groups were asked for the following details and surrounding context of their online food order: how they received their order (whether through pick-up or delivery), what they ordered, who they were with (by themselves, with romantic partner, work colleagues, friends, or family), where they were (home, workplace, or university), how far away they were from the food outlet in walking minutes, if any promotions or discount codes were used, their hunger levels, and cravings and stress levels at the time of ordering. All instances were timestamped to record when they consumed their online food order. Please see Figure 3 for a screenshot of the EMA survey questions delivered through the mEMASense app.

Figure 3. Selection of questions from the survey delivered through the mEMASense app to capture context surrounding an online food delivery event.



Incentives

Participants were offered gift vouchers up to AUD 45 (US $29) for their completion of the entire pilot study. If participants completed one-third of the study, they were awarded 1 AUD 15 (US $9) voucher for their time, and participants who had completed two-thirds were compensated with AUD 30 (US $19) worth of gift vouchers.

Feasibility

The compliance rate was used to measure the feasibility of the study. This was calculated differently for the different sampling approaches. In the Prompts group, the compliance rate per participant was the number of answered prompts divided by 18, which was the total number of expected prompts. The mean compliance rate for the study was calculated to represent the sample in the Prompts group.

Similarly, for the No Prompts group, the compliance rate per participant was the number of answered end-of-day reminders divided by the total number of expected end-of-day reminders. These end-of-day reminders were sent daily over the 1-week study period, resulting in 7 total reminders. Furthermore, the percentage of online food ordering events captured by the end-of-day reminders was also analyzed. This indicated how likely participants were to forget their recording of each online food ordering event using the No Prompts approach.

Acceptability

Participants were administered a final end-of-study questionnaire composed of 5 items to gather insight into their acceptability of the EMA methodology, the mobile app used to send the EMA prompts, and the perceived participant burden from the 2 different sampling approaches. These 5-point Likert scale questions were adapted from a previous study [29] with an additional closing open-ended question to gather general qualitative feedback.

Study Measurements and Outcomes

Demographic

Age, gender, ethnicity, education status, working status, number of dependents, residential postcode, and car access were obtained from the demographic survey. Residential postcodes were matched to quintiles on the Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD), an area-based socioeconomic measure, where a higher index value reflects an area of greater socioeconomic advantage compared with people residing in other areas [35]. Participants were also asked to report their height (cm) and weight (kg) so that their BMI could be determined.

Lifestyle Chronic Disease Risk Factors

Measures from the Australian Eating Survey generated the Australian Recommended Food Score, a diet quality index that reflects alignment with the Australian Dietary Guidelines [36]. Responses to the International Physical Activity Questionnaire allowed participants to be categorized into low, moderate, or high physical activity levels using scoring guidelines [37]. Similarly, a score was produced to categorize participants' perceived stress levels into low, moderate, or high [38]. Sleep quality was reflected by the number of hours of sleep reported by the participants and compared with recommended sleep guidelines [39].

Data collected from the EMA app were used as the outcome variables of interest, including the number of OFD events, OFD order details, and related contextual data, as shown in Figure 1.

Statistical Analysis

Overview

The study's analytic sample comprised participants who met the minimum requirements outlined in the sampling protocol. This entailed completing at least one-third of the study period, which translated to a minimum of 4 prompts for adults or 2 prompts for adolescents within a single study day for those in the signal-contingent group. For participants in the event-contingent group, it meant answering at least 2 end-of-day reminders. In addition, those who had finished the end-of-study acceptability questionnaire were also included in this sample. Data sets from REDCap surveys, the Australian Eating Survey, and the EMA app were downloaded and linked together via participants' unique mobile code for the EMA app and REDCap ID number. Data were checked and cleaned before analysis. All statistical analyses were performed using R statistical software (version 2023.03.0+386; R Core Team).

Descriptive statistics, including means, SDs, and frequencies, were used to report the baseline characteristics of the sample and the OFD behaviors and context captured via the EMA app. Other statistical analyses were conducted as described below in the Feasibility Analysis, Acceptability Analysis, Comparison of Sampling Approaches, and Exploratory Analysis sections.

Feasibility Analysis

Compliance rates were the main outcome to determine the feasibility of the EMA protocol. In the Prompts group (signal-contingent), this was calculated as the total number of prompts answered over the expected total number of prompts answered. These compliance rate calculations slightly differed for adolescent participants who received fewer prompts per study day. In the No Prompts group, as the event of interest may not have occurred daily, compliance was measured by participant's responses to the end-of-day study reminders. This was similarly measured as the total number of end-of-day reminders answered above the expected total of 7.

To investigate factors associated with compliance, a 2-level multilevel logistic regression analysis was conducted on participant data from the Prompts group. The outcome variable used for analysis was whether the participant had answered the delivered prompt (coded as "TRUE") or had not answered (coded as "FALSE"). Level 1 factors were factors at the observation level such as time of the day (ie, morning, afternoon, or evening) and day of the week (weekday or weekend). Level 2 factors were at the person or participant level such as their demographics including age, gender, and IRSAD decile. The intraclass correlation coefficients were calculated to determine the proportion of variance that is because of between-person differences.

Acceptability Analysis

Median scores and IQRs were reported to summarize responses to the 5-item acceptability questionnaire in both the Prompts and No Prompts groups. Open-ended feedback was analyzed using content analysis methods. Concepts were developed based on the occurrence of terms.

Comparison of Sampling Approaches

A multivariable logistic regression model was used to analyze whether the sampling approach (ie, Prompts vs No Prompts) was associated with capturing OFD events during the study period. The outcome variable was whether an OFD event was recorded during the study period. The model was adjusted for age, gender, and frequency of OFD ordering reported at baseline.

Exploratory Analysis on Demographic and Lifestyle Risk Factors and Their Associations With OFD Use

A zero-inflated negative binomial regression model was built to analyze associations between participant demographic characteristics, lifestyle chronic disease risk factors, and number of OFD events in a week. Incidence rate ratios were the effect size from this model used to indicate how often OFD events occurred over the study period in relation to participants' demographic characteristics and lifestyle chronic disease risk factors. A zero-inflated model was deemed appropriate as our data had an excess of "zeros" for the number of OFD events recorded by the participants. This model was adjusted for age, gender, and socioeconomic status as indicated by the IRSAD quintile. An influential point, which is an outlier that affects the slope of the regression, was identified through a bar plot of the

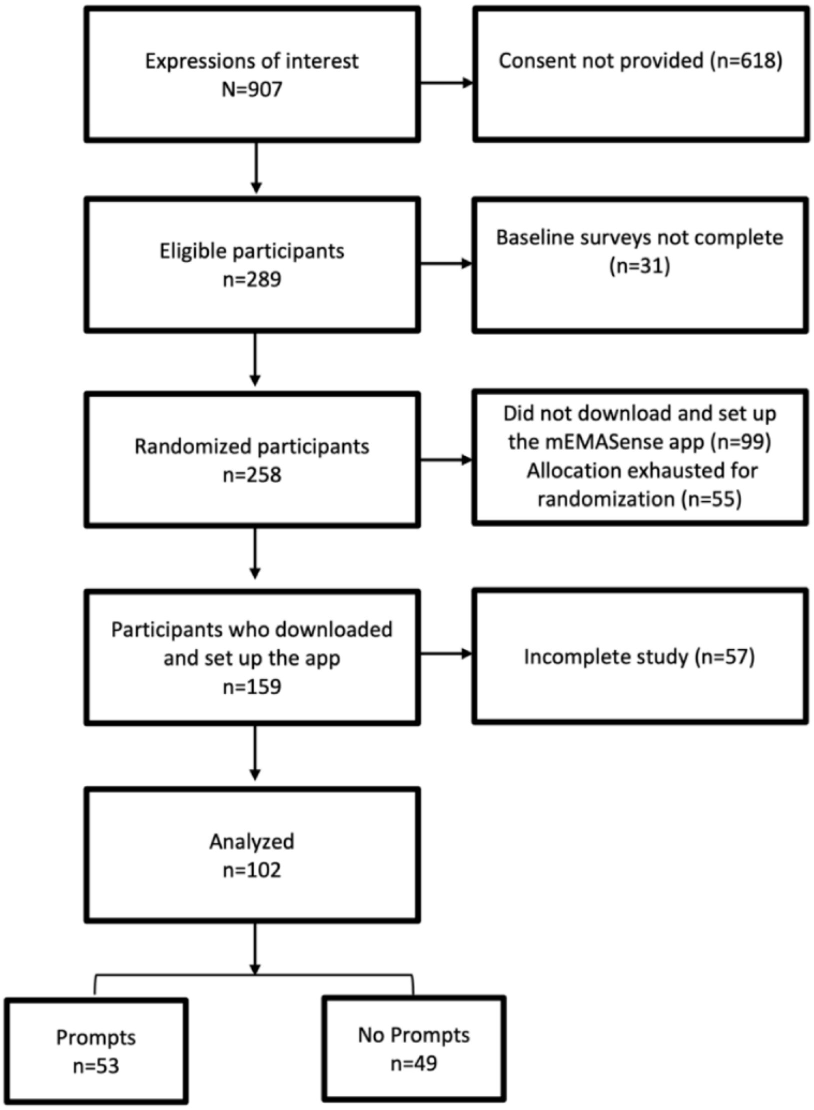
Cook distance and subsequently removed as it had a significant impact on the coefficients.

Results

Participants

As shown in Figure 4, of the 907 expressions of interest received on REDCap, 289 participants provided consent and continued to complete the baseline surveys. However, only 258 (89.2%) of the 289 participants completed all baseline surveys and were subsequently randomized into either the Prompts or No Prompts group. Only 159 (61.6%) of the 258 participants proceeded to the next stage of the study by downloading and setting up the mEMASense app. Of the 159 participants, 102 (64.2%) were part of the final analytical sample, as 57 (35.8%) did not meet the minimum requirements for study completion. Of the 102 participants, there were 53 (52.0%) in the Prompts group and 49 (48.0%) in the No Prompts group.

Figure 4. Participant recruitment process and attrition.



Participant Characteristics

The mean age of all participants was 24.4 (SD 5.4) years. Women comprised 70.6% (72/102) of the study sample, men comprised 28.4% (29/102) of the study sample, and 1.0% (1/102) of the participants identified as nonbinary. In total, 63.7% (65/102) of the participants had a normal BMI range, whereas 28.4% (29/102) were overweight and obese. Approximately half of the sample (48/102, 47%) identified their ethnicity as White, approximately a quarter (24/102, 23.5%) identified as Chinese, and 14.7% (15/102) reported their ethnicity as Asian and South-East Asian (not further specified). Most of the participants (89/102, 87.3%) resided in areas of high socioeconomic advantage in the upper quintiles of the IRSAD, and most participants (58/102, 56.9%) were highly educated (Table 1).

Table 1. Demographics of participants (N=102).

Demographics	Value
Age (years), mean (SD)	24.4 (5.4)
Age (years), n (%)	
16-18	17 (16.7)
19-24	34 (33.3)
25-29	33 (32.3)
30-35	18 (17.6)
Gender identity, n (%)	
Woman	72 (70.6)
Man	29 (28.4)
Nonbinary	1 (1)
BMI (kg/m^2), n (%)	
Underweight (<18.5)	7 (6.9)
Normal weight (18.5-24.9)	65 (63.7)
Overweight (25.0-29.9)	20 (19.6)
Obese (>30.0)	9 (8.8)
N/A[a]	1 (1)
Ethnicity, n (%)	
Asian (not specified)	11 (10.8)
Bangladeshi	1 (1)
Chinese	24 (23.5)
Indian	3 (2.9)
Japanese	1 (1)
Jewish	1 (1)
Southeast Asian (not further specified)	4 (3.9)
Vietnamese	1 (1)
White	48 (47)
Multiracial[b]	7 (6.9)
N/A[a]	1 (1)
IRSAD[c] quintiles (1 representing most disadvantaged and 5 representing most advantaged), n (%)	
Quintile 1	5 (4.9)
Quintile 2	2 (2)
Quintile 3	6 (5.9)
Quintile 4	17 (16.7)
Quintile 5	72 (70.6)
Education level, n (%)	
Current high school student	13 (12.7)
Completed high school	2 (2)
Currently studying for degree or diploma	28 (27.4)
Completed trade or technical qualification	0 (0)
Completed a degree or diploma	33 (32.3)
Completed a postgraduate degree	25 (24.5)

Demographics	Value
N/A[a]	1 (1)
Car access, n (%)	
Yes	61 (59.8)
No	41 (40.2)
Number of dependents, n (%)	
0	78 (76.5)
1	8 (7.8)
2	15 (14.7)
3	1 (1)

[a]N/A: not applicable (owing to missing values).

[b]Including White and Chinese, White and Japanese, White and Pacific Islander, White and Asian, and nonspecified.

[c]IRSAD: Index of Relative Socio-Economic Advantage and Disadvantage.

Usual OFD Behaviors and Lifestyle Behavioral Risk Factors

As shown in Table 2, in total, 52% (53/102) of the participants reported using an OFD service 1 to 2 times a month, and 27.5% (28/102) of the participants reported using it once a week at baseline. Approximately 89.2% (91/102) of the sample had used Uber Eats, which was followed by DoorDash as the next highly used service by 59.8% (61/102) of the participants. Only 23.5% (24/102) of the participants had membership with an OFD service, which was commonly a free 6-month trial.

The mean Australian Recommended Food Score for participants was 31 (SD 11.6) out of a maximum score of 73 indicating optimal nutrition. The median percentage of energy intake contributed by discretionary foods was 35% (IQR 26-44) among participants . Analyses from the International Physical Activity Questionnaire indicated that a large proportion of participants (64/102, 62.7%) had high physical activity levels, and 54.9% (56/102) of the sample experienced moderate levels of stress. The mean hours of sleep reported was 6.5 (SD 1.0) in adolescents and 7.2 (SD 1.1) in adults (Table 2).

Table 2. Online food delivery (OFD) behaviors at baseline and self-reported lifestyle behavioral risk factors (N=102).

Lifestyle risk factor	Value
OFD behaviors	
Frequency of using an OFD service, n (%)	
Rarely or never	0 (0)
1-2 times a month	53 (52)
Once a week	28 (27.4)
A few times a week	13 (12.7)
Once a day	4 (3.9)
More than once a day	4 (3.9)
OFD service (those electing "Yes, I use this delivery service"), n (%)	
Uber Eats	91 (89.2)
MenuLog	58 (56.9)
Deliveroo	60 (58.8)
Doordash	61 (59.8)
EASI	18 (17.6)
Others[a]	12 (11.8)
Fantuan	3 (2.9)
HungryPanda	1 (1)
Panda	3 (2.9)
PandaFresh	2 (2)
Restaurant-specific delivery service	1 (1)
N/A[b]	3 (2.9)
OFD services membership, n (%)	
Yes	24 (23.5)
No	78 (76.5)
OFD membership (those electing "Yes, I have a membership subscription")[a], n (%)	
Dashpass	6 (5.9)
Deliveroo	3 (2.9)
Uber Eats	10 (9.8)
N/A[b]	5 (4.9)
Monthly cost of OFD membership for those with a subscription (US $), n (%)	
Free trial for 6 months	6 (5.9)
Dashpass (9.99)	2 (2)
Uber Eats (9.99)	3 (2.9)
Deliveroo plus (12.99)	3 (2.9)
N/A[b]	10 (9.8)
Usual dietary behaviors[c]	
Australian Recommended Food Score (maximum 73), mean (SD)	31 (11.6)
Percentage of energy from noncore foods, median (IQR)	35% (26-44)
Out-of-home food consumption (general, nonspecific to OFD), n (%)	
<1 per week	20 (19.6)
1-2 per week	51 (50)

Lifestyle risk factor	Value
3-4 per week	24 (23.5)
5-6 per week	3 (2.9)
Once a day	2 (2)
1 or more per day	2 (2)
Physical activity levels[d] n, (%)	
Low	3 (2.9)
Moderate	35 (34.3)
High	64 (62.7)
Perceived stress levels[e] n, (%)	
Low	30 (29.4)
Moderate	56 (54.9)
High	16 (15.7)
Hours of sleep, mean (SD)	
Adolescents aged 16-18 years (n=17)	6.5 (1.0)
Adults aged 18-35 years (n=85)	7.2 (1.1)

[a]Participants were allowed to enter >1 other food delivery service or membership they used.

[b]N/A: not applicable (owing to invalid online food delivery service).

[c]Obtained from the Australian Eating Survey results.

[d]Physical activity levels were calculated according to the International Physical Activity Questionnaire analysis guidelines [37].

[e]Perceived stress levels were calculated according to guidelines for the Perceived Stress Scale [38].

Feasibility

Compliance rate was used as a measure of feasibility for the EMA study. Among the analytic sample, compliance with the study protocol was compared between the Prompts and No Prompts groups.

Prompts Group

Adolescents Aged 16-18 Years

The compliance rate for the adolescent Prompts group was 70.4% (76/108), with an average of 8 (out of 12) prompts answered per adolescent participant.

Adults Aged 18-35 Years

The compliance rate for the adult Prompts group was 72.5% (574/792), with an average of 13 (out of 18) prompts answered per adult participant.

No Prompts Group

Compliance in the No Prompts group was measured by the response to all end-of-day study reminders. The compliance rate for the No Prompts group was 73.2% (251/343). Furthermore, of 49 participants, 36 (73%) followed the protocol and logged instances of using an OFD service during their study period, unprompted, as per the "event-contingent" sampling protocol.

Factors Influencing Missing Prompt Data

Results from a multilevel logistic regression model showed that compliance or response to a prompt did not vary by age, gender, socioeconomic status, or whether it was a weekday or weekend (Table 3). However, compared with the evening between 6 PM and 12 AM, participants were 1.47 times more likely to answer prompts in the afternoon between 12 PM and 6 PM (95% CI 1.02-2.12). The intraclass correlation coefficients between participants were low (0.27), indicating that multilevel modeling was appropriate.

Table 3. Results from the multilevel logistic regression model to examine observation level and within-person level factors and their influence on a participant's likelihood to respond to prompts (n=53).

Predictors	Response to prompt (yes or no)	
	Odds ratios (95% CI)	*P* value
Days (weekend)	0.81 (0.59-1.11)	.18
Time of day (afternoon)	1.47 (1.02-2.12)	.04[a]
Time of day (morning)	0.84 (0.58-1.22)	.36
Age (years)	1.02 (0.95-1.08)	.6
Gender (woman)	2.07 (1.00-4.31)	.05
IRSAD[b] decile	1.19 (0.99-1.42)	.06

[a]Intraclass correlation coefficient: 0.27.

[b]IRSAD: Index of Relative Socio-economic Advantage and Disadvantage.

Acceptability

Prompts

Responses from participants in the Prompts group reflected an overall opinion that the number of prompts sent during the study period was an appropriate amount, indicated by a median score of 3. The times of day that the prompts were sent were generally agreed upon by participants as being appropriate and not burdensome, with a high median score of 4 (Multimedia Appendix 2). Participants were ambivalent in their response to the app being easy to use, with a median score of 3 indicating that they neither agreed nor disagreed with the statement. The EMA study did not appear to make a majority of participants more conscious of their diet and eating behaviors, with more responses skewed toward disagreeing with the statement. The length of study for the Prompts group, however, was distributed more toward being *too short* (Multimedia Appendix 2).

No Prompts

In the No Prompts group, participants' responses also showed that the number of prompts sent during the study period was appropriate, with a median score of 3. Many participants agreed that end-of-day reminders were necessary to remember to log an OFD order, as shown by the median score of 4 and with a distribution skewed toward "strongly agree" (Multimedia Appendix 2). However, more participants disagreed that the mobile app was easy to use. This was indicated by a median score of 3, with more scores indicating strong disagreement. Furthermore, responses from participants varied on whether the surveys made them more conscious of their diet and eating behaviors, as shown by the median score of 3. Responses on length of study for the No Prompts group were also distributed more toward being *too short*.

Open Feedback

Open feedback was obtained from 45 participants. Responses from 38% (17/45) of the participants indicated the difficulty of navigating the app, with many stating that it was "clunky," "not user-friendly," and "cumbersome." Moreover, 27% (12/45) of the responses indicated that the timing of the prompts could be improved. Participants in the group that received prompts suggested receiving prompts closer to mealtimes. Some participants in the No Prompts group commented on the timing of the "end of day" reminders as being "too late" and after they had gone to sleep.

Moreover, 4 participants suggested further clarity of the training instructions and to have these instructions available and accessible to them throughout the study period. There were variable open responses on study length, although more participants wrote that a longer study period would be more representative and able to capture their OFD behaviors.

OFD Events

Accuracy of Using EMA to Capture OFD Events

As shown in Table 4, this study captured 124 OFD events from 56.9% (58/102) unique participants. The actual frequency of OFD ordering throughout the study period did not match the reported frequency at baseline in 52% (53/102) of the participants and exactly matched in 48% (49/102) of the participants. For instance, if a participant had reported using an OFD service once a week in the baseline survey and the study did not capture an OFD event during the monitoring period, this would be a case where the reported frequency did not match the actual frequency. In the 53 participants where actual food ordering frequency differed, 26 (49%) had more OFD events during the study compared with 27 (51%) who had ordered less than what was reported at the start of the study.

Table 4. Comparison of the number of online food delivery (OFD) events captured between varying sampling approaches (event-contingent vs signal-contingent; n=58 unique participants).

Sampling approach	OFD events captured, n	OFD orders from end-of-day reminders, n	Total, n
No Prompts (event-contingent)	57	22	79
Prompts (signal-contingent)	41	4	45
Total	98	26	124

As shown in Table 4, there appeared to be more OFD events (79/124, 63.7%) captured from the event-contingent group compared with the signal-contingent group (45/124, 36.2%). Results from a multivariable logistic regression showed that there was a strong level of evidence to support a significant association between the sampling approach and the capture of an OFD event during the study period after adjusting for age and gender. Compared with the Prompts group, participants in the No Prompts group were 3.53 times more likely (95% CI 1.52-8.17) to have logged an OFD event.

Contextual Data Surrounding Online Food Ordering Events

A total of 124 instances of OFD were captured from 58 unique participants. More than half of all orders (65/124, 52.4%) were placed in the evening (between 6 PM and 12 AM) and were less common at night or morning, comprising 9.7% (12/124) of ordering events each (Multimedia Appendix 3). As shown in Multimedia Appendix 3, most orders were made on a Saturday (25/124, 20.2%), followed by Tuesday (22/124, 17.7%) and Friday (21/124, 16.9%). The most common food delivery orders were pizza (23/124, 18.5%), fried chicken (18/124, 14.5%), Chinese food (18/124, 14.5%), and burgers (17/124, 13.7%). However, 28.2% (35/124) of the orders were from a category other than what was provided, including Mexican; bubble tea; fast-food franchises (Red Rooster, KFC, and McDonald's); cakes, pastries, and bakery items; groceries; and Vietnamese food. The following food categories were mentioned only once from all OFD events: hot chips and Caesar salad, breakfast platter, Malaysian, Nepalese, Afghan, modern Australian, Japanese, fried vegetables, Korean (fried chicken, chips, and seafood pancake), salad bowl, and sandwich.

For 50% (62/124) of the OFD orders captured, participants used Uber Eats, followed by DoorDash (23/124, 18.5% of the orders). In 54.8% (68/124) of the orders, participants ordered for just themselves, 15.3% (19/124) ordered for family, and 13.7% (17/124) ordered for their romantic partner.

The prevailing reason for using an OFD service was convenience, which contributed to 37.1% (46/124) of the responses. Taste and cravings attributed to 20.9% (26/124) of the orders. Most participants (98/124, 79%) ordered when they were at home, and 8.1% (10/124) of the participants ordered when they were at university or technical and further education. In 54.8% (68/124) of the online food ordering events, participants had ordered from a food outlet situated within a 15- to 30-minute walking distance, and 21.8% (27/124) of the participants ordered from an outlet that was between a 30-minute to 1-hour walk away (Multimedia Appendix 3).

A promotional code or offer was used for approximately one-third (37/124, 29.8%) of all orders. A wide variety of promotions and discounts were reported by the participants, from free delivery to receiving 10% or 50% off on their orders. The following discounts and promotions were only mentioned once from all OFD events: first-order discount, free additional food, referral discount, ShopBack, 30% off, AUD $16 off, AUD $20 off, and AUD $30 off (AUD $1 = USD $0.66). For 50.8% (63/124) of recorded OFD events, participants described the hunger levels as "quite hungry" or "really hungry." In 72.6% (90/124) of the OFD orders, participants responded that the order satisfied their cravings. Participants reported neutral to high levels of stress, ranging from not stressed or relaxed to very stressed in 72.6% (90/124) of all the OFD events.

Associations

In a zero-inflated negative binomial regression model adjusted for age, gender, and socioeconomic status, there was evidence that age and dependents were significantly associated with the number of OFD events in a week (Table 5). With every 1-year increase in age, the incidence rate of OFD events in a week decreased by 5% (95% CI 0.91-0.99). Compared with participants with no dependents, those with dependents had an incidence rate of OFD events 2.01 times higher (95% CI 1.16-3.49). Stress levels, age, gender, and socioeconomic status were not statistically significant factors associated with OFD events.

Table 5. Incidence rate ratios (IRRs) of demographic and lifestyle risk factors in a zero-inflated negative binomial regression model (N=101).

Factors	IRR (95% CI)	*P* value
Dependents		
No dependents (reference)	N/A[a]	N/A
Dependents	2.01 (1.16-3.49)	*.01*
Diet quality score	0.98 (0.96-1.00)	.1
Stress levels		
High stress (reference)	N/A	N/A
Moderate stress	0.60 (0.36-1.00)	.05
Low stress	0.51 (0.26-1.00)	.05
Age (years)	0.95 (0.91-0.99)	.03
Gender		
Man (reference)	N/A	N/A
Woman	0.85 (0.47-1.54)	.59
Socioeconomic status (IRSAD[b])	0.97 (0.84-1.13)	.73

[a]N/A: not applicable.

[b]IRSAD: Index of Relative Socio-Economic Advantage and Disadvantage.

Discussion

Principal Findings

This pilot study tested the feasibility and acceptability of using EMA to capture OFD events and contextual behaviors in young people aged 16 to 35 years. Compliance rates, as an indicator of feasibility, to signal-contingent and event-contingent protocols were similar at 72% and 73%, respectively, in adults. However, participants in the event-contingent group were more likely to have captured an OFD event during the study period. The EMA protocols for both groups were acceptable; however, most participants agreed that the study could be improved by an app that was easier to navigate. Contextual data revealed that OFD orders consisted largely of unhealthy foods. These orders were commonly placed for evening meals, typically after 6 PM, were mostly for a single person, and were ordered at home. Moreover, the orders often originated from an outlet within a 15- to 30-minute walking distance. Analyses showed that there was a significant association between age, number of dependents, and number of OFD events. These formative findings warrant further investigation with refinement of the EMA methodology to increase acceptability and improve the OFD data captured from participants.

High compliance rates indicated a feasible and acceptable EMA protocol for the study population and events captured. In the guidelines proposed by Stone and Shiffman [40] for EMA protocols, compliance rates of ≥80% are considered optimal. This study observed compliance rates between 70% and 73%, which fell short of the recommendations yet aligned with the findings from systematic reviews of similar dietary EMA studies conducted in young people. In a review by Liao et al [30], an average compliance rate of 71% was reported among studies. However, this is in contrast with the review by Battaglia et al [28], where compliance rates were reported to be >80% in more than half of the included studies. Stone et al [41] suggested that noncompliance is mainly attributed to monitoring burden and participants forgetting to record. For this study, the 3 additional reminders set 15 minutes apart and 1 end-of-day prompt included in the protocol aimed to eliminate instances of noncompliance from forgetfulness. Therefore, the monitoring burden may have influenced compliance. Feedback from participants showed that 27% (12/45) of the participants thought that the timing of prompts could have been improved. Data from the multilevel regression analysis supported this as prompts were more likely responded to in the afternoon compared with the evening. Similarly, another feasibility study on a mobile EMA intervention reported that several participants disliked receiving alerts at night, which may have influenced response rates to prompts [42]. These are important considerations for future EMA study designs aiming to collect data from participants during evening hours.

Participants who were assigned to the event-contingent sampling approach were 3 times more likely to have had an OFD event captured during the study compared with the signal-contingent group. This indicates that event-contingent sampling appeared to be the better approach to capturing OFD events. It is likely that event-contingent sampling was more appropriate because of OFD being a specific and sporadic event that, among study participants, varied in the frequency of use. Thus, unprompted records of any instance of OFD over the 1-week monitoring period were more effective than answering prompts delivered on 3 random sampling days. However, an additional end-of-day reminder proved to be critical for the event-contingent group as almost one-third (22/79, 28%) of the OFD orders were captured through this reminder, which otherwise would have been forgotten.

Another possible key factor affecting participants' experience of the EMA study and compliance was the perceived ease of use of the mobile app used to capture data. Many participants found that the EMA app was difficult to navigate and not user friendly, which may have negatively affected their experience of this study. Another EMA study also found participants commenting on the "difficult and bothersome" nature of EMA apps, although this was specific to the Visual Analog Slider used for one of the momentary assessments [29]. The Technology Acceptance Model is a highly influential information systems theory that suggests consumer acceptability is largely attributed to the perceived ease of use and usefulness of a given technological tool [43]. Consequently, the layout and technical aspects of EMA app themselves are critical to improving acceptability and compliance by participants.

This study revealed that orders were commonly placed for evening meals, typically after 6 PM, were mostly for a single person, and were ordered at home. Moreover, the orders often originated from an outlet within a 15- to 30-minute walking distance. According to physical activity research, a 1-km distance typically reflects a 15- to 20-minute walk for an average adult [44]. As such, this study revealed that a large proportion of customers are ordering food from potentially >1-2 km away and beyond what is typically considered their walkable neighborhood food environment [45] or the 20-minute neighborhood concept that is growing in popularity [46]. Furthermore, the results from this study are consistent with an annual market report from Uber Eats Canada, which revealed that 6 PM, typically considered "dinner time," was the most popular time of day for orders [47]. An Australian study on the food preparation location context of meals and snacks consumed by young adults showed that dinners prepared from outside of the home are predominately discretionary (20%) compared with the predominately 5-food group (13%) [48]. In addition, pizza was the most popular food category that was ordered, followed by fried chicken, Chinese food, and burgers. These findings also align with the annual report from Uber Eats Australia [49], which indicated that chicken burgers and chicken burritos were in the top 10 most popular orders list from local small businesses. Altogether, this provides preliminary evidence that most foods ordered by OFD services are not healthy and supplements previous studies that have analyzed online menu items and showed that a high proportion of offerings are poor in nutritional quality [13-15,50].

The results from a zero-inflated negative binomial regression model showed that dependents were positively associated with the number of OFD events. This finding is consistent with a previous study conducted by Keeble et al [51], which determined that the odds of any OFD service use were greater for those living with children (OR 2.71, 95% CI 2.44-3.01). Similarly, another study established that the positive relationship between access to the greatest number of food outlets online and use of OFD services in the previous week was specific to those living with children [12]. It is possible that dependents may increase their levels of financial and time stress [52], both of which present barriers to the preparation of healthy home-cooked meals [53]. The use of OFD services may help alleviate time constraints and hence explain the higher frequency of use among those with dependents.

Convenience prevailed as the selected reason for an OFD order (46/124, 37.1%), whereas "being busy" accounted for 12.1% (15/124) of the orders. Price promotions were also used in approximately one-third of all recorded events. These findings align with a qualitative study from the United Kingdom that investigated customer experiences of using OFD services [54]. Keeble et al [54] identified "less effort for more convenience" as a key theme behind customer's use of OFD. In addition, price promotions often influenced and justified the use of these services. Individuals with limited time may be higher consumers of out-of-home food, as research from the United States showed that in a sample of busy young adults, working >40 hours a week was associated with time-related barriers to healthful eating [55]. This perceived lack of time is linked to lower fruit and vegetable intake and greater intake of convenience and fast foods [56]. Of concern, a previous study has shown that ordering meals online for home delivery was significantly associated with higher levels of sugary drinks consumption (P=.003) and fast-food restaurant patronage (P<.001) [57]. A survey conducted in China similarly showed that young adults aged 18 to 30 years often ignored the nutritional value of their OFD orders; however, concerning trends emerged, as these young consumers reported potential physical health changes such as weight gain, elevated blood lipids, and gastrointestinal discomfort as a result of long-term OFD use [58]. However, this study did not find a significant association between diet quality score and the number of OFD events. Despite this, further research is needed to clarify the relationship between unhealthy diets and OFD use.

Strengths and Limitations

To the best of our knowledge, this is the first study to use EMA to capture OFD behaviors and context in near real time. Real-time monitoring is critical to ensuring ecological validity and is a major advantage of the EMA methodology, given the sporadic nature of OFD use. Moreover, this study gathered data on the use of OFD that is independent of market research or company reports and is an important addition to the public health literature that currently lacks evidence on the consumption of OFD.

However, several limitations of this study must also be acknowledged. This was a pilot study where 2 varying EMA protocols were tested for their feasibility and acceptability. Hence, caution is warranted when interpreting the associations found in this study between demographics, lifestyle risk factors, and OFD events. In addition, this study has yet to validate the capture of OFD consumption against gold standard dietary assessment methods such as the 24-hour recall. An important next step would be to examine the percentage match for each occasion of OFD consumption between a 24-hour recall and the EMA protocol.

All data gathered and analyzed in this study were self-reported by the participants, which is inherently prone to response bias. Furthermore, the EMA app used in this study was limited to iOS or Android users and appeared to cause more issues with Android users. Owing to convenience sampling, the study sample was highly educated females residing in areas of high socioeconomic advantage, which accordingly may limit the generalizability of the findings.

In addition, it is noted that the monitoring period of 7 days may not have adequately or accurately captured OFD events, as more than half of the study sample reportedly used OFD services 1 to 2 times a month. Future studies could use the event-contingent sampling approach and potentially extend the monitoring period to 14 days instead, alongside incorporating end-of-day reminders.

Moreover, although a training guide was provided online to participants, the study would also benefit from an in-person training day, as recommended by previous EMA studies [59,60]. This would reinforce what is required of the participants for the study, reduce missing data, and improve adherence to the protocol. A training day can also provide an opportunity to obtain objective measurements for weight, height, and waist circumference.

Conclusions

This study showed that mobile EMA is a viable method to capture OFD events and behaviors of its users in near real time, with better results obtained from event-contingent sampling. However, further amendments to the study protocol are necessary to improve compliance and acceptability among participants. In addition, validation of the EMA protocol against gold standard dietary assessment methods such as the 24-hour recall is an important next step. Further research is also essential to support the significant associations found between higher OFD use and those of younger age and those with dependents. This formative work may inform future EMA protocols to capture OFD orders and contextual factors and assist in further understanding OFD use and its impact on diet and health.

Acknowledgments

SJ was supported by a Research Training Program Stipend Scholarship from the Australian Government and the King and Amy O'Malley Trust Foundation Postgraduate Research Scholarship. SRP was supported by a National Heart Foundation Future Leader Fellowship (grant 106646). AAG was supported by an Investigator Grant from the Australian National Health and Medical Research Council (AAP1173784). These sources of support had no involvement in, or restrictions on, publication. The authors would like to express their gratitude to Ursula Power Vudrag, Janne Veija, Kimaya Jambhale, and John Ormerod from the School of Mathematics and Statistics, University of Sydney, for their assistance with the statistical analyses used for this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adapted Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist for Reporting EMA Studies (CREMAS).
[PDF File (Adobe PDF File), 150 KB - mhealth_v11i1e49135_app1.pdf]

Multimedia Appendix 2
Acceptability of the mobile-based ecological momentary assessment protocol comparing the 2 sampling approaches.
[PNG File , 148 KB - mhealth_v11i1e49135_app2.png]

Multimedia Appendix 3
Table of online food delivery (OFD) behaviors and contextual data captured from 58 unique participants.
[DOCX File , 23 KB - mhealth_v11i1e49135_app3.docx]

References

1. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
2. Adams J, Goffe L, Brown T, Lake AA, Summerbell C, White M, et al. Frequency and socio-demographic correlates of eating meals out and take-away meals at home: cross-sectional analysis of the UK national diet and nutrition survey, waves 1-4 (2008-12). Int J Behav Nutr Phys Act 2015 Apr 16;12(1):51 [FREE Full text] [doi: 10.1186/s12966-015-0210-8] [Medline: 25889159]
3. Saksena M, Anekwe T. America's eating habits: food away from home. U.S. Department of Agriculture, Economic Research Service. 2018. URL: https://www.ers.usda.gov/publications/pub-details/?pubid=90227 [accessed 2023-11-16]
4. Venn D, Banwell C, Dixon J. Australia's evolving food practices: a risky mix of continuity and change. Public Health Nutr 2017 Oct;20(14):2549-2558 [FREE Full text] [doi: 10.1017/S136898001600255X] [Medline: 27652992]

5. Ouyang Y, Wang H, Su C, Du W, Wang Z, Zhang B. Why is there gender disparity in the body mass index trends among adults in the 1997-2011 China health and nutrition surveys? Asia Pac J Clin Nutr 2015;24(4):692-700 [FREE Full text] [doi: 10.6133/apjcn.2015.24.4.06] [Medline: 26693755]
6. Zeng Q, Zeng Y. Eating out and getting fat? A comparative study between urban and rural China. Appetite 2018 Jan 01;120:409-415. [doi: 10.1016/j.appet.2017.09.027] [Medline: 28964905]
7. Popkin BM, Reardon T. Obesity and the food system transformation in Latin America. Obes Rev 2018 Aug 24;19(8):1028-1064 [FREE Full text] [doi: 10.1111/obr.12694] [Medline: 29691969]
8. Jaworowska A, Blackham T, Davies IG, Stevenson L. Nutritional challenges and health implications of takeaway and fast food. Nutr Rev 2013 May 11;71(5):310-318. [doi: 10.1111/nure.12031] [Medline: 23590707]
9. Nago ES, Lachat CK, Dossa RA, Kolsteren PW. Association of out-of-home eating with anthropometric changes: a systematic review of prospective studies. Crit Rev Food Sci Nutr 2014 Feb 05;54(9):1103-1116. [doi: 10.1080/10408398.2011.627095] [Medline: 24499144]
10. Mills S, Brown H, Wrieden W, White M, Adams J. Frequency of eating home cooked meals and potential benefits for diet and health: cross-sectional analysis of a population-based cohort study. Int J Behav Nutr Phys Act 2017 Aug 17;14(1):109 [FREE Full text] [doi: 10.1186/s12966-017-0567-y] [Medline: 28818089]
11. Online food delivery - Australia. Statista. 2021. URL: https://www.statista.com/outlook/dmo/online-food-delivery/australia [accessed 2023-11-16]
12. Keeble M, Adams J, Vanderlee L, Hammond D, Burgoine T. Associations between online food outlet access and online food delivery service use amongst adults in the UK: a cross-sectional analysis of linked data. BMC Public Health 2021 Oct 31;21(1):1968 [FREE Full text] [doi: 10.1186/s12889-021-11953-9] [Medline: 34719382]
13. Jindarattanaporn N, Suya I, Lorenzetti L, Kantachuvesiri S, Thamarangsi T. Nutritional content of popular menu items from online food delivery applications in Bangkok, Thailand: are they healthy? Int J Environ Res Public Health 2023 Feb 23;20(5):3992 [FREE Full text] [doi: 10.3390/ijerph20053992] [Medline: 36901000]
14. Partridge SR, Gibson AA, Roy R, Malloy JA, Raeside R, Jia SS, et al. Junk food on demand: a cross-sectional analysis of the nutritional quality of popular online food delivery outlets in Australia and New Zealand. Nutrients 2020 Oct 12;12(10):3107 [FREE Full text] [doi: 10.3390/nu12103107] [Medline: 33053705]
15. Mahawar N, Jia SS, Korai A, Wang C, Allman-Farinelli M, Chan V, et al. Unhealthy food at your fingertips: cross-sectional analysis of the nutritional quality of restaurants and takeaway outlets on an online food delivery platform in New Zealand. Nutrients 2022 Oct 30;14(21):4567 [FREE Full text] [doi: 10.3390/nu14214567] [Medline: 36364829]
16. Poelman MP, Thornton L, Zenk SN. A cross-sectional comparison of meal delivery options in three international cities. Eur J Clin Nutr 2020 Oct 24;74(10):1465-1473. [doi: 10.1038/s41430-020-0630-7] [Medline: 32332863]
17. Meal delivery services Uber Eats, Menulog, Deliveroo and DoorDash experienced rapid growth during 2020 – a year of lockdowns & work from home. Roy Morgan. 2021. URL: https://www.roymorgan.com/findings/meal-delivery-services-uber-eats-menulog-deliveroo-and-doordash-experienced-rapid-growth-during-2020-a-year-of-lockdowns-work-from-home [accessed 2023-11-16]
18. Online food delivery - worldwide. Statista. 2012 Aug 01. URL: https://www.statista.com/outlook/dmo/online-food-delivery/worldwide [accessed 2023-11-16]
19. Truesdale KP, Stevens J, Lewis CE, Schreiner PJ, Loria CM, Cai J. Changes in risk factors for cardiovascular disease by baseline weight status in young adults who maintain or gain weight over 15 years: the CARDIA study. Int J Obes (Lond) 2006 Sep 14;30(9):1397-1407 [FREE Full text] [doi: 10.1038/sj.ijo.0803307] [Medline: 16534519]
20. Growing up in Australia: the Longitudinal Study of Australian Children (LSAC) annual statistical report 2018. Australian Institute of Family Studies. 2018. URL: https://growingupinaustralia.gov.au/research-findings/annual-statistical-reports-2018 [accessed 2023-11-16]
21. Zhang Y, Hou F, Yang S, Li J, Zha X, Shen G. Beyond emotion: online takeaway food consumption is associated with emotional overeating among Chinese college students. Eat Weight Disord 2022 Mar;27(2):781-790. [doi: 10.1007/s40519-021-01224-2] [Medline: 34052988]
22. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014 Jul 22;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
23. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
24. Bucher Della Torre S, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1):76 [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
25. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A mobile ecological momentary assessment tool (devilSPARC) for nutrition and physical activity behaviors in college students: a validation study. J Med Internet Res 2016 Jul 27;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
26. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

27. de Vries LP, Baselmans BM, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021 Oct 23;22(5):2361-2408 [FREE Full text] [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
28. Battaglia B, Lee L, Jia SS, Partridge SR, Allman-Farinelli M. The use of mobile-Based Ecological Momentary Assessment (mEMA) methodology to assess dietary intake, food consumption behaviours and context in young people: a systematic review. Healthcare (Basel) 2022 Jul 18;10(7):1329 [FREE Full text] [doi: 10.3390/healthcare10071329] [Medline: 35885855]
29. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013 Sep 24;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
30. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
31. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
32. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
33. Himmelstein PH, Woods WC, Wright AG. A comparison of signal- and event-contingent ambulatory assessment of interpersonal behavior and affect in social situations. Psychol Assess 2019 Jul;31(7):952-960 [FREE Full text] [doi: 10.1037/pas0000718] [Medline: 30958026]
34. Collins CE, Boggess MM, Watson JF, Guest M, Duncanson K, Pezdirc K, et al. Reproducibility and comparative validity of a food frequency questionnaire for Australian adults. Clin Nutr 2014 Oct;33(5):906-914. [doi: 10.1016/j.clnu.2013.09.015] [Medline: 24144913]
35. Census of population and housing: socio-economic indexes for areas (SEIFA). Australian Bureau of Statistics. 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001#:~:text=SEIFA%202016%20has%20been%20created,of%20Economic%20Resources%20(IER) [accessed 2023-11-16]
36. Collins C, Burrows T, Rollo M, Boggess M, Watson J, Guest M, et al. The comparative validity and reproducibility of a diet quality index for adults: the Australian Recommended Food Score. Nutrients 2015 Jan 23;7(2):785-798 [FREE Full text] [doi: 10.3390/nu7020785] [Medline: 25625814]
37. Scoring the international physical activity questionnaire. University of Dublin. 2018. URL: https://ugc.futurelearn.com/uploads/files/bc/c5/bcc53b14-ec1e-4d90-88e3-1568682f32ae/IPAQ_PDF.pdf [accessed 2023-11-16]
38. Perceived Stress Scale (PSS-10). Child Outcomes Research Consortium. URL: https://www.corc.uk.net/outcome-experience-measures/perceived-stress-scale-pss-10/ [accessed 2023-11-16]
39. Sleep Needs Across Lifespan. Sleep Health Foundation. 2015. URL: https://www.nzgp-webdirectory.co.nz/site/nzgp-webdirectory2/Sleep-Needs-Across-Lifespan.pdf [accessed 2023-11-16]
40. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
41. Stone A, Shiffman S, Atienza A, Nebeling L. The Science of Real-Time Data Capture: Self-Reports in Health Research. Oxford, UK: Oxford University Press; 2007.
42. Beres LK, Mbabali I, Anok A, Katabalwa C, Mulamba J, Thomas AG, et al. Acceptability and feasibility of mobile phone-based ecological momentary assessment and intervention in Uganda: a pilot randomized controlled trial. PLoS One 2022 Aug 26;17(8):e0273228 [FREE Full text] [doi: 10.1371/journal.pone.0273228] [Medline: 36018846]
43. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319 [FREE Full text] [doi: 10.2307/249008]
44. Murtagh EM, Mair JL, Aguiar E, Tudor-Locke C, Murphy MH. Outdoor walking speeds of apparently healthy adults: a systematic review and meta-analysis. Sports Med 2021 Jan 08;51(1):125-141 [FREE Full text] [doi: 10.1007/s40279-020-01351-3] [Medline: 33030707]
45. Weinstein Agrawal A, Schlossberg M, Irvin K. How far, by which route and why? A spatial analysis of pedestrian preference. J Urban Des 2008 Feb;13(1):81-98. [doi: 10.1080/13574800701804074]
46. Thornton LE, Schroers R, Lamb KE, Daniel M, Ball K, Chaix B, et al. Operationalising the 20-minute neighbourhood. Int J Behav Nutr Phys Act 2022 Feb 12;19(1):15 [FREE Full text] [doi: 10.1186/s12966-021-01243-3] [Medline: 35151334]
47. Uber Canada's 2022 cravings report is in: you've got great taste, Canada!. UberEats Canada. 2022. URL: https://www.uber.com/en-CA/newsroom/2022-uber-canada-cravings-report/#:~:text=%E2%80%9CThis%20year%2C%20Canadians%20wanted%20spice,Uber%20Canada%20Cravings%20Report%20below [accessed 2023-11-16]
48. Chan V, Wellard-Cole L, Davies A, Watson W, Hughes C, Chapman K, et al. The association of social and food preparation location context with the quality of meals and snacks consumed by young adults: findings from the MYMeals wearable camera study. Eur J Nutr 2022 Oct 06;61(7):3407-3422 [FREE Full text] [doi: 10.1007/s00394-022-02891-2] [Medline: 35523897]

49. Uber Eats. Onions on the outer: Uber's Cravings report reveals how Australians were eating and drinking in 2022. UberEats Australia. 2022. URL: https://www.uber.com/en-AU/newsroom/cravings2022/#:~:text=Onions%20on%20the%20outer%3A%20Uber's,eating%20and%20drinking%20in%202022&text=Australia's%20infatuation%20with%20Asian%20cuisine,ahead%20of%20Thai%20and%20Chinese [accessed 2023-11-16]
50. Wang C, Korai A, Jia SS, Allman-Farinelli M, Chan V, Roy R, et al. Hunger for home delivery: cross-sectional analysis of the nutritional quality of complete menus on an online food delivery platform in Australia. Nutrients 2021 Mar 11;13(3):905 [FREE Full text] [doi: 10.3390/nu13030905] [Medline: 33799532]
51. Keeble M, Adams J, Sacks G, Vanderlee L, White CM, Hammond D, et al. Use of online food delivery services to order food prepared away-from-home and associated sociodemographic characteristics: a cross-sectional, multi-country analysis. Int J Environ Res Public Health 2020 Jul 17;17(14):5190 [FREE Full text] [doi: 10.3390/ijerph17145190] [Medline: 32709148]
52. Buddelmeyer H, Hamermesh DS, Wooden M. THE stress cost of children on moms and dads. Eur Econ Rev 2018 Oct;109:148-161. [doi: 10.1016/j.euroecorev.2016.12.012]
53. Mills S, White M, Wrieden W, Brown H, Stead M, Adams J. Home food preparation practices, experiences and perceptions: a qualitative interview study with photo-elicitation. PLoS One 2017 Aug 30;12(8):e0182842 [FREE Full text] [doi: 10.1371/journal.pone.0182842] [Medline: 28854196]
54. Keeble M, Adams J, Burgoine T. Investigating experiences of frequent online food delivery service use: a qualitative study in UK adults. BMC Public Health 2022 Jul 16;22(1):1365 [FREE Full text] [doi: 10.1186/s12889-022-13721-9] [Medline: 35842625]
55. Escoto KH, Laska MN, Larson N, Neumark-Sztainer D, Hannan PJ. Work hours and perceived time barriers to healthful eating among young adults. Am J Health Behav 2012 Nov;36(6):786-796 [FREE Full text] [doi: 10.5993/AJHB.36.6.6] [Medline: 23026037]
56. Welch N, McNaughton SA, Hunter W, Hume C, Crawford D. Is the perception of time pressure a barrier to healthy eating and physical activity among women? Public Health Nutr 2009 Jul;12(7):888-895 [FREE Full text] [doi: 10.1017/S1368980008003066] [Medline: 18647424]
57. Dana LM, Hart E, McAleese A, Bastable A, Pettigrew S. Factors associated with ordering food via online meal ordering services. Public Health Nutr 2021 Dec;24(17):5704-5709 [FREE Full text] [doi: 10.1017/S1368980021001294] [Medline: 33762026]
58. Dai X, Wu L, Hu W. Nutritional quality and consumer health perception of online delivery food in the context of China. BMC Public Health 2022 Nov 19;22(1):2132 [FREE Full text] [doi: 10.1186/s12889-022-14593-9] [Medline: 36403009]
59. Dunton GF, Dzubur E, Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016 Jun 01;18(6):e106 [FREE Full text] [doi: 10.2196/jmir.5398] [Medline: 27251313]
60. Degroote L, DeSmet A, De Bourdeaudhuij I, Van Dyck D, Crombez G. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2020 Mar 10;17(1):35 [FREE Full text] [doi: 10.1186/s12966-020-00932-9] [Medline: 32151251]

Abbreviations

EMA: ecological momentary assessment
IRR: incidence rate ratio
IRSAD: Index of Relative Socio-Economic Advantage and Disadvantage
OFD: online food delivery
REDCap: Research Electronic Data Capture

Edited by G Eysenbach, L Buis; submitted 22.05.23; peer-reviewed by M Keeble, S Wang; comments to author 29.09.23; revised version received 18.10.23; accepted 05.11.23; published 29.11.23.

Please cite as:
Jia SS, Allman-Farinelli M, Roy R, Phongsavan P, Hyun K, Gibson AA, Partridge SR
Using Mobile Ecological Momentary Assessment to Understand Consumption and Context Around Online Food Delivery Use: Pilot Feasibility and Acceptability Study
JMIR Mhealth Uhealth 2023;11:e49135
URL: https://mhealth.jmir.org/2023/1/e49135
doi:10.2196/49135
PMID:38019563

Stakeholders' Perceptions Regarding Digital Therapeutics Reimbursement in South Korea: Qualitative Study

Boram Sim[1], MPH, PhD; Jin Han Ju[1], MPH; Byungsoo Kim[2], BP, MPEP; Jin Yong Lee[3,4], MHA, MD, PhD

1

2

3

4

Corresponding Author:
Jin Yong Lee, MHA, MD, PhD

Abstract

Background: Digital therapeutics (DTx) are therapeutic interventions driven by software and directly provided to patients, allowing them to manage their health with ease in any setting. A growing interest in DTx has spurred a discussion concerning their reimbursement pathways. However, DTx are still at a premature stage, with insufficient evidence on effectiveness, efficiency, and safety. Currently, although industries desire to quickly enter the market, especially by getting their products reimbursed by the National Health Insurance (NHI) fund, the NHI is cautious about DTx due to their uncertainties. Thus, public discussion and social consensus are crucial in deciding whether to reimburse DTx by the NHI fund.

Objective: This study examined multiple stakeholders' awareness and attitudes toward DTx and perceptions of regulatory pathways for adopting DTx.

Methods: In-depth interviews were conducted with 11 stakeholders in South Korea (industry: n=4, health care: n=3, academia: n=2, and consumer: n=2) using semistructured guidelines. They were purposively sampled to identify individuals with expertise in DTx and NHI policies. The interviews were conducted either in person or via a videoconference for 45-70 minutes. Qualitative data were analyzed using directed content analysis, which uses interview guidelines as an analytical framework.

Results: Findings were divided into three categories: (1) awareness and attitude toward DTx, (2) perception of whether DTx are worth entering the market and being reimbursed by the NHI fund, and (3) perception of how to enter the market and how to reimburse DTx by the NHI fund if they are worth it. Although consumer stakeholders were not familiar with the basic concept of DTx, the other stakeholders understood it thoroughly. However, all participants showed positive attitudes and acceptance of DTx. Most of them responded that DTx are worth entering the market, but they could not reach an agreement on the pathways for DTx to enter the market. Although participants were in favor of the reimbursement of DTx in principle, they responded that a conservative approach is required due to insufficient clinical evidence for DTx.

Conclusions: We found that stakeholders in South Korea had positive attitudes toward DTx, perceived them as worth using, and agreed to allow them to enter the market. The main issue was not the problem of the technology itself but the difference in opinion as to the pathways for reimbursement. Therefore, this study concluded that the NHI fund, which is operated very conservatively, is insufficient to quickly adopt and implement DTx. Various reimbursement methods, including tax-based financing, raising innovation funds for new technologies, and pilot studies using the NHI fund, should be used to rapidly generate clinical evidence and reduce the uncertainties of DTx to secure a stable market.

(JMIR Mhealth Uhealth 2023;11:e47407) doi:10.2196/47407

KEYWORDS

digital therapeutics; digital technology; mobile applications; internet-based intervention; reimbursement mechanism; qualitative research; perception; reimbursement; software; intervention; stakeholder; implementation; opinion; reimbursement method

Introduction

Digital therapeutics (DTx) refer to software as a medical device that is independently operated from hardware. They are used to prevent, manage, and treat various illnesses [1]. The most acclaimed DTx product is the *reSET* mobile app developed by Pear Therapeutics. Approved by the US Food and Drug Administration in 2017, the *reSET* mobile app provides cognitive behavioral therapy for patients with substance use disorder. It has garnered global attention as studies proved its efficacy to treat diseases, unlike several previously launched health care apps [2]. Subsequently, DTx have gained momentum as a new treatment modality that may substitute or supplement conventional medical practices. Notably, as the COVID-19 pandemic prompted individuals to recognize the advantages of telemedicine, in which health care services could be provided remotely, the potential of DTx to deliver therapeutic interventions through software has gained significant attention [3]. Considering this recent growing interest in DTx, their market has been expanding rapidly [4,5].

The emergence of the DTx market has spurred discussions concerning reimbursement pathways, as securing them guarantees a stable market for DTx developers upon their release. In South Korea, where almost the majority of the population is included in the National Health Insurance (NHI), the fact of whether DTx is covered under the NHI has been directly linked to the future viability of DTx developers. However, the decision on whether medical technologies are eligible for reimbursement from the NHI requires careful consideration, given its implications on the population's health and limited available financial resources. Thus, the clinical and economic aspects of medical technology are evaluated through several rounds of assessments after obtaining approval from the Korean Ministry of Food and Drug Safety (MFDS), including the New Health Technology Assessment (nHTA) and NHI benefit coverage determination. Reaching a final decision regarding these processes requires time, which can cause advanced technologies with short market cycles to abandon research and development due to delayed market entry. Thus, in March 2019, the South Korean government introduced a new track for the nHTA called Innovative Health Technology Assessment (IHTA), which allows technologies to collect evidence after rapidly entering the market, provided that it has substantial potential value. Moreover, some technologies designated as IHTA subjects can be reimbursed by the NHI temporarily. In November 2021, the South Korean government announced that they would consider DTx to be reimbursed via this track [6].

However, contrary to other innovative health technologies, DTx have not yet been launched in South Korea. Although 2 DTx for insomnia obtained approval from the MFDS in February and April 2023, they have not been released to the market yet. Furthermore, there are limited cases of development and use of DTx in overseas markets, which obscure their practicality. Moreover, it is challenging to ascertain whether physicians and patients will accept the novel technology. The therapeutic effectiveness of DTx hinges on sufficient patient engagement. Nevertheless, there is no guarantee that patients will consistently adhere to their treatment protocol [7,8]. Therefore, discussing DTx reimbursement currently constitutes a NHI fund investment with an uncertain future value. Hence, it is imperative to arrive at a consensus regarding how the principles of NHI reimbursement must be upheld and the opportunities that should be allocated to innovative health technologies. Nonetheless, most discussions are driven by only industry stakeholders, with no existing studies integrating diverse group opinions, including those of physicians and patients.

To our knowledge, the health care industry is urgently advocating for DTx reimbursement to secure a stable market after its development and establish a foothold for overseas expansion. Conversely, although the health care community acknowledges the potential of digital health technologies, there are increasing concerns that the value of DTx is currently overestimated [9,10]. Interestingly, in the academic community, some researchers are demanding appropriate regulatory reforms for these new technologies [11,12]. However, consumer perspectives are predominantly lacking in the existing body of literature. Therefore, this study aimed to examine diverse stakeholders' awareness and attitudes toward DTx and their perceptions of regulatory pathways for adopting them. To the best of our knowledge, this is the first study to comprehensively analyze the perceptions of various stakeholder groups regarding DTx reimbursement in South Korea.

Methods

Overview

For this study, we conducted in-depth interviews to explore the stakeholders' perceptions about DTx reimbursement. In-depth interviews are a qualitative research methodology wherein focused interviews are conducted with participants for an extensive exploration of their perspectives on a specific topic [13]. Thus, the interviews are conducted with a small sample size of participants, typically from 10 to 30 [13,14]. Data collection using individual interviews was deemed to be appropriate for this study, as the participants may have had varying levels of understanding about DTx and the health insurance reimbursement decision process.

Ethical Considerations

This study was approved by the institutional review board of the Health Insurance Review and Assessment Service (2021-116-003). All participants were contacted by email or phone and agreed to be interviewed. They were reimbursed with KRW 200,000 (approximately US $ 148) for their participation in this study.

Recruitment

In all, 11 academic, health care, industry, and consumer experts were included as participants to explore various group opinions using purposive sampling. Individuals who are currently developing DTx, have participated in their development process, or are conducting relevant research were included. We collected news published on this topic over the past year and identified experts with opinions on both DTx and NHI policies. Individuals who have participated in the NHI reimbursement decision-making were selected for the consumer group, as only a limited number of people have had exposure to DTx. Representatives of a civic group and a consumer group belonging to the Health Insurance Policy Deliberation Committee were selected for this study.

Development of Interview Guidelines

We developed the interview guidelines following the latest government policy announcements [6]. Subsequently, these were finalized in collaboration with a qualitative research expert. The critical questions for the interview were about perception and attitude regarding DTx; whether it has high value to warrant market entry and reimbursement through the NHI fund; and if so, how they must be given market access (Table 1).

Table . Structure of the interview guidelines.

Categories	Questions
Perceptions about DTx[a]	• Do you know what DTx are? What do you think the scope of DTx is? • What benefits do you anticipate when DTx are adopted in health care? • Conversely, what concerns do you have?
Perceptions about the reimbursement of DTx	• Do you think that DTx should be covered by health insurance?
Perceptions about pathways to adopt DTx	• Given that there is inadequate clinical evidence for DTx, what do you think about permitting the release of DTx products to the market before a health insurance coverage decision is made? • In that case, what do you think about reimbursing it provisionally from the NHI[b] fund (selective benefit[c], where the NHI reimburses 10%)? • What criteria should be used to assess products subject to the NHI reimbursement decision? Do you think it is appropriate to compare DTx with the standard of care that they are replacing or complementing? • In this case, do you think that DTx could still be accepted based on their other benefits even if their effectiveness falls short of that of the standard of care?

[a]DTx: digital therapeutics.

[b]NHI: National Health Insurance.

[c]The selective benefit is a policy applying to medical services that do not have sufficient evidence yet, but the need for reimbursement is recognized. They are reimbursed by lowering the percentage paid by the National Health Insurance.

Interview Procedures

The interviews were conducted either in person or via a videoconference from January 25 to February 15, 2022, by a female researcher (BS). Before the interview, the interviewer discussed the interview methods and contents with a qualitative research expert. The interview took place at a closed meeting room of Health Insurance Review and Assessment Service or at the participant's workplace. Aside from the interviews with the 2 industry participants, nobody else was present during the interview apart from the participants and researchers. The 2 industry participants were accompanied by observers from their companies. There was no prior relationship between the interviewer and the participants.

The interview was conducted using semistructured guidelines. The interview topics were sent to the participants beforehand for them to organize their thoughts and opinions. They were also informed on the background and goals of the study, excluding the characteristics of the interviewer that could lead to bias. For the consumer group, the participants were asked whether they knew about DTx during the screening process, and it was found that DTx awareness was deficient among the participants. Thus, a brief explanation of the concept and overseas cases of DTx were provided to them.

The interviews lasted for 45-70 minutes, and the entire interview was recorded and transcribed. The researchers also took field notes during the interviews. All interviews were conducted in Korean and were subsequently translated to English for this paper. Interviews were conducted until saturation was reached.

Analysis

This study performed a qualitative data analysis called directed content analysis, which involves analyzing data via a structured process using existing theories or study frameworks [15], using interview guidelines as the analytical framework.

For accuracy, the first author (BS), who conducted the interview, along with another researcher independently reviewed the interview recordings. After reading the recordings thoroughly to get an overall impression, the reviewers repeatedly read the transcriptions to grasp its meaning and highlighted the relevant statements. Furthermore, they made notes to identify critical themes and categorized them into predefined categories based on the guidelines. Unclassifiable content that featured a shared concept or category was established as a new category. For example, the perception of the reimbursement of DTx was divided into the perception of market entry and that of reimbursement by the NHI fund.

To improve the reliability and validity of the data analysis, the critical themes were repeatedly verified by 2 reviewers. Subsequently, all researchers discussed the content and names of the critical themes and arrived at a consensus on the primary findings. Additionally, we sought feedback on the findings and interpretations from experts with practical experience in health insurance reimbursement decisions not involved in this study. Finally, quotes were selected to illustrate each theme.

Results

Participant Characteristics

A total of 11 participants were included in this study (academia: n=2, industry: n=4, health care: n=2, and consumer: n=2). In all, 9 (82%) participants were male, and the participants were aged between their 30s to 60s (Table 2). The academia and health care groups were found to have experience and expertise on DTx. The industry group consisted of chief executive officers of companies that are in the progress of clinical trials approved by the MFDS. Thus, they had views and opinions on the adoption pathways of DTx. For the consumer group, even though they did not have experience related to DTx, they were well aware of the adoption pathways of new technologies in general.

Table . Participant profile.

Group and participant number	Sex	Age group (y)	Experiences
Academia			
1	Male	30s	• Experience in DTx[a] policy development
2	Male	50s	• Experience in developing DTx product
Industry			
3	Male	40s	• CEO[b] of a DTx company • MFDS[c]-approved clinical trials in progress
4	Male	30s	• CEO of a DTx company • MFDS-approved clinical trials in progress
5	Male	50s	• CEO of a DTx company • MFDS-approved clinical trials in progress
6	Male	50s	• CEO of a DTx company • MFDS-approved clinical trials in progress
Health care			
7	Male	60s	• Psychiatrist • Experience in developing DTx product
8	Male	40s	• Psychiatrist • Experience in developing DTx product
9	Male	50s	• Psychiatrist • Experience in developing DTx product
Consumer			
10	Female	60s	• Representative of a consumer organization • Experience on NHI's[d] benefit coverage determination
11	Female	50s	• Representative of a civil society • Experience on NHI's benefit coverage determination

[a]DTx: digital therapeutics.

[b]CEO: chief executive officer.

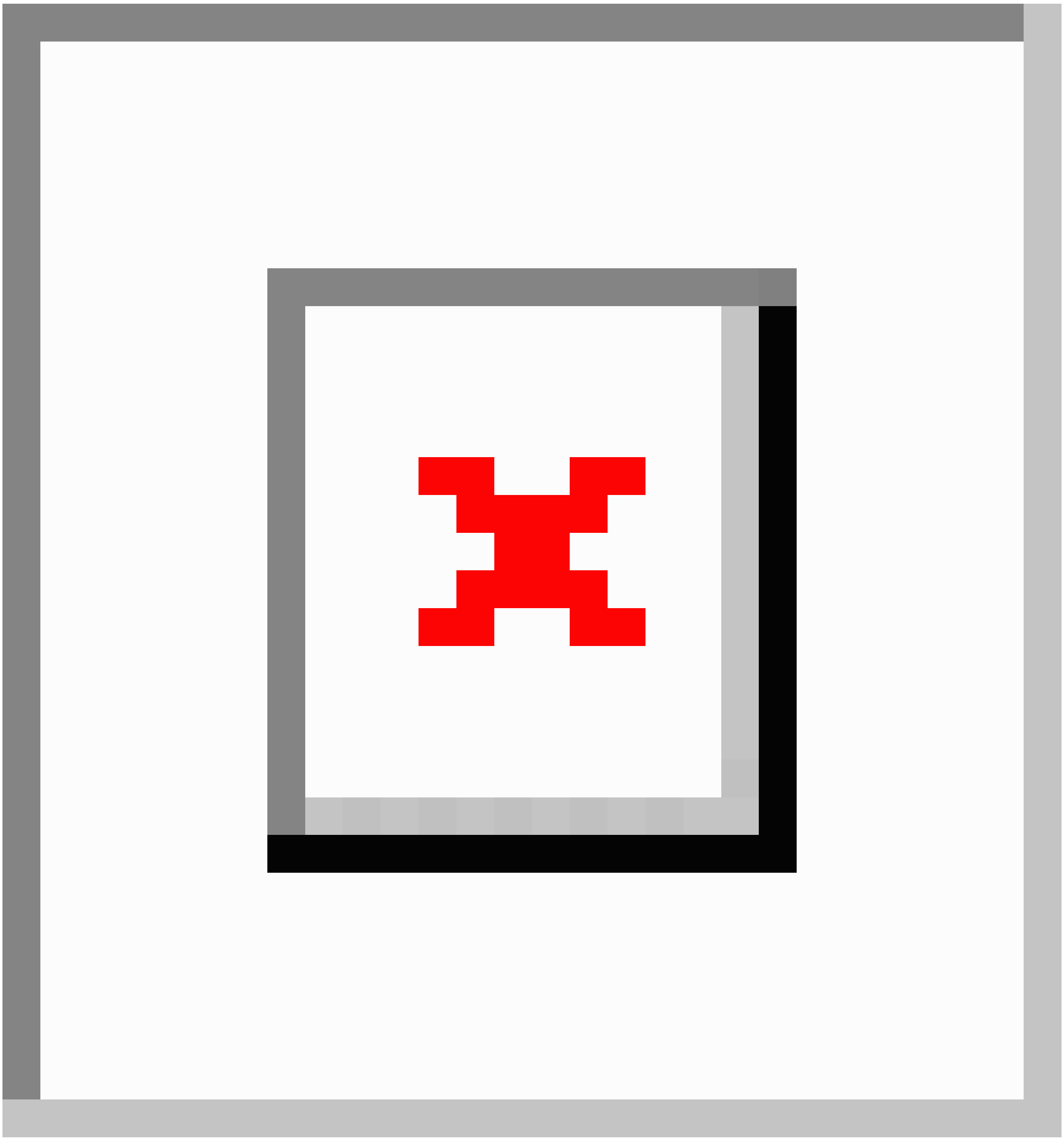
[c]MFDS: Ministry of Food and Drug Safety.

[d]NHI: National Health Insurance.

Overview

The findings were divided into three categories: (1) awareness and attitude toward DTx, (2) perception of whether DTx are worth entering the market and being reimbursed by the NHI fund, and (3) perception of how to enter the market and how to reimburse DTx by the NHI fund if they are worth it (Figure 1).

Figure 1. Framework of the study. DTx: digital therapeutics; NHI: National Health Insurance.

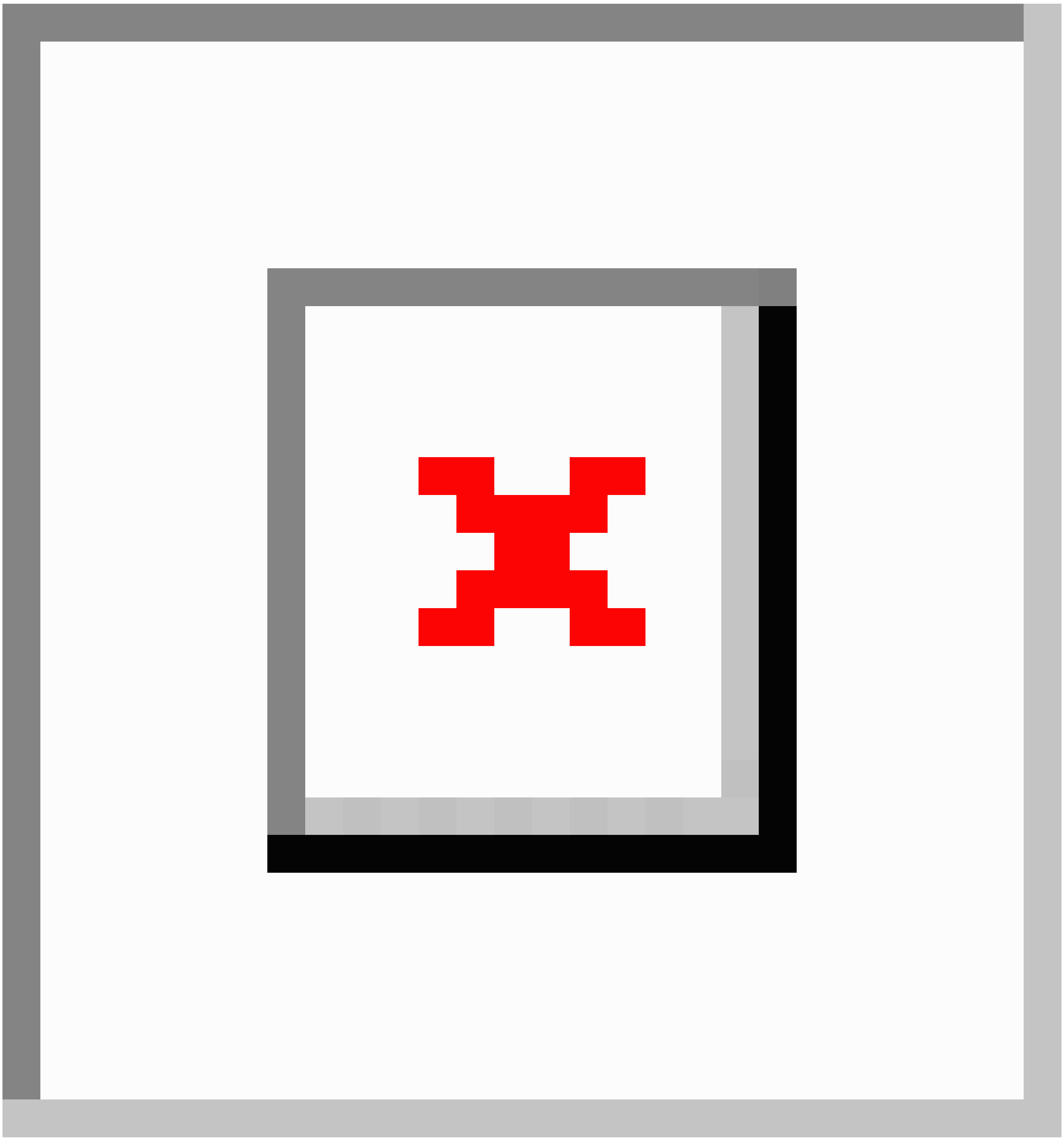


Awareness and Attitudes Toward DTx

DTx Concept

It was found that all participants of the industrial, academic, and health care communities had a good understanding of the DTx concept as a “medical software that provides evidence-based therapeutic interventions” (Figure 2). However, the consumer group participants had limited knowledge of DTx. Participants with DTx knowledge emphasized that it must be distinguished from wellness products that aim to prevent the occurrence of diseases or maintain the well-being of people who are healthy or at health risk. Nevertheless, there were variances in their perceptions about its details, such as whether a physician’s prescription should be required, whether DTx include prevention before disease onset, and whether the places of product use should be differentiated (such as hospitals and patients’ homes).

Figure 2. Awareness and perceptions about DTx and its market release and reimbursement. DTx: digital therapeutics; NHI: National Health Insurance.



Before, it was hard to come across information about DTx. So, I would say that my knowledge about DTx is just about the same as an average person's. I have heard some explanations from a company that wanted to expand its business of a wearable [device] a while ago, so I have some basic understanding about it. However, when it comes to present-day DTx, I do not really know much about it. [Consumer stakeholder, participant 10]

This is my first time learning about DTx. So, I searched the news before coming here. [Consumer stakeholder, participant 11]

Just like the level of evidence differs between prescription drugs and health functional foods, there is bound to be confusion between DTx and healthcare products. It is crucial to differentiate them, especially in the early stages. The key to this is to have therapeutic evidence (clinical trials) just like prescription drugs. [Academia stakeholder, participant 2]

The definition provided by the Digital Therapeutics Alliance does not actually mention the need for approval from the Korean Food and Drug Administration or a prescription from a doctor. Therefore, some [can] argue that we should perceive the category of DTx more broadly. In fact, in other countries, there is already a distinction between

> *Prescription DTx (PDT) and Non-PDT.* [Health care stakeholder, participant 8]

> *It does not make sense to call anything that can be used like a general application as DTx. I think we should limit the scope of DTx to things that are actually prescribed by physicians and actually benefit patients when they use it.* [Health care stakeholder, participant 7]

> *Although there is no set business model for the DTx currently being developed, it could be used in hospitals or homes.* [Industry stakeholder, participant 6]

> *I personally think that anything prescribed by a doctor or applied to a patient should be considered DTx. When there is a prescription from a doctor, it could be applied both to the patient and to those who are at risk of disease. I believe that DTx should remain applicable to people in the pre-hypertension stage who does not need medication yet. If we treat people in the pre-hypertension stage, it prevents them from progressing to hypertension. Therefore, the scope of DTx should include prevention and management, not just treatment.* [Industry stakeholder, participant 5]

Expectations and Concerns Regarding DTx

The industry sector placed a high value on the fact that DTx can provide treatment that could not be adequately provided in the existing health care system. The industry experts specifically believed that DTx can reduce drug dependence by inducing behavioral changes and will ultimately reduce pharmaceutical spending and other societal costs eventually. Academic and health care experts emphasized that DTx help patients manage their condition in their daily lives and allows doctors to monitor the progression of treatment of their patients. Moreover, some expect DTx to create an opportunity for payers to pursue value-based care because they can collect treatment progress and outcome data through DTx. The consumer group participants also hoped that DTx would enable patients to assess and manage their conditions more objectively. As DTx provide a noninvasive therapeutic intervention, the participants expected the software to have few safety concerns.

> *To treat patients with insomnia properly, you need to use cognitive behavioral therapy, but it is underutilized in hospitals, right? DTx can fill this gap, providing proper treatment, reducing social costs, and improving the quality of life. It would have these kinds of benefits.* [Industry stakeholder, participant 5]

> *In the medical field, there are many gaps, especially in improving self-management and lifestyle habits to enhance health and overcome diseases in daily life. I think there could be a new paradigm shift in these areas.* [Health care stakeholder, participant 8]

> *Doctors also want to see patient data. I think DTx provides a reliable channel for doctors to access patient data, and through this, an agenda should be set on how the quality of care can be improved.* [Academia stakeholder, participant 1]

> *You know, we use several smartwatches these days. I think this is like a pre-DTx phase. This is why I think it is better from a consumer's perspective because it clearly shows my status in numbers. But first, I think the patient must be willing.* [Consumer stakeholder, participant 11]

> *The Ministry of Food and Drug Safety (MFDS) strictly manages clinical trials and examines whether DTx are effective or harmful. However, they do that because they are worried that some treatments are going to be claimed effective, although they are not, not because there are safety issues.* [Health care stakeholder, participant 7]

Perception on Whether DTx Deserve to Enter the Market

Although most participants perceived that DTx are suitable to be allowed market entry, they had concerns regarding their effective use. Specifically, they raised concerns about the "prescribing physicians' lack of knowledge on using DTx," "possible low therapeutic benefits from the inherent limitations of DTx," and "uncertainty surrounding patient engagement." Additionally, the consumer group participants expressed concerns regarding DTx use for commercial purposes.

> *DTx is a field where the necessity itself is clear. From the perspective of chronic patient management, it is certain that we will go in this direction in the future. There is no reason for not using it that has already been developed, and it can definitely be helpful in managing health. It would be great to introduce it quickly, but the problem is whether it has been validated and whether it really has that much effect.* [Health care stakeholder, participant 7]

> *When DTx is introduced, the level of education of prescribing physicians is one of the crucial factors that needs to be considered. There are some concerns that physicians may not be able to distinguish between general digital health devices and DTx and may not know how to use them properly and say things like, "How can I use this in treatment? How much do I need to use?"* [Health care stakeholder, participant 9]

> *Do older patients adapt well to DTx? Although companies developing and doctors prescribing these may be extremely enthusiastic, how actively will patients participate? Even after the pharmacist explains about a medicine thoroughly, the patients just decide not to follow the instruction the moment they turn away…I am also concerned that companies may push the product anyway despite predicting the side effects or results that fall short of expectations.* [Consumer stakeholder, participant 10]

Perception on Whether DTx Deserve to Be Reimbursed

The participants responded that the reimbursement of DTx is necessary to improve patients' health and enhance the financial stability of the NHI fund. However, some participants were cautious about reimbursement due to uncertainties about DTx's effectiveness and acceptability. Nevertheless, the industry participants considered reimbursement important for securing a stable market and recognizing DTx to be distinct from general wellness products. Nonetheless, some industry participants state that they may choose different strategies that vary by product, because once a product gets reimbursed, its price will be controlled by the NHI, and it may be more advantageous as a noncovered item.

> *Wouldn't reimbursement be needed if its effectiveness is proven and it is believed to be helpful for patients' health?* [Industry stakeholder, participant 6]

> *Reimbursement from the NHI fund is necessary. I prefer HTA after its market release but if that's difficult in our country, then I think market release after HTA is also good. (Reimbursement is needed) for sound finances of the NHI.* [Academia stakeholder, participant 1]

> *If the DTx is prescribed by a doctor in a hospital and used on a patient, then having it reimbursed allows more patients access to it, so we have that model in mind for now.* [Industry stakeholder, participant 5]

> *Having our product reimbursed is a particularly important issue for us because it means that the government acknowledges its effectiveness and officially compensates for it. This will help people distinguish therapeutics that they really need to use from health functional products they can just buy at a market.* [Industry stakeholder, participant 4]

> *It would be nice to have DTx reimbursed, but in terms of practicality, how willingly will the patients accept it?* [Consumer stakeholder, participant 11]

> *I'm a little cautious about the reimbursement. Reimbursing it as an independent therapeutic should be decided with more deliberation. This is the national budget we are talking about here, and I believe that if there is a better option with more concrete effectiveness, then it is right to give more money to that option.* [Health care stakeholder, participant 8]

> *From the company's perspective, reimbursement is not necessarily all sunshine and rainbows because the price will be controlled by the NHI. It might be better to leave it as a non-covered item. So, I think I could say that companies do not necessarily want reimbursement of all DTx items they are developing.* [Industry stakeholder, participant 3]

Perception on Adopting Pathways of DTx

Early Market Entry

Currently, DTx follow the IHTA track; hence, the products may be released in the market to accumulate clinical evidence before being subject to the nHTA [6]. The participants positively viewed approving its early market entry, considering the "rapid development cycle of DTx," "low risk," and "easier data collection for assessment."

> *DTx is much easier to collect RWD (real-world data) from, and due to its digital nature, safety issues are of much less concern. This is why we are trying it on the field to collect data.* [Academia stakeholder, participant 2]

> *Considering the rapidly changing pace of technological development, we could keep falling behind if it takes too long to apply DTx to clinical settings.* [Industry stakeholder, participant 5]

> *Many companies are eager to enter the market early. The development is short, and the development cost is low, so even if they get MFDS approval, they are afraid that other companies will copy the form or algorithm before they are listed in the NHI.* [Health care stakeholder, participant 9]

Provisional Reimbursement From the NHI Fund

It was noted that the participants had conflicting opinions regarding the provisional reimbursement from the NHI fund for the DTx products with an early market entry. First, the industry group participants stated that provisional reimbursement is essential, considering that the NHI shares the risk taken by physicians and patients in trying a new treatment modality. The consumer group participants agreed to reimburse temporarily to alleviate patients' financial burden and help effectively adopt DTx in the health care system. The academia and health care group participants considered provisional reimbursement positive for supporting the development of innovative technologies; however, they emphasized that products with potential must be adequately screened. Conversely, some participants stated that DTx should be released in the market as a noncovered item at first to examine the market's reaction to this novel treatment modality.

> *Provisional reimbursement means that the government is sharing the risk, even if it is only 10%. I agree with that way from the perspective that it grants doctors some comfort when prescribing a new treatment.* [Industry stakeholder, participant 4]

> *If the purpose of the provisional reimbursement is to provide even the slightest support until they can show that the product is effective (I think it is needed), and of course, it would be great if all DTx products can be reimbursed, but if the financial resources do not allow it, then I think it would be right to apply it first to the products that have greater potential…Initiating reimbursement quickly may not be the best option. I hope a slower approach is taken so that both doctors and patients can experience the benefits of DTx through successful cases.* [Health care stakeholder, participant 8]

> *The concept is still relatively new and unfamiliar, so it may be necessary for it to start as a non-covered item to assess how much it is used. This is often the reason why non-covered items exist. However, from my viewpoint, regardless of whether it is provisionally reimbursed, it is not a big issue as long as it can be used.* [Health care stakeholder, participant 7]

> *I think it would be more rational to allow innovative medical technologies to enter the market quickly but with the price competitiveness being determined within the functions of users and suppliers. If it is effective, and there is demand among patients, they would be willing to pay for it, right?* [Industry stakeholder, participant 6]

Criteria for Reimbursement Decision of DTx

As the financial resources of the NHI are limited, new medical procedures, devices, or medicine are required to demonstrate their value through comparison with their alternatives. Most participants considered it to be suitable to compare the effectiveness of DTx with medical practices that they will substitute or complement, specifically in terms of the standard of care. However, there were conflicting opinions regarding the effectiveness level to be proven for DTx to be eligible for reimbursement. Some presented conservative opinions that suggested that DTx must have therapeutic benefits equivalent to an existing treatment to be eligible for NHI reimbursement. Others presented more modern perspectives, stating that even if DTx have lower therapeutic benefits than existing treatments, they should still be considered for reimbursement based on their unique benefits. The unique DTx benefits proposed were "increased access to treatment," "improved patient experience and convenience," "reduced adverse drug reactions," and "lower societal costs."

> *I think there should be a convincing criterion for reimbursement decisions. Of course, there may be criticism that the standard of care is not an equivalent comparison, but I do not think we should just see it that way, as there may be no other alternatives…* [Health care stakeholder, participant 8]

> *I think the basic condition is that even if the effectiveness is lower than expected, it should at least be as effective as the care originally provided in the hospital. However, if the difference in effectiveness is clinically acceptable, and there are other advantages like saving patients' travel time, cost, and other social costs, we can consider reimbursement.* [Consumer stakeholder, participant 11]

> *If compared to treatment as usual (TAU), DTx would never be adopted. However, there are cases where DTx treatment is necessary, such as people who refuse to go to a mental health clinic, people who do not have time to go to a hospital, and so on, even if its effectiveness is somewhat lower.* [Academia stakeholder, participant 1]

> *I hope that DTx can be highly valued for its ability to provide treatments that are recommended in domestic and international clinical guidelines but are not actually available.* [Industry stakeholder, participant 4]

> *However, we should think about what level of effectiveness we are aiming for. There will surely be people who try to bring groundless things with absolutely no evidence for therapeutic potential and call it DTx.* [Health care stakeholder, participant 9]

Discussion

Principal Findings

The development of DTx is currently at an early stage globally; hence, their effectiveness, safety, efficiency, and other aspects have not been adequately proven. Thus, sufficient societal discussion and consensus are required to decide whether DTx deserve to be reimbursed through the NHI fund. As per our knowledge, this is the first study to explore various stakeholders' opinions regarding the awareness and attitudes toward DTx and their perception on the regulatory pathways through which DTx should be adopted.

We found that the industry, academic, and health care experts possess a firm understanding of the fundamental concepts of DTx, whereas consumer experts have limited knowledge. Hence, it is highly probable that the general public remains entirely unfamiliar with DTx or has a limited understanding of them. This lack of awareness causes difficulty in accurately appraising the value of DTx [16]. Therefore, the public must be informed regarding what are DTx and their advantages and disadvantages to foster informed societal discourse.

It was found that all participants displayed favorable and receptive attitudes toward DTx. Most perceived that the DTx technology deserves market access and agreed to the quick adoption of this novel technology to prevent it from becoming obsolete. However, some participants expressed doubts about its clinical effectiveness, the appropriate use of the technology by health care providers, and patient engagement. In other words, the participants were optimistic about the transition in the health care system that DTx can bring and simultaneously recognized the high level of uncertainty associated with them. These findings align with prior research that cautioned to avoid overestimating DTx value without sufficient evidence [10,17].

This perception was also apparent in the participants' attitudes toward the reimbursement of DTx. Notably, the industry experts sought reimbursement from the NHI fund to ensure a stable market, whereas the other stakeholders, despite recognizing the need for reimbursement, were cautious about the decision. Specifically, along with insufficient clinical evidence, a conservative approach is required rather than supporting lowering the financial burden of patients. Therefore, although several regulatory improvements to the market entry process were suggested, it is still crucial to ensure there is robust evidence to assess the eligibility of the technology for NHI coverage [18].

Nevertheless, the participants recognized the potential need for greater flexibility in the determination of reimbursement for DTx once the evidence of their effectiveness is established. This suggests that traditional criteria assessing clinical and economic value may not adequately capture the value of DTx. Notably, the participants of this study expected DTx to improve patient health and have wider benefits for the health care system, such as reducing unmet medical needs, lowering treatment and societal costs, and providing data on treatment progress and outcome. Therefore, going forward, the DTx value should be measured broadly and from multiple stakeholders' perspectives, including those of patients, physicians, and payers, and the reimbursement criteria must reflect this broader DTx value perspective [19].

Taken together, this study highlights the necessity of an interim phase to address uncertainties regarding DTx, such as their therapeutic benefits and user acceptance, before DTx are covered by the NHI. Thus, it is reasonable to permit early market entry through the IHTA track but initially adopt DTx as a noncovered item to examine the market's response to the technology. However, the South Korean government has announced its commitment to promoting the digital health care industry as a future growth engine [20], captivating several developers to delve into DTx development. Given the advanced information and communications technology infrastructure and medical technologies in South Korea, the country may produce globally competitive DTx products. Thus, the government must increase market predictability for these companies to develop quality DTx products. Various measures that reduce the financial burden of patients will facilitate the rapid introduction and expansive adoption of DTx technology. Several countries worldwide are using government subsidies or separate funds when introducing innovative technologies [21], which may provide some options for the South Korean government. Nevertheless, if DTx should be reimbursed by the NHI fund (or if the NHI is the best option for the reimbursement of DTx), it is necessary to clarify the purpose of paying through the NHI fund: to manage the use of DTx so that they are incorporated properly into the health care system, to accumulate robust real-world evidence that is acceptable for all stakeholders, and to contribute to the soundness of the NHI fund in long-term perspectives.

Strengths and Limitations

This is the first study to comprehensively analyze the perceptions of various stakeholder groups on DTx reimbursement in South Korea. We succeeded in including diverse stakeholders of DTx in academic, industry, health care, and consumer experts who work on the front line of DTx and NHI policies; therefore, the information we collected was rich in content. However, these findings cannot be generalized to all stakeholders. Particularly, considering that this is an early stage of DTx, the participants in this study were mostly directly or indirectly related to DTx and may have greater expectations than the general population; thus, the results may seem to be favorable to the DTx industry. Therefore, the perceptions of a larger population pertaining to the reimbursement for DTx should be quantitatively explored after its widespread awareness. Nevertheless, this study is notable as a developmental study that extensively explored the DTx reimbursement issue and can be the basis for social consensus.

Conclusions

The most important concern among stakeholders in South Korea was contrasting opinions regarding the pathway to reimbursement, as opposed to problems with the technology itself. The conservative NHI fund may be insufficient to quickly adopt and use this novel technology. Thus, concurrently using various pathways such as government subsidies and innovative funds for reimbursement to accumulate clinical evidence rapidly and eliminate uncertainties may be required to ensure that the technology does not fall behind in the market.

Acknowledgments

We would like to thank all the study participants for their time and valuable insights.

Conflicts of Interest

None declared.

References

1. Understanding DTX: what is a DTX? Digital Therapeutics Alliance. URL: dtxalliance.org/understanding-dtx/what-is-a-dtx/#understand [accessed 2023-01-06]
2. De novo classification request for reSET. US Food and Drug Administration. 2016 May. URL: www.accessdata.fda.gov/cdrh_docs/reviews/DEN160018.pdf [accessed 2023-10-04]
3. Chung JY. Digital therapeutics and clinical pharmacology. Transl Clin Pharmacol 2019 Mar;27(1):6-11. [doi: 10.12793/tcp.2019.27.1.6] [Medline: 32055575]
4. Dang A, Dang D, Rane P. The expanding role of digital therapeutics in the post-COVID-19 era. The Open COVID Journal 2021 May 21;1(1):32-37. [doi: 10.2174/2666958702101010032]
5. Digital therapeutics market by application (prediabetes, nutrition, care, diabetes, CVD, CNS, CRD, MSD, GI, substance abuse, rehabilitation), sales channel (B2C (patient, caregiver), B2B (providers, payer, employer, pharma)) & region - global forecasts to 2027. Markets and Markets. 2022 Oct. URL: www.marketsandmarkets.com/Market-Reports/digital-therapeutics-market-51646724.html [accessed 2023-01-06]
6. The 25th (2021) Health Insurance Policy Deliberation Committee meeting (11.25). Article in Korean. Korea Ministry of Health and Welfare. 2021 Nov 25. URL: www.korea.kr/briefing/pressReleaseView.do?newsId=156482847 [accessed 2023-10-04]
7. Yan K, Balijepalli C, Druyts E. The impact of digital therapeutics on current health technology assessment frameworks. Front Digit Health 2021 Jun 9;3:667016. [doi: 10.3389/fdgth.2021.667016] [Medline: 34713140]

8. AMCP partnership forum: digital therapeutics-what are they and where do they fit in pharmacy and medical benefits. J Manag Care Spec Pharm 2020 May;26(5):674-681. [doi: 10.18553/jmcp.2020.19418] [Medline: 32175784]
9. Park KM, Lee S, Lee E. Can digital therapeutics open a new era of sleep medicine? Chronobiol Med 2021 Dec;3(4):142-148. [doi: 10.33069/cim.2021.0028]
10. Kim HS. Apprehensions about excessive belief in digital therapeutics: points of concern excluding merits. J Korean Med Sci 2020 Nov 23;35(45):e373. [doi: 10.3346/jkms.2020.35.e373] [Medline: 33230984]
11. Shin SY. Current status and future direction of digital health in Korea. Korean J Physiol Pharmacol 2019 Sep;23(5):311-315. [doi: 10.4196/kjpp.2019.23.5.311] [Medline: 31496868]
12. Kim SY, Moon JY, Shin J, Sim JY, Kim M, Jang J. Survey for government policies regarding strategies for the commercialization and globalization of digital therapeutics. Yonsei Med J 2022 Jan;63(Suppl):S56-S62. [doi: 10.3349/ymj.2022.63.S56] [Medline: 35040606]
13. Boyce C, Neale P. Conducting In-Depth Interviews: A Guide for Designing and Conducting In-Depth Interviews for Evaluation Input 2nd Ed: Pathfinder International; 2006.
14. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General practitioners' perspective on eHealth and lifestyle change: qualitative interview study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88. [doi: 10.2196/mhealth.8988] [Medline: 29666045]
15. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
16. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health-attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14(1):109. [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
17. Krumholz HM. 12th Korea Healthcare Congress 2021; 김치국부터 마시지 말라; the time for digital health is almost here. Yonsei Med J 2022 May;63(5):493-498. [doi: 10.3349/ymj.2022.63.5.493] [Medline: 35512753]
18. Rhee SY, Kim C, Shin DW, Steinhubl SR. Present and future of digital health in diabetes and metabolic disease. Diabetes Metab J 2020 Dec;44(6):819-827. [doi: 10.4093/dmj.2020.0088] [Medline: 33389956]
19. Kolasa K, Kozinski G. How to value digital health interventions? a systematic literature review. Int J Environ Res Public Health 2020 Mar 23;17(6):2119. [doi: 10.3390/ijerph17062119] [Medline: 32209988]
20. Fully support DTX as a future growth engine. Article in Korean. Korea Ministry of Trade, Industry, and Energy. 2022 Feb 24. URL: motie.go.kr/motie/ne/presse/press2/bbs/bbsView.do?bbs_cd_n=81&bbs_seq_n=165332&from_brf=brf&brf_code_v=13 [accessed 2023-10-04]
21. Ju JH, Sim B, Lee J, Lee JY. Reimbursement of digital therapeutics: future perspectives in Korea. Korean Circ J 2022 Apr;52(4):265-279. [doi: 10.4070/kcj.2022.0014] [Medline: 35388995]

Abbreviations

DTx: digital therapeutics
IHTA: Innovative Health Technology Assessment
MFDS: Ministry of Food and Drug Safety
NHI: National Health Insurance
nHTA: New Health Technology Assessment

Edited by L Buis; submitted 20.03.23; peer-reviewed by H Lee, JY Shin, S Hinder, T Kim; revised version received 27.08.23; accepted 08.09.23; published 30.10.23.

Please cite as:
Sim B, Ju JH, Kim B, Lee JY
Stakeholders' Perceptions Regarding Digital Therapeutics Reimbursement in South Korea: Qualitative Study
JMIR Mhealth Uhealth 2023;11:e47407
URL: https://mhealth.jmir.org/2023/1/e47407
doi:10.2196/47407

Research Letter

The Effects of Providing a Connected Scale in an App-Based Digital Health Program: Cross-sectional Examination

Lisa A Auster-Gussman[1], PhD; Mohit Rikhy[1], MS; Kimberly G Lockwood[1], PhD; OraLee H Branch[1], PhD; Sarah A Graham[1], PhD

Lark Health, Mountain View, CA, United States

Corresponding Author:
Sarah A Graham, PhD
Lark Health
2570 El Camino Real
Mountain View, CA, 94040
United States
Phone: 1 650 300 1755
Email: sarah.graham@lark.com

(JMIR Mhealth Uhealth 2023;11:e40865) doi:10.2196/40865

KEYWORDS

engagement; retention; scales; self-monitoring; mobile app; digital health; AI; smartphone; platform; app; application; health program; program

Introduction

Self-monitoring technologies (eg, digital scales) have been shown to improve health outcomes [1,2], such as weight loss, when combined with additional interventions such as coaching [3]. This may be because they facilitate increased self-weighing, which has been shown to be related to better health outcomes [4,5]. The purpose of this study was to examine whether the provision of a digital body weight scale as part of one's digital health program was related to increased self-weighing and longer retention. The primary hypothesis was that members provided with a scale would weigh more frequently and remain in the program for longer than those not provided with a scale.

Methods

Study Design

We conducted an observational study of members enrolled in an artificial intelligence (AI)–powered digital health program available via smartphone on a platform called Lark. Information about Lark is published elsewhere [5,6]. We examined differences in self-weighing and retention between members with and without scales provided by their commercial insurance provider. Members received their scales immediately after the completion of enrollment.

Ethical Considerations

The study received exemption status from Advarra Institutional Review Board (protocol #Pro00047181) for retrospective analyses of previously collected and deidentified data.

Participants, Program Description, and Inclusion Criteria

We conducted an analysis of 3488 members enrolled from 2019 to 2021 in a yearlong digital program focused on general health and well-being; weight loss was not a specific target. The program included automated personalized coaching via in-app messaging using conversational AI, weekly lessons related to healthy lifestyle choices, meal logging, and weekly weight logging. The inclusion criteria were age ≥18 years, ≥1 full year has passed since enrollment date, completion of ≥1 educational lesson in the first 6 months, and ≥1 weigh-in during the first 6 months.

Outcome Measures

We examined two key outcome variables. Weigh-ins included the total number of days with recorded weigh-ins during each member's first 6 months in the program. This included both weigh-ins from the provided scales that sync with the app automatically and manually entered weights. Scales not provided by Lark do not pair directly with the app, so weigh-ins on non-Lark scales would need to be entered manually in the app. We analyzed the first 6 months because this is the active weight loss phase of the program for members who set a goal to lose weight. The second 6 months is the maintenance phase. Active retention was the total number of days from the day the individual enrolled to the last day that they used in-app functions, such as conversation or meal logging, up to 365 days.

Analysis

We calculated descriptive statistics for the overall sample and examined differences in weigh-ins and active retention between members provided with versus not provided with scales using analysis of covariance (ANCOVA).

Results

Participants

Participants were 68.3% (2384/3488) female with a mean age of 45.19 (SD 11.45) years. Of the 3314 members who reported a starting weight, the mean starting BMI was 31.1 (SD 6.93) kg/m^2, and 48.6% (1611/3314) had obesity. Race/ethnicity data were not available for more than half of the sample and therefore not reported. Approximately 43.5% (1519/3488) of members received the insurance-provided scale, and the pairing rate was 93.7% (1423/1519).

Associations Between Provision of a Digital Scale and Descriptive Statistics

Although the groups were similar in mean starting BMI (no scale mean 31.52, SD 7.25 kg/m^2; scale mean 30.55, SD 6.44 kg/m^2) and age (no scale mean 44.63, SD 11.35 years; scale mean 45.91, SD 11.55 years), the differences were statistically significant (starting BMI t_{3312}=4.05, P<.001; age t_{3816}=–3.22; P=.001). There was also a greater proportion of women among those not provided with a scale (1411/1969, 71.7%) than among those provided with a scale (973/1519, 64.1%; χ^2_1=22.93; P<.001). Therefore, we controlled for starting BMI, age, and sex in the ANCOVAs.

Associations Between Provision of a Digital Scale, Weigh-ins, and Active Retention

The ANCOVAs revealed that members provided with versus not provided with a scale had significantly more days with weigh-ins and days of active retention (see Table 1); their mean last day with a weigh-in also occurred further into the program (no scale mean 72, SE 2 days; scale mean 138, SE 3 days). On average, members not provided with a scale weighed themselves 1-2 days per month and were retained for approximately 4 months, whereas members provided with a scale weighed themselves 1 day per week and were retained for almost 6 months.

Table 1. Analysis of covariance table of mean differences based on the provision of digital scale[a].

Engagement features	No scale, mean (SE)	Scale, mean (SE)	*F* test (*df*)	*P* value
Days with weigh-ins during first 6 months	7.35 (0.24)	21.56 (0.66)	276.80 (1, 3202)	<.001
Days of active retention in first 12 months	120.44 (2.28)	172.45 (2.77)	15.74 (1, 3202)	<.001

[a]Analyses of covariance include age, sex, BMI, and days of active retention as control variables; 3208 members included in the analysis who had no missing data needed for analyses; 1822 members with no scale and 1386 members with scale.

Discussion

These findings demonstrate that members provided with a scale recorded 3 times more days with weigh-ins and were retained almost 2 months longer than those not provided with a scale. The direct digital transfer of weights from the scale to the app greatly streamlined the weigh-in process, improving the member experience, which may have led to the observed longer retention. This research was limited by the fact that the provision of the scale was not individually randomized but the result of insurance providers' choice to provide or not provide scales. In addition, although findings from past research show that both self-weighing and retention are associated with weight loss [4,5,7], suggesting that the provision of a scale might also be related to weight loss, this question was outside the scope of this analysis. This preliminary study is timely, given the increasing importance of understanding and attenuating the high rates of attrition in digital health [8-10], and might assist insurance providers to weigh the cost of provisioning a scale to the benefits of increased retention in lifestyle behavioral programs.

Conflicts of Interest

MR, KGL, and SAG are employees of Lark Technologies, Inc.

References

1. Carpenter CA, Eastman A, Ross KM. Consistency with and disengagement from self-monitoring of weight, dietary intake, and physical activity in a technology-based weight loss program: exploratory study. JMIR Form Res 2022 Feb 18;6(2):e33603 [FREE Full text] [doi: 10.2196/33603] [Medline: 35179513]
2. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
3. Carpenter CA, Ugwoaba UA, Cardel MI, Ross KM. Using self-monitoring technology for nutritional counseling and weight management. Digit Health 2022;8:20552076221102774 [FREE Full text] [doi: 10.1177/20552076221102774] [Medline: 35663238]
4. Hori JH, Sia EX, Lockwood KG, Auster-Gussman LA, Rapoport S, Branch OH, et al. Discovering engagement personas in a digital diabetes prevention program. Behav Sci (Basel) 2022 May 24;12(6):159 [FREE Full text] [doi: 10.3390/bs12060159] [Medline: 35735369]

5. Graham SA, Pitter V, Hori JH, Stein N, Branch OH. Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. Digit Health 2022;8:20552076221130619 [FREE Full text] [doi: 10.1177/20552076221130619] [Medline: 36238752]
6. Branch OH, Rikhy M, Auster-Gussman LA, Lockwood KG, Graham SA. Relationships between blood pressure reduction, weight loss, and engagement in a digital app-based hypertension care program: observational study. JMIR Form Res 2022 Oct 27;6(10):e38215 [FREE Full text] [doi: 10.2196/38215] [Medline: 36301618]
7. Behr H, Earl S, Ho AS, Lee J, Mitchell ES, McCallum M, et al. Changes in health-promoting behaviors and their association with weight loss, retention, and engagement on a digital program: prospective study. Nutrients 2022 Nov 14;14(22):4812 [FREE Full text] [doi: 10.3390/nu14224812] [Medline: 36432498]
8. Schroé H, Van der Mispel C, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. A factorial randomised controlled trial to identify efficacious self-regulation techniques in an e- and m-health intervention to target an active lifestyle: study protocol. Trials 2019 Jun 10;20(1):340 [FREE Full text] [doi: 10.1186/s13063-019-3456-7] [Medline: 31182147]
9. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
10. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]

Abbreviations

AI: artificial intelligence
ANCOVA: analysis of covariance

Edited by L Buis; submitted 07.07.22; peer-reviewed by S Amagai, K Ross, M Patel; comments to author 09.09.22; revised version received 23.11.22; accepted 18.01.23; published 03.02.23.

Please cite as:
Auster-Gussman LA, Rikhy M, Lockwood KG, Branch OH, Graham SA
The Effects of Providing a Connected Scale in an App-Based Digital Health Program: Cross-sectional Examination
JMIR Mhealth Uhealth 2023;11:e40865
URL: https://mhealth.jmir.org/2023/1/e40865
doi:10.2196/40865
PMID:36735288

Research Letter

Uptake of Remote Physiologic Monitoring in the US Medicare Program: A Serial Cross-sectional Analysis

Jeffrey R Curtis[1], MD, MS, MPH; James Willig[2], MD, MSPH

[1]Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Jeffrey R Curtis, MD, MS, MPH
Division of Clinical Immunology and Rheumatology
University of Alabama at Birmingham
1825 University Blvd
Shelby 121H
Birmingham, AL, 35233
United States
Phone: 1 2059370585
Email: jcurtis@uab.edu

(JMIR Mhealth Uhealth 2023;11:e46046) doi:10.2196/46046

KEYWORDS

remote patient monitoring; remote therapeutic monitoring; digital health; mobile technology; mHealth; mobile health; remote monitoring; patient monitoring; cost; economic; payment; Medicare; insurance; health coverage; outpatient

Introduction

The opportunity to provide continuous care to patients between office visits using digital technologies holds tremendous potential to improve health care quality and patient outcomes. In 2019, the Center for Medicare and Medicaid Services (CMS) launched the remote physiologic monitoring (RPM) program that provided reimbursement for using technology to monitor patients between visits [1]. RPM delivers continuous or periodic digital data to a central location. These data typically are reviewed by clinical staff (eg, nurses, medical assistants) whose time is billed “incident to” the supervising physician. RPM offers an intuitive complement to remote care delivered via telehealth. In part related to the COVID-19 public health emergency, RPM subsequently was expanded by CMS to improve coverage and reduce barriers to access.

The RPM program requires that a biosensor be used to monitor patients between visits, often but not always in conjunction with a smartphone app. For many health conditions, a biosensor device is a logical component to chronic disease management. Examples include continuous glucose monitoring (diabetes), daily weights via a smart scale (heart failure), dysrhythmia detection (cardiac conditions), and ambulatory blood pressure monitoring (hypertension). Early evidence supporting RPM use appears favorable [2]. Outside of isolated published examples that have largely been confined to a single chronic illness, the extent to which RPM has been deployed on a national scale is unknown. Using US Medicare data, we examined the uptake of RPM in the United States from 2019 (its inception year) to 2021.

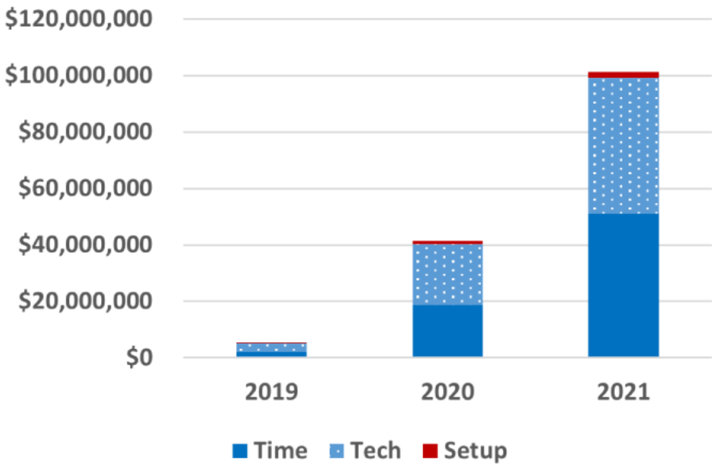
Methods

We examined publicly available Medicare Part B National Summary Data File data from January 2019 to December 2021 [3]. We extracted Medicare payment amounts and the associated services allowed based on relevant Current Procedural Terminology (CPT) codes; individual patient information is not available in this data source. RPM services were grouped as setup (CPT 99453), data transmission (CPT 99454), and monitoring time, which is billed in 20-minute increments (CPT 99457,99458). Results were stratified by calendar year and analyzed in R version 4.2.3 (R Foundation for Statistical Computing).

Results

In 2019, the total amount paid by CMS for RPM was US $5.5 million. In 2020, RPM payments increased almost 9-fold to US $41.5 million, followed by a further 2.5-fold increase in 2021, totaling more than US $101 million annually (Figure 1).

Figure 1. Medicare payments for remote patient monitoring services, 2019-2021.



Assuming providers initiate the program and bill setup fees only once per patient, the number of new patients increased from 20,640 (2019) to 90,149 (2020) and further increased to 123,476 (2021). The total payments made by CMS for the technical service (data transmission) were comparable to the payment for the time spent monitoring. Most (69%) monthly reimbursement for patient monitoring was for 20 minutes; only 31% was for monitoring beyond 20 minutes.

Discussion

Based on national data from the Medicare program, RPM grew approximately 19-fold over 3 years, suggesting rapid uptake. However, some have raised concerns about the potential for overuse of RPM without clinical benefit [3]. Moreover, the use of RPM appears to be confined to a small group of physicians, predominantly primary care providers focused on hypertension or diabetes management [4]. In addition to Medicare, both commercial insurance programs and many states' Medicaid programs also cover RPM services [4,5]. Importantly, in 2022 CMS further expanded remote monitoring for certain medical specialties (musculoskeletal [rheumatology, orthopedics], respiratory medicine). Under this new program called remote therapeutic monitoring (RTM), a software app alone can be used for monitoring, and patients provide data through the app without a biosensor [6]. The software itself is the medical device and would be registered and cleared by the US Food and Drug Administration as a class 1 (or higher) device.

Thus beginning in 2022, RTM widens the spectrum of health domains available for monitoring, since any patient-reported outcome (eg, disease activity) or clinically relevant information (eg, medication adherence) now is reimbursable. We eagerly await the evaluation of the impact on patient outcomes offered by RPM and RTM, recognizing that much work to optimize their use remains.

Acknowledgments

This study was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (P30AR072583). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability

Data can be shared and are publicly available (see manuscript for source).

Authors' Contributions

JRC contributed toward the concept and design of the study; drafting the manuscript; statistical analysis; administrative, technical, or material support; and supervision. JW made critical revisions to the manuscript for important intellectual content. JRC and JW contributed toward the acquisition, analysis, or interpretation of data and obtained funding for the study.

Conflicts of Interest

JRC receives consulting fees from and owns stock in TNacity Blue Ocean. JRC is a part-time employee of Illumination Health (a health care research organization). JW has no conflicts of interest.

References

1. Payment policies for CY 2019. Federal Register. URL: https://www.federalregister.gov/documents/2018/11/23/2018-24170/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-and-other-revisions#h-81 [accessed 2022-10-26]

2. Taylor ML, Thomas EE, Snoswell CL, Smith AC, Caffery LJ. Does remote patient monitoring reduce acute care use? A systematic review. BMJ Open 2021 Mar 02;11(3):e040232 [FREE Full text] [doi: 10.1136/bmjopen-2020-040232] [Medline: 33653740]
3. Mecklai K, Smith N, Stern AD, Kramer DB. Remote patient monitoring - overdue or overused? N Engl J Med 2021 Apr 15;384(15):1384-1386. [doi: 10.1056/NEJMp2033275] [Medline: 33853209]
4. Tang M, Mehrotra A, Stern AD. Rapid growth of remote patient monitoring is driven by a small number of primary care providers. Health Aff (Millwood) 2022 Sep;41(9):1248-1254. [doi: 10.1377/hlthaff.2021.02026] [Medline: 36067430]
5. Remote patient monitoring. Center for Connected Health Policy. URL: https://www.cchpca.org/topic/remote-patient-monitoring/ [accessed 2023-04-03]
6. Medicare Program; CY 2022 payment policies under the physician fee schedule and other changes to part B payment policies; Medicare Shared Savings program requirements; provider enrollment regulation updates; and provider and supplier prepayment and post-payment medical review requirements. Federal Register. URL: https://public-inspection.federalregister.gov/2021-23972.pdf [accessed 2022-10-25]

Abbreviations

CMS: Center for Medicare and Medicaid Services
CPT: Current Procedural Terminology
RPM: remote physiologic monitoring
RTM: remote therapeutic monitoring

Edited by L Buis; submitted 27.01.23; peer-reviewed by D Fletcher, J Sharp; comments to author 31.03.23; revised version received 04.04.23; accepted 10.04.23; published 01.05.23.

Please cite as:
Curtis JR, Willig J
Uptake of Remote Physiologic Monitoring in the US Medicare Program: A Serial Cross-sectional Analysis
JMIR Mhealth Uhealth 2023;11:e46046
URL: https://mhealth.jmir.org/2023/1/e46046
doi:10.2196/46046
PMID:37040464

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org